Termination fees revisited*

Bruno Jullien[†], Patrick Rey[‡] and Wilfried Sand-Zantman[§]

March 2012

Abstract

We reconsider the question of the optimal level of termination fees between communication networks in the context of heterogeneous usage and elastic participation. The interaction between these two features yields insights; in particular: i) The profit maximizing reciprocal termination fee is above marginal cost; ii) The welfare maximizing termination fee is also above cost, but below the former.

1 Introduction

In most communication networks, users expect to be able to interact regardless of which network they subscribe to. To achieve this, operators enter into interconnection agreements, which not only cover technical aspects, but also stipulate access fees compensating the terminating network for the cost of communications originated from another network. These so-called termination fees have been the center of many investigations, and the question of the privately and socially optimal levels of those fees is still hotly debated in many communication industries (fixed and mobile telephony, Internet...). In this paper we revisit this question by considering the impact of heterogeneous demands for both calls and subscriptions. We show that, when the consumers who call less have also a more elastic demand for subscription:

*We thank Stefan Behringer, Doh-Shin Jeon and seminar participants at IESE (Barcelona), Leuven, ARCEP, CREST-LEI, Ecole Polytechnique and ESWC (Shanghai) for helpful comments. We are also very grateful to two referees and to the editor Benjamin Hermalin.

[†]Toulouse School of Economics (GREMAQ and IDEI).

[‡]Toulouse School of Economics (GREMAQ and IDEI).

[§]Toulouse School of Economics (GREMAQ and IDEI).

i) the profit-maximizing reciprocal termination fee is above the marginal cost of termination;

ii) the welfare-maximizing reciprocal termination fee is also above cost, but below the profit-maximizing level.

A key element of our analysis is the negative correlation between the traffic originated and the elasticity of participation, a feature that is present in many networks. A good illustration is mobile telephony, which exhibits a considerable heterogeneity in usage patterns. This heterogeneity is reflected to some extent in the large variety of post-paid contracts targeting different customer categories, as well as in the differences between pre-paid and post-paid users. It is a source of traffic imbalance at the customer level, since some customers call more than they receive while others receive more than they call. Genakos and Valletti (2011b) note or example that "anecdotal evidence seems to suggest [...] that pre-paid consumers predominantly use their phone for incoming calls". Another illustration of the difference between pre-paid and post-paid clients is given by a change in the collection of data on mobile traffic by the French regulator that occurred during the year 2005.[1] During the first semester (Q1 and Q2 in Table 1 below), volumes included the minutes of calls emitted, along with fixed-to-mobile termination and roaming. Afterwards, the volumes also included the number of minutes of off-net mobile-to-mobile calls received. Using these data, for each quarter of 2005 we computed average volumes for pre-paid and post-paid customers:

Volume per subscriber 2005 (mn)	$Q1$	$Q2$	$Q3$	$Q4$	$Q3/Q2$
Post-pay	786	798	837	867	+5%
Pre-pay	156	159	205	201	+30%

Table 1

The data from the first two quarters confirms that pre-paid customers call much less than post-paid ones. But the difference between the third and second quarters, representing the volume of calls received from other mobile networks, also shows that the proportion of calls received is much higher for pre-paid customers.[2]

[1] France moved away from bill-and-keep for mobile to mobile termination fees in January 1, 2005. The regulatory regime was thus stable throughout 2005, but the change in the statistics published by the *Observatoire des mobiles* was only introduced in the third quarter.

[2] Using the first two quarters to account for dynamic trends does not affect the 5% and 30% figures for post-pay and pre-pay. Assuming that the proportions of outgoing calls within the data perimeter of the first quarters, and of on-net calls among incoming mobile calls, are stable from one quarter to another, this means that the ratio of calls received to calls emitted is 6 times higher for pre-pay users.

Other examples include Internet, since most of the traffic originates from websites but the extensive margin of the market includes many content users, and the convergence between fixed and mobile services, as fixed line customers still call more mobile customers than the reverse.

Our paper is concerned with termination fees between networks that are active on all segments of the market, a situation that is often referred to as a two-way access problem. When callers and receivers do not belong to the same network, a network terminating a communication enjoys market power as it can hardly be bypassed and the receiving customer is not necessarily sensitive to the price paid by those who call him. One of the main conclusions of the existing literature on termination fees is that network operators should collectively favor low fees, which is somewhat at odds with the observation that, in practice, network operators often resist reducing those fees. We aim at reconciling theory and practice and show why firms may favor above cost termination fees. We show that the socially optimal fee is also above cost.

Formally, we use the framework of Laffont, Rey and Tirole (1998a) – hereafter LRT – in which we introduce user heterogeneity – and also account for the utility of receiving calls. Our model is based on the above observation that the willingness to pay for a subscription is related to the volume of calls. Customers with very large volumes of calls are infra-marginal customers, who may switch between operators when prices increase but always subscribe to one operator; marginal customers are instead those who also call less. We thus distinguish two types of customers: heavy and light users; the latter not only call less often, but their demand for subscription is also more elastic.

To keep things simple, we assume the following:

- light users only receive calls;[3] we moreover first consider a benchmark model where their utility from receiving calls is fixed; later on, we account for endogenous reception utility;
- network operators can offer different two-part tariffs, each including a subscription fee and a unit price for calls, to heavy and light users;later on, we also allow the operators to charge different prices for on-net and off-net calls (termination-based price discrimination).

In each situation, we analyze the impact of reciprocal termination fees on subscription and usage prices, as well as on profits and welfare. In equilibrium, usage prices are equal to perceived costs and there is no profit from

[3]In our 2010 working paper, we show that our results extend to the case where light users have a small demand for calls.

origination; network operators' profit is thus driven by termination profit and by subscription fees. We identify two new effects:

Raising termination profit weakens the competition for heavy users: introducing light users reduces competition for heavy users when the termination fee is above cost, since the operators then obtain more profit from terminating off-net calls than on-net calls; losing a heavy user to the competitor thus raises the termination profit on light users – without generating an equivalent cost, as light users call less than they are called.

Raising termination profit intensifies the competition for light users: since light users generate a positive termination balance, they become more profitable when the termination mark-up increases, hence a reduction in the equilibrium price due to increased retail competition. In our setting this "waterbed" effect[4] is however modified, due to the fact that losing light users to the competing network generates a termination deficit, since light users are mainly receivers; this additional cost further intensifies competition for light users.

In the case of uniform pricing, the former effect dominates for profit while the latter dominates for welfare. As a result, both profit and welfare are maximal for termination fees that are above cost. The operators prefer a positive mark-up because the extra revenue from termination by heavy users is not fully competed away through subscription fees. Adopting a positive termination mark-up also increases welfare because it generates a market expansion that benefits all customers – in contrast, in the absence of any scope for demand expansion, welfare would be maximized for cost-based termination fees. A conflict arises, however, since network operators favor excessively high termination fees.

When on-net pricing is allowed, the market exhibits tariff-mediated network effects: with a positive termination mark-up, the off-net price is above the on-net price so a customer is better off joining a larger network; these network effects in turn intensify competition, as pointed out by Laffont, Rey and Tirole (1998b) and Gans and King (2001). In our setting, while network effects mostly concern heavy users, the operators compete more fiercely for both heavy and light users, and we show that welfare is still maximized for a termination rate that lies above cost; the operators also prefer an above-cost termination fee when the size of the demand from light users is not too small.

Finally when the receivers' utility is not fixed but depends on the number of calls received, the usage price is distorted downward to generate more calls, more utility and thus higher revenues for the firms. By way of numerical simulations, we show that the main insights – profit-maximizing and welfare-

[4]The term was coined by Paul Geroski. See Schiff (2008) for a formal analysis.

maximizing termination fees are above cost – extend to this more general model.

Starting with the work of Armstrong (1998) and Laffont, Rey and Tirole (1998a,b), a body of literature has analyzed the role of termination fees in industries with two-way access.[5] In particular, LRT found that the termination fee had no impact on equilibrium profits when networks compete in two-part tariffs and subscription demand is inelastic, a result extended to heterogeneous calling patterns by Dessein (2003) and Hahn (2004). Subsequent work suggests that network operators should favor *below-cost* termination fees: see e.g. Gans and King (2001) for competition in two-part tariffs with termination-based price discrimination, Berger (2004, 2005) taking into account call externalities,[6] or Dessein (2003) and (2004) for, respectively, elastic but homogenous demand and heterogenous but inelastic demand. This paper shows that allowing instead for both heterogeneity and elastic demand drastically changes the previous conclusion, leading to above-cost private and social optimal termination fees.

DeGraba (2004) and Hermalin and Katz (2010) compare different interconnections pricing schemes when operators can charge both callers and receivers;[7] they focus on the role of cost sharing in achieving optimal usage while we focus on participation. Atkinson and Barnekov (2000) consider instead the allocation of investment costs in a setting where usage demand is fixed.

In the case of mobile telephony, Armstrong and Wright (2009) explain the opposition of network operators to a reduction in termination fees by a link (due to arbitrage possibilities) between fixed and mobile rates along with the importance of fixed-to-mobile termination revenue.[8] We show that accounting for demand heterogeneity yields similar insights even in the absence of fixed-to-mobile termination – and that socially optimal termination fees are then also above cost. The same thus applies to any communication network with the above pattern of demand.[9] In the case of mobile telephony, Genakos and Valletti (2011a and 2011b) estimate the impact of the regula-

[5] See Armstrong (2002) for an overview of this literature.

[6] Note that adding constraints on two-part tariffs, such as participation constraints as in Poletti and Wright (2004), may lead to non-neutrality.

[7] Bolt and Tieman (2006) discuss the conflict between social efficiency and cost recovery in this context.

[8] Calzada and Valleti (2008) and Lopez and Rey (2009) points to the possibility that high termination fees may act as a barrier to entry.

[9] Convergence in communication markets favors integration so that the distinction between fixed and mobile operators may not be relevant in the future. In France, for instance, all four operators now propose fixed and mobile services.

tion of termination fees on retail tariffs, confirming the existence of a partial waterbed effect, which is in line with an elastic participation. Their results for post-paid and pre-paid contracts are also consistent with the predictions of our model.[10] Recently, several papers have explored other features that may explain mobile operators' attitude towards termination fees. Hoernig, Inderst and Valletti (2011) account for heterogeneity in the destination of calls and show that, with termination-based price discrimination, the profit-maximizing termination fee is above cost when calling patterns are sufficiently concentrated on on-net calls. Hurkens and Lopez (2010) suppose that consumers' expectations about networks' market shares do not react to changes in prices, and show that, with on-net price discrimination, this also results in profit-maximizing termination fee above cost. We focus instead on heterogeneity in the volume of calls, and rational expectations about market shares, and our conclusions are valid both with and without on-net price discrimination. In a very recent paper, Tangeras (2012) shows that accounting for wealth effects in the demand function may also help explaining why profit increases with termination fees above cost. We do not consider wealth effects here and focus instead on quasi-linear utilities.

The paper is organized as follows. Section 2 presents our model. Section 3 develops the main insights of our analysis in a simplified framework. Section 4 extends the results to the general model. Section 5 concludes.

2 The model

Two mobile operators 1 and 2 compete for two types of customers: *heavy users* wishing to call as well as to receive calls, and *light users* who are only interested in being reached.

We will denote by c the cost of a call.[11] In the case of an off-net call, the calling network pays an access fee to the receiving network, which is assumed to be reciprocal and non-negative;[12] denoting by m the termination mark-up (i.e., the difference between the access fee and the actual cost of terminating the call) earned by the receiving network, the originating network thus bears

[10] We show in our 2010 working paper that the waterbed effect of FTM rates is stronger for heavy users than for light users, which is confirmed by Genakos and Valletti (2011b).

[11] For the sake of exposition, we will ignore here fixed costs (per network as well as per customer), as they do not affect the qualitative analysis (in what follows, subscription fees can be interpreted as net of per subscriber fixed costs, and profits as gross of network fixed costs).

[12] Negative termination charges could generate abuses.

a cost $c+m$. To study the impact of this access fee on network competition, we consider the following timing:

- first, the reciprocal access mark-up m is set (more on this below);
- second, the two operators compete in retail prices.

We moreover assume that the two types of users are sufficiently different that each network $i = 1, 2$ can discriminate between them by offering two contracts: a two-part tariff for heavy users, which consists of a subscription fee F_i^H and a price p_i (per unit) for originating a call, and a simple fixed fee F_i^L for light users (together with a high usage price, say). In practice, light users' contracts do not include many of the services offered to heavy users, making these contracts unattractive to heavy users. Conversely, heavy users' contracts often involve quantity forcing elements which make them unattractive to light users.[13]

Networks are differentiated and face symmetric demands:

- Light users' demand for network i is given by $L_i = D^L(F_i^L, F_j^L; V_i^L, V_j^L)$ (for $i \neq j = 1, 2$), where V_i^L represents light users' utility from receiving calls from network i.
- Heavy users are uniformly distributed on an Hotelling line of length 1, the network operators being located at the two ends of the segment.

Heavy users moreover have a balanced calling pattern and thus call all subscribers (heavy and light) with equal probability. A volume of calls q gives them a utility $u(q)$ per subscriber; a usage price p generates a volume of calls per subscriber given by

$$q(p) \equiv \arg\max_{q \geq 0} \{u(q) - pq\},$$

where we assume that $q(p)$ is differentiable; we will denote by $v(p) \equiv \max_{q \geq 0} \{u(q) - pq\}$ the surplus so achieved. Letting H_T and L_T denote, respectively, the total number of (connected) heavy and light users, by subscribing to network i a heavy user located at distance x obtains a net utility

$$w_i + u_0 - \frac{x}{2\sigma},$$

[13] In our setting, where heavy users' contracts are efficient but take the form of two-part tariffs, they may attract light users if competition were to drive heavy users' subscription fees below light users' fees. This however never occurs in the simulations of section 4.2 – and thus would a fortiori not occur either if the same offers were implemented through quantity forcing contracts.

where

$$w_i \equiv (H_T + L_T)\, v(p_i) - F_i^H \tag{1}$$

denotes the variable net surplus from placing calls, whereas the other terms reflect a fixed utility from being connected (including the utility from receiving calls), the parameter σ measuring the degree of substitution between the two networks.

We first provide below a complete analysis of the case where light users' utility from receiving calls is not sensitive to the volume of calls. We then provide a characterization of the equilibrium in the more general case, and check the robustness of our insights by way of simulations.

3 Main insights

In this section, we present our main insights in a simplified framework that neglects the impact of call volumes on light users' utilities: that is, V_i^L does not depend on heavy users' actual calling behavior; for the sake of exposition, we will then denote by $D\left(F_i^L, F_j^L\right) \equiv D^L(F_i^L, F_j^L; V_i^L, V_j^L)$ the resulting subscription demand. We will moreover assume that demand for network i is bounded, twice continuously differentiable with bounded derivatives, decreasing with its own price[14] ($D_1 < 0$) and nondecreasing with the rival's price ($D_2 \geq 0$), and that the aggregate demand, $L_T = D_T(F^L) \equiv 2D(F^L, F^L)$, is decreasing ($D_1 + D_2 < 0$).

3.1 Retail equilibrium

It is well-known that departing from cost-based termination fees (i.e., $m \neq 0$) may introduce non-concavity problems. However, building on the analysis of LRT, it can be checked that a unique symmetric equilibrium exists as long as the termination mark-up is not too large and/or networks are sufficiently differentiated. Thus, throughout the paper we will assume the following:

Assumption A

1. u_0 is large enough to ensure that heavy users always subscribe; normalizing their total mass to 1 (so that $H_T = 1$), operator i's subscription demand (and market share) from heavy users is then given by:

$$H_i = \frac{1}{2} + \sigma\left(w_i - w_j\right). \tag{2}$$

[14] In the following, D_i denotes the partial derivative of the demand function D with respect to its i^{th} argument.

2. $v(0)$ and $q(0)$ are bounded and the two networks are sufficiently differentiated that there always exists a unique, pure strategy equilibrium, in which the two networks share the market equally.[15]

Assumption A ensures that, for any given access mark-up m, there exists a unique, symmetric equilibrium, moreover characterized by first-order conditions. We will thus focus on these conditions in what follows.

For given prices $\left(p_i, F_i^H, F_i^L\right)_{i=1,2}$ and given subscription demands from heavy and light users $(H_i, L_i,)_{i=1,2}$, network i's profit, for $i \neq j = 1, 2$, is:

$$\begin{aligned}\Pi_i &= H_i \left[(1 + L_T)(p_i - c)\, q(p_i) - (H_j + L_j)\, mq(p_i) + F_i^H\right] \\ &\quad + (H_i + L_i)\, H_j mq(p_j) + L_i F_i^L. \end{aligned} \tag{3}$$

A first and by now standard step consists in optimizing with respect to the usage price p_i, adjusting the fee F_i^H so as to maintain the surplus $w_i = (1 + L_T)\, v(p_i) - F_i^H$ (that is, $\partial F_i^H / \partial p_i\big|_{w_i} = -(1 + L_T)\, q(p_i)$); this keeps market shares constant and yields a marginal gain:

$$\begin{aligned}\left.\frac{\partial \Pi_i}{\partial p_i}\right|_{w_i, F_i^L} &= H_i \left[(1 + L_T)\left((p_i - c)\, q'(p_i) + q(p_i)\right) - (H_j + L_j)\, mq'(p_i) - (1 + L_T)\, q(p_i)\right] \\ &= H_i q'(p_i) \left[(1 + L_T)(p_i - c) - (H_j + L_j)\, m\right],\end{aligned}$$

which, evaluated at a symmetric equilibrium ($H_j = 1/2, L_j = L_T/2$), leads to:

$$p_1 = p_2 = p^* \equiv c + \frac{m}{2}. \tag{4}$$

As in the previous literature, the networks price usage at the average perceived marginal cost. Given these equilibrium prices, network i's profit is equal to:

$$\Pi_i = H_i \left[(H_i + L_i)\frac{mq^*}{2} - (H_j + L_j)\frac{mq^*}{2} + F_i^H\right] + (H_i + L_i)\, H_j mq^* + L_i F_i^L,$$

where $q^* \equiv q(p^*)$ denotes the equilibrium volume of calls per subscriber and:

$$H_i = 1 - H_j = \frac{1}{2} - \sigma\left(F_i^H - F_j^H\right).$$

[15] Laffont, Rey and Tirole (1998a) show that, in the case of homogeneous users, a symmetric shared-market equilibrium exists when m and/or σ is small enough. The argument can easily be extended here (as in Dessein (2003), who considers the case of implicit discrimination among heterogenous users); in particular, the bound on $v(0)$ puts a limit on non-concave terms in profit expressions for $m > 0$, while the restriction to non-negative termination charges puts a similar limit for $m < 0$. Lopez and Rey (2009) provide a detailed analysis of the existence of shared-market and cornered market equilibria for m and/or σ large.

Differentiating with respect to the subscription fee F_i^H yields, at a symmetric equilibrium:

$$\begin{aligned}\left.\frac{\partial \Pi_i}{\partial F_i^H}\right|_{F_1^H=F_2^H=F^{H*},L_1=L_2=L_T/2} &= -\sigma F^{H^*}+\frac{1}{2}\left(-\sigma mq^*+1\right)+\left(-\sigma\frac{1}{2}+\sigma\frac{1+L_T}{2}\right)mq^* \\ &= \frac{1}{2}-\sigma F^{H^*}+\sigma\left(L_T-1\right)\frac{mq^*}{2}.\end{aligned}$$

Therefore, the equilibrium fixed fee F^{H^*} is given by

$$F^{H^*}=\frac{1}{2\sigma}+\left(L_T-1\right)\frac{mq^*}{2}. \tag{5}$$

Condition (5) is similar to that obtained by LRT, except for the term in L_T. To understand this condition, consider first the profits earned by network i on the calls made or received by a heavy user. If the user subscribes to network i, his own calls generate no profit since the usage price reflects the average variable cost, taking into account the termination mark-up paid on the proportion of off-net calls; the calls received from the same network generate however a *retail* profit equal to $H_i^*\left(p^*-c\right)q^*=mq^*/4$, while the calls received from the rival network generate a *termination* profit equal to $H_j^* mq^*=mq^*/2$. If the user subscribes instead to the rival network, his calls to network i's heavy users generate a termination profit of $H_i^* mq^*=mq^*/2$, while the calls he receives from network i generate a retail profit $H_i^*\left(p^*-c-m\right)q^*=-mq^*/4$, due to the difference between the price and the cost of an off-net call. On the whole, attracting the user generates a net gain equal to $mq^*/2$, as in LRT, and equilibrium fees are reduced by this amount.

The existence of light users mitigates this first impact. While the calls placed by network i's subscribers to light users still generate no profit, calls from the rival network generate a termination profit equal to $L_i mq^*=L_T mq^*/2$. *Losing* a heavy user to the rival thus generates an additional net gain of $L_T mq^*/2$, which increases equilibrium fees by the same amount.

Conditions (4) and (5) characterize the retail equilibrium prices for heavy users, for a given number of light users. Setting heavy users' prices to their equilibrium values (which yields $H_1^*=H_2^*=1/2$), and letting

$$t\equiv mq\left(c+\frac{m}{2}\right)$$

denote the termination profit, network i's total profit becomes:

$$\Pi_i=\frac{1}{2}\left[\left(\frac{1}{2}+L_i\right)\frac{t}{2}-\left(\frac{1}{2}+L_j\right)\frac{t}{2}+F^{H^*}\right]+\left(\frac{1}{2}+L_i\right)\frac{t}{2}+L_iF_i^L.$$

Maximizing this profit with respect to F_i^L amounts to maximizing

$$(F_i^L + \frac{3t}{4}) D\left(F_i^L, F_j^L\right) - \frac{t}{4} D\left(F_j^L, F_j^L\right). \tag{6}$$

In the first term, $3t/4$ reflects the profit attached to the calls received by network i's light users, namely, a retail profit $H_i^*\left(p_i^* - c\right) q\left(p_i^*\right) = t/4$ from on-net calls and a termination profit $H_j^* m q\left(p_j^*\right) = t/2$ from incoming off-net calls; the last term represents the termination deficit generated by the calls to the rival network, $H_i^*\left(p_i^* - c - m\right) q\left(p_i^*\right) = -t/4$.

Let $\gamma(F^L) \equiv -D_2(F^L, F^L)/D_1(F^L, F^L)$ denote the "replacement ratio" , which is such that $0 \leq \gamma(F^L) < 1$,[16] and $\mu(F^L) \equiv -D(F^L, F^L)/D_1(F^L, F^L)$ denote the "market power" associated with the demand from light users, in the terminology of Weyl and Fabinger (2009). The first-order condition that characterizes the unique, symmetric equilibrium, $F_1^L = F_2^L = F^{L^*}$, is:

$$F^{L^*} + \frac{3 + \gamma\left(F^{L^*}\right)}{4} t = \mu(F^{L^*}), \tag{7}$$

Differentiating (7) yields

$$\frac{dF^{L^*}}{dt} = -\frac{\frac{3+\gamma\left(F^{L^*}\right)}{4}}{1 - \mu'\left(F^{L^*}\right) + \gamma'\left(F^{L^*}\right)\frac{t}{4}}. \tag{8}$$

It follows that

$$-1 < \frac{dF^{L^*}}{dt} < 0 \tag{9}$$

whenever

$$\mu'\left(F^L\right) < \left(1 - \gamma\left(F^L\right) + \gamma'\left(F^L\right) t\right)/4. \tag{10}$$

We will assume that this is indeed the case:

Assumption B $\mu'(.)$ and $|\gamma'(.)|$ are small enough, so that (9) is satisfied for any t in the relevant range.[17]

Assumption B implies that, for any termination mark-up m, (7) has a unique solution, which moreover satisfies (9). This assumption holds for example when (i) $\mu' \leq 0$, which ensures that a reduction in the opportunity cost of servicing consumers is (at least partially) passed on to them through

[16] The limit case $\gamma = 1$ would correspond to the case of fixed participation.

[17] Since the termination charge cannot be negative, t is lowest for "bill and keep" ($t = -c_T q\left(c - c_T/2\right)$ for $m = -c_T$, where c_T denotes the actual cost of terminating a call); conversely, t cannot exceed the "monopoly level" $\max_m mq\left(c+m\right)$.

a reduction in the price (see Weyl and Fabinger (2009)) and is satisfied by many demand functions, and (ii) γ' is close to zero, so that the opportunity cost of letting consumers join the rival's network (the second term in (6)) does not alter this partial pass-through property. Assumption B holds for many usual demand specifications. For example, $\gamma' = 0$ and $\mu' = -1$ for a linear demand ($D = \alpha - \beta F_i^L + \gamma\beta F_j^L$) and $\gamma' = \mu' = 0$ for a log-linear demand ($\log D = \alpha - \beta F_i^L + \gamma\beta F_j^L$), so that (10) holds for any $\gamma < 1$. For a constant elasticity demand ($D = L\exp\left(-\varepsilon F_i^L + \gamma\varepsilon F_j^L\right)$), $\gamma' = 0$ but $\mu' = 1/\varepsilon > 0$; yet, (10) holds when the elasticity ε is large enough (namely, $\varepsilon > 4/(1-\gamma)$).[18]

3.2 Choosing the termination rate

Let us now derive the private and social optimal values for the termination mark-up m. Evaluating (3) for $F_i^H = F^{H^*}$ and $F_i^L = F^{L^*}$ yields each operator's equilibrium profit:

$$\Pi^* = \frac{1}{4\sigma} + \left(F^{L^*} + t\right)\frac{L_T^*}{2}, \tag{11}$$

where $L_T^* \equiv D_T(F^{L^*}(t))$ denotes the equilibrium number of light users. This equilibrium profit depends on m only through $t = mq(c + m/2)$: this termination profit affects the subscription revenue from light-users ($F^{L^*}L_T^*/2$) as well as the last term ($tL_T^*/2$), which captures two effects. First, the presence of light users reduces the intensity of competition for heavy users; as a result, the profit on heavy users increases by $tL_T^*/4$. Second, light users generate a termination profit $tL_T^*/4$ (which is however partially granted back through a reduction in F^{L^*}).

Similarly, light users' equilibrium surplus, S^{L^*}, depends only on F^{L^*} and thus on t. As for heavy users, their surplus can be written as:

$$\begin{aligned} S^{H^*} &= (1 + L_T^*)\, v(p^*) - F^{H^*} - \frac{1}{8\sigma} \\ &= (1 + L_T^*)\, v(p^*) + \frac{t}{2}(1 - L_T^*) - \frac{5}{8\sigma}. \end{aligned} \tag{12}$$

The termination mark-up m thus affects their surplus both through the access profit t and through the equilibrium price $p^* = c + m/2$.

[18] For a logit demand ($D = \frac{\alpha\exp\left(-\varepsilon F_i^L\right)}{\beta + \exp\left(-\varepsilon F_i^L\right) + \exp\left(-\varepsilon F_j^L\right)}$), $\mu' < 0$ and $\gamma' > 0$; Assumption B thus holds for $t \geq 0$. By continuity, it holds as well in the range $t \in [-c_T q\,(c - c_T/2)\,, 0]$ when c_T is small, as it is indeed the case in practice.

Let us now define the "monopoly" termination mark-up,

$$m^M \equiv \arg\max_m t\left(m\right),$$

which, for the sake of exposition, is assumed to be unique.[19] We can note a useful preliminary result:

Proposition 1 *For any $m > m^M$, there exists $\tilde{m} < m^M$ that Pareto dominates m.*

Proof. Take any candidate $m > m^M$. Since $t(0) = 0$ and $t(.)$ is continuous (by the continuity of demand), there exists $\tilde{m} \in [0, m^M]$ such that $t(\tilde{m}) = t(m)$, and:

- the profit is the same for m and $\tilde{m}$ since it only depends on t;
- for the same reason, light users' surplus is also the same for m and $\tilde{m}$;
- heavy users' surplus is higher with $\tilde{m}$ than with m since p^* is lower for $\tilde{m}$.

■

We now show that the monopoly rate maximizes networks' equilibrium profit:

Proposition 2 *The profit-maximizing termination mark-up is positive and equal to the monopoly termination mark-up m^M.*

Proof. The impact of t on total profits is given by:

$$\frac{d\left(2\Pi^*\right)}{dt} = D'_T \frac{dF^{L^*}}{dt}\left(F^{L^*} + t\right) + L^*_T\left(\frac{dF^{L^*}}{dt} + 1\right). \tag{13}$$

For $m \geq 0$, $t \geq 0$ and thus

$$F^{L^*} + t \geq F^{L^*} + \frac{3 + \gamma\left(F^{L^*}\right)}{4} t = \mu(F^{L^*}) > 0,$$

where the equality stems from (7). Since $D'_T < 0$ and $0 > \frac{\partial F^{L^*}}{\partial t} > -1$, total profits therefore increase with t. Hence, in the range $m \geq 0$, the profits are maximal for $m = m^M$.

[19] If the monopoly rate m^M is not uniquely defined, the same analysis applies to its lowest value.

For $m < 0$, a similar reasoning applies as long $F^{L^*} + t \geq 0$, since $\frac{dt}{dm} = \frac{mq'}{2} + q > 0$ for $m < 0$; if instead $F^{L^*} + t < 0$, then from (11) the equilibrium profit is lower than $\frac{1}{4\sigma}$; but for $t = 0$, $F^{L^*} = \mu\left(F^{L^*}\right) > 0$ from (7), and thus $\Pi^* > \frac{1}{4\sigma}$ from (11). Therefore, $m < 0$ cannot yield more profit than $m = 0$, which implies that the profit-maximizing termination mark-up is m^M. ■

That operators favour a high termination mark-up is not entirely surprising: if only light users were present (and receiving calls from some external source, say), the operators would indeed favour maximal termination profit since, with an elastic demand, only part of the termination profit would be passed-on to consumers through lower prices; conversely, if only heavy users were present, LRT shows that operators would be indifferent as to the level of the termination fee. To be sure, when both categories of users are present, their interaction affects the intensity of competition for each segment; yet, our analysis shows that these new effects also induce the operators to favour maximal termination profit $\left(m = m^M\right)$.

Let us now turn to users. Light users' equilibrium surplus is of the form $S^{L^*} = S^L\left(F^{L^*}, F^{L^*}\right)$, where $S^L\left(F_1^L, F_2^L\right)$ is such that $\frac{\partial S^L}{\partial F_i^L} = -L_i$. Therefore:

$$\frac{dS^{L^*}}{dt} = \left(\frac{\partial S^L}{\partial F_1^L} + \frac{\partial S^L}{\partial F_2^L}\right)\frac{dF^{L^*}}{dt} = -L_T^* \frac{dF^{L^*}}{dt} > 0.$$

As light users' equilibrium surplus increases with t, it is maximal for $m = m^M$. As for heavy users, we show in the appendix that, at $m = 0$:

$$\left.\frac{dS^{H^*}}{dm}\right|_{m=0} = L_T^* v\left(c\right)\left(\left|\frac{D_T'\left(F^{L^*}\right)}{D_T\left(F^{L^*}\right)}\right|\left|\left.\frac{dF^{L^*}}{dm}\right|_{m=0}\right| - \left|\frac{v'\left(c\right)}{v\left(c\right)}\right|\right). \tag{14}$$

Therefore:

Proposition 3 *Light users' surplus increases with the termination profit and is thus maximal for $m = m^M$; furthermore, for m small, increasing m raises heavy users' surplus if light users' subscription demand is very elastic or if heavy users' usage surplus is not very elastic.*

Proof. See Appendix A.1. ■

The effect on heavy users is two-fold. First, raising the termination mark-up reduces the net surplus from usage, which in the presence of light users is no longer fully compensated by a reduction in subscription fees. Second, heavy users benefit from an increase in light users' participation, due to intensified competition on this customer segment. The latter effect dominates if the subscription demand of light users is sufficiently elastic.

Finally, total welfare can be written as:

$$W^* = \left[(1+L_T^*)\left(v^* + \frac{mq^*}{2}\right) - \frac{1}{8\sigma}\right] + \left[S^L(F^{L^*}, F^{L^*}) + L_T^* F^{L^*}\right]. \qquad (15)$$

The first term within bracket represents the joint surplus generated with heavy users, including call termination profits. The second term represents the joint surplus generated with light users (excluding termination profits). We then obtain:

Proposition 4 *The welfare-maximizing termination mark-up is positive and strictly less than m^M.*

Proof. Using $p^* = c + \frac{m}{2}$ and $\frac{dS^{L^*}}{dm} = -L_T^* \frac{dF^{L^*}}{dm}$, we have:

$$\frac{dW^*}{dm} = (1+L_T^*)\frac{mq'(p^*)}{4} + \left(v^* + \frac{mq^*}{2} + F^{L^*}\right) D_T'\left(F^{L^*}\right)\frac{dF^{L^*}}{dm}. \qquad (16)$$

For $m \leq 0$, (i) the first term is non-negative, (ii) $D_T'\left(F^{L^*}\right)\frac{dF^{L^*}}{dm} > 0$ (since $\frac{dF^{L^*}}{dm} = \frac{dF^{L^*}}{dt}t'(m)$, where $\frac{dF^{L^*}}{dt} < 0$ and $t'(m) = q^* + \frac{mq'(p^*)}{2} > 0$), and (iii) from (7), $F^{L^*} \geq -\frac{3+\gamma^*}{4}mq^*$; therefore:

$$\frac{dW^*}{dm} \geq \left(v^* - \frac{1+\gamma^*}{4}mq^*\right) D_T'\left(F^{L^*}\right)\frac{dF^{L^*}}{dm} > 0.$$

Since $\frac{dW^*}{dm} > 0$ for $m \leq 0$, from Proposition 1 the socially optimal termination mark-up lies in the range $\left]0, m^M\right]$. To conclude the proof, it suffices to note that, at $m = m^M$ (> 0), $\frac{dF^{L^*}}{dm} = \frac{dF^{L^*}}{dt}\frac{dt}{dm} = 0$ and thus:

$$\left.\frac{dW^*}{dm}\right|_{m=m^M} = (1+L_T^*)\frac{mq'(p^*)}{4} < 0.$$

Hence, the welfare-maximizing level is strictly below m^M. ■

Therefore, the presence of light users, whose participation is elastic,[20] leads to favoring a positive termination mark-up. Note that the above analysis puts the same weight on both categories of users. If a regulator wanted

[20] In the case of a fixed participation (i.e., D_T constant), $\frac{dW^*}{dm} = (1+D_T)\frac{mq'(p^*)}{4}$ and thus welfare is maximal for $m = 0$.

to promote the participation of light users, thus placing a higher weight on those users, the optimal termination mark-up would be even higher. Note moreover that raising the termination fee above cost may benefit here *all* categories of agents. In particular, if the participation of light users is quite elastic, heavy users are also better off with a positive mark-up, as this increases their calling opportunities.

3.3 On-net pricing

We now allow networks to set different prices for on-net and off-net calls. We keep the same notation as before except that p_i and $\hat{p}_i$ now denote the prices that network i charges for on-net and off-net calls. Network i's profit becomes, for $i \neq j = 1, 2$:

$$\begin{aligned}\Pi_i &= H_i\left[(H_i + L_i)(p_i - c)\, q(p_i) + (H_j + L_j)(\hat{p}_i - c - m)\, q(\hat{p}_i) + F_i^H\right] \\ &\quad + (H_i + L_i)\, H_j m q(\hat{p}_j) + L_i F_i^L,\end{aligned}$$

where:

$$\begin{aligned}H_i &= \frac{1}{2} + \sigma(w_i - w_j), \\ w_i &= (H_i + L_i)v(p_i) + (H_j + L_j)v(\hat{p}_i) - F_i^H.\end{aligned}$$

This profit can also be written as a function of w_i rather than F_i^H:

$$\begin{aligned}\Pi_i &= H_i[(H_i + L_i)((p_i - c)\, q(p_i) + v(p_i)) \\ &\quad (H_j + L_j)((\hat{p}_i - c - m)\, q(\hat{p}_i) + v(\hat{p}_i)) - w_i] \\ &\quad + (H_i + L_i)\, H_j m q(\hat{p}_j) + L_i F_i^L\end{aligned}$$

Differentiating with respect to usage prices p_i and $\hat{p}_i$ while adjusting the subscription fee F_i^H so as to keep constant the surplus w_i (and thus the market shares) yields:

$$p_1 = p_2 = c \text{ and } \hat{p}_1 = \hat{p}_2 = \hat{p} = c + m.$$

Using the notation $\hat{q} \equiv q(c + m)$, $v \equiv v(c)$ and $\hat{v} \equiv v(c + m)$, network i's profit can be written as:

$$\Pi_i = H_i F_i^H + H_j(H_i + L_i)m\hat{q} + L_i F_i^H, \tag{17}$$

where the market shares can be expressed as a function of the fixed fees:

$$\begin{aligned}H_i &= \frac{1}{2} + \sigma(w_i - w_j) \\ &= \frac{1}{2} + \sigma[(2H_i - 1 + L_i - L_j)(v - \hat{v}) - (F_i^H - F_j^H)],\end{aligned}$$

and thus:

$$H_i - \frac{1}{2} = \sigma \frac{(L_i - L_j)(v - \hat{v}) - (F_i^H - F_j^H)}{1 - 2\sigma(v - \hat{v})}. \tag{18}$$

Differentiating (17) with respect to F_i^H then yields, at a symmetric equilibrium:

$$\left.\frac{\partial \Pi_i}{\partial F_i^H}\right|_{L_1 = L_2 = \frac{L_T}{2}, F_1^H = F_2^H = F^H} = \frac{1}{2} - \frac{\sigma}{1 - 2\sigma(v - \hat{v})} \left(F^H - \frac{1 + L_T}{2} m\hat{q} + \frac{m\hat{q}}{2} \right),$$

which leads to:

$$F_1^H = F_2^H = \hat{F}^H = \frac{1}{2\sigma} + L_T \frac{m\hat{q}}{2} - (v - \hat{v}). \tag{19}$$

To understand this characterization of the equilibrium fees, it is useful to decompose again the profits generated by the calls made or received by a heavy user. If he subscribes to network i, as in the absence of on-net pricing, his calls generate no profit, since usage prices reflect again marginal costs, including the termination mark-up in the case of off-net calls; as for the calls received, those originating off-net still generate an termination profit $m\hat{q}/2$, whereas those originating on-net no longer generate any profit since the price of these calls now reflect their actual cost. If instead the user switches to the rival network, then his off-net calls generate a termination profit $(1 + L_T)\, m\hat{q}/2$ whereas the calls received from network i no longer generate any profit, since the price of off-net calls now reflects their actual cost. On the whole, *losing* the user to the rival network yields a net gain of $L_T m\hat{q}/2$, which increases the equilibrium fee by the same amount. Compared to the situation without on-net pricing, the net gain of attracting this user is however reduced by $m\hat{q}/2$, which tends to raise the equilibrium fee.

This first effect is mitigated by a tariff-mediated network effect. As in Laffont, Rey and Tirole (1998b), on-net pricing increases competition between networks: since attracting an additional user raises the value of a network by $v - \hat{v}$, networks compete more fiercely for subscribers; and the higher the difference between the utilities generated by on-net and off-net calls, the more intense the competition and the lower the fixed fee.

Consider now the subscription fees for light users. Setting heavy users' prices to their equilibrium values, network i's profit becomes:

$$\Pi_i = H_i \hat{F}^H + H_j (H_i + L_i) m\hat{q} + L_i F_i^L, \tag{20}$$

where F_1^L and F_2^L affect the market shares in both segments: $L_i = D\left(F_i^L, F_j^L\right)$ and

$$H_i = \frac{1}{2} + \frac{\sigma(L_i - L_j)(v - \hat{v})}{1 - 2\sigma(v - \hat{v})}.$$

Differentiating (20) with respect to F_i^L then yields, at a symmetric equilibrium:

$$\left.\frac{\partial \Pi_i}{\partial F_i^L}\right|_{F_1^L=F_2^L=\hat{F}^L} = \frac{\sigma\left(v-\hat{v}\right)\left(D_1-D_2\right)}{1-2\sigma\left(v-\hat{v}\right)}\left(\hat{F}^H - L_T\frac{m\hat{q}}{2}\right)+D_1\left(\frac{m\hat{q}}{2}+\hat{F}^L\right)+\frac{L_T}{2},$$

which, using $D_2 = -\gamma D_1$ and (19), can be rewritten as:

$$\hat{F}^L + \frac{\left(1+\gamma(\hat{F}^L)\right)\left(v-\hat{v}\right)+m\hat{q}}{2} = \mu(\hat{F}^L). \tag{21}$$

We now study the impact of on-net pricing on the prices offered to light users:

Proposition 5 *For termination fees close to the marginal cost, on-net pricing leads to lower (resp., higher) subscription fees for both heavy and light users when m is positive (resp., negative).*

Proof. See Appendix A.2. ■

As mentioned above, on-net pricing generates two conflicting effects on the fees charged to heavy users. On the one hand, the opportunity cost of losing a heavy user is reduced, since there is less cross-subsidy between different types of calls; this first effect tends to decrease competition. On the other hand, tariff-mediated network effects tend to increase competition. The proposition shows that, for small termination mark-ups, the second effect dominates, so that on-net pricing therefore benefits heavy users.

On-net pricing also induces a decrease in the price for light users when the termination fee lies above cost. This is again partly driven by network effects: adding an additional light user renders a network comparatively more attractive for heavy users, which encourages networks to compete more fiercely for light users. In addition, while on-net calls to light users no longer generate any net revenue, off-net incoming calls still generate a termination profit equal to $m\hat{q}/2$, which contributes again to reduce prices.

Using the same decomposition as before, total welfare now becomes:

$$\hat{W} = \left[\left(1+\hat{L}_T\right)\frac{v+\hat{v}+m\hat{q}}{2} - \frac{1}{8\sigma}\right] + \left[\hat{S}^L + \hat{L}_T\hat{F}^L\right], \tag{22}$$

where $\hat{S}^L = S^L(\hat{F}^L, \hat{F}^L)$. It is then again socially desirable to raise the termination fee above cost:

Proposition 6 *With on-net pricing, the welfare-maximizing termination mark-up is positive.*

Proof. Using $\frac{\partial \hat{S}^L}{\partial \hat{F}^L} = \frac{\partial S^L}{\partial F_1^L} + \frac{\partial S^L}{\partial F_2^L}\Big|_{F_1^L=F_2^L=\hat{F}^L} = -\hat{L}_T$, we have

$$\frac{\partial \hat{W}}{\partial m} = \left(1+\hat{L}_T\right)\frac{mq'\left(\hat{p}\right)}{2} + \frac{\partial \hat{L}_T}{\partial m}\left(\frac{v+\hat{v}+m\hat{q}}{2}+\hat{F}^L\right),$$

where, from (21):

$$\frac{v+\hat{v}+m\hat{q}}{2}+\hat{F}^L = \mu(\hat{F}^L)+\hat{v}-\gamma\left(\hat{F}^L\right)\frac{v-\hat{v}}{2}.$$

Since $\hat{v} \geq v$ for $m \leq 0$, this implies $\frac{\partial \hat{W}}{\partial m} > 0$ for $m \leq 0$. ■

Unsurprisingly, the result on profit is more ambiguous. Indeed, while the competition weakening effect described in the case without on-net pricing is still present, it is now lower than before and moreover mitigated by the impact of tariff-mediated network effects. Total profit can be written as:

$$2\hat{\Pi} = \left(\frac{1}{2\sigma}+\frac{m\hat{q}}{2}+\hat{v}-v\right)+\hat{L}_T\left(\hat{F}^L+m\hat{q}\right). \tag{23}$$

The term within parentheses is maximal for a negative value of m, as shown by Gans and King (2001) and Dessein (2003). The last term is more complex. Still, we can establish:

Proposition 7 *If at $m=0$,*

$$L_T^*\left(1-\frac{\gamma^*\left(1+\frac{\gamma^*}{2}\right)}{1-\mu'\left(F^{L*}\right)}\right) > \frac{1}{2}, \tag{24}$$

then total profit increases with m for m close to zero.

Proof. See Appendix A.3. ■

The right-hand side of condition (24) is proportional (by a factor $1/q\left(c\right)$) to the reduction in profit due to tariff mediated network effects. In the absence of light users, there is no other effect: the condition is then violated and the operators favour a termination fee below cost. The left-hand side captures the additional effects arising from the presence light users: increasing the termination fee above cost *i*) has a positive impact on the subscription fee $\hat{F}^H$ through the competition-softening effect (see equation (19)); *ii*) boosts the termination profits generated by light users; *iii*) but intensifies the competition for light users. As long as γ^* is not too large, the increase in termination profits is not fully competed away on the retail market, so that the overall additional effect is positive. It then dominates the negative impact of tariff-mediated network effects when there are enough light users.

Note that, if $\mu' > -1/2$, the left-hand side becomes negative when the participation of light users is quite inelastic (γ^* large). In this case, an increase in termination profits triggers a strong intensification of the retail competition for light users, which outweighs the positive effects. This may occur with on-net pricing and not without because, as pointed in proposition 5, the reduction in light users' subscription fee is stronger with on-net pricing than without.

The result does not extend easily to larger departures from cost-based termination fees, due to the impact on the volume of traffic. We can however extend it when the latter is not too sensitive to usage prices. To see this, suppose that:

- the individual usage demand is inelastic:

$$q(p) = \begin{cases} \bar{q} \text{ if } p \leq \bar{p}, \\ 0 \text{ if } p > \bar{p}, \end{cases}$$

  where $\bar{p} > c$ and $\bar{q} > 0$; and

- light users subscription demand is linear:

$$D\left(F_1^L, F_2^L\right) = \alpha - \beta F_1^L + \gamma\beta F_2^L,$$

where $\beta > 0$ and $0 \leq \gamma < 1$.

We show in Appendix A.4 that the operators' total profit is then equal to:

$$2\hat{\Pi} = \frac{1}{2\sigma} - \frac{m\bar{q}}{2} + \frac{2}{(2-\gamma)^2\beta}\left(\alpha + \beta(1-\gamma)\left(1+\frac{\gamma}{2}\right)m\bar{q}\right)\left(\alpha + \beta\left(1-\frac{3\gamma}{2}\right)m\bar{q}\right).$$

This profit is concave for $\gamma > 2/3$, in which case condition (24), which (using $L_T^* = 2\alpha/(2-\gamma)$ for $m = 0$ and $\mu' = -(1-\gamma)$) boils down here to

$$\frac{2\alpha}{2-\gamma}\frac{4-4\gamma-\gamma^2}{2-\gamma} > 1, \tag{25}$$

is necessary and sufficient for the operators to favour a positive termination mark-up.

For $\gamma < 2/3$, the profit is convex and it is then optimal for the operators to favour either bill-and-keep ($m = -c_T$, where $c_T < c$ is the termination cost) or the maximal sustainable termination mark-up, $m = \bar{p} - c$ (which coincides here with m^M); we show in the Appendix that, if $\bar{p} - c > c_T$, then condition (25) ensures that the latter option dominates, which leads to:

Corollary 1 *If individual usage is inelastic, with $\bar{p} - c > c_T$, and light users' demand is linear, then the welfare-maximizing termination fee exceeds the profit-maximizing one, which is above cost under condition* (25).

Proof. See Appendix A.4. ■

Proposition 6 already shows that, when light users' participation is elastic, insisting on cost-based termination fees is no more socially desirable with on-net pricing than without. The above corollary shows moreover that, when on-net pricing is prevalent, any price cap regulation is socially detrimental (if it is binding) when usage is inelastic and light users' participation demand is linear.

Condition (25) moreover ensures that the profit-maximizing termination fee is also above cost. This condition is satisfied as long as light users' participation is elastic (γ not too large),[21] and their population is large enough.[22] When for example the operators have a local monopoly over their own clientele of light users ($\gamma = 0$), the condition is satisfied if the equilibrium proportion of light users at $m = 0$ exceeds one third of the total customer base.

Finally, for $\gamma < 2/3$, an additional mild condition (namely, $\bar{p} - c > c_T$, which is likely to be satisfied as the marginal cost of termination is very small in practice) then ensures that the operators favour the maximal termination fee ($\bar{p} - c$, which coincides here with m^M); we show in appendix that this maximal termination fee coincides in that case with the socially optimal one.

4 Endogenous utility from reception

We now extend the analysis to the case where light users' utility varies with the volume of calls received. We first derive analytically the equilibrium fees for light and heavy users. We then use numerical simulations to show that the main insights derived in the benchmark case remain valid in this extended setting.

[21] The left-hand side of (25) is positive as long as $\gamma < 2\sqrt{2} - 2 \simeq 0.83$.

[22] Since scaling the demand D by a multiplicative factor does not affect the equilibrium prices, the condition is indeed easier to satisfy when there are many light users.

4.1 Market equilibrium

We now assume that receiving a volume of calls q from a particular subscriber gives light users a utility $u_L(q)$. Hence, in the absence of on-net pricing, the total reception utility of a light user is $V^L = H_i u_L(q(p_i)) + H_j u_L(q(p_j))$; the demand from light users for network i is then given by $L_i = D(F_i^L, F_j^L, V^L)$,[23] with $D_3 \equiv \partial D/\partial V^L \geq 0$. Heavy users are modeled as before; however, the usage price p_i now influences the surplus w_i offered by network i not only through the direct effect on $v(p_i)$, but also indirectly through the the impact of the number of calls on the utility and thus the participation of light users.

As in the benchmark model, operator i's profit is equal to

$$\begin{aligned}\Pi_i &= H_i \left[(1 + L_T)(p_i - c)\, q(p_i) - (H_j + L_j)\, mq(p_i) + F_i^H\right] \\ &\quad + (H_i + L_i)\, H_j mq(p_j) + L_i F_i^L.\end{aligned}$$

We will again first optimize with respect to p_i, adjusting F_i^H so as to keep the market shares H_1 and H_2 constant, but accounting now for the impact of calls on light users' participation:

$$\begin{aligned}\left.\frac{\partial \Pi_i}{\partial p_i}\right|_{H_i, F_i^L} &= H_i \Bigg[\left(1 + L_T + \frac{\partial L_T}{\partial p_i}(p_i - c) - m\frac{\partial L_j}{\partial p_i}\right) q(p_i) \\ &\quad + \left((1 + L_T)(p_i - c) - (H_j + L_j)m\right) q'(p_i) + \frac{\partial F_i^H}{\partial p_i}\Bigg] \\ &\quad + H_j \frac{\partial L_i}{\partial p_i} mq(p_j) + \frac{\partial L_i}{\partial p_i} F_i^L,\end{aligned}$$

where $\frac{\partial L_j}{\partial p_i} = \frac{\partial L_i}{\partial p_i} = \frac{1}{2}\frac{\partial L_T}{\partial p_i}$ (since L_1 and L_2 depend on p_i only through V^L). Furthermore, since

$$H_i = \frac{1}{2} + \sigma(w_i - w_j) = \frac{1}{2} + \sigma(1 + L_T)\left[v(p_i) - v(p_j)\right] - \sigma\left(F_i^H - F_j^H\right),$$

keeping market shares constant now requires

$$\frac{\partial F_i^H}{\partial p_i} = -(1 + L_T) q(p_i) + \frac{\partial L_T}{\partial p_i}[v(p_i) - v(p_j)].$$

Therefore, the above first-order condition, evaluated at a symmetric equilibrium ($p_1 = p_2 = p$, $F_1^L = F_2^L = F^{L^*}$ and $H_1 = H_2 = 1/2$), yields

$$p^* = c + \frac{m}{2} - \frac{\partial L_T}{\partial p_i}\frac{(p^* - c)q(p^*) + F^{L^*}}{(1 + L_T^*)q'(p^*)},$$

[23] Under uniform prices, light users' utility from reception does not depend on the network to which they subscribe: $V_1^L = V_2^L = V^L$; we thus add only one argument in the demand function.

or, using $\frac{\partial L_T}{\partial p_i} = D_3 u_L' q'(p^*)$,

$$p^* = c + \frac{m}{2} - D_3 u_L' \frac{(p^* - c)q\left(p^*\right) + F^{L^*}}{1 + L_T}. \tag{26}$$

Compared to the benchmark model, the last term is new and accounts for the negative impact of usage prices on calls and thus on light users' participation; in particular, the numerator is the profit derived from light users, which includes the fixed fee F^{L^*} they pay as well as the retail profit they generate through the on-net calls they receive.[24]

For $p_1 = p_2 = p^*$, $H_i = 1/2 + \sigma\left(F_j^H - F_i^H\right)$; differentiating Π_i with respect to F_i^H then yields, at a symmetric equilibrium:

$$\begin{aligned}
\frac{\partial \Pi_i}{\partial F_i^H} &= -\sigma\left[(1+L_T)(p^*-c)q(p^*) - (\frac{1}{2} + L_j)mq(p^*) + F_i^H\right] \\
&\quad + \frac{1}{2}\left[-\sigma mq(p^*) + 1\right] + \sigma[\frac{1}{2} + L_i]mq(p^*) - \frac{1}{2}\sigma mq(p^*).
\end{aligned}$$

The equilibrium fee is therefore such that

$$F^{H^*} = \frac{1}{2\sigma} + (L_T^* - 1)\frac{mq^*}{2} - (1 + L_T^*)(p^* - c - \frac{m}{2})q^*. \tag{27}$$

This condition is similar to the one obtained in the benchmark model, except for the last term: the retail price now departs from the average marginal cost, which alters the gain from attracting heavy users; if for example the price is below average marginal cost, then each call emitted generates a loss and, as a result, the equilibrium fee increases.

We now turn to light users; differentiating Π_i with respect to F_i^L leads to

$$\begin{aligned}
\frac{\partial \Pi_i}{\partial F_i^L} &= H_i\left[\frac{\partial L_T}{\partial F_i^L}(p^*-c)q^* - \frac{\partial L_j}{\partial F_i^L}mq^*\right] \\
&\quad + H_j\frac{\partial L_i}{\partial F_i^L}mq^* + \frac{\partial L_i}{\partial F_i^L}F_i^L + L_i.
\end{aligned}$$

As $\frac{\partial L_i}{\partial F_i^L} = D_1$ and $\frac{\partial L_j}{\partial F_i^L} = D_2 = -\gamma D_1$, the symmetric equilibrium fee for light users, $F_1^L = F_2^L = F^L$, is such that

$$F^{L^*} = \mu^* - \frac{(1-\gamma^*)(p^*-c)q^*}{2} - \frac{(1+\gamma^*)mq^*}{2}, \tag{28}$$

[24] In a symmetric equilibrium, the termination profits associated with the off-net calls received by light users cancel each other.

where $\mu^* = -D/D_1$ and $\gamma^* = -D_2/D_1$ are evaluated at the equilibrium. Finally, plugging (28) into (26) leads to

$$p^* = c + \frac{m}{2} + \frac{1}{2}\frac{D_3}{D_1}u_L' \frac{L_T^*}{1+L_T^*}\left[1 + \frac{1+\gamma^*}{2\mu^*}(p^* - c - m)q^*\right]. \qquad (29)$$

This condition allows us to show that, at least for m small, the equilibrium usage price lies below cost:

Proposition 8 *For termination fees close to the marginal cost, when light users' demand depends on the volume of calls received the equilibrium usage prices lie below cost (i.e., $p^* < c$).*

Proof. For $m = 0$, condition (29) boils down to:

$$p^* = c - \lambda\left[1 + \frac{1+\gamma^*}{2\mu^*}(p^* - c)q^*\right],$$

where $\lambda \equiv -\frac{1}{2}\frac{D_3}{D_1}u_L' \frac{L_T^*}{1+L_T^*} > 0$ (as $D_1 < 0$ and $D_3, u_L' > 0$), and thus

$$p^* - c = \frac{-\lambda}{1 + \frac{1+\gamma^*}{2\mu^*}\lambda q^*} < 0.$$

By continuity, $p^* < c$ for m close to 0. ■

4.2 Private and social optimal values of m

We now show that the main insights derived in the benchmark model remain valid. As it is not possible to obtain analytical results, we illustrate this through numerical simulations. To do so, we adopt the following linear functional forms:

1. $q(p) = b\,(a - p)$;
2. $D\left(F_1^L, F_2^L, V^L\right) = \alpha - \beta F_1^L + \gamma\beta F_2^L + (1-\gamma)\,\beta V^L$, and $u_L\,(q) = u_L q$.[25]

We calibrate the model to be consistent with french data from second quarter of 2005, interpreting heavy users as post-paid and light users as pre-paid subscribers. The number of post-paid and pre-paid contracts were respectively 28,55 millions and 16,84 millions, which, normalizing the population of heavy users to 1, gives $L_T^* = 16,84/28,55 = 0.59$. The monthly

[25] This corresponds to $D_i = \alpha - \beta(F_i^L - u_L q) + \gamma\beta(F_j^L - u_L q)$, where $F_i^L - u_L q$ can be interpreted as an hedonic price.

volume of mobile-to-mobile calls were 5094 millions of minutes (mn), which yields a volume of calls per user $(1 + L_T^*)q^* = 178$ mn/month and thus on average $q^* = 112$ mn/month received by a subscriber. Following de Bilj and Peitz (2004) and ARCEP (2008), we assume a cost $c = 2$ euro cents (cts) equally divided between origination and termination.

To determine the parameters of the demand for calls $q(.)$, we follow the methodology of Harbord and Hoernig (2011). The termination fee for Orange and SFR were 12.5 cts, which corresponds in our model to a price $p^* = 2 + (12.5 - 1)/2 = 7,75$ cts. Using $q(p^*) = q^*$ and assuming a price-elasticity of usage (equal to $-p^* q'(p^*)/q(p^*) = bp^*$) of 0.5, we can determine the coefficients a and b, which yields (expressing the price in euros) $q(p) = 722(0.23 - p)$.

For the parameters of the demand $D(.)$, we first note that, for $u_L = 0$, the elasticity of the aggregate demand of light users to the price, evaluated at symmetric prices, is $\varepsilon = 2(1-\gamma)\beta P^L / L_T^*$, where P^L is the total price.[26] We thus adjust the parameter β as a function of the replacement ratio γ and of the elasticity ε, evaluated at a total price $P^L = 10$ euros and $L_T^* = 0.59$:

$$\beta = \frac{0.59}{20} \frac{\varepsilon}{1-\gamma}.$$

For instance, assuming an elasticity of 1 and a replacement ratio $\gamma = 0.5$ yields a value of $\beta = 0.059$. Finally, we set the parameter α so as to maintain $L_T^* = 0.59$ in the benchmark case without reception utility; that is, using (7) and the linear specification of the demand:

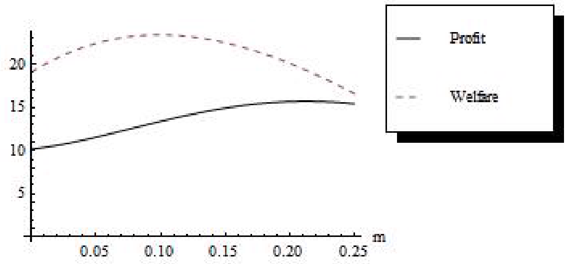
$$\alpha = (2-\gamma)\frac{L_T^*}{2} - (1-\gamma)\beta \frac{3+\gamma}{4} m q^*,$$

where L_T^*, m and q^* are set at the calibrated values.

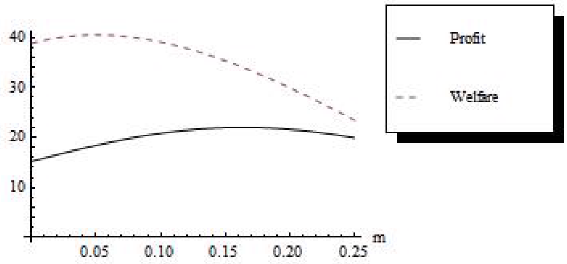
In the simulation we use several values of the elasticity of the aggregate demand of light users, ε, and of the replacement ratio γ. As for the range of u_L, note first that it is the monetary equivalent of 1 mn of call. For the calibrated values, the surplus that heavy users derive from calling can be written as $v_H = (1 + L_T^*)(q^*)^2/2b = 13.80$ euros, whereas light users' surplus from reception is $112u_L$. To maintain things comparable while allowing for a substantial utility from reception, we vary u_L from 0 to 0.15 euro/mn; note that for $u_L = 0.15$ euro/mn, light users's utility from reception is 16.8 euros, and is thus larger than the utility that heavy users derive from calling.

[26] As we normalized fixed costs to zero, in our model F^L represents in fact the margin and not the price.

We first present the impact of the termination mark-up on the equilibrium profit and welfare, both in the benchmark case where light users' reception utility is fixed (Figure 1, where $u_L = 0$) and in the case where it varies with the volume of calls received (Figure 2, where $u_L = 0.1$).



Profit and welfare for $u_L = 0$.



Profit and welfare for $u_L = 10$ cts/mn.

These Figures show that the main insights of the benchmark model carry over to the extended setting: the equilibrium profit and welfare are both maximal for a positive termination mark-up ($m > 0$), although the optimal values are smaller when light users' utility is sensitive to the volume of calls received. When $u_L = 0$, increasing m makes light users very valuable, inducing a reduction in their subscription fees and an increase in their welfare. But increasing m has also an impact on the usage price p, and therefore on the number of calls received, which harms light users when u_L becomes positive. This is why the privately and socially optimal values of m are smaller for $u_L = 10$ cts.

In the next table, we show how the privately and socially optimal levels of the termination mark-up m (in cents) varies with u_L for different values of ε and γ.

Adding the termination cost of 1 ct/mn to the mark-up, the profit-maximizing termination fee ranges from the monopoly level 22 cts/mn to 14.3 cts/mn and thus remains quite large even for a significant marginal utility from reception. The welfare-maximizing termination fee is consistently

	$\varepsilon = 1$ and $\gamma = 0.5$		$\varepsilon = 0.5$ and $\gamma = 0.5$		$\varepsilon = 1$ and $\gamma = 0.7$	
	Profit max m	Welfare max m	Profit max m	Welfare max m	Profit max m	Welfare max m
$u_L = 0$	21.0	9.9	21.0	8.7	21.0	9.6
$u_L = 5\ cts$	18.8	7.6	19.0	6.5	18.6	7.2
$u_L = 10\ cts$	16.4	5.0	16.8	4.1	15.9	4.4
$u_L = 15\ cts$	13.9	2.1	14.6	1.5	13.3	1.3

below the level of 12.5 cts/mn that was effective in 2005 in France, although it remains significantly above cost.

The table confirms that it is beneficial (both for the operators and society) to decrease m when light users react to the number of calls received. The table also shows that the socially optimal value of m decreases when the aggregate light users demand becomes less elastic. As mentioned above, raising the termination fee has both a positive impact on participation and a depressing effect on usage. When the participation of light users is only slightly elastic, the negative impact of a price increase on usage tends to prevail and the socially optimal termination fee is smaller. This reasoning does not hold for the profit-maximizing level, because firms care about light users' participation rather than about their utility. Thus, a lower elasticity of light users' participation may result into a higher profit-maximizing termination mark-up, due to a lower effect of usage on participation.

For a given value of the elasticity ε, the replacement ratio γ can be interpreted as a measure of competition. As more competition implies that a larger share of termination revenue is transferred to light users, firms have less incentive to raise the termination fee when γ increases. Moreover, as the difference between the private and the socially optimal light users subscription prices is smaller, the negative impact of a price increase on usage prevails and the socially optimal termination fee also decreases with γ.

5 Conclusion

This paper revisits the theoretical analysis of termination fees in communication networks. We show that the insights of the existing literature, which suggest profit-maximizing fees at or below cost, rely critically on the related assumptions of fixed participation and full pass-through, as well as on the homogeneity of calling patterns. When instead the elasticity of subscription and the intensity of usage are negatively correlated across users, as empirical observation suggests, then the profit-maximizing reciprocal termination fee is always above cost in the absence of on-net pricing, and can still be so with on-net pricing; in addition, the welfare-maximizing termination fee is also above cost, although it is below the former one in the absence of termination-based price discrimination.[27]

Our results thus imply that, while some cap on termination fees is desirable, the regulated cap should be above termination costs. The optimal cap depends on factors such as the proportion of light users and their demand elasticity. Thus local market conditions matter, suggesting for instance that, in the context of the European regulation of mobile telephony markets, some discretion should be left to national regulators.

By stressing the critical role of demand heterogeneity for regulatory debates, the analysis also points to the need for better empirical facts on the composition of the demand and on the participation elasticities of the various categories of users of telecommunication services.

[27] In our 2010 working paper, we also study the robustness of these insights when taking fixed-to-mobile termination into consideration, or when accounting for a small demand for calls by light users.

References

[1] ARCEP (2008), *Les référentiels de coût des opérateurs mobiles en 2008,* Paris.

[2] Armstrong, M. (1998), "Network Interconnection in Telecommunications", *Economic Journal*, 108:545-564.

[3] Armstrong, M. (2002), "The Theory of Access Pricing and Interconnection", in *Handbook of Telecommunications Economics*, eds. M. Cave, S. Majumdar and I. Vogelsang, North-Holland, 295-384.

[4] Armstrong, M., and J. Wright (2009), "Mobile Call Termination", *Economic Journal*, 119(538):270-307.

[5] Atkinson, J., and C. Barnekov (2000), "A Competitively Neutral Approach to Network Interconnection", OPP working Paper n°34, Federal Communications Commission.

[6] Berger, U. (2004), "Access Charges in the Presence of Call Externalities", *Contribution to Economic Analysis and Policy*, 3(1), Article 21.

[7] Berger, U. (2005), "Bill-and-Keep vs. cost-based access pricing revisited", *Economics Letter*, 86(1):107-112.

[8] Bolt, W., and W. Tieman (2006), "Social Welfare and Cost Recovery in Two-Sided Markets", *Review of Network Economics*, 5(1):7.

[9] Calzada, J. and T. Valletti (2008),"Network Competition and Entry Deterrence", *Economic Journal*, 118:1223-1244.

[10] de Bilj P. and M. Peitz (2004), "Dynamic regulation and entry in telecommunications markets: a policy framework", *Information Economics and Policy*", 16(3):411–437.

[11] DeGraba, P. (2004), "Efficient Intercarrier Compensation for Competing Networks when Customers Share the Value of a Call," *Journal of Economics & Management Strategy*, 12:207-230.

[12] Dessein, W. (2003), "Network Competition in Nonlinear Pricing", *Rand Journal of Economics*, 34(4):593-611.

[13] Dessein, W. (2004), "Network Competition with Heterogeneous Customers and Calling Patterns", *Information Economics and Policy*, 16:323-345.

[14] Gans, J., and S. King (2001), "Using 'Bill and Keep' Interconnect Agreements to Soften Network Competition", *Economics Letters*, 71(3):413-420.

[15] Genakos, C., and T. Valletti (2011a), "Testing the "Waterbed" Effect in Mobile Telephony", *Journal of the European Economic Association*, 9(6):1114-1142.

[16] Genakos, C., and T. Valletti (2011b), "Seesaw in the air: Interconnection regulation and the structure of mobile tariffs", *Information Economics and Policy*, 23:159-170.

[17] Hahn, J.-H. (2004), "Network competition and interconnection with heterogeneous subscribers", *International Journal of Industrial Organization*, 22(5):611-631.

[18] Harbord, D., and S. Hoernig (2011), "Welfare Analysis of Regulating Mobile Termination Rates in the UK with an application to the Orange/T-Mobile Merger", mimeo.

[19] Hermalin, B., and M. Katz (2010), "Customer or Complementor? Intercarrier Compensation with Two-Sided Benefits ", *Journal of Economics and Management Strategy*, 20(2):379-408.

[20] Hoernig, S., R. Inderst and T. Valletti (2011) "Calling Circles: Network Competition with Non-Uniform Calling Patterns", CEIS Working Paper 206, Tor Vergata University.

[21] Hurkens, S., and A. Lopez (2010) "Mobile Termination, Network Externality, and Consumers Expectations, IESE Research Papers D/850, IESE Business School. "

[22] Jullien, B., Rey, P. and W. Sand-Zantman, "Mobile Call Termination Revisited", IDEI Working Paper, n°551, 2010.

[23] Laffont, J.-J., P. Rey and J. Tirole (1998a), "Network Competition: I. Overview and Nondiscriminatory pricing", *Rand Journal of Economics*, 29(1):1-37.

[24] Laffont, J.-J., P. Rey and J. Tirole (1998b), "Network Competition: II. Price Discrimination", *Rand Journal of Economics*, 29(1):37-56.

[25] Lopez, A., and P. Rey (2009), "Foreclosing Competition through Access Charges and Price Discrimination", IDEI Working Paper n°570.

[26] Poletti, S., and J. Wright (2004), "Network interconnection with participation constraints", *Information Economics and Policy*, 16(3):347-373.

[27] Schiff, A. (2008), "The Waterbed Effect and Price Regulation", *Review of Network Economics*, 7(3):392-414.

[28] Tangeras, T. (2012), "Network Competition with Income Effects ", IFN working paper 903.

[29] Weyl, G., and M. Fabinger (2009), "Pass-Through as an Economic Tool", mimeo, Harvard University.

A Appendix to section 3

A.1 Proof of proposition 3

Differentiating (12) at $m = 0$ yields (using $\frac{dp^*}{dm} = \frac{1}{2}$ and $\frac{dt}{dm}\big|_{m=0} = q(c) = -v'(c)$):

$$\begin{aligned}\left.\frac{dS^{H^*}}{dm}\right|_{m=0} &= (1+L_T^*)\frac{v'(c)}{2} + D_T'(F^{L^*})\left.\frac{dF^{L^*}}{dm}\right|_{m=0} v(c) - (1-L_T^*)\frac{v'(c)}{2} \\ &= L_T^* v(c)\left(\frac{D_T'\left(F^{L^*}\right)}{D_T\left(F^{L^*}\right)}\left.\frac{dF^{L^*}}{dm}\right|_{m=0} + \frac{v'(c)}{v(c)}\right),\end{aligned}$$

from which the expression (14) follows since $\frac{D_T'\left(F^{L^*}\right)}{D_T\left(F^{L^*}\right)}$, $\left.\frac{dF^{L^*}}{dm}\right|_{m=0} = \left.\frac{dF^{L^*}}{dt}\right|_{t=0} q(c)$ and $\frac{v'(c)}{v(c)}$ are all negative.

A.2 Proof of proposition 5

By construction, for $m = 0$ there is no difference in prices: $p^* = \hat{p} = c$, and thus $\hat{F}^H = F^{H^*}$ and $\hat{F}^L = F^{L^*}$. Furthermore, from (5) and (19) we have:

$$F^{H^*} - \hat{F}^H = \frac{L_T^* - 1}{2} m q^* - \frac{\hat{L}_T}{2} m \hat{q} + v - \hat{v},$$

where $\hat{L}_T$ denotes the equilibrium number of users under price discrimination, and thus:

$$\left.\frac{\partial(F^{H^*} - \hat{F}^H)}{\partial m}\right|_{m=0} = \frac{L_T^* - 1}{2} q(c) - \frac{L_T^*}{2} q(c) + q(c) = \frac{q(c)}{2} > 0.$$

Consider now the fees charged to light users. From (8) we have (using $\frac{dt}{dm}\big|_{m=0} = q(c)$) $\left.\frac{\partial F^{L^*}}{\partial m}\right|_{m=0} = -\left.\frac{3+\gamma\left(F^{L^*}\right)}{1-\mu'\left(F^{L^*}\right)}\right|_{m=0} \frac{q(c)}{4}$, whereas differentiating (21) yields:

$$\left.\frac{\partial \hat{F}^L}{\partial m}\right|_{m=0} = -\left.\frac{1 + \frac{\gamma(F^{L^*})}{2}}{1 - \mu'(F^{L^*})}\right|_{m=0} q(c). \tag{30}$$

Therefore,

$$\left.\frac{\partial\left(F^{L^*} - \hat{F}^L\right)}{\partial m}\right|_{m=0} = \left.\frac{1 + \gamma(F^{L^*})}{1 - \mu'(F^{L^*})}\right|_{m=0} \frac{q(c)}{4},$$

which is positive under Assumption B.

A.3 Proof of proposition 7

Using $L_T = 2D\left(\hat{F}^L, \hat{F}^L\right)$, we have:

$$\left.\frac{\partial\left(2\hat{\Pi}\right)}{\partial m}\right|_{m=0} = -\frac{q}{2} + \left(L_T + 2\left(1-\gamma\right) D_1 F^L\right) \frac{\partial \hat{F}^L}{\partial m} + L_T q,$$

where $q = q\left(c\right)$ and L_T and $\gamma = \gamma(F^L)$ are evaluated at $F^L = F^{L^*}\left(0\right) = \hat{F}^L\left(0\right)$. F^L is characterized by (21):

$$F^L = \mu\left(F^L\right) = -\frac{D}{D_1} = -\frac{L_T}{2D_1}.$$

This in turn yields:

$$L_T + 2\left(1-\gamma\right) D_1 F^L = \gamma L_T.$$

From (30), we have:

$$\left.\frac{\partial \hat{F}^L}{\partial m}\right|_{m=0} = -\frac{\left(1+\frac{\gamma}{2}\right) q}{1-\mu'(F^L)},$$

and thus:

$$\frac{1}{q}\left.\frac{\partial\left(2\hat{\Pi}\right)}{\partial m}\right|_{m=0} = -\frac{1}{2} + L_T\left(1 - \frac{\gamma\left(1+\frac{\gamma}{2}\right)}{1-\mu'(F^L)}\right).$$

A.4 On-net pricing with inelastic usage for heavy users and linear demand for light users

From (23), and using $q\left(c\right) = q\left(c+m\right) = \bar{q}$ and $v - \hat{v} = m\bar{q}$, we have:

$$\begin{aligned} 2\hat{\Pi} &= \frac{1}{2\sigma} + \frac{m\bar{q}}{2} - m\bar{q} + \hat{L}_T\left(\hat{F}^L + m\bar{q}\right) \\ &= \frac{1}{2\sigma} - \frac{m\bar{q}}{2} + \hat{L}_T\left(\hat{F}^L + m\bar{q}\right), \end{aligned}$$

where

$$\hat{L}_T = 2\left(\alpha - \left(1-\gamma\right)\beta\hat{F}^L\right)$$

and $\hat{F}^L$ is characterized by (21), which yields:

$$\hat{F}^L = \frac{\frac{\alpha}{\beta} - \left(1+\frac{\gamma}{2}\right) m\bar{q}}{2-\gamma}. \tag{31}$$

This leads to:

$$
\begin{aligned}
2\hat{\Pi} &= \frac{1}{2\sigma} - \frac{m\bar{q}}{2} + \frac{2}{\beta}\left(\frac{\alpha + \beta(1-\gamma)\left(1+\frac{\gamma}{2}\right)m\bar{q}}{2-\gamma}\right)\left(\frac{\alpha+\beta\left(1-\frac{3\gamma}{2}\right)m\bar{q}}{2-\gamma}\right) \\
&= \frac{1}{2\sigma} - \frac{m\bar{q}}{2} + \frac{2}{(2-\gamma)^2\beta}\left(\alpha + \beta(1-\gamma)\left(1+\frac{\gamma}{2}\right)m\bar{q}\right)\left(\alpha+\beta\left(1-\frac{3\gamma}{2}\right)m\bar{q}\right).
\end{aligned}
$$

We thus have:

$$
\frac{d^2\left(2\hat{\Pi}\right)}{dm^2} = \left(\frac{2}{3}-\gamma\right)\frac{(1-\gamma)(2+\gamma)}{(2-\gamma)^2}3\beta\bar{q}^2,
$$

so that this profit is concave if $\gamma > 2/3$, in which case (24) ensures that the operators favor a positive termination mark-up.

For $\gamma < 2/3$, the profit is convex and is thus maximal at $m = \bar{p} - c$ or $m = -c_T$. The difference between the value at $\bar{p} - c$ and at $-c_T$ is equal to:

$$
\begin{aligned}
&- \frac{(\bar{p}-c+c_T)\bar{q}}{2} \\
&+2\frac{\alpha\left(\frac{4-4\gamma-\gamma^2}{2}\right)(\bar{p}-c+c_T)\bar{q} + \beta(1-\gamma)\left(\frac{2+\gamma}{2}\right)\left(\frac{2-3\gamma}{2}\right)\left[(\bar{p}-c)^2 - c_T^2\right]\bar{q}^2}{(2-\gamma)^2} \\
&= (\bar{p}-c+c_T)\frac{\bar{q}}{2}\left[\frac{2\alpha}{2-\gamma}\frac{4-4\gamma-\gamma^2}{2-\gamma} - 1 + 2\beta\frac{(1-\gamma)(2+\gamma)}{(2-\gamma)^2}\left(\frac{2-3\gamma}{2}\right)(\bar{p}-c-c_T)\bar{q}\right],
\end{aligned}
$$

which is positive for $\gamma < 2/3$ if $\bar{p} - c - c_T > 0$ and $\frac{2\alpha}{2-\gamma}\frac{4-4\gamma-\gamma^2}{2-\gamma} > 1$.

Using (22), we have:

$$
\hat{W} = \left[\left(1+\hat{L}_T\right)v - \frac{1}{8\sigma}\right] + \left[\hat{S}^L + \hat{L}_T\hat{F}^L\right],
$$

where $\frac{\partial \hat{S}^L}{\partial \hat{F}^L} = -\hat{L}_T$, $\hat{L}_T = 2\left(\alpha - (1-\gamma)\beta\hat{F}^L\right)$ and $\hat{F}^L$ is given by (31). Therefore:

$$
\frac{d\hat{W}}{dm} = \left(v + \hat{F}^L\right)\frac{d\hat{L}_T}{dm},
$$

where $v = (\bar{p} - c)\bar{q}$ and

$$
\frac{d\hat{L}_T}{dm} = (1-\gamma)\frac{2+\gamma}{2-\gamma}\beta\bar{q} > 0.
$$

It follows that total welfare is concave in m ($\frac{d^2\hat{W}}{dm^2} = \frac{d\hat{F}^L}{dm}\frac{d\hat{L}_T}{dm} < 0$) and that it is maximal for $m^W = \min\left\{\hat{m}^W, \bar{p} - c\right\}$, where

$$\hat{m}^W \equiv 2\frac{\beta\left(2-\gamma\right)\left(\bar{p}-c\right)\bar{q}+\alpha}{\beta\left(2+\gamma\right)\bar{q}} > 0.$$

$\hat{m}^W$ decreases with γ and, for $\gamma = 2/3$, we have:

$$\hat{m}^W\Big|_{\gamma=\frac{2}{3}} > 2\frac{\beta\left(2-\gamma\right)\left(\bar{p}-c\right)\bar{q}}{\beta\left(2+\gamma\right)\bar{q}}\Bigg|_{\gamma=\frac{2}{3}} = \bar{p} - c.$$

If follows that, for $\gamma \leq 2/3$, total welfare is maximal for $m = \bar{p} - c$.

Profits and welfare are thus both maximal for $m = \bar{p} - c$ when $\gamma \leq 2/3$; we now focus on the case $\gamma > 2/3$. Total profit is then concave in m and, under condition (25) is maximal for $m^\Pi = \min\left\{\hat{m}^\Pi, \bar{p} - c\right\}$, where

$$\hat{m}^\Pi \equiv \frac{2\alpha\left(4-4\gamma-\gamma^2\right)-\left(2-\gamma\right)^2}{\left(1-\gamma\right)\left(3\gamma-2\right)\left(2+\gamma\right)2\beta\bar{q}} > 0.$$

We first show that $\hat{m}^W > \bar{p} - c$ whenever $\hat{m}^\Pi \geq \bar{p} - c$; indeed, the latter condition implies

$$\alpha \geq \frac{\left(1-\gamma\right)\left(3\gamma-2\right)\left(2+\gamma\right)2\beta\bar{q}\left(\bar{p}-c\right)+\left(2-\gamma\right)^2}{2\left(4-4\gamma-\gamma^2\right)} > \frac{\left(1-\gamma\right)\left(3\gamma-2\right)\left(2+\gamma\right)}{4-4\gamma-\gamma^2}\beta\left(\bar{p}-c\right)\bar{q},$$

and thus:

$$\begin{aligned}\hat{m}^W - \left(\bar{p}-c\right) &> \left(2\frac{\beta\left(2-\gamma\right)\bar{q}+\frac{\left(1-\gamma\right)\left(3\gamma-2\right)\left(2+\gamma\right)}{4-4\gamma-\gamma^2}\beta\bar{q}}{\beta\left(2+\gamma\right)\bar{q}}-1\right)\left(\bar{p}-c\right)\\ &= \frac{\gamma\left(3\gamma-2\right)\left(2-\gamma\right)}{\left(2+\gamma\right)\left(4-4\gamma-\gamma^2\right)}\left(\bar{p}-c\right),\end{aligned}$$

which is positive since $\hat{m}^\Pi \geq \bar{p} - c\,(> 0)$ implies $4 - 4\gamma - \gamma^2 > 0$.

We now show that $\hat{m}^W > \hat{m}^\Pi$ whenever $\hat{m}^\Pi < \bar{p} - c$; indeed, the latter condition implies

$$\frac{2\alpha\left(4-4\gamma-\gamma^2\right)-\left(2-\gamma\right)^2}{2\beta\left(2+\gamma\right)\left(1-\gamma\right)\left(3\gamma-2\right)} < \left(\bar{p}-c\right)\bar{q},$$

and thus:

$$\begin{aligned}\hat{m}^W - \hat{m}^\Pi &> 2\frac{\beta\left(2-\gamma\right)\frac{2\alpha\left(4-4\gamma-\gamma^2\right)-\left(2-\gamma\right)^2}{2\beta\left(2+\gamma\right)\left(1-\gamma\right)\left(3\gamma-2\right)}+\alpha}{\beta\left(2+\gamma\right)\bar{q}} - \frac{1}{2\beta\left(2+\gamma\right)\bar{q}}\frac{2\alpha\left(4-4\gamma-\gamma^2\right)-\left(2-\gamma\right)^2}{\left(1-\gamma\right)\left(3\gamma-2\right)}\\ &= \frac{2-\gamma}{\left(1-\gamma\right)\left(2+\gamma\right)^2}\frac{2-\gamma+2\alpha\gamma}{2\beta\bar{q}}\\ &> 0.\end{aligned}$$

Since $\hat{m}^W > 0$, and $\hat{m}^\Pi < \bar{p} - c$ by assumption, it follows that $m^W = \min\left\{\hat{m}^W, \bar{p} - c\right\} > m^\Pi = \max\left\{\hat{m}^\Pi, -c_T\right\}$.

To summarize, we find that:

- If $\gamma \leq \frac{2}{3}$: $m^W = \bar{p} - c \geq m^\Pi$ and, under condition (25), $m^\Pi = \bar{p} - c$;
- If $\gamma > \frac{2}{3}$: $m^W = \min\left\{\hat{m}^W, \bar{p} - c\right\} \geq m^\Pi$ and, under condition (25), $m^\Pi = \min\left\{\hat{m}^\Pi, \bar{p} - c\right\} \geq 0$.

B Details of the numerical simulations

With the linear functional forms, profits and surpluses are given by:

$$\begin{aligned}
2\Pi^* &= (1 + L_T^*)(p^* - c)\, q^* + F^{H^*} + L_T^* F^{L^*}, \\
S^{L^*} &= \frac{\left(\alpha + (1-\gamma)\beta u_L q^* - (1-\gamma)\beta F^{L^*}\right)^2}{(1-\gamma)\beta} = \frac{(L_T^*)^2}{4(1-\gamma)\beta} \\
S^{H^*} &= (1 + L_T^*)\frac{q^{*2}}{2b} - F^{H^*} - \frac{1}{8\sigma}, \\
W^* &= (1 + L_T^*)\left(\frac{q^{*2}}{2b} + (p^* - c)q^*\right) + \frac{(L_T^*)^2}{4(1-\gamma)\beta} + L_T^* F^{L^*} - \frac{1}{8\sigma}.
\end{aligned}$$

Conditions (27) and (28) allow us to express the equilibrium subscription fees as functions of the equilibrium price p^* and participation L_T^*:

$$\begin{aligned}
F^{H^*} &= \frac{1}{2\sigma} - (1 + L_T^*)(p^* - c)\, q^* + L_T^* m q^*, \\
F^{L^*} &= \frac{L_T^*}{2\beta} - \frac{1-\gamma}{2}(p^* - c)\, q^* - \frac{1+\gamma}{2} m q^*,
\end{aligned}$$

where $q^* = b(a - p^*)$. The latter expression moreover allows us to express L_T^* as a function of p^*:

$$\begin{aligned}
L_T^* &= 2\alpha + 2(1-\gamma)\beta u_L q^* - (1-\gamma)\beta 2F^{L^*} \\
&= 2\alpha + 2(1-\gamma)\beta u_L q^* - (1-\gamma)\left[L_T^* - (1-\gamma)\beta(p^* - c)\, q^* - (1+\gamma)\beta m q^*\right] \\
&= \frac{2\alpha + 2(1-\gamma)\beta u_L q^* + (1-\gamma)^2\beta(p^* - c)\, q^* + (1-\gamma^2)\beta m q^*}{2-\gamma} \qquad (32)
\end{aligned}$$

Finally, multiplying (29) by $(1 + L_T^*)$ yields:

$$(1 + L_T^*)\left(p^* - c - \frac{m}{2}\right) = -\frac{1-\gamma}{2} u_L \left[L_T^* + (1+\gamma)\beta(p^* - c - m)q^*\right],$$

which, using (32), provides a cubic equation characterizing the equilibrium price p^*. When u_L is close to zero, this cubic equation has three solutions: one close to $c + m/2$ and the others close to the solutions to $L_T = -1$. As participation should be positive, the only acceptable solution is the one that is close to $c + m/2$, which corresponds to the second root of the cubic equation. It is this solution that is used in the simulations.