RECOVERING FIXED AND COMMON COSTS FOR MOBILE NETWORKS IN EUROPE

JANUSZ A. ORDOVER[*]

AUGUST 4, 2008

I. INTRODUCTION

Mobile termination rates ("MTRs"), the rates charged by mobile networks for terminating calls on their networks, have long been debated by economists, policymakers and industry participants in the European Community ("EC") and elsewhere. Under the commonly used Calling-Party-Pays ("CPP") pricing scheme, MTRs are paid by the network originating the call – that is, by its subscribers – to the network terminating the call. In the past, the apparently high levels of unregulated MTRs for fixed-to-mobile ("F2M") calls, and the resulting high F2M retail prices, raised concerns that Mobile Network Operators ("MNOs") were setting these rates at "monopoly" levels despite the evident competition among the MNOs.[1] This outcome was attributed to the "bottleneck monopoly" held by each network over its customers for calls terminating on that network. Similarly, the regulators were concerned that MNOs were using high MTRs for mobile-to-mobile calls as means of implementing elevated off-net mobile-to-mobile ("M2M") rates. These concerns resulted in the regulation of MTRs in many countries.

---

[1] Mark Armstrong aptly described this situation as that of "competitive bottlenecks." *See, e.g.,* M. Armstrong, "Competition in Two-Sided Markets," *RAND J. Econ.*, vol. 37 (2006).

In the EC countries, national regulators are required to set MTRs based on costs. While many different cost standards are possible, the general approach has been to set the mobile termination rates based on long run average incremental cost ("LRAIC").[2] Costs are calculated by first identifying the efficient costs that are incremental to traffic and then allocating a proportion of these costs to termination based on network routing factors. Termination costs also have included a mark-up over LRAIC of termination to recover costs that are not incremental to traffic, such as common costs relating to non-network business overheads and to the provision of minimum network coverage as well as spectrum costs. These non-incremental costs are allocated to termination costs on a proportional basis.

The European Commission is currently consulting on a Recommendation which proposes an alternative approach regarding the recovery of fixed and common costs.[3] Under this approach, MTRs would continue to be set by national regulators based on the LRAIC model but MTRs would not be used to contribute to the recovery of fixed and common costs. Instead, MNOs would have to recover these costs solely from the unregulated retail side of the market, essentially from prices charged to their own subscribers. The Commission claims that this approach will improve efficiency and decrease competitive distortions in the mobile telecommunications market.

In this paper, we examine the proposed regulatory scheme and whether it constitutes an efficient solution for the recovery of fixed and common costs in the European mobile telecommunications market. We find that the proposed regulatory scheme errs by assuming that efficiency dictates that prices for inputs, such as

---

[2] *See,* "EC Recommendation on Interconnection in a liberalised telecommunications market," Recommendation (98/195/EC) and IRG, "Principles of implementation and best practice regarding FL-LRIC cost modelling," Best Practice Paper (2000).

[3] *See,* "Draft Commission Recommendation on the Regulatory Treatment of Fixed and Mobile Termination Rates in the EU," European Commission (2008) and "Draft Commission Staff Working Document Explanatory Note: Accompanying Document to the Commission Recommendation on the Regulatory Treatment of Fixed and Mobile Termination Rates in the EU," European Commission (2008).

termination, should be set at marginal cost (or LRAIC). This paper first recalls that setting retail prices equal to marginal cost does not result in welfare maximizing outcomes in one-sided markets for either single product or multi-product firms when total revenues calculated at marginal cost prices do not cover total costs.[4] Such a pricing scheme does not necessarily result in welfare maximizing outcomes in two-sided markets either. We then extend this discussion to the pricing of termination in such settings and also conclude that economic efficiency calls for marking up of inputs sold to rivals, here the input being termination, in a variety of settings.

Given these results, we find that some portion of fixed and common costs should continue to be recovered from both the retail and the wholesale termination sides of the market as in the current regulatory scheme. We find that in order for the recovery of all fixed and common costs of an MNO from its retail prices to possibly be an efficient pricing outcome, call externalities must be relatively larger than access externalities, in the sense that the aggregate benefits to mobile subscribers from receiving calls on the network (call externalities) outweigh the aggregate benefits to call originators from being able to reach a larger base of mobile subscribers on the network (access externalities) — that is, in the sense that the un-internalized call externality effect dominates the access externality effect. We do not think that this is likely to be true in the mobile sector and it is inconsistent with existing regulatory views on the relative magnitudes of the call and access externality effects.

This paper is structured as follows: Section II discusses the economic theory of efficient retail pricing for single product firms in one-sided markets, multi-product firms in one-sided markets, and firms in two-sided markets. Section III considers efficient pricing for mobile termination services provided in either a one-sided or a two-sided market. Section IV analyses the efficiency of the Commission's proposed regulatory scheme. Section V concludes.

---

[4] This is a standard definition of increasing returns to scale for single- and multi-product firms.

II. EFFICIENT PRICING RULES FOR SINGLE- AND MULTI-PRODUCT FIRMS: THE BASICS

It appears that a recommendation to set MTRs at the incremental or marginal cost of providing terminating service is based on the view that marginal cost pricing of inputs ("termination") is the efficient price level. While this is certainly correct in the first-best welfare sense, i.e., where any revenue shortfalls from marginal cost pricing can be made up via lump-sum taxes, this is certainly not the case where such lump-sum taxes are unavailable and the operators must break even. Consequently, in order to examine the appropriate regulatory treatment of termination fees, we need to clearly describe the environments in which the actual retail and wholesale prices are set, as well as consider the fixed and common costs that have to be recovered for the MNOs to remain viable over the long-haul.

Before we proceed with a more detailed discussion, we note that the major reason against marking-up input prices – here the price of termination – does not directly apply in the current context. This aversion stems from the fact that marking-up input prices above their marginal costs leads to *production* inefficiency because buyers of the input will substitute away from the marked-up input to an input that is provided at undistorted (or relatively less distorted) prices.[5] In the context we focus on, the possibilities for such substitution are non-existent inasmuch as one minute of termination is required for every minute of an F2M call, or an M2M call, and access-seekers have no ability to bypass the terminating network's facility. Put another way, marking-up the price of termination on a mobile network does not induce inefficient substitution to another termination technology on a given mobile network and is not likely to induce construction of another access

---

[5] *See, e.g.,* P. A. Diamond and J. A. Mirrlees, "Optimal Taxation and Public Production I: Production Efficiency," *Am. Econ. Rev.*, vol. 61(1) (1971) and "Optimal Taxation and Public Production II: Tax Rules," *Am. Econ. Rev.*, vol. 61(3) (1971). Of course, when the producer of the input cannot break-even at marginal cost prices, it will have to mark up input prices appropriately. *See, e.g.,* J. A. Ordover and J. C. Panzar, "On the non-linear pricing of inputs," *Intl. Econ. Rev.*, vol. 23 (1982).

facility to the network's subscribers. Consequently, inefficiencies, if any, from marking-up termination costs must come from the suppression of overall F2M (or M2M) call volumes that such above-LRAIC pricing must induce. If shortfalls in total costs instead have to be made up in mobile retail prices, this would result in a suppression of retail activity either in terms of fewer calls being made by mobile network subscribers or from fewer subscribers or both. It is not clear *a priori* why one form of suppression of calling volumes is preferable to another.

In the next section we look at the efficient pricing rules for three market scenarios: single-product firms in one-sided markets, multi-product firms in one-sided markets, and firms in two-sided markets. Importantly, we find that the simple rule that price equals marginal cost does not apply in any of these market settings.

A. SINGLE- AND MULTI-PRODUCT FIRMS IN ONE-SIDED MARKETS

The simple economic precept that price ought to equal marginal cost only holds in the benchmark setting when there are not scale or scope economies or when any budgetary deficits for market participants resulting from marginal cost pricing can be covered with non-distortionary (lump-sum) taxes. Given that cost functions for telecommunications services are generally characterized by scale and scope economies and given that lump-sum taxes are not available to recover the shortfalls that would result from deviations from such pricing, marginal cost pricing is simply not feasible in the mobile telecommunications industry. This general point is well-recognized in economic literature and serves as a foundation for the theory of second-best (Ramsey-Boiteux) pricing.[6]

---

[6] *See, e.g.,* W. J. Baumol and J. G. Sidak, *Toward Competition in Local Telephony,* MIT Press (1994) for a non-technical discussion. *See also,* W. J. Baumol, J. C. Panzar, and R. D. Willig, *Contestable Markets and the Theory of Industry Structure,* Harcourt Brace Jovanovich, rev. ed. (1988) and J.-J. Laffont and J. Tirole, *Competition in Telecommunications,* MIT Press (2000) for more technical expositions.

For a single-product firm operating in a one-sided market, the second-best pricing formula simply leads to the prescription that price must equal average total cost. For multi-product firms operating in a one-sided market, optimal retail prices deviate from their underlying marginal costs in a manner that is determined by the elasticities of demand for each of the products as well as by the cross-elasticities of demand among the products of the multi-product firm. In the simplest case where these cross-elasticities are zero, individual product mark-ups are simply inversely related to the own elasticities of demand. This is the standard Ramsey-pricing formula under which products that have low elasticities of demand ought to bear higher percentage mark-ups above marginal cost as compared to products that have higher demand elasticities (where all these elasticities are calculated at the Ramsey-optimal vector of retail prices).[7] The obvious intuition is that such a pricing scheme minimizes the dead-weight cost of recovery of fixed and common costs because it minimizes the necessary suppression of output from the first-best levels while ensuring cost recovery. What this formulation also reveals is that exempting some products or services from such mark-ups merely puts an additional burden of responsibility for generating contributions toward the recovery of the joint and common costs on the remaining products. Unless there are sound economic or public policy reasons, no products or services should thus be exempt from bearing some responsibility for the recovery of total costs.

One way by which firms (or regulators) can ameliorate the inefficiencies from marking-up usage exclusively above marginal cost is by means of multi-part pricing, say a typical two-part tariff comprising a fixed entry (subscription) fee and a uniform usage fee. Such two-part tariffs do not solve the inefficiencies resulting from pricing above marginal cost, unless all potential customers are identical. This is because the positive fee for the right to purchase the product (i.e., entry or subscription fee) discourages some

[7] W. J. Baumol and D. Bradford, "Optimal Departures from Marginal Cost Pricing," *Am. Econ. Rev.*, vol. 60 (1970) as well as papers cited above. For a path-breaking analysis of Ramsey pricing in industries with network externalities, *see*, R. D. Willig, "The Theory of Network Access Pricing," in H.M. Trebing (ed.), *Issues in Public Utility Regulation,* Michigan State U.P. (1979).

customers from buying the service in the first place. Consequently, *ceteris paribus,* the higher the entry fee the smaller number will subscribe. On the other hand, a positive entry fee enables a lower usage fee, if overall profits are held constant, so that usage is less repressed by those who join the network. Hence, the optimal structure of a two-part tariff must inevitably reflect a Ramsey-like trade off (driven by the respective elasticities) between distortions on the extensive margin (participation) and the intensive margin (usage).[8] Put another way, whatever the set of instruments used to raise sufficient revenues to recover the total costs and earn the permissible profit, the structure of prices will always reflect, at least in part, the underlying elasticities of demand for the service. In fact, in telecommunications markets, we observe extensive reliance on complex pricing schemes that reflect these tradeoffs (as well as other considerations as we note below).

In sum, regulators have extensive experience in analyzing prices for multi-product firms when there are significant joint and common costs including markets in which these costs are common to both regulated and non-regulated services. Indeed, the principles of Ramsey pricing have been endorsed by regulators in various industries, at least as a proper basis for setting rates. However, despite its theoretical acceptance, in many cases the informational requirements for Ramsey pricing are too formidable for it to be implemented in practice[9] and rule-of-thumb allocation principles such as EPMU are applied instead.[10] Even with this caveat, it is important to keep the Ramsey-like

---

[8] *See, e.g.,* R. D. Willig, "Pareto-Superior Nonlinear Outlay Schedules," *Bell J. Econ.,* vol. 9 (1978) and J. A. Ordover and J. C. Panzar, *supra* note 5. The latter paper provides a model of input pricing with negative inter-customer externalities which shows that in the presences of such externalities, the input is always sold at above marginal cost.

[9] Other practical impediments to using Ramsey prices include the political unattractiveness of implementing the skewed distribution of prices across products that often results from Ramsey principles and the inability to use Ramsey prices to make comparisons between countries or operators. *See, e.g.*, M. Canoy, P. de Bijl and R. Kemp, "Access to Telecommunications Networks," in P.-A. Buigues and P. Rey (eds.), *The Economics of Antitrust and Regulation in Telecommunications*, Edward Elgar Publishing (2004), p. 157.

[10] The EPMU (or Equal Proportionate Mark-up) differs from Ramsey pricing in that the mark-up for common costs is proportional to the incremental costs of the product or service and does not include the willingness to pay

perspective when considering the soundness of the proposals for the reform of MTRs since these principles also inform the rules for the recovery of fixed and common costs in the market scenarios we examine next.

B. FIRMS IN TWO-SIDED MARKETS

We now consider the matter of access pricing and joint and common cost recovery in a more complicated setting, namely where firms operate in two-sided markets, that is, a market in which a firm supplies a product (or products) to two separate but interrelated groups of consumers—one on each side of the market. Firms operating in two-sided markets face additional considerations when determining the efficient prices for the products they supply on the two sides of the market. In particular, pricing in two-sided markets considers not only the total level or price charged by the two-sided platform to the two sides but also the ratio of the two prices. As we shall see, these considerations do not generally cause us to deviate from the overall conclusion that termination should be priced at above the incremental cost of termination (defined as LRAIC). In fact, the two-sided considerations can, by themselves, potentially enhance such policy conclusions.

Although there are several definitions of two-sided markets, they all boil down to the presence of inter-side externalities where the two separate groups of consumers generate membership externalities (e.g., by joining a network or buying a product) and usage externalities (e.g., by participating in the network after joining or using the product after buying it) on each other. In the base case, these two groups are unable to negotiate with one another to set prices that internalize these externalities. As a result, a platform (or intermediary) is needed in order to bring the two sides together and to set an efficient

---

(...continued)
component. While this method results in lower allocative efficiency than Ramsey pricing, the distortion is minor when common costs are small relative to incremental costs. *See, e.g.*, *id.*

8

price structure so as to maximize the utility of both groups of consumers subject to (at least) breaking even.[11]

Two-sided markets differ fundamentally from one-sided markets due to the volume of traffic being determined by both the structure of prices between the groups of consumers and the overall price level. Profit-optimizing and welfare-optimal prices (subject to break-even constraints) are determined by the magnitude of the price elasticities as well as the strength of inter-side externalities emanating on both sides of the market. Suppliers determine prices by balancing demand on the two sides of the market. Where the platform facilitates transactions between the two groups of customers, the per-unit cost of a transaction is joint and common to the two groups. Consequently, its recovery should be guided by typical Ramsey-type considerations. Indeed, in setting equilibrium prices, platform (network) operators are thus led – again consistent with general Ramsey-pricing principles – to recover more of the total network costs (including platform fixed and common costs) on the side of the market on which the consumers are less elastic and/or the side which values the participation of the other side more, so that the side that experiences higher external benefits contributes more, *ceteris paribus*. As a result, the prices which the interrelated groups of consumers face can depart very substantially from any conventional view of 'cost.'

This is the key insight from the economic research into two-sided markets. For example, in many two-sided markets, one set of consumers may obtain services for 'free,' whilst the entire cost of the platform is funded by another set of consumers. Such pricing occurs, for example, with advertising-funded newspapers, or search engines on the internet. These pricing structures – in which one side appears to bear most (or all) of the cost burden – can nonetheless be highly efficient and welfare maximizing.

---

[11] J.-C. Rochet and J. Tirole, "Two-Sided Markets: A Progress Report," *RAND J. Econ.*, vol. 37(3) (2006).

III.SETTING EFFICIENT MTRS

Mobile telecommunications firms operate in two-sided markets and provide multiple services to consumers in those markets. In addition, mobile telecommunications firms face significant fixed and common costs due to the need to invest heavily in infrastructure in order to achieve a minimum scale of operations. As evident from the above discussion, these features pose critical challenges for regulators trying to develop efficient pricing schemes for these markets. In particular, the two-sided market structure has specific implications for the efficient price level of mobile call termination that differ from those which would arise if mobile call termination were provided by a multi-product firm in a one-sided market.

To illustrate, assume that mobile termination services are provided in a one-sided market. The discussion above can be easily extended to the pricing of termination in a simple scenario in which the incumbent firm sells termination to its customers to rivals offering potentially differentiated products, such as long-distance service or various vertical services. In this case, often termed a "one-way" access scenario, termination is just another service (albeit a wholesale service) offered by the incumbent firm. The incumbent firm uses net revenues from all of its services to fund its fixed and common costs. From the discussion above, it readily follows that if termination is priced at marginal cost and the rivals divert sales from the incumbent, their activities will render the profit-constrained incumbent non-viable, unless a replacement source of revenues is found. Termination fees are one such source of revenues.

It is easy to show that in the benchmark case where all the products are substitutes, termination should be provided at rates exceeding the marginal cost of termination.[12] This policy prescription makes sense since if the incumbent could directly

---

[12] If rivals' products are complementary to those of the incumbent, termination may be priced below cost so as to stimulate usage of the incumbent's services which enhances revenues and allows smaller mark-ups on these products. Again, this is entirely consistent with the Ramsey-Boiteux framework in which cross-elasticity effects modify the typical Ramsey mark-up rule.

set the rivals' retail prices, it would set them above the marginal costs of production and use the mark-up to defray some portion of its fixed costs. The actual deviation from marginal cost will depend on the bypass possibilities as well as the range of mechanisms available to the regulator but the logic underlying these results is always the same: if some deviations from marginal cost pricing are necessary, then the proper approach is to minimize the welfare losses from the necessary mark-up. "Taxing" termination may be (and generally will be) a part of such policy.[13]

Thus, there is sound economic reason to impose a markup on termination charges to facilitate recovery of fixed and common costs incurred by the incumbent network spanning all services, including the provision of termination, because, as noted above, it is not possible for callers to substitute to another input; the inefficiency comes from the fact that the mark-up is reflected in a higher retail price, which causes a suppression of demand. As a result, while there is no production inefficiency from a mark-up on termination costs as a source of revenue for defraying a portion of fixed and common costs, the inefficiency arises on the usage side. Consequently, regulators will have to respect the Ramsey principles and gauge the relative magnitudes of the pertinent elasticities in determining the relative mark-ups on retail services (including those arising from mark-ups on termination).[14]

In a two-sided market, numerous factors – which add another layer of complexity to the already challenging considerations noted above – must be accounted for in

---

[13] Armstrong shows that where there is a possibility of some bypass, termination should be priced at cost provided the incumbent can recover the shortfall by means of a tax on the competitors' output. This result is consistent with the notion alluded to earlier that when input taxation causes productive inefficiency, it should be avoided if there are enough instruments at the disposal of the regulator. *See,* M. Armstrong, "The theory of access pricing and interconnection," chap. 8, in M. Cave, et al., *Handbook of Telecommunications Economics*, vol. 1, North Holland Elsevier (2002).

[14] It may appear that a first-best solution to the recovery of fixed and common costs could be implemented by means of a multi-part pricing imposed on users, with a fixed fee component and marginal-cost-based volume pricing. This is so only when consumers are homogenous, however. When consumers are not homogenous, the fixed fee component will create distortions by discouraging some consumers from subscribing to the network, which necessitates the careful assessment of the pertinent elasticities.

designing rules for the recovery of network costs, including the cost of termination. In the context of mobile termination, the two-sided nature of the market comes from the fact that a mobile network facilitates communications between two sets of users – the callers and those who are called. It is clear that in such a setting the rules for cost recovery can be rather complex. Moreover, the existence of competition between mobile platforms further complicates matters because it provides an independent source of constraint on retail prices.

The economic literature that addresses the issue of welfare maximizing MTRs shows that relationship of such MTRs to the marginal cost of termination is fundamentally determined by the existence and size of the access externality and the call externality, amongst other factors. Generally speaking, the welfare maximizing level of the MTR, in the presence of access externalities and no call externalities, is above the cost of providing termination. When call externalities are introduced into the analysis, the welfare maximizing level of the MTR falls back towards cost.[15] Indeed, Baranes and Flochel have examined the impact of call externalities on two-way access and shown that where subscribers care about the volume of calls they receive – i.e., subscribers prefer a network where they will receive a higher volume of incoming calls – and networks use multi-part pricing, call externalities will be internalized by the networks through below cost termination fees which enables the networks to charge higher fixed subscription fees.[16] Consequently, in a two-way access situation, the ability of networks to use termination fees to internalize call externalities removes the reason for cost-based

---

[15] Armstrong, Wright, Gans and King, and Hausman and Wright all conclude that in the presence of access externalities, in the context of one-way access, the welfare maximizing MTR is above the cost of termination. *See,* Armstrong, *supra* note 13; J. Wright, "Access Pricing under Competition: An Application to Cellular Networks," *J. Industrial Econ.,* vol. 50 (2002); J. Gans and S. King, "Mobile Network Competition, Consumer Ignorance and Fixed-to-mobile Call Prices," *Information Econ. & Policy,* vol. 12 (2000); and J. Hausman and J. Wright, "Two Sided Markets with Substitution: Mobile Termination Revisited," ms (2006). The existence of a call externality ameliorates this effect. For example, Armstrong, *supra* note 13, and Wright, *supra* note 15, show that allowing for call externalities lowers the welfare maximizing MTR.

[16] *See*, E. Baranes and L. Flochel, "Competition in Networks with Call Externalities," ms (2004). *See also*, B. E. Hermalin and M. L. Katz, "Customer or Complementor? Intercarrier Compensation with Two-Sided Benefits," ms (2006), who model benefits to both callers and calling parties in a one-way access setting.

regulation of MTRs because, at least in theory, such termination fees would be *below* LRAIC.

In practice, any intervention to set prices in a two-sided market must involve decisions as to how the recovery of various categories of costs should be distributed among different customers and services. Here, the magnitude of the two externality effects, alongside other factors such as the magnitude of the price elasticities for services on the respective sides of the market, determine from which side of the market the majority of the costs should be recovered. For example, all else equal, if the access externality were more significant than the call externality, then the majority of total network costs should be recovered from the termination side via MTRs whereas if the call externality were greater, then a greater proportion of costs should be recovered via retail prices.[17] Given these possibilities, the two-sided nature of mobile telecommunications markets could therefore suggest a structure of prices in which one or the other side of the market – either the mobile subscribers or those wishing to call the mobile subscribers – bears a larger proportion of the overall cost of the call and of the underlying network on which these calls are enabled. The question of which side of the market should bear the larger proportion of costs is ultimately an empirical question of whether the access externality effect or the call externality effect dominates. There are also additional considerations, such as those alluded to earlier, namely the extent of subscriber heterogeneity as gauged in terms of their overall benefit from joining a mobile network. As we have seen, such heterogeneity reduces the efficiency of the recovery of fixed costs

---

[17] That is, the more relative value that consumers making calls place on being able to contact mobile network subscribers on their mobile phones, the greater MTRs tend to be in equilibrium. Higher MTRs tend to result in lower retail prices and/or higher subsidies to subscribers, enticing more consumers to join the mobile network and further increasing value of the network to calling parties. On the other hand, when called parties receive relatively more value from receiving calls, the equilibrium and welfare maximizing termination charges to fixed and off-net mobile callers tend to be lower. Lower MTRs incentivize these callers to make more calls to mobile consumers, which benefits the mobile consumers receiving these calls. *See, e.g.*, Armstrong, *supra* note 13 and Wright, *supra* note 15.

by means of lump-sum subscriber fees. This effect is especially important if, as discussed below, the access externalities are significant.

In evaluating the extent of the access and call externalities, many regulators have concluded that it is not possible to estimate the extent of these externality effects with the certainty required to ensure that any resulting cost allocation is more efficient than one based purely on efficiently incurred costs.[18] We are not aware of any regulators outside the UK who have attempted to quantify empirically the externality effects present in the mobile telecommunications industry.[19]

The UK competition regulator (UK Competition Commission) and the UK telecommunications regulator (Ofcom) have examined in detail the application of, and the size of, externalities applicable to the mobile industry.[20] Both of these regulators have concluded that the access externality effect is more significant than the call externality effect.[21] That is not to say that the called party receives no benefit,[22] but rather that the

---

[18] For example, ARCEP concluded in its 2007 MTR decisions (decision nos. 04-937 & 04-938) that the complexity of measuring the level of the externality did not justify the *a priori* low impact it would have on the MTR (based on Ofcom's MTR mark-up due to the externality surcharge) and the potential to distort competition if applied incorrectly.

[19] However, a higher MTR due to network externality effect is allowed by the Greek NRA through a mark-up on LRAIC.

[20] *See*, Competition-Commission, *Vodafone, O2, Orange and T-Mobile: Reports on References Under Section 13 of Telecommunications Act 1984 on Charges Made by Vodafone, Orange, O2 and T-Mobile for Terminating Calls Made by Fixed and Mobile Networks*, HMSO, London (2003); Ofcom, *Wholesale Mobile Voice Call Termination: Statement*, Office of Communications, London (2004); and Ofcom, *Mobile Call Termination: Statement*, Office of Communications, London (2007).

[21] Ofcom estimated the size of the network externality by identifying through market research the number of non-subscribers likely to join a mobile network if subsidized, and the number of marginal subscribers who were not willing to pay the full cost of subscribing to a network. Ofcom calculated the necessary subsidy – using a net externality factor – in order to ensure that these marginal subscribers would still receive subscription subsidies. This approach was approved by the UKCC in 2003. The UK Competition Commission (2003) and Ofcom (2004 and 2007) also examined the applicability of other types of externalities – including the call externality. These externalities were discounted on the basis that (a) they are likely to be internalized and (b) the value is likely to be small and because the UKCC "did not think that [the UKCC] would be able to measure them accurately." *See,* Competition-Commission, *supra* note 20, p. 226.

[22] Ofcom market research indicates that the benefit of receiving a call is far less significant than the benefit of making a call. Market research, conducted in 2003, shows that 28% of respondents spontaneously stated that cost of making a call was the main determinant in determining a network. This compares to 2% who said that the cost to

un-internalized benefit received by existing subscribers from the addition, or retention, of a subscriber to mobile networks, is larger than the un-internalized benefits received by mobile subscribers from being able to be called by other fixed and mobile subscribers.

In fact, while it held that the majority of call externalities would be internalized, Ofcom noted that there is likely to be some un-internalized call externality arising from calls outside of repeat calling relationships.[23] Importantly, the effect of this was taken into account in the economic modeling used by the UKCC in 2003 and Ofcom in 2004 and 2007 to determine the level of the optimal externality surcharge. As such, the externality surcharge implemented by Ofcom in 2004 and 2007 should be interpreted as being the net externality surcharge taking into account the level of access externalities and un-internalized call externalities.[24]

The regulators' finding that the access externality is more significant than the call externality is not surprising, since it is easy to see that call externalities are easier to internalize between calling parties than are access externalities which can be generated by subscribers joining the network with which the existing subscriber may have only a very weak personal link.[25] That is, call externalities arise from a two-person relationship, in which the ability to negotiate and internalize the external call benefits is high, while

---

(...continued)

others to call them was the main factor. *Available at* http://www.ofcom.org.uk/consult/condocs/wholesale/wholesale.pdf.

[23] Evidence was presented that the majority of calls occur within repeat calling relationships. As such, any call externality will be internalized through the ongoing relationship between the parties. *See,* Competition-Commission, *supra* note 20, p. 255. It must also be acknowledged that calls from unknown parties also contain a probability that the call results in negative utility for the called party – i.e., nuisance calls. Calls from unknown parties, therefore, create the possibility that the call externality is negative.

[24] The key factor determining the level of the externality surcharge is the R-G Factor. Rohlfs adjusted cross-elasticities of demand estimates to take into account non-internalized call externalities. He recommended that a gross R-G Factor of 1.3 to 1.7, equating to a net factor of less than 1.1, be used. *See,* J. H. Rohlfs, "A Model of Prices and Costs of Mobile Network Operators," Report prepared at the request of Oftel (2002). Ofcom used a gross R-G Factor of 1.5 in calculating the optimal externality surcharge. *See,* Ofcom (2007), *supra* note 20, p. 342. The surcharge should therefore be interpreted as being the externality surcharge net of un-internalized call externalities.

[25] We are not aware of any regulator that has concluded – based on empirical studies or for other reasons – that the call externality is greater than the access externality.

access externalities are generated not only by persons in a relationship with each other but also by parties who are not in a relationship or even known to one another. In this situation, the welfare maximizing allocation of costs between the two sides is one that allocates a larger proportion of costs to the origination side of the mobile call market.

IV. ANALYSIS OF THE PROPOSED REGULATORY SCHEME

The regulatory approach adopted in Europe to date is consistent with pricing in a two-sided market in which the access externality is more significant than the call externality. Specifically, regulators have applied a 'Calling-Party-Pays' model to the division of costs. In this model, the network costs incurred in delivering calls are met by the initiating or calling party, being the mobile subscriber in the case of calls from the mobile network and the subscriber on the other network in the case of calls to the mobile network. In addition, regulators also have decided, to date, that the common costs of providing mobile services should be shared amongst both calling and called parties through a mark-up on all services. Other costs of the platform – marketing costs, subscriber acquisition and retention costs, and other costs associated with gaining mobile subscribers who can be called – are in principle borne entirely by mobile subscribers (the called party) rather than by those benefiting from the opportunity to call them.[26] Of course, when all costs need to be recovered by the network, it really does not matter in the end whether a particular element of cost is placed in one bucket versus another for the purposes of analyzing cost recovery. The allocation matters only insofar as it puts a cap on the quantum of recovery from one service versus another, e.g., retail services versus wholesale services.

As discussed above, the proposed regulatory scheme would change this approach to recover fixed and common costs on the retail side of the market only rather than from both the retail and the wholesale termination sides. The European Commission claims

---

[26] Although the two-sided nature of the market does not mean this result is obvious – the presence of externalities in the platform means that it may be efficient for calling parties to contribute to these costs – this is explicitly provided for in Ofcom's adoption of a 'network externality surcharge' on MTRs.

that the proposed scheme is preferable to apply a LRAIC approach to MTRs without any mark-up for fixed and common costs because this approach will improve efficiency and reduce competitive distortions in the mobile telecommunications market. As we have already indicated and further explain below, it is not clear that it will accomplish either of these objectives.

A. EFFICIENCY

In this context, improving efficiency involves reallocating costs in order to achieve more efficient retail prices in the market.[27] As discussed above, MTRs are not set in isolation but in conjunction with retail prices and total network costs are recovered from both sides of the market based on the relative magnitudes of the externalities on each side and other considerations. We have already discussed how difficult it is to benchmark any particular pricing proposals against an 'optimally efficient' benchmark for a two-sided market in which potent, but difficult to measure, externalities are key drivers of efficient pricing.

In the absence of rigorous assessment of these effects, concluding that the call externality is larger than the access externality effect, it is impossible for the Commission to recommend at this time that it is more efficient to recover all of the joint and common costs from the retail side of the market. We are not aware of any such assessment by the Commission. Thus, the Commission's proposal to reallocate the burden of common cost recovery entirely to one side of the market can easily reduce efficiency, given that such allocation of recovery responsibilities is not a necessarily welfare-maximizing public policy in a variety of realistic market scenarios. Since it is clear that, on average, both callers and called parties derive utility from the making and receiving of calls, but existing research shows that the access externality is larger than any call externality, the

---

[27] The issue here is not about whether inefficiencies should be captured in regulated prices, but about how (i.e., from which customers) efficiently incurred and unavoidable costs should be recovered by MNOs. We note that we strongly agree that inefficiently incurred costs should not be recovered from regulated prices – in fact, we do not think they should or would be recovered at all in a competitive market.

least distortive solution to the recovery of common costs would clearly require that they be recovered from both callers and called parties, with a potentially greater proportion allocated to the party making the call. In other words, there is no reason to believe that the proposed regulatory scheme is more efficient than the current scheme.

B. COMPETITIVE DISTORTIONS

There are two potential competitive distortions that generally concern regulators with respect to cost recovery in mobile telecommunications markets. The first is said to arise from a (partial) transfer of responsibility for cost recovery between fixed and mobile markets, where proponents argue the effect of this transfer "has been to injure fixed customers and their operators…" and that "the transfer has also distorted competition between fixed and mobile operators."[28] However, even from a narrow perspective of cost analysis, this so-called transfer is highly unlikely to constitute a cross-subsidy based on standard definitions. For example, according to Faulhaber, a service cross-subsidizes other services when the revenues from the service exceed the stand-alone cost of providing that service.[29] Temin extends this definition and defines a cross-subsidized price as being below the incremental cost of the associated service. Temin further notes that in the presence of joint costs, there exists a range of prices lying between the stand-alone cost and the incremental cost that are neither cross-subsidizing nor cross-subsidized. Such prices are possible because the multi-product firm realizes the benefits of scale and scope economies and is able to pass them on to consumers in form of favorable prices.[30] If the current MTRs entail some cross-subsidy from fixed to mobile customers, then according to these definitions, the revenues based on the current MTRs

---

[28]   *See,* O. Bomsel, M. Cave, G. Le Blanc, and K.-H. Neumann, "How mobile termination charges shape the dynamics of the telecom sector," University of Warwick and WIK Consult (2003), *available at* http://www.cerna.ensmp.fr/Documents/OB-GLB-F2M-FinalReport.pdf, p. 7.

[29] *See,* G. R. Faulhaber, "Cross-Subsidization: Pricing in Public Enterprises," *Am. Econ. Rev.*, vol. 65 (1975). Of course, this definition assumes that firm's profits are constrained.

[30] *See*, P. Temin, "Cross Subsidies in the Telephone Network after Divestiture," *J. Regulatory Econ.*, vol. 2 (1990). Note that Ramsey prices can entail some cross-subsidy because these are benchmarked against marginal costs and not incremental costs (which may include some product-specific fixed costs).

will be greater than the stand-alone cost to an FNO of establishing its own mobile network to provide its fixed subscribers with mobile termination on F2M calls.

As noted above, the MTR under the current scheme is set at LRAIC plus a proportional mark-up for fixed and common costs. In asking whether fixed subscribers are cross-subsidizing mobile subscribers one could, following a standard practice, ascertain whether the revenues from mobile termination exceed the stand-alone cost of the mobile network. This test reflects the fact that fixed subscribers derive benefits from having a mobile network with subscribers that can be reached. Since the stand-alone cost of providing mobile termination would include all fixed and common costs associated with the operation of the network, rather than just some notional share of these costs, the current revenues calculated at current MTRs are likely less than the stand-alone cost and thus, the current MTRs are not a cross-subsidizing price.[31] Of course, because current MTRs exceed the incremental cost of providing mobile termination (i.e., LRAIC), MTRs are not a cross-subsidized price either. This result places the current MTRs in the range of prices that result from operating efficiencies in a multi-product firm.

As this analysis shows, the difference in FNO and MNO fixed and common costs is irrelevant with respect to whether fixed network callers are cross-subsidizing the mobile network subscribers who they call. Moreover, given that there is no evidence that current MTRs entail a cross-subsidy to the mobile side, proponents of the proposed scheme will be hard-pressed to argue that the proposed scheme will reduce a competitive distortion resulting from such a cross-subsidy.

We would further add that, as noted above, efficient pricing schemes in two-sided markets may in fact involve an allocation of costs based on elasticities and externalities that results in one side paying very little or nothing or even being paid to participate. For

---

[31] Of course, in this hypothetical situation, the fixed network would not charge itself LRAIC for termination. It would book termination on its mobile network at marginal cost but it would have to mark-up the retail price of an "on-net" F2M call sufficiently to recover all the fixed and common costs.

example, men generally subsidize women in on-line dating clubs and newspaper subscribers are usually cross-subsidized by advertisers. This is true even if we assume away all the fixed costs of a platform. Thus, even if a cross-subsidy were proven to exist between fixed and mobile consumers, the existence of such a cross-subsidy would not be sufficient to establish that it caused a competitive distortion to exist as well.

The second claimed competitive distortion is said to occur within the mobile market itself. In this case, the customers of smaller MNOs are said to cross-subsidize the customers of larger MNOs due to the higher proportion of off-net calls, and thus the higher proportion of MTRs paid to other networks, incurred by the smaller networks. However, there is no reason to believe that the existence of different-sized MNOs results in a cross-subsidy from the smaller to the larger MNOs. The size of the MNOs does not matter as long as MTRs are equal and calling patterns between the networks are balanced—i.e., proportional to each network's size.[32] In any case, in the case of two-way access, it has been shown, again consistent with the precepts embodied in the principles of the efficient component pricing rule (ECPR) or modified ECPR (MECPR) that termination fees between mobile networks should include an element of mark-up towards the recovery of mobile network fixed and common costs.[33]

Moreover, applying the same definitions of a cross-subsidizing price as above returns the same result as above—namely, that the current MTRs (set at LRAIC plus a proportional mark-up for fixed and common costs) are less than the stand-alone cost of providing mobile termination service to other MNOs and greater than the incremental cost of providing this service. Once again, there is no cross-subsidy causing a

---

[32] For example, a network that has 10% of subscribers will have 90% of its subscribers' calls be off-net calls and will receive 90% of its incoming calls as off-net calls (under a broad range of symmetry assumptions). *See*, P. de Bijl and M. Peitz, *Regulation and Entry into Telecommunications Markets*, Cambridge University Press (2002) for more on balanced calling patterns.

[33] *See, e.g.,* Laffont and Tirole, *supra* note 6. For discussions of termination fees that reflect the MECPR principles, *see,* Doh-Shin Jeon "A Simple Access Pricing Rule to Achieve the Ramsey Outcome for Interconnected Networks" ms (2005) and Sue H. Mialon, "Pricing Access in Network Competition," *J. Regulatory Econ.*, vol. 31(1) (2007).

competitive distortion to be reduced by the proposed scheme and therefore, no reason to implement the proposed scheme based on such a claimed benefit.[34]

v. CONCLUSION

The above discussion demonstrates that there is no compelling support for the proposed regulatory scheme of recovering fixed and common costs from retail prices only instead of also through MTRs. There is abundant support in the theoretical literature for setting MTRs above cost. Moreover, the existing evidence shows that the network externality is more significant than the call externality, supporting greater cost recovery from termination side in this case. However, in the reverse case of a more significant call externality, the literature has shown that MTRs will be below cost, eliminating the need for regulation at all.

The Commission has yet to offer a convincing case regarding the benefits of the scheme. As we showed above, the Commission's key justifications, improved efficiency and decreased competitive distortions, fail under closer examination. In addition, it is unclear what effects the proposed scheme would have on welfare. As we have discussed, in two-sided markets with externalities that we cannot measure well, setting prices at cost (including fixed and common costs) is the least distortive approach and therefore the optimal second best solution. The current scheme follows this approach. Thus, there is no reason to believe that departing from this scheme would enhance welfare and in fact, given that the proposed scheme will generate welfare-decreasing as well as welfare-increasing effects, there is reason to believe that it could produce the opposite effect.

---

[34] The real distortions between MNOs arise because regulators have allowed asymmetric MTRs, with mark-ups over cost, between MNOs for years. Elimination of these asymmetries would reduce competitive distortions. *See, e.g.*, Armstrong, *supra* note 13.