

Regulation of fixed and mobile termination charges



Amsterdam, 27 February 2013
Commissioned by UPC Nederland and Ziggo

Regulation of fixed and mobile termination charges

A comparison of *pure* BULRIC and BULRIC *plus* as regulating principles

Viktória Kocsis
Bert Tieben
Rob van der Noll



seo economic research

“The science of knowing”

SEO Economic Research carries out independent applied economic research on behalf of the public and private sectors. This research makes a major contribution to the decision-making processes of our clients. SEO Economic Research is affiliated to the University of Amsterdam. This gives us access to the latest scientific methods. Operating on a not-for-profit basis, SEO continually invests in the intellectual capital of its staff by arranging for them to pursue graduate studies, publish scientific works and participate in academic networks and conferences.

SEO-report nr. 2013-13

Summary

In this report, SEO Economic Research investigates the welfare consequences of different regulatory models applied in the wholesale call termination market. Wholesale call termination is a service that network operators provide to each other when a call from one network needs to terminate in the other network. As each call terminates in the network of the called party, this network has a monopoly position on the market for call termination to its subscribers. Consequently, every network is a relevant market with significant market power (SMP). In the Netherlands, fixed and mobile call terminations are seen as markets, where each network has SMP. Therefore, fixed termination rates (FTRs) and mobile termination rates (MTRs) are subject to ex ante price cap regulation.

The caps for FTRs and MTRs are based on methodologies determined by the National Regulatory Authorities, OPTA in the Netherlands. In 2010, following the recommendations of the European Commission, OPTA decided to impose termination rates based on the pure BULRIC model. However, the Dutch Trade and Industry Appeals Tribunal (*College van Beroep voor het bedrijfsleven*, CBB) found this model not proportionate to fulfill the requirements set in the Dutch Telecommunications Act and required the adaptation of the BULRIC plus model. The main difference between the models lies in which cost elements are considered as variable with call termination. A BULRIC plus termination rate includes several fixed costs elements in favor of network operators and hence implies a higher tariff.

The report addresses the following questions:

- Why is the pure BULRIC approach necessary to remedy the problems in the termination markets? This is the question of *effectiveness*. Is pure BULRIC more effective than plus BULRIC as a remedy to curtail market power in call termination? What explains the difference in effectiveness?
- Why does the pure BULRIC approach lead to improved welfare / consumer surplus over the BULRIC plus approach? This is the issue of *proportionality*. According to the Dutch Telecommunications Act, regulation must be proportional to the problem at hand, in this market power. A plus BULRIC method can be proportional if it turns out to be effective *and* welfare improving. This means that the evidence for the impact of plus BULRIC on welfare needs to be critically assessed.

The analysis is based on an extensive review of theoretical and empirical studies (see empirical studies marked with * in the bibliography). To critically compare different regulatory methods for call termination, economic arguments are systematically organized and presented around the concept of welfare. *Welfare is the sum of consumer and producer surplus*. The above mentioned concepts of effectiveness and proportionality are defined in the context of welfare economics.

Effectiveness

From a welfare perspective, it is desirable that tasks are performed efficiently. Efficiency can be distinguished in the short and the long term. Short-term (or static) efficiency is defined as the

welfare level, at which all firms are on their production possibility frontier (i.e., productive or cost efficiency). Long-term (or dynamic) efficiency is defined as expected future welfare that includes investments and innovations. A regulatory method is effective if it reduces market power to the minimum in the wholesale market (static efficiency) and still provides incentives to invest in call termination (dynamic efficiency).

The most important factor for effectiveness is which cost elements need to be considered as incremental in cost calculations. In particular, the focus lies on unavoidable fixed costs as these costs are claimed to be necessary for dynamic efficiency. Based on economic studies and the characteristics of Dutch telephony markets, the justification for including unavoidable fixed costs in cost calculation fails for the following reasons. Firstly, network externalities are necessary to include overhead and common costs in cost calculation. In a mature market, such as the telephony markets in the Netherlands, network externalities are negligible. Secondly, common and fixed cost can be more effectively recovered in the retail market than in the wholesale market, in particular when non-linear retail prices are set, which is currently the case in mobile markets. The fixed or the flat fee is an effective measure to recover these costs.

As including unavoidable costs is not justified, both BULRIC methods provide sufficient investment incentives, but BULRIC plus includes a markup above incremental costs. Therefore, *BULRIC plus cannot reduce market power to the minimum and hence it is not effective*. Higher termination rates based on BULRIC plus have two further negative consequences in the call termination market. First, higher termination rates distort competition between asymmetric mobile operators. Second, the application of asymmetric cost methodologies in European countries creates competitive distortions in the internal market at the cost of countries applying the pure BULRIC model. These negative effects are not discussed in the CBB decision.

Proportionality

A regulatory method is proportional if it increases welfare (volume effect) and if the increase in welfare does not influence parties disproportionately (distribution effect: no large differences appear between consumers and firms in the same market and between fixed and mobile markets).

In the Netherlands, termination rates are paid by the network of the calling party (*calling party network pays* principle). Therefore, call termination has effects on retail prices. In this context, two relevant mechanisms can be distinguished: the termination rate is a cost element for calling party network and the termination rate is a revenue element for the called party network.

Termination rate as cost: As a result of higher termination rates, the costs of call termination increase for the calling party network. This network can be a fixed network or a competing mobile network. Higher costs are then translated into higher per minute prices (see Table S.1).

Termination rate as revenue: Call termination is a two-sided market. In the two sides of the market are the customers of the called network and the calling party network. They make contact with each other via the platform, which is the called network. Decisions taken in one side (retail prices) influence decisions to be taken on the other side (termination rates). As call termination is regulated, the called network has incentives to balance revenues from call termination by retail

revenues. This is the so-called *waterbed effect*. As a consequence of higher termination rates, a “reversed” waterbed effect applies. The higher the termination rates, the higher the revenues the called party network may receive. This network operator then sets lower subscription prices to keep customers in and receive the termination revenues for calls to these customers.

Table S.1 Due to higher mobile termination rates per minute prices increase and fixed fees reduce

Changes in	Mobile telephony Linear prices (pre-paid packages)	Non-linear prices (post-paid packages)	Fixed telephony (in particular, off-net prices)
Per minute price	↑	↑	↑
Fixed fee	n.a.	↓	n.a.
Number of called minutes	↓	↓	↓
Number of subscriptions	n.a.	No significant effect ¹	No significant effect ¹ or ↓ ²

Source: SEO Economic Research; ↑: higher, ↓: lower; ¹ In mature markets; ² Higher per minute prices in fixed telephony may reduce the value of fixed-to-mobile connectivity for mobile subscribers (↓).

Changes in retail prices have consequences on welfare (see Tables S.1 and S.2).

Table S.2 The achieved welfare is higher by pure BULRIC than by BULRIC plus

	Pure BULRIC (P)	relation	BULRIC plus (+)
Mobile market with linear prices			
Consumer surplus (CS)	CS(P)	>	CS(+)
Total producer surplus (PS)	PS(P)	<	PS(+)
Producer surplus (PSr): retail	PSr(P)	<	PSr(+)
Producers surplus (PSw): wholesale	PSw(P)	<	PSw(+)
Deadweight loss (DWL)	DWL(P)	<	DWL(+)
Welfare (W)	W(P)	>	W(+)
Mobile market with non-linear prices			
Consumer surplus (CS)	CS(P)	=	CS(+)
Total producer surplus (PS)	PS(P)	≥	PS(+)
Producer surplus (PSr): retail	PSr(P)	>	PSr(+)
Producers surplus (PSw): wholesale	PSw(P)	<	PSw(+)
Deadweight loss (DWL)	DWL(P)	<	DWL(+)
Welfare (W)	W(P)	>	W(+)
Fixed telephony			
Consumer surplus (CS)	CS(P)	>	CS(+)
Total producer surplus (PS)	PS(P)	>	PS(+)
Producer surplus (PSr): retail	PSr(P)	<	PSr(+)
Producers surplus (PSw): wholesale	PSw(P)	>>	PSw(+)
Deadweight loss (DWL)	DWL(P)	<	DWL(+)
Welfare (W)	W(P)	>>	W(+)

Source: SEO Economic Research based on finding in the literature.

Deadweights loss is created hence welfare reduces. As a result of higher per minute prices, consumers call less (see Table S.1). This reduced volume creates a lost value in the total market, called deadweight loss. In a mature market, such as in the Netherlands, it is unlikely that the number of subscribers significantly increases as a result of lower fixed fees. Due to deadweight loss, BULRIC plus leads to lower welfare than pure BULRIC.

Consumer are never better off with BULRIC plus. In the case of linear prices, consumers benefit less from a higher termination rate: consumer surplus reduces as a result of less called minutes (Table S.2). Furthermore, call externalities cannot be fully utilized. However, this result is more relevant in a growing market. In the case of non-linear prices, consumer surplus remains zero.

Firms only benefit from BULRIC plus if linear (pre-paid) prices are set. However pre-paid packages account for only a small proportion of firms' revenues. Total producer surplus is the sum of surplus in the wholesale and retail market (see Table S.2). In the mobile market, total surplus increases in the case of linear prices due to increased termination costs and, as network operators are left with market power in the wholesale market, due to double marginalization. In the case of non-linear prices, the profit neutrality rule holds for incumbent firms: termination rates leave firms' surplus unaffected as a result of the waterbed effect. However, entrants are worse off with a higher termination rate, such as determined by BULRIC plus. In the fixed telephony market, firms are always worse off at a higher MTR as FTRs are substantially lower than MTRs.

As Table S.2 and the analysis of OPTA show, BULRIC plus distributes welfare disproportionately between market parties. The surplus of mobile firms can only increase at a disproportionately larger decrease of both consumer surplus and of the surplus created in fixed telephony. These effects are not taken into consideration in the CBb ruling. As a conclusion, BULRIC plus is not a proportional regulatory measure.

Conclusions

To conclude, the research questions are recalled:

- Why is the pure BULRIC approach necessary to remedy the problems in the termination markets?
- *Although termination rates based on both BULRIC models provide sufficient investment incentives, BULRIC plus cannot reduce market power to the minimum the way pure BULRIC does. Therefore, BULRIC plus cannot be seen as an effective measure.*
- Why does the pure BULRIC approach lead to improved welfare / consumer surplus over the BULRIC plus approach?
- *Even in the presence of the "reversed" waterbed effect, BULRIC plus termination rates lead to higher per minute prices and less called minutes. Less volume creates a lost value for the total market, which reduces total welfare. In some cases, BULRIC plus makes mobile firms better off (mainly due to the market power in the wholesale market) but it happens disproportionately at the cost of consumers in both markets and fixed telephony firms. Therefore, BULRIC plus cannot be seen as a proportional measure.*

Table of contents

Summary	i
1 Introduction	1
2 Termination rates need to be regulated	3
2.1 What is call termination?	3
2.2 Justification for regulating termination rates	3
2.3 Recommendation in the economic literature.....	5
2.4 Recommendation of the EC.....	6
2.5 Discussion in the Netherlands	7
2.6 What is effectiveness and proportionality?	10
2.7 Conclusions	11
3 Effectiveness of BULRIC models	13
3.1 Which cost elements may play a role?	13
3.2 Justification for including fixed costs fails	13
3.3 BULRIC+ leaves firms with market power.....	16
3.4 Negative effects of BULRIC+ in wholesale markets.....	19
3.5 Conclusions	20
4 Proportionality of BULRIC models	23
4.1 Welfare as a basis for proportionality	24
4.2 Retail tariff structures	26
4.3 BULRIC plus is not a proportional measure.....	27
4.4 Conclusions	36
Literature	37

1 Introduction

Wholesale call termination is a service that network operators provide to each other when a call from one network needs to terminate in the other network. According to the European regulation, the network operator of the calling part pays for this service, which payment is then passed onto the customers in the per minute retail price. As each call terminates in the network of the called party, this network has a monopoly position on the market for call termination to its subscribers. Consequently, every network is a relevant market with significant market power (SMP). In the Netherlands, fixed and mobile call terminations are seen as markets (Markets 3 and 7 in Article 6 of the Dutch Telecommunications Act),¹ where each network has SMP. Therefore, fixed termination rates (FTRs) and mobile termination rates (MTRs) are subject to ex ante price cap regulation.

The caps for FTRs and MTRs are based on methodologies determined by the National Regulatory Authorities, OPTA in the Netherlands. Based on the recommendation of the European Commission (2009/396/EC), OPTA decided that these two markets need to be considered as separate markets and the related market analysis needs to be conducted separately. Yet, the methodologies that determine the level of termination charges need to be comparable due to the expected converge of fixed and mobile communication markets. Furthermore, OPTA also followed the recommendation of the EC and chose the pure BULRIC (bottom-up long run incremental cost) method to determine the cap on termination rates (OPTA, 2010). However, the Dutch Trade and Industry Appeals Tribunal (*College van Beroep voor het bedrijfsleven*, CBb) rejected the application of the pure BULRIC method on 31 August 2011 (CBb, 2011). According to the decision of CBb, a plus BULRIC method satisfies all the legal demands of the Dutch Telecommunications Act but imposes lower costs on the industry, which means that from a legal point of view it should be preferred over a pure BULRIC method. OPTA is expected to publish its view on the appropriateness of the plus BULRIC model to the regulation of the termination rates in fixed and mobile telecommunications. The aim of this research is to prepare economic analysis, which allows the commission parties for this research to quickly respond to OPTA's revised view on plus BULRIC *vis-à-vis* pure BULRIC.

In particular UPC Nederland and Ziggo are looking for economic motivations for using a pure BULRIC-method. The following questions are of particular interest:

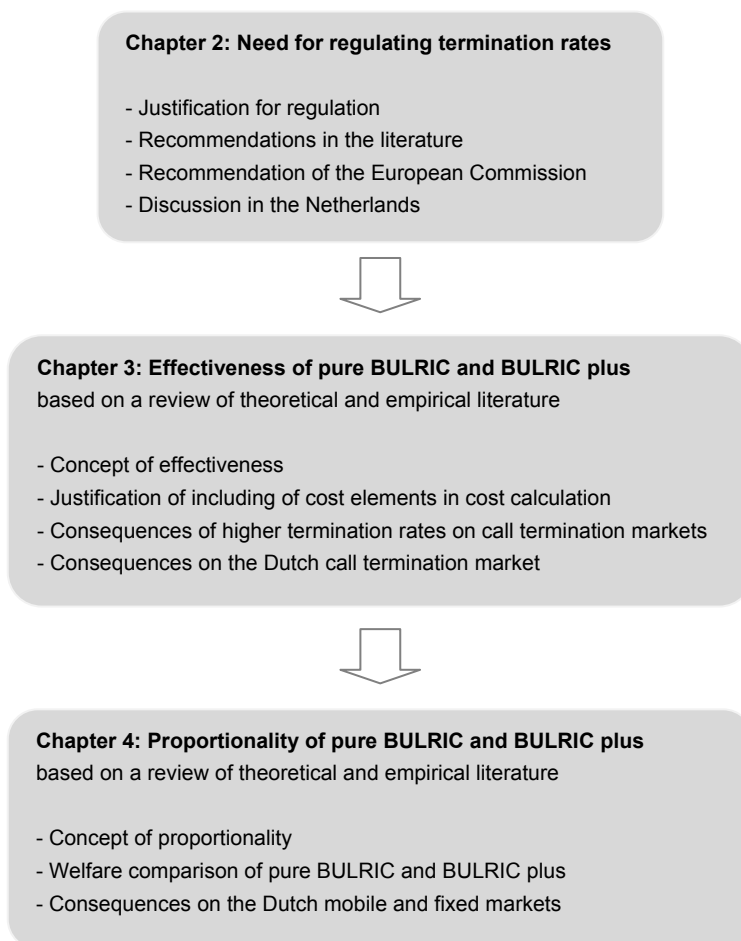
- Why is the pure BULRIC approach necessary to remedy the problems in the termination markets? This is the question of *effectiveness*. Is pure BULRIC more effective than plus BULRIC as a remedy to curtail market power in call termination? What explains the difference in effectiveness?
- Why does the pure BULRIC approach lead to improved welfare / consumer surplus over the BULRIC plus approach? This is the issue of *proportionality*. According to the Dutch Telecommunications Act, regulation must be proportional to the problem at hand, in this

¹ Among others the 088, 084/087 and 112 numbers belong to Market 3; 06 numbers fall under the scope of Market 7.

market power. A plus BULRIC method can be proportional if it turns out to be effective *and* welfare improving. This means that the evidence for the impact of plus BULRIC on welfare needs to be critically assessed.

The research has three main steps, presented in three chapters of this report (Figure 1.1). In Chapter 2, the background for regulating call termination is presented. The recommendations in the literature are compared with the European recommendations. In the light of these recommendations, the discussion in the Netherlands – in particular between OPTA and CBB – is introduced. In Chapter 3, the effectiveness of the regulatory models is analyzed with a special focus on the effects in the wholesale market. In Chapter 4, the regulatory methods are compared based on their proportionality. Welfare consequences are analyzed in both the mobile and fixed telephony markets. In Chapter 5, conclusions are drawn.

Figure 1.1 The structure of the report



2 Termination rates need to be regulated

As unregulated network operators have market power for terminating calls, they set high termination charges that differ from the cost-efficient price. Therefore, wholesale call termination is subject to regulation. But what is an effective and proportional regulatory measure? This chapter provides the background for the discussion.

Several arguments underlie the fact that unregulated network operators have incentives and market power to increase termination rates above costs. Regulated termination rates are based on bottom-up incremental costs, as suggested by the literature. There is, however, a discussion about which cost elements need to be included in order to provide an effective and proportional measure to reduce market power and maintain investment incentive at the benefit of consumers and firms. Chapter 2 presents this discussion about the different regulatory models, in particular pure BULRIC and BULRIC plus.

In Section 2.1, the concept of call termination is explained, followed by argument for regulation in Section 2.2. The recommendations in the literature and of the European Commission are presented in Sections 2.3 and 2.4. In the light of these recommendations, the discussion in the Netherlands – in particular between OPTA and CBb – is introduced in Section 2.5. Finally in Section 2.6, definitions, consistent with the economic literature, are given on effectiveness and proportionality. Section 2.7 concludes the results.

2.1 What is call termination?

Wholesale call termination is a service that network operators provide to each other when a call from one network needs to terminate in the other network. According to the European regulation, the network operator of the calling part pays for this service, which payment is then passed onto the customers in the per minute retail price. Each network has a monopoly for calls terminating in its network. Consequently, every network is a relevant market with significant market power (SMP). In the Netherlands, fixed and mobile call termination are seen as markets (Markets 3 and 7 in Article 6 of the Dutch Telecommunications Act),² where each network operator has SMP. Therefore, fixed termination rates (FTRs) and mobile termination rates (MTRs) are subject to ex ante price cap regulation.

2.2 Justification for regulating termination rates

To critically compare different regulatory methods for call termination, economic arguments will be systematically organized and presented around the concepts of efficiency and welfare. Efficiency and welfare are two related concepts. *Welfare is the sum of consumer and producer surplus.* From a welfare perspective, it is desirable that tasks are performed efficiently. Efficiency can be distinguished in the short and the long term.

² Among others the 088, 084/087 and 112 numbers belong to Market 3; 06 numbers fall under the scope of Market 7.

Short-term (or static) efficiency is defined in the standard economics literature as the welfare level, at which all firms are on their production possibility frontier (i.e., productive or cost efficiency).³ *Long-term (or dynamic) efficiency* is defined as expected future welfare that includes investments and innovations. More reliable products (i.e., increased capacity in communication networks or new innovative services, such as VoIP) positively affect these expected future revenues.

As a result of short-term efficiency, firms that produce at low costs can also set prices at this low cost level. This efficient price is equal to the marginal cost of production, from which users benefit the most. However, this low price may provide insufficient incentives for firms to invest, thus lowering long-term welfare.

A well-known result in the economic literature on call termination is that as unregulated network operators have market power for terminating calls, they set high termination charges that differ from the cost-efficient price. These high tariffs are paid by the network of the calling party (*calling party pays principle*). Several arguments lie behind high termination rates:⁴

- Due to *calling party pays principle*, as the called party does not pay for having the call terminated in its network, he has no incentives to respond to (changes in) the level of termination rates by switching to another provider. Consequently, his network sets an extensive termination rate (Armstrong 2002, Bijlsma & van Dijk 2007).
- Symmetric firms, with by and large the same number of calling as called customers, have incentives to charge high termination rates in order to soften competition by *raising a rival's termination costs*. This monopolistic behavior, which is present if firms set linear retail prices, also leads to high termination charges (Armstrong 1998, Laffont et al. 1998, Carter & Wright 1999).⁵
- Customers are *heterogeneous* and therefore *calling patterns* are hardly symmetric. There are subscribers that are more price sensitive than others. The extent of price sensitivity for subscription is often correlated with the number of called minutes. Those consumers that call less as the price per minute increases also more easily change subscription. These customers often call less than receive calls (*light users*). Subscribers that are less price sensitive and change their calling volume to a lesser degree as an effect of price changes are called *heavy users*. Networks have incentives to set termination rates above costs as higher rates weaken competition for heavy users (it is beneficial to lose a heavy user to increase termination profits) and strengthen competition for light users (it is beneficial not to lose a light user to avoid termination deficits). These results are valid also in the presence of *third degree price discrimination* (i.e. different retail prices for light and heavy users) and *second degree price discrimination* (i.e. pre- and post-paid packages). See Jullien et al. (2010).

Arguments also exist for below-cost unregulated termination rates. Service providers may set different retail prices for on-net and off-net calls (so called termination-based price discrimination). In this case, incumbent network operators prefer collectively setting termination

³ See e.g., De Bijl & Kocsis (2007).

⁴ Hoering & Valletti (2012) provide an extensive review of these arguments.

⁵ Under two-part and non-linear retail tariffs, tacit collusion is not present any more as profits do not react to any change in termination rates (so called profit neutrality; see Laffont et al. 1998, Dessein 2003). Even consumer heterogeneity does not reverse this result (Dessein 2003, Hahn 2004).

rates below costs. Benefits are created via two channels. First, network operators can compensate for the lost termination revenue due to lower termination rates by a higher subscription fee (see waterbed effect in Section 4.1). Moreover, firms then compete for new customers less intensively. Second, the lost termination revenue can also be compensated by a higher off-net tariff. Furthermore, the firm can keep the on-net price low. Consumers then have an incentive to join the largest available network as calling within the network is much cheaper than calling someone connected to another network (so called tariff-mediated network effects). Therefore, firms are able to benefit from a higher difference between on-net and off-net retail prices. This behavior is however disadvantageous for potential entrants: in the presence of termination-based price discrimination and thus tariff-mediated network effects, entrants are less able to build up market shares. To conclude, both mechanisms create problems for consumers as competition reduces. This result holds under two-part tariffs (Gans & King 2001) and in the presence of strong call externalities, that is, when consumers value high the possibility to be called (Berger 2004).

However, in practice, above-cost tariffs are observed. According to Hoernig & Valletti (2012), network operators charge above-cost termination rates because they mainly set prices unilaterally as profit maximizing monopolies and even if they negotiate, the earlier mentioned economic effects outweigh the benefits of lower termination rates.

Sufficiently (and not extensively) high termination rates are necessary to provide investment incentives (higher long-term welfare). However, termination rates above cost level contain a mark-up thus raising the costs of calling party networks (lower short-term welfare). In addition, termination rates are passed on to consumers in the retail prices, which may result in higher retail prices, thus substantially reducing consumer surplus.⁶ Therefore, termination rates are subject to regulation.

2.3 Recommendation in the economic literature

The literature on wholesale call termination shows that the efficient regulated termination rate, in the case of symmetric firms, needs to be based on bottom-up long run incremental costs (Laffont & Tirole 2000, Jullien & Rey 2008). The following arguments underlie this statement:

- *Incremental costs*: Prices set at the level of incremental (marginal) costs that vary with traffic flow lead to the highest short-term welfare (static efficiency).
- *Long-run costs*: Investments in next generation networks guarantee higher long-term welfare (dynamic efficiency). Consequently, investment costs need to be included in cost calculations.
- *'Bottom-up' consideration*: Historical cost data, thus 'top-down' consideration, do not reflect the future developments of telecommunications markets. Therefore, current (including hypothetical) costs need to be taken into consideration in cost calculations.

⁶ The effect of higher termination rates on retail prices will be discussed in the following step of the research.

This reasoning corresponds to the standard economic arguments for cost-based pricing in the absence of external effects and fixed costs.⁷ Whether external effects and fixed costs provide a convincing argument for setting regulated termination rates above costs, will be discussed in Chapters 3 and 4. Box 2.1 summarizes the arguments discussed later on.

Box 2.1 The impact of externalities on termination rates

An important result of this report is that externalities form insufficient economic reason for a cost-plus approach such as BULRIC plus to termination rates. In particular, two types of externalities are relevant.

Network externalities mean that a subscriber benefits more as the network size increases, that is, if more users connect to the network. These benefits are not necessarily taken into account when consumers decide to join a growing telephony network. This argument applies for new subscribers within the mobile market (i.e. mobile-to-mobile) but also between fixed and mobile markets (i.e. fixed-to-mobile, FTM). Network externalities are, therefore, commonly seen as underpinning the need for above-cost termination rates to internalize the external network effect. But this externality only exists in markets that have not yet reached full coverage or penetration. In the mature Dutch telecommunications market, the argument of network externalities no longer applies.

Call externalities arises under the calling party network pays principle because only the calling party needs to pay for a call. Call externalities imply that (new) subscribers benefit from being called by either mobile or fixed customers but this benefit is not taken into account by the calling party. In other words, customers call less as they do not account for the additional benefits of receivers. If the per minute price is higher due to higher termination rates, the number of calls reduces further. This problem is particularly relevant in the light of the different level of fixed and mobile termination rates. As FTRs are substantially lower than MTRs, (potential) mobile customers benefit less from the presence of the FTM interconnection and thus the possibility to be called by fixed customers. Consequently, call externalities can be better internalized under lower termination rates. Call externalities are however less relevant in a network with full penetration.

2.4 Recommendation of the EC

The recommendations of the European Commission are in line with the previously described literature. According to EC (2009, Points 1 and 2):

“1. When imposing price control and cost-accounting obligations in accordance with Article 13 of Directive 2002/19/EC on the operators designated by National Regulatory Authorities (NRAs) as having significant market power on the markets for wholesale voice call termination on individual public telephone networks (hereinafter referred to as ‘fixed and mobile termination markets’) as a result of a

⁷ Other relevant characteristics of telecommunications have effects on other elements of regulating termination rates. For instance, *call externalities* play a role when ‘calling party network pays’ principle is applied in comparison to a ‘bill-and-keep’ system. *Heterogeneous calling patterns* and *tariff-mediated network effects* in a not fully penetrated market play a role in the discussion about symmetric versus asymmetric regulation.

market analysis carried out in accordance with Article 16 of Directive 2002/21/EC, NRAs should set termination rates based on the costs incurred by an efficient operator. This implies that they would also be symmetric. In doing so, NRAs should proceed in the way set out below.

2. It is recommended that the evaluation of efficient costs is based on current cost and the use of a bottom-up modelling approach using long-run incremental costs (LRIC) as the relevant cost methodology.”

However, there are several methods that are based on bottom-up long run incremental costs. The main discussion between stakeholders relates to which costs should be included as incremental costs in the wholesale call termination market. In particular, the current debate in the European as well as the Dutch regulatory platform concerns two models, namely the pure BULRIC and BULRIC plus methods. According to the *pure BULRIC* model, only avoidable costs relating to call termination need to be considered as incremental. Avoidable costs of a service are only those costs that would disappear if the network operator stops providing that service; in this case, call termination. According to the *BULRIC plus* method, other, unavoidable fixed costs also need to be included as traffic related costs to provide sufficient incentives to invest.

According to the European Commission, tariff regulation based on pure BULRIC is the only effective measure to counteract the competition problem due to market power for call termination and to provide investment incentives (EC 2009, Point 6):

“Within the LRIC model, the relevant increment should be defined as the wholesale voice call termination service provided to third parties. This implies that in evaluating the incremental costs NRAs should establish the difference between the total long-run cost of an operator providing its full range of services and the total long-run costs of this operator in the absence of the wholesale call termination service being provided to third parties. A distinction needs to be made between traffic-related costs and non-traffic-related costs, whereby the latter costs should be disregarded for the purpose of calculating wholesale termination rates. The recommended approach to identifying the relevant incremental cost would be to attribute traffic-related costs firstly to services other than wholesale voice call termination, with finally only the residual traffic-related costs being allocated to the wholesale voice call termination service. This implies that only those costs which would be avoided if a wholesale voice call termination service were no longer provided to third parties should be allocated to the regulated voice call termination services.”

The Body of European Regulators for Electronic Communications (BEREC),⁸ as well as the Dutch regulator, OPTA,⁹ agrees with this recommendation.

2.5 Discussion in the Netherlands

In line with the European recommendation, OPTA imposed the pure BULRIC method in its 2010 decision on the following grounds:¹⁰

⁸ BEREC (2011), Section 4.3.1.

⁹ OPTA (2010).

¹⁰ OPTA (2010).

- Tariff regulation based on the argument that pure BULRIC is the only effective measure to counteract the observed competition problem due to market power for call termination causing high termination tariffs and margin squeeze.
- An important economic cost of the competition problem for call termination is the inefficient retail tariff structure: High termination rates are costs that are imputed in retail prices. This distorts efficient operation of the retail market. A more efficient tariff structure implies a lower tariff per minute offset by a higher price for the subscription. This will have an impact on the demand for fixed and mobile telecommunication services;
- Only a pure BULRIC tariff will fix the problem of an inefficient retail tariff structure. It also prevents the problem of double marginalization: the double counting of fixed costs, which are first included in the wholesale tariff and subsequently imputed in retail tariffs, which means that the end user pays twice for the same cost. A pure BULRIC tariff generates the highest net surplus in terms of economic welfare;
- A pure BULRIC tariff, therefore, complies with the requirements of the Telecommunications Act. It is directly aimed to solve the competition problem of high tariffs for call termination. Further the measure is a proportional solution:
 - It serves the interests of end users and, in fact, provides maximum benefit for consumers. This is in agreement with the demand of European regulation, according to OPTA;
 - The alternative, a plus BULRIC model, does not solve the problem of double marginalization;
 - A pure BULRIC tariff does not impose unreasonable costs on the industry.

The Dutch Trade and Industry Appeals Tribunal (*College van Beroep voor het bedrijfsleven*, CBb) systematically assessed these arguments on the basis of the legal obligations of the Dutch Telecommunications Act.¹¹

Box 2.2 Legal framework: The Dutch Telecommunications Act

The Telecommunications Act defines the goal for intervention and the conditions which must be met by the intervening authority, OPTA.

Article 1.3.1

OPTA ensures that its decisions contribute to the realization of the goals as formulated in (...) Directive nr. 2002/21/EG by:

- a. Promoting competition in the supply of electronic communication networks, electronic communication services (...) amongst others by supporting efficient infrastructure investment and the promotion of innovation;
- b. The development of the internal market;
- c. The promotion of the interests of end users in terms of choice, price and quality.

Article 6a.2/1

If an investigation as intended in article 6a.1.3/4 shows that the relevant market is not competitive, OTPA will determine which companies, supplying electronic communication

¹¹ College van Beroep voor het bedrijfsleven (CBb), LJN: BR6195, 31 August 2011.

networks and electronic communication services, should be classified as an undertaking with significant market power, and (a) imposes on these firms an appropriate obligation (...).

Article 6a.2/3

An obligation is appropriate if it is based on the nature of the competitive problem obstructing this market and can be considered proportional and justified given the objectives of article 1.3 Telecommunications Act.

Article 6a.7

1. On the basis of article 6a.2/1, for specific forms of access, OPTA can impose an obligation to control access tariffs and costs, if a market analysis shows that the lack of competition allows the undertaking in question to maintain prices at an excessive level and squeeze the margins of competitors, in both cases to the disadvantage of end users.
2. This obligation can involve using a cost-oriented tariff for access or using a method for determining cost-oriented tariffs for access, to be approved by OPTA.

The CbB agrees with OPTA that the legal conditions of articles 6a.2/1 and 6a.7 Telecommunications Act are met (see Box 2.2). According to CbB, OPTA has sufficiently demonstrated the existence of a competitive problem for call termination: without regulation, prices are excessive and margin squeeze is an acute problem. CbB also acknowledges that tariff regulation is an appropriate measure in the sense that it constitutes a remedy for the competition problem at hand, thus satisfying in part the conditions of Article 6a.2/3. CbB states that alternative and less imposing remedies would not have been sufficiently effective measures. Furthermore, CbB argues that both pure BULRIC and BULRIC plus provide *effective tariffs* as both tariffs are cost-oriented and hence satisfy all legal obligations.

A critique of CbB is whether the type of tariff regulation proposed by OPTA can also be considered a *proportional* remedy given the objectives of Article 1.3 of the Telecommunications Act. This is the main issue at stake, as explicitly stated by the CbB in its ruling.¹²

At this point, a disagreement between OPTA and CbB surfaces. OPTA argues that its regulation should maximize consumer benefit, a point also mentioned in Directive 2002/21/EG. But CbB does not consider consumer welfare an absolute objective. According to CbB, OPTA should strive to balance the interest of all parties concerned: consumers *and* different types of producers. For CbB this forms an important precondition for the goal to further the benefits of consumer. In sum, consumer welfare forms a relative goal and should not be maximized at all costs.

This disagreement about the objectives of regulation explains the final ruling of CbB. According to CbB, the welfare calculations of OPTA show that a plus BULRIC tariff also improves consumer welfare, while generating lower costs to the industry.¹³ In the eyes of the Tribunal, it also forms a cost-oriented tariff and hence satisfies all legal obligations. In that sense it should be preferred over a pure BULRIC methodology, because it forms a less demanding type of regulation. In other words, as a remedy pure BULRIC is *not* proportional to the problem at hand.

¹² See Point 4.8.3.1 of the CbB ruling.

¹³ These are the calculations in the market analyses of OPTA (2010), pp. 174-175.

The final outcome is that OPTA is called upon to study the appropriateness of plus BULRIC as a remedy for the competitive problems for call termination.

On the grounds of proportionality given the objectives of Article 1.3 of the Telecommunications Act, CBB has required from OPTA to use the BULRIC plus model (see point 4.8.3.1 of the CBB ruling). Arguments on proportionality will be discussed in the next step of the research.

2.6 What is effectiveness and proportionality?

As can be seen from the discussion between OPTA, EC, and BEREC on the one hand and the CBB on the other hand, there is a disagreement about what an effective and proportional measure means. In the following sections, definitions are provided that are consistent with the standard economic literature.

2.6.1 Definition of effectiveness

From a welfare perspective, it is desirable that regulated termination rates reduce market power to the minimum and provide sufficient investment incentives. To eliminate market power in wholesale call termination, regulated tariffs should be as low as possible. At the same time, regulated tariffs need to provide investment incentives in the call termination market. A cost calculation model is seen effective if it achieves these goals. In Chapter 3, arguments for and against the effectiveness of pure BULRIC and BULRIC plus are presented.

2.6.2 Welfare consideration as a basis for proportionality

A termination rate based on a BULRIC model is proportional if the following two welfare criteria are fulfilled. First, a termination rate needs to create the highest possible level of welfare, which is the sum of consumer surplus and producer surplus. Total welfare increases if no value is lost in the market. In that sense, an increase in welfare can be seen as a volume effect: if customers call less, for instance as a result of higher prices, then the value of this lost minutes for consumers and firms is a pure welfare loss (also called deadweight loss).

Second, a BULRIC model is proportional if it distributes welfare proportionally between consumers and the providers of telephony services in the same market and between the mobile and fixed market. The underlying notion is that market power in call termination generates a welfare cost to society. The reduction of this market power by means of regulation should not cost more in terms of economic welfare than its pay-off in terms of increased consumer welfare. In case of a positive benefit-cost ratio this report considers the regulation proportional. The cost involved is clearly the reduction in producer surplus (profit) for mobile network operators. Consumers and some providers gain in terms of welfare when regulation enforces lower termination rates.

This notion of proportionality is derived from the Hicks compensation criterion in welfare economics. This criterion takes into account the distribution of welfare gains and losses. A strict welfare gain benefiting all consumers and producers is not always feasible and forms a very

stringent condition on the regulator.¹⁴ For a Hicks-efficient welfare improvement it suffices to generate a net welfare gain summing up gains and losses for the different parties involved.

In short, a proportional regulatory measure needs to take the level of total welfare *and* the distribution of welfare between market players into consideration. In Chapter 4, arguments around the proportionality of pure BULRIC and BULRIC plus are discussed.

2.7 Conclusions

Regulating termination rates is necessary as network operators have market power over their infrastructure when a call terminates in their network. In Europe, price cap regulation is applied. However, there is a discussion between stakeholders which costs should be included as incremental costs in the wholesale call termination market. In particular, the current debate in the European as well as the Dutch regulatory platform concerns two models, namely the pure BULRIC and BULRIC plus methods. According to the *pure BULRIC* model, only avoidable costs relating to call termination need to be considered as incremental. Avoidable costs of a service are only those costs that would disappear if the network operator stops providing that service; in this case, call termination. According to the *BULRIC plus* method, other, unavoidable fixed costs also need to be included as traffic related costs to provide sufficient incentives to invest.

According to the European Commission, tariff regulation based on pure BULRIC is the only effective measure to counteract the competition problem due to market power for call termination and to provide investment incentives. The Body of European Regulators for Electronic Communications (BEREC), as well as the Dutch regulator, OPTA, agrees with this recommendation. The Dutch Trade and Industry Appeals Tribunal (*College van Beroep voor het bedrijfsleven*, CBb) argues that both pure BULRIC and BULRIC plus provide *effective tariffs* as both tariffs are cost-oriented and hence satisfy all legal obligations. According to CBb, the welfare calculations of OPTA show that a BULRIC plus tariff also improves consumer welfare, while generating lower costs to the industry. As BULRIC plus also forms a cost-oriented tariff and hence satisfies all legal obligations, it should be preferred over a pure BULRIC methodology. According to CBb, as a remedy pure BULRIC is *not* proportional to the problem at hand.

What is an effective and proportionate measure? Both concepts relate to the welfare implication of different tariffs. For effectiveness, it is desirable that regulated termination rates reduce market power to the minimum (static efficiency) and provide sufficient investment incentives (dynamic efficiency). In addition, a termination rate is proportional if achieves the highest possible welfare (volume effect) and distributes welfare proportionally between consumers and the providers of telephony services in the same market and between the mobile and fixed market (distribution effect).

The economic literature extensively analyzes why termination rates need to be regulated and provides a guideline for welfare analysis. In this chapter, the reasons for and the general principles of regulation are mapped. In the presence of calling party network pays principle,

¹⁴ In the economic literature this type of welfare gain would be called Pareto efficient.

several arguments support the idea that unregulated network operators have incentives to charge termination rates above costs:

- Mobile network operators are monopolies in the call termination market and can charge extensively high termination tariffs.
- Furthermore, they collude in order to raise each other's termination costs. This claim holds in the case of linear prices.
- Network operators can avoid heavy users with high termination costs and attract light users with high termination revenues. This claim holds when prices are not discriminated based on where calls terminate (i.e. on-net vs. off-net prices).
- Some arguments are presented why network operators would choose below-cost termination rates. However, this behavior has not been observed in practice.

For these reasons, regulation is justified. The literature on wholesale call termination sets the basic principles for regulation. According to that, the efficient regulated termination rate – in short term as well as long term – needs to be based on bottom-up long run incremental costs. Any deviation from this method needs to be further justified. The presence of externalities might be a reason for that. The following chapters provide arguments why for the Dutch market this deviation is not justified.

3 Effectiveness of BULRIC models

From a welfare perspective, it is desirable that regulated termination rates reduce market power to the minimum and provide sufficient investment incentives. Termination rates based on both BULRIC models provide investment incentives, but BULRIC plus cannot reduce market power to the minimum the way pure BULRIC does. Therefore BULRIC plus is not an effective regulatory measure.

To eliminate market power in wholesale call termination, regulated tariffs should be as low as possible. At the same time, regulated tariffs need to provide investment incentives in the call termination market. A cost calculation model is seen effective if it achieves these goals. In this chapter, arguments for and against the effectiveness of pure BULRIC and BULRIC plus are presented.

In Section 3.1, cost elements are briefly presented that need to be considered in regulation. Then in Section 3.2, justification is given for which cost elements are correct in such a cost calculation. Sections 3.3. and 3.4 present the negative consequences of a higher termination rate such as determined by BULRIC plus. Section 3.5 concludes the results.

3.1 Which cost elements may play a role?

The main discussion in the policy debate is about which cost elements need to be included as incremental in cost calculations. According to the pure BULRIC model, only avoidable costs relating to call termination need to be considered as incremental. Avoidable costs of a service are only those costs that would disappear if the network operator stops providing that service; in this case, call termination. For instance, the marginal cost of terminating a call, costs of investment in capacity to transport additional wholesale traffic or unavoidable business overhead costs fall into this category. According to the *BULRIC plus* method, other, unavoidable fixed costs also need to be included as traffic related costs to provide sufficient incentives to invest (EC 2009).

For OPTA, Analysys Mason compiled a model that calculates these costs (Analysys Mason 2010a,b). Cost elements that relate to call termination – for instance, the costs of interconnection, maintenance, wholesale billing systems – have effects on capital costs (CAPEX), operational costs (OPEX), and gross replacement costs. The exceptions are costs that relate to retail services, but also several fixed cost-elements, such as overheads or licenses. The discussion between stakeholders relates in particular to the latter group of costs. Section 3.2 provides economic arguments under which circumstances these costs can be included in cost calculations.

3.2 Justification for including fixed costs fails

Regulatory measures that are based on incremental costs are often seen as the most effective instrument to reduce market power to the minimum. This standard result of the economic literature holds if fixed, common or coverage costs and externalities can be covered by other than regulated prices or are simply not present. Otherwise, these costs might be part of regulated rates.

As it is shown in the next sections, no evidence can be presented for the inclusion of fixed, common and coverage costs (i.e. in favor of BULRIC plus) in the Netherlands and only a model that is based on incremental costs (i.e. pure BULRIC) can be seen as effective.

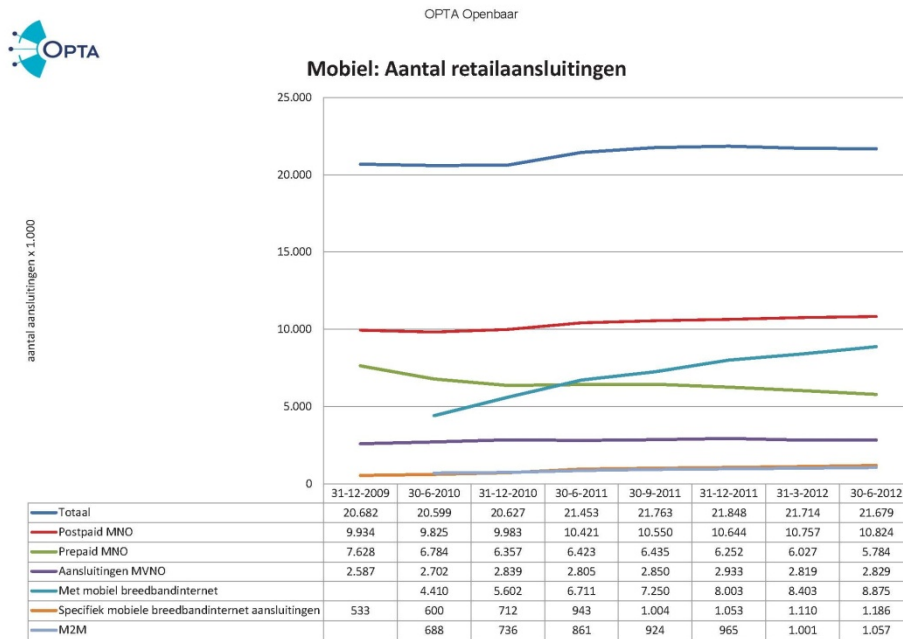
When fixed costs are present, tariffs cannot only take variable costs into consideration, otherwise these costs are not recovered and the firm makes a loss. The discussion between regulators and network operators is about which tariff needs to cover these costs. Regulators believe that these costs are not traffic dependent and need to be taken into consideration only in the retail prices (e.g. Ofcom 2011). On the contrary, network operators argue that in addition, common costs, particularly overhead and coverage costs increase as networks become larger (ETNO 2010). In particular, these costs can be part of termination rates when externalities are present. Furthermore, fixed costs need to be shared between services according to their price sensitivity: the less price sensitive a product is (e.g. wholesale in comparison to retail or voice in comparison to data), the more it can bear from the fixed costs. Consequently, these cost elements should be seen as incremental costs and be considered in the termination rate. In the following sections, economic arguments – taking the specifications of the Dutch market into account – are presented to put an end to this discussion.

3.2.1 Inclusion of overhead and coverage costs is not justified

Coverage and overhead costs can be seen as incremental costs as long as these costs increase with the network size. As long as networks can expand, customers can benefit from a larger subscriber-base; network externalities are present. Network operators can internalize these external effects by setting a higher termination rate. How does this argument work? Is this a relevant argument for the Dutch telecommunication markets?

Network externalities mean that a subscriber's benefits increase if more users are connected to the network (see also Box 2.1). For termination rates, these effects are particularly relevant in the context of fixed-to-mobile termination. In this case, network externalities reflect the benefits of both fixed and mobile subscribers as mobile penetration increases because subscribers are able to call more people. These benefits are not necessarily taken into account when consumers decide to join a growing mobile network. Therefore, from a welfare perspective, mobile operators that build up their network coverage can be allowed to benefit by charging a higher termination rate (Armstrong 2002, Valletti & Houpis 2005). However, if network externalities are not present, cost-based tariffs provide maximum welfare (Armstrong 2002). Network externalities are present until the market reaches full penetration. As penetration has reached 100 percent and hardly increases in the Netherlands (126 percent in Q3 2012, see Figure 3.2), network externalities can be assumed to no longer exist and therefore cannot form an argument for the inclusion of coverage and overhead costs.

Figure 3.2 Mobile penetration in the Netherlands has reached 100 and stopped increasing



Op basis van gegevens van KPN, T-MOBILE en VODAFONE. Op basis van vragen 1_A_2_1-2-4-5-6-7 en 1_B_5_2-7-8 van de SMM.

Source: OPTA (Openbare rapportage mobiel Q2 2012)

3.2.2 Inclusion of common costs and other fixed costs is not justified

When fixed and common costs are present, tariffs cannot only take variable costs into consideration, otherwise these costs are not recovered and the firm makes a loss. When such tariffs are regulated, the so-called Ramsey-pricing principle leads to the socially most desirable prices (Laffont & Tirole 2000). According to Ramsey-pricing, a multi-product firm, such as mobile operators, which provide wholesale and retail services, may contribute a part of its fixed and common costs to products that are not price sensitive. Wholesale call termination is seen as a price inelastic service as it cannot be avoided if a subscriber calls another subscriber (Ofcom 2011). Therefore, the Ramsey principle may provide an argument for including fixed and common costs in termination rates.

However, several problems arise with Ramsey pricing. First of all, Ramsey-pricing is not practical as it requires regulators to obtain precise *information* about the demand and elasticities of all services. In practice, regulators generally do not have this information. Secondly, the Ramsey principle leads to efficient regulated tariffs if all tariffs are *linear* (Laffont & Tirole 2000). In practice, this is again not the case, as network operators use several retail pricing model, such as two-part tariffs (e.g. post-paid packages). The reason for this pricing scheme is exactly to recover fixed costs by charging fixed fees while allowing the retail prices to stay as low as possible and thus close to the efficient level (Jeon & Hurkens 2008). Consequently, wholesale services are provided efficiently, while competition for customers is driven by fixed fees. A similar argument holds for non-linear tariffs (e.g. packages for different amount of called minutes at decreasing per minute prices).

To conclude, implementing Ramsey-prices is difficult due the necessary information. And even if Ramsey-pricing is implementable, it is not applicable, as two-part and non-linear tariffs can more effectively recover fixed costs than termination rates. Therefore, Ramsey-pricing, based on different elasticities of services, is not an argument for including these costs in cost calculation.

Note: Data services are seen as a relevant cost-driver of future telecommunication networks (Jullien et al. 2010). Furthermore, it seems a complement, not a substitute of voice services. Therefore, common and fixed costs need to be related to both products. As voice is less price elastic than data (Ofcom 2011), network operators have a preference to direct more costs to voice services. However, as discussed just before, this should not be an argument for loading these costs on wholesale call termination. Instead, it can be an argument for innovative retail pricing schemes for voice and data that balance well the difference in price elasticities and may subsidize the necessary investments for data services. Moreover, it needs to be clarified by regulators how data traffic, in particular some online services such as VoIP, affects the competitiveness of mobile networks. These topics are, however, beyond the scope of this research.

3.3 BULRIC+ leaves firms with market power

Cbb argues that both pure BULRIC and BULRIC plus are effective measures. However, including fixed costs, particularly unjustified, such as proposed by BULRIC plus, shifts the focus from a marginal (or incremental) cost consideration towards an average cost consideration. This switch gives rise to a mark-up above the efficient cost level in the long run and thus causing excess producer surplus for network operators.

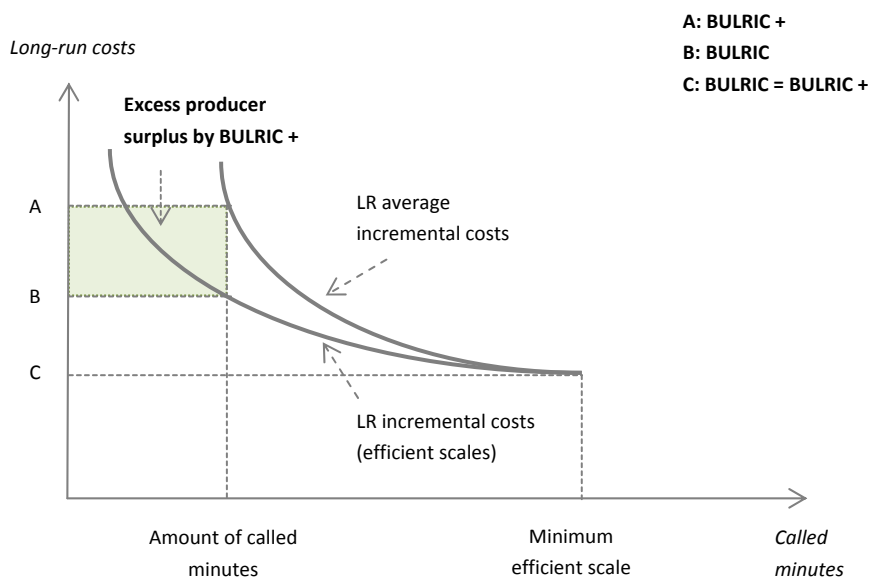
This situation is explained by Figure 3.3. In the figure, long-run incremental (marginal) and long-run average incremental costs are illustrated as a function of called minutes. Both cost measures are expressed as per minute costs. Long-run average incremental costs include avoidable costs *and* fixed costs. The long-run incremental costs include *only* avoidable costs. Furthermore, the long-run incremental cost function represents the efficiency, which can be achieved at a given amount of called minutes, that is, at a given level of investment. Termination rates corresponding to this cost level are the lowest possible, resulting in the highest welfare in the termination market.

As a result of investments in network extension or upgrade, increasing efficiency can be expected in the long run. Therefore, both curves are decreasing (i.e., represent lower marginal and average costs) as the number of called minutes increases. In the very long run, however, production reaches the minimum achievable efficient scale (point C) and these two costs will be equal to each other.¹⁵ Increasing the number of called minutes further to this point by means of investments is not profitable any more (the curves present only those situations where efficiency gains can be achieved). As the figures show, in the range of investments where higher efficiency can be achieved, average costs are higher than marginal costs.

¹⁵ This reasoning partly reflects the argument of Anlys Mason (2010c).

These figures can be translated into the two cost calculation methods. Point A reflects tariffs according to BULRIC plus and point B reflects tariffs according to pure BULRIC. The cost difference between points A and B is clearly a benefit for network operators. It can also be seen as an additional mark-up on the efficient cost level, implying that the market power of firms is not reduced to the minimum achievable level. Given the current amount of called minutes,¹⁶ this cost difference or mark-up results in excess producer surplus in the wholesale market, illustrated by the green rectangle in the figure. Therefore, with BULRIC plus, the market power of network operators and thus producer surplus cannot be reduced to the minimum in the call termination market. As a consequence, BULRIC plus is not an effective regulatory measure.

Figure 3.3 Producer surplus in the call termination market cannot be reduced to the minimum by BULRIC plus



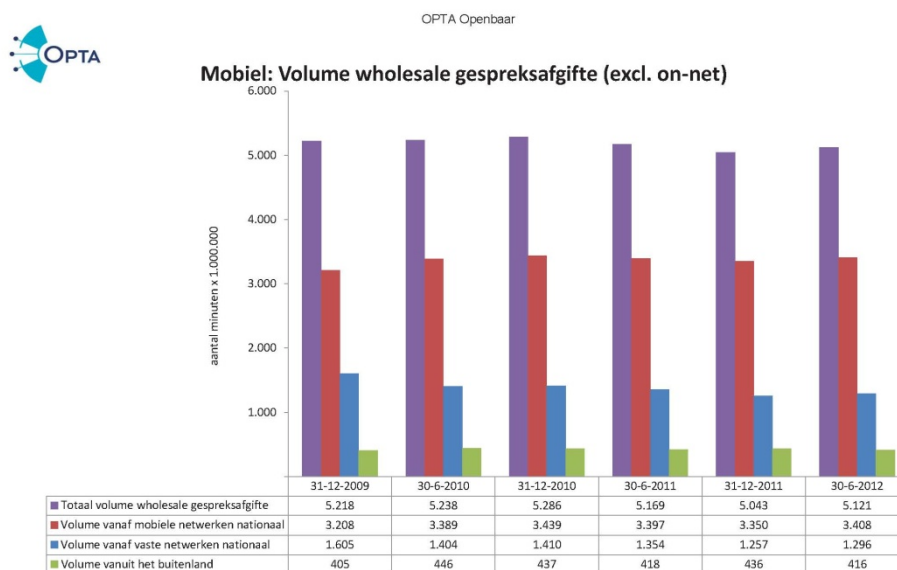
Source: SEO Economic Research; based on Varian (2006)

Three things need to be noted here. First, the green rectangle represents the excess producer surplus above pure BULRIC in the wholesale call termination market. Producer surplus is not necessarily undesirable, particularly, when it provides finances for the monopolist to invest. However, there is no justification that all investments need to be recovered in the wholesale market (see Section 3.2). In this case, the mark-up above efficient costs may lead to competition distortions in the wholesale markets as it will be explained in the following section.

Second, the producer surplus and markup in the wholesale market may be carried on to the retail market and consequently imply higher prices and producer surplus there (double marginalization). The exact link between wholesale and retail prices will be analyzed in Chapter 4.

¹⁶ As it will be argued in Section 4.3.2, higher termination rates lead to less called minutes. This decrease in called minutes influences how much surplus network operators make in the wholesale market. However, independently of the change in call minutes, BULRIC plus creates surplus in the wholesale market as it deviates from incremental costs.

Figure 3.4 The volume of terminated minutes has been stable since 2011



Op basis van gegevens van KPN, T-MOBILE en VODAFONE. Op basis van vragen 1_C_8_1 t/m 1_C_8_3 van de SMM.

Source: OPTA

Finally, when the minimum efficient scale is achieved by investments (point C), pure BULRIC and BULRIC plus result in the same costs, thus termination rates. In the Netherlands, the number of called minutes has recently stopped increasing (see Figure 3.4). Theoretically, it could mean that it is not possible to profitably increase the number of called minutes by additional investments. Therefore, this may indicate that the market of call termination has reached or is close to its minimum efficient scale. In that sense, pure BULRIC and BULRIC plus provide the same termination rate. However, this equal termination rate reflects an ideal situation when stakeholders, including the regulator, have complete information about costs and demand. In practice, this is not the case. Regulators have information disadvantage about the costs of network operators. This asymmetric information often provides incentives not to report the true costs of operation but higher costs. Consequently, regulated termination rates can be higher than under complete information. This is an argument why tariffs based on the two methods also differ from each other. For BULRIC plus tariffs, more cost elements are included that are unknown to the regulator than for pure BULRIC. As a higher mark-up is more attractive, firms have incentives to report higher costs for all cost elements. Consequently, the difference between the actual tariff and the theoretical one is higher for BULRIC plus than for pure BULRIC. Pure BULRIC has the advantage for a regulator that the chance for the abuse of information asymmetry is smaller.

3.4 Negative effects of BULRIC+ in wholesale markets

If cost elements that are not justified are included in cost calculations, such as might be the case for BULRIC plus, termination rates will be higher than if only avoidable costs are seen as incremental. Therefore, BULRIC plus contains a mark-up above the efficient cost level and hence cannot reduce market power to the minimum. Consequently, BULRIC plus is not an effective instrument.

There are two negative consequences of this mark-up and thus higher termination rates in call termination markets:

- Higher termination rates distort competition between asymmetric mobile operators.
- Asymmetric cost measures between European countries create competitive distortions in the internal market at the cost of countries applying the pure BULRIC model.

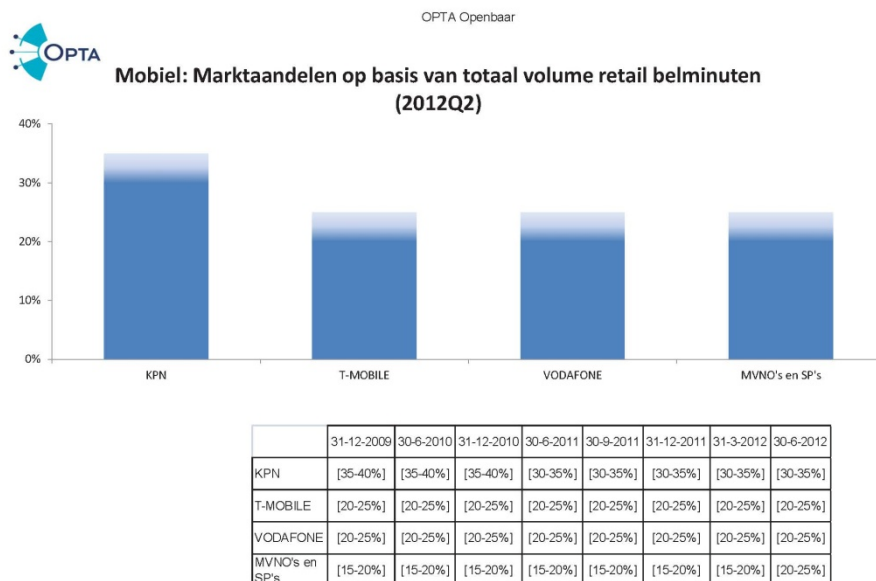
These are two disadvantages of the BULRIC plus methodology not included in the decision of the CBB. In the following sections, the economic reasoning behind these negative effects will be explained in detail.

3.4.1 BULRIC+ creates competitive distortion between service providers

Higher termination rates lower competition between services providers that have a larger subscriber-base (incumbent network operators) and those that have fewer subscribers (entrants; see Ofcom 2011). This asymmetry of subscriber base is present in the Dutch mobile market (see Figure 3.5).

Mobile service providers with fewer subscribers can compete less effectively with larger operators and may not be able to reach the minimum efficient scale. The reason for this is that they bear the costs of off-net calls that are larger in proportion to their on-net calls. Consequently, incentives to invest in innovative services by service providers without networks might be eroded. The overall effects of higher MTRs on competition between service providers depend on how these higher rates influence retail prices. This mechanism will be discussed later.

Figure 3.5 The market share of each network operator approximates the total market share of other service providers



Op basis van gegevens van KPN, T-MOBILE en VODAFONE. Op basis van vragen 1_A_3_2, 1_B_5_4, 1_C_9_1-2-3, 1_D_3_1, 1_B_6_4 en 1_D_4_1-2-3 van de SMM.

Source: OPTA

3.4.2 Dutch BULRIC+ creates competitive distortion in internal markets

Among others, the EC considers the evolution of internal markets as a future goal when providing European regulatory recommendations. Deviating from the recommendation and applying different measures in the member states can create competitive distortions between European network operators.

For similar reasons as before, network operators in member states that can set termination rates by BULRIC plus are allowed to charge a mark-up over efficient costs. This implies higher termination costs to those operators that set prices according to pure BULRIC. Concerns about such competitive distortions between network operators in member states are presented by the EC and BEREC (BEREC 2012a). Again, the extent of distortions depends on the asymmetries in terms of costs, market shares, and volume in the European markets.

3.5 Conclusions

This chapter analyzed the effectiveness of regulatory models for determining termination rates. A regulatory method is effective if it reduces market power to the minimum in the wholesale market and still provides incentives to invest in call termination.

The most important discussion point relates to which cost elements need to be considered as incremental in cost calculations. According to the pure BULRIC model, only avoidable costs

relating to call termination need to be considered as incremental. Avoidable costs of a service are only those costs that would disappear if the network operator stops providing that service; in this case, call termination. According to the BULRIC plus method, other, unavoidable fixed costs also need to be included as traffic related costs to provide sufficient incentives to invest.

Justification for including fixed and common costs in cost calculation fails for the following reasons:

- Firstly, network externalities are necessary to include overhead and common costs in cost calculation. In a mature market, such as the telephony markets in the Netherlands, network externalities are negligible.
- Secondly, common and fixed cost can be more effectively recovered in the retail market than in the wholesale market. In particular, the fixed fee is an effective measure to recover these costs.

It implies that both BULRIC methods provide sufficient investment incentives but BULRIC plus contains costs which are not justifiable. Therefore, it cannot reduce market power to the minimum and hence it is not effective. Furthermore, higher termination rates, such that are determined by BULRIC plus, have two further negative consequences in the call termination market:

- Higher termination rates distort competition between asymmetric mobile operators.
- Asymmetric cost measures between European countries create competitive distortions in the internal market at the cost of countries applying the pure BULRIC model.

These negative effects are not discussed in the Cbb decision.

4 Proportionality of BULRIC models

Even in the presence of the “reversed” waterbed effect, BULRIC plus termination rates lead to higher per minute prices and, in a mature market, less called minutes. Less volume creates a lost value for the total market, which reduces total welfare. In some situations, BULRIC plus makes firms better off but it happens disproportionately at the cost of consumers in both markets and fixed telephony firms. Therefore, BULRIC plus cannot be seen as a proportional measure.

A termination rate is proportional if it creates the highest possible level of welfare, which is the sum of consumer surplus and producer surplus. Furthermore, it is proportional if it distributes welfare proportionally between consumers and the providers of telephony services in the same market and between the mobile and fixed market. In short, a proportional regulatory measure needs to take total welfare *and* this distribution effect into consideration.

In the analysis of the proportionality of regulatory models, welfare changes and its distribution between market players are analyzed when termination rates increase. The following questions are raised:

- How does a higher termination rate affect retail prices?
- How does the number of subscriptions and called minutes change as an effect of modified retail prices? Is there deadweight loss?
- How does consumer surplus change?
- How does producer surplus change?
- What is the effect of a higher termination rate on total welfare?
- How is any change in welfare distributed between consumers and firms in the same market and between mobile and fixed telephony?

The analysis is based on theoretical and empirical studies about the effects of varying termination rates on retail prices. Several theoretical studies and a few empirical papers assess the presence and the size of the waterbed effect (e.g. Schiff 2009, Genakos & Valletti 2011a,b, 2012, and Baranes et al. 2012). These studies have several limiting assumptions for a robust analysis. To our knowledge, only one study with model calibration (Harbord & Hoernig 2012) exists that releases many of these assumptions and assesses extensively the welfare aspect of higher termination rates.¹⁷ Furthermore, there is no systematical empirical welfare analysis related to termination rates. Even the recommendation of the EC is based primarily on qualitative analysis. The following analysis is based on the review of this stream of literature.

Following the definition of proportionality, changes in the level of termination rates need to be translated welfare concepts. In Section 4.1, the most relevant mechanisms behind welfare changes are explained. As welfare implications strongly depend on retail tariff structures, the most common pricing models are presented in Section 4.2. In Section 4.3, an extensive welfare analysis follows. In Section 4.4, conclusions are drawn.

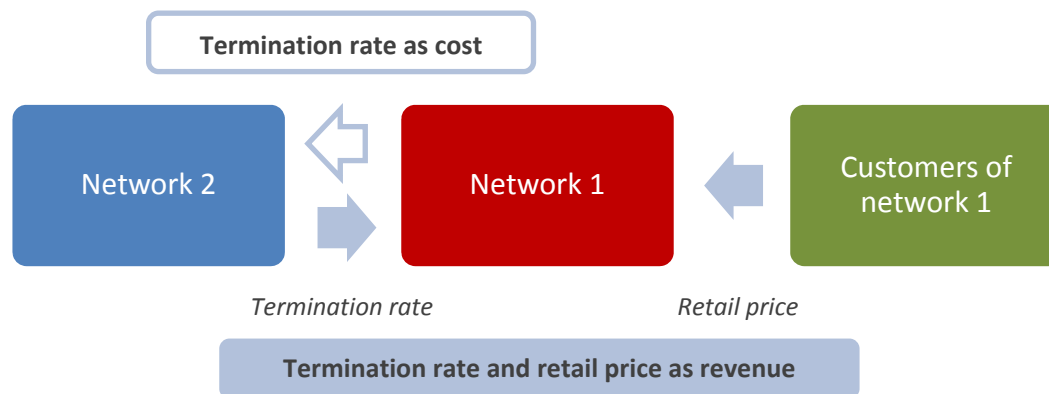
¹⁷ No empirical results, only theory on heterogeneous calling patterns (Jullien et al. 2009) and tariff-based network externalities (Gans & King 2001, Armstrong & Wright 2009). These studies are discussed in Chapter 3. Furthermore, no studies exist that analyze the effects on termination rates on retail prices and welfare in the presence of investments. In the existing studies, fixed costs are in general seen as unchanged.

4.1 Welfare as a basis for proportionality

To determine which method provides a proportional regulatory measure, the welfare consequences of higher termination rates set by BULRIC plus need to be analyzed, both in the mobile and the fixed telephony markets. In standard economic and policy literature, *welfare is defined as the sum of consumer and producer surplus*. Consumer surplus as well as producer surplus is directly linked to retail prices. As retail prices are affected by the level of termination rates, call termination indirectly affects welfare.

In the Netherlands, termination rates are paid by the network of the calling party (*calling party network pays principle*). Therefore, call termination has effects on retail prices. In well-functioning markets, there is a direct link between costs and prices. The lower the costs, the lower the prices become. If there is market power or firms provide services inefficiently, lower costs and retail prices increase welfare. The closer the price is to the marginal cost, the higher the welfare can be. Indeed, termination rates are a cost element for the calling party network (see the perspective of Network 1 in Figure 4.6). Network 1 needs to pay the per minute termination rate to Network 2 after every minute call of its customers that terminates in Network 2. Having the number of called minutes unchanged, the lower the termination rate, the lower costs the calling party network needs to bear and the lower retail prices it will charge at the benefit of consumers.

Figure 4.6 Termination rate functions as a cost and revenue element



Source: SEO Economic Research

Wholesale call termination is, however, characterized by market failures that need to be considered in determining welfare improving measures. Market failures influence the efficient performance of markets. One example of market failure is market power. On the one hand, network operators are monopolies over their infrastructure when calls terminate in their network. Therefore, termination rates are subject to regulation. This type of market power is discussed in the chapter about efficiency. On the other hand, the number of network operators has been limited by the available spectrum. Before networks were open for competition, network

operators had market power also in the retail market. It implied above cost retail prices. The obligation to provide access to services providers (i.e. facilities-based competition) pushed retail prices closer to costs. However, mobile markets are still not fully competitive and have a few operators that for instance in the Netherlands account for 85 percent of the total market (see Figure 3.5). Consequently, retail prices are expected to contain a markup over marginal costs (including termination costs). The analysis of market power in the retail market is however not in the scope of this research.

Another relevant example is related to the two-sided nature of telecommunications (Genakos & Valletti 2012). In a *two-sided market*, two sets of agents meet and interact through a platform, which is the network in telecommunications. In two-sided markets, decisions taken in one side influence decisions to be taken on the other side. In that sense, each side exerts externalities on the other side. This translates to the following mechanism in the call termination market (see again the perspective of Network 1 in Figure 4.6). The two agents are the calling party network (Network 2) and the customers of the called party network, and the platform is the called network itself (Network 1). If the level of termination rates changes, it influences retail prices and though that revenues made in both sides of the market, i.e. termination *and* retail revenues. Why is it so? Termination rates are revenues for the called party network. The lower the termination rates, the lower the revenues Network 1 may receive. As a consequence, Network 1 sets higher retail prices to soften competition for customers and to compensate for the lost termination revenues. This is the so-called *waterbed effect*.

In addition to the waterbed effect, other market failures, such as *call externalities* (or receiver benefits) or (tariff-mediated) *network externalities* also influence how exactly retail prices and welfare change in relation to termination rates (see Box 2.1). In the following section, the functioning of these market failures will also be explained.

How network operators eventually balance between termination rates and retail prices, determines exactly the effects of changed termination rates on *producer surplus*. Producer surplus is created in the wholesale *and* retail markets. Producer surplus in the retail market is shown in Figure 4.7. If retail prices are above marginal costs of providing telephony services (see high price in Figure 4.7), then a higher level of producer surplus is created (see green rectangle). On top of that, network operators can make producer surplus in the wholesale market if termination rate is set above the marginal termination costs. Due to the waterbed effect, a decrease in the surplus in the wholesale (i.e. as a result of lower termination rates) may be compensated by an increase in the surplus in the retail market. The two effects, as is shown in several studies, may eventually cancel each other out.

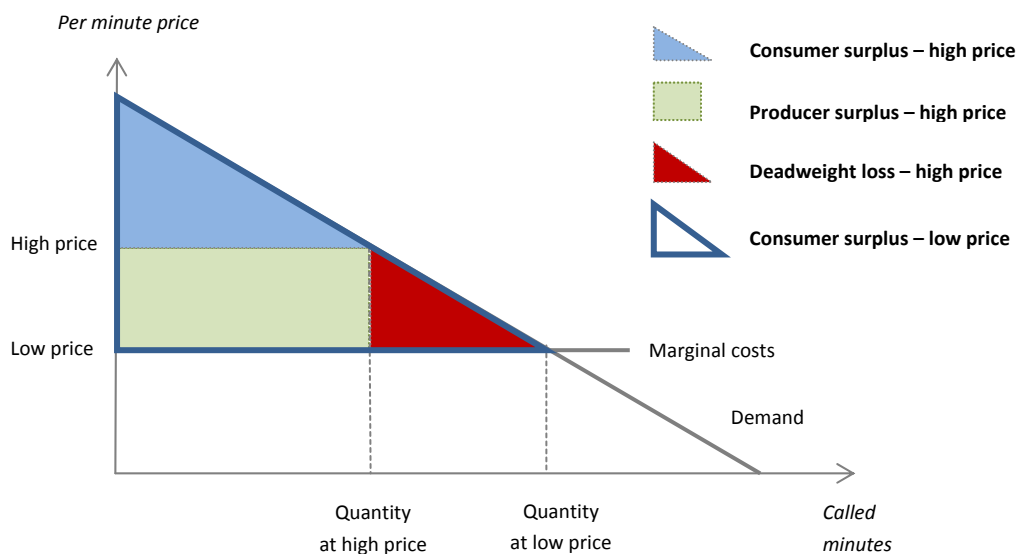
In addition, mobile termination rates have effects on producer surplus in the fixed telephony market. By changing the level of mobile termination rates, the surplus of mobile and fixed network operators may change in different direction and in different proportion. This is a distribution effect between the two markets.

If retail prices increase, consumers (in both markets) buy less telephony services – subscription and called minutes. In this case, *consumer surplus* decreases (in Figure 4.7, from big triangle to small

blue triangle). Similarly to firms, it may occur that consumers in mobile and fixed telephony bear the costs of prices differently. This is also a distribution effect.

Furthermore, the value of a decreased number of called minutes or number of subscriptions is lost, often for both consumers *and* firms. This is the so called *deadweight loss* (red triangle in Figure 4.7). Due to deadweight loss, *welfare* always decreases.

Figure 4.7 Schematic illustration of welfare changes in retail market as a result of higher prices



Source: SEO Economic Research; In the figure, it is assumed that efficient firms do not set mark-up above marginal costs.

In short, a measure that reduces welfare – that is consumer surplus and producer surplus – cannot be seen proportional. Moreover, an instrument is also not proportional if it distributes welfare in a way that some players, consumers or firms in other markets, need to bear more costs.

4.2 Retail tariff structures

Welfare effects strongly depend on the retail tariff structure network operators pursue. In practice, two main tariff structures are applied:

- *Linear or pre-paid tariffs*: In the case of pre-paid packages, customers pay only after called minutes. The monthly number of called minutes and thus the retail revenue of firms can vary substantially between subscribers. This is typical example of *linear prices*. Under linear tariffs, the per minute price is the only tool with which firms compete for customers.
- *Non-linear or post-paid tariffs*: Until recently, post-paid packages contained monthly fixed fees and per minute charges after every called minute. It is the so called *two-part tariff*. If on-net and off-net calls are differentiated, then the tariff structure is called multi-part. More recently, customers pay a monthly flat fee which contains a certain amount of

called minutes.¹⁸ Independently of whether they use these minutes, they pay the fixed fee: by signing a contract, customers commit to a certain amount of payment per month. This is the so called *general non-linear tariff*. Once a customer reaches the number of called minutes in the package, she needs to pay a substantially increased per minute charge afterwards. The common point in two-part and general non-linear tariffs is that the relation between quantity and total payment is non-linear. Under non-linear tariffs, firms have more instruments to compete for customers: monthly payments and per minute prices.

Different retail price structures fulfill different goals of network operators. Linear tariffs serve light user that are usually more often called than they call. They are users that provide, in the first place, termination revenues for networks. Non-linear tariffs aim at heavy users that often call with larger volume. These customers provide higher retail revenues (see Table 4.3).

Table 4.3 Post-paid customers account for almost the total retail revenues

	Post-paid	Pre-paid	Total	% post-paid
The Netherlands: Vodafone only ¹				
Number of subscriber (thousands)	3,538	1,750	5,288	66.9%
Average monthly retail revenue per subscriber (€)	39,20	6,10	-	-
Total monthly retail revenue (€ thousand)	116,035	9,451	125,487	92.5%
The UK: total market ²				
Number of subscriber (thousands)	42,744	39,365	82,109	52.1%
Average monthly retail revenue per subscriber (£)	24.55	6.23	15.67	-
Total monthly retail revenue (£ thousand)	1,049,365	245,244	1,294,609	81.1%

Source: ¹ Vodafone (2013), Vodafone has somewhat more than 25 percent market share within the Dutch MNOs, ² Ofcom (2013), Q3 2012.

These tariff structures have different welfare consequences. Therefore, the welfare effects of higher termination rates in the case of linear and non-linear prices are considered separately.

4.3 BULRIC plus is not a proportional measure

Even though network operators compete for customers by lower fixed fees, higher termination rates lead to higher per minute prices. As a consequence, the number of called minutes decreases, creating the lost values of these unused minutes. Independently of the tariff structure chosen by network operators and the distribution of surplus between consumers and firms, this deadweight loss leads to lower welfare under the BULRIC plus method (see Table 4.4). Therefore, BULRIC plus is seen as a less proportional model than pure BULRIC.

¹⁸ These packages are currently bundles of different services, such called minutes *and* sms, mms, and data. In the analysis, the focus lies on voice services.

Table 4.4 Effects of higher mobile termination rates as a consequence of BULRIC plus in retail prices and welfare

Changes in	Mobile telephony Linear prices (pre-paid packages)	Non-linear prices (post-paid packages)	Fixed telephony (in particular, off-net prices)
Per minute price	↑	↑	↑
Fixed fee	n.a.	↓	n.a.
Number of called minutes	↓	↓	↓
Number of subscriptions	n.a.	No significant effect ¹	No significant effect or ↓ ²
Consumer surplus	↓	↓	↓
Producer surplus	↑	↓	↓↓
Deadweight loss	↑	↑	↑
Welfare	↓	↓	↓↓

Source: SEO Economic Research; ↑: higher, ↓: lower, ↓↓: substantially lower; ¹ In mature markets; ² In mature markets, however, higher per minute prices in fixed telephony may reduce the value of fixed-to-mobile connectivity for mobile subscribers (↓).

4.3.1 Per minute prices increase while fixed fees reduce

In the Netherlands, termination rates are paid by the network of the calling party (*calling party network pays* principle). Therefore, call termination has effects on retail prices. Two relevant mechanisms can be distinguished: the termination rate is a cost element for calling party network and the termination rate is a revenue element for the called party network. Figure 4.6 gives an illustration for these mechanisms.

- *Termination rate as cost:* As a result of higher termination rates, the costs of call termination increase for the calling party network. This network can be a fixed network or a competing mobile network. Higher costs are then translated into higher retail prices. This is the so-called *cost pass-through effect* (see the above relation between Networks 1 and 2 in Figure 4.6).
- *Termination rate as revenue:* Call termination is a two-sided market (Genakos & Valletti 2012; see the perspective of Network 1 below in Figure 4.6). In the two sides of the market are the customers of the called network and the calling party network (Network 2), that make contact with each other via the platform, the called network (Network operator 1). Decisions taken in one side influence decisions to be taken on the other side. In that sense, each side exerts externalities on the other side. As call termination is regulated, Network operator 1 has incentives to balance revenues from call termination by retail revenues. This is the so-called *waterbed effect* (Schiff 2009). As a consequence of higher termination rates, a “reversed” waterbed effect applies. The higher the termination rates, the higher the revenues the called party network may receive. Network operator 1 then sets lower retail prices to keep customers in and receive those termination revenues after these customers.

At the first sight, these mechanisms work opposite to each other. The resultant of these two effects depends, however, on the retail price structure. First, the total effects in the mobile market are analyzed and then the same effects in the fixed market.

In the mobile market, for *linear tariffs*, such as pre-paid packages, the per minute price is the only instrument with which the network operator can balance between termination costs and

revenues. Even though the “reversed” waterbed effect would drive down the retail price to attract more customers, the network operator increases the per minute price slightly to be able to recover termination costs (see empirical evidence in Growitsch et al. 2010).

Furthermore, the per minute price can also be used as an instrument to recover fixed costs. As termination rates based on BULRIC plus have already partly contained fixed costs, influencing retail prices further by fixed costs may lead to double marginalization (see argument of OPTA 2010).

In the case of *non-linear tariffs*, such as post-paid packages, the network operator has two instruments to balance termination costs and revenues by retail revenues. A well-known result in the literature of telecommunications is that network operators set the per minute price as low as possible, which is often the cost-efficient level, and charge fixed fee as high as possible to extract total consumer surplus (Laffont & Tirole 2000). As termination rates are costs, the per minute price will include these cost. The higher the termination rate, the higher the per minute price will be (empirical evidence: Andersson & Hansen 2009, Veronese & Pesendorfer 2009).

However, due to the “reversed” waterbed effect the called party network has incentives to ease competition and hence gain as many customers as possible by a lower fixed fee (see empirical evidence of Genakos & Valletti 2011a, b). By doing so, it can guarantee higher termination revenues. This result holds even when firms differentiate per minute prices based on call termination (Jullien et al. 2010). In this case, a network operator can achieve higher termination revenues by a larger market share (tariff-based network externalities).

In addition, since the fixed fee is used to cover fixed costs, the per minute price and hence the termination rate can be held as low as possible (see arguments in Chapter 3 about effectiveness). In this way, double marginalization can be avoided. In other words, the impact of the waterbed effect is differentiated depending on the tariff structure.

For a fixed telephony company, call termination is also a cost and a revenue element. The firm needs to pay the mobile termination rate (MTR) after every minute terminating in mobile networks and receives the fixed termination rates (FTR) after every called minute from mobile phones terminating in its fixed network. Even though the same regulatory model applies for both telephony firms (i.e. reciprocity), MTRs are currently more than six times higher than FTRs (see Table 4.5; the difference decreases over time). Therefore, a fixed telecom company needs to set a much higher per minute retail price than a mobile company to cover termination costs. If termination-based price discrimination is allowed, it implies a substantially higher fixed off-net price than on-net price.

It also needs to be noted that a shift from pure BULRIC to BULRIC plus implies a proportionally smaller increase in termination rates for fixed companies (from 0.36 to 0.37 eurocent) than for mobile companies (from 1.2 to 2.4 eurocent). Consequently, the disadvantage of fixed companies also increases from a shift to BULRIC plus. Under the pure BULRIC method, MTRs are “only” three times higher than FTRs implying also somewhat lower fixed-to-mobile retail prices than under BULRIC plus.

Table 4.5 Mobile termination rates (MTR) and fixed termination rates (FTR) in the Netherlands

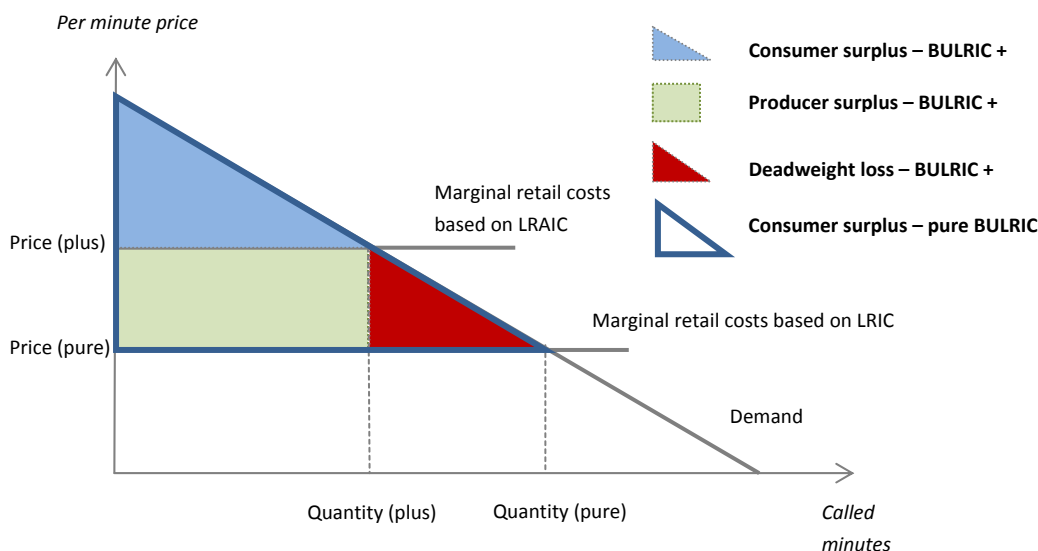
Eurocent per minute	2011 H2	1-1-2012 ¹ /1-9-2011 ² to 1-9-2012	As of 1-9-2012 (BULRIC plus)	As of 1-9-2012 (pure BULRIC)
	Source: OPTA (2012)	OPTA (2012)	OPTA (2012)	OPTA (2010)
FTR	0.72	0.72	0.37	0.36
MTR	4.2	2.7	2.4	1.2

Source: OPTA (2010, 2012); ¹ For fixed telephony; ² For mobile telephony

4.3.2 Called minutes reduce and deadweight loss is created

As the per minute price increases, the *number of called minutes* decreases. This is a reasonable result as demand for called minutes is well-behaving. However, demand for voice services is relatively inelastic. According to the most recent estimate (Growitsch et al. 2010), 1 percent increase in per minute price reduces the number of called minutes by 0.52 to 0.61 percent. Dewenter & Haucap (2007) estimated -0.47 to -1.1 for the similar elasticity in Austria. This range can be explained by differences between service providers. Ofcom (2011) refers to a study of Credit Suisse from 2010 that estimated -0.75 elasticity. According to an earlier statement of Ofcom (2007), equal own-price elasticity was estimated for both mobile and the fixed telephony. In its last decision, OPTA used -0.5 for own-price elasticity for mobile-to-mobile calls, -0.3 for fixed-to-mobile, -0.35 for mobile subscriptions and -0.15 for fixed subscription.

Figure 4.8 Linear tariffs (Pre-paid): By switching to BULRIC plus, producer surplus increases but consumer surplus reduces and deadweight loss is created in the retail market

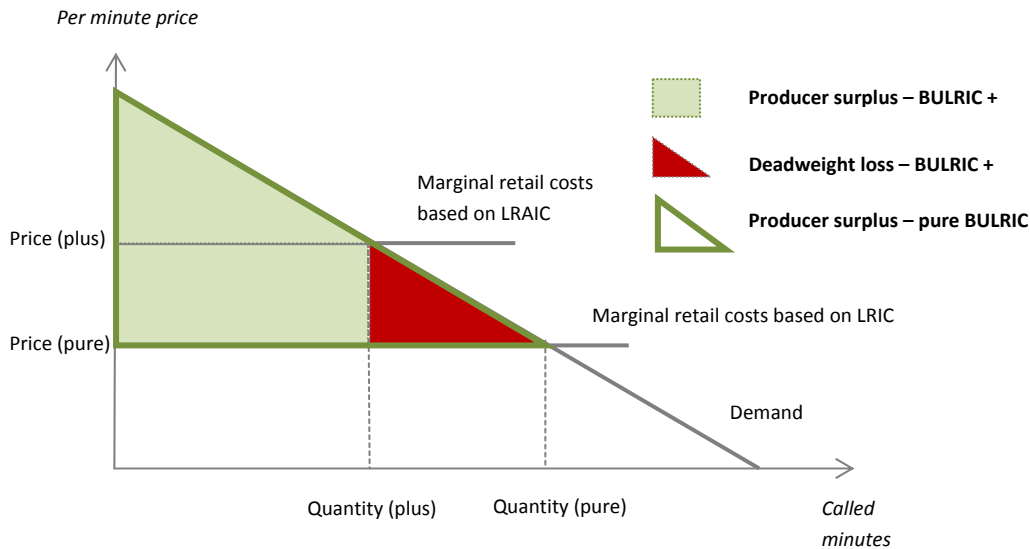


Source: SEO Economic Research; The figure does not take potential market power in the retail market into consideration.

Due to the reduction of called minutes, the value of these minutes for consumers and network operators is lost: consumers cannot derive utility from calling and being called, and network operators do not receive payments for these lost called minutes. The value, which is not achieved,

but would be available at lower prices, is called *deadweight loss*. Deadweight loss is created by a higher termination rate, independently of the retail price structure. The red triangles in Figure 4.8 and Figure 4.9 illustrate this lost value.

Figure 4.9 Multi-part tariffs (Post-paid): By switching to BULRIC plus, consumer surplus remains zero but producer surplus reduces due to deadweight loss in the retail market



Source: SEO Economic Research; The figure does not take potential market power in the retail market into consideration.

What happens to the *number of subscription* as a result of higher termination rates is unclear. Due to the “reversed” waterbed effect, higher termination rates imply more intensive competition for consumers by lower fixed fees. At the first glance, lower fixed fees would then lead to more subscriptions. This result is confirmed in the case of mobile-to-mobile (MTM) call termination (Cunningham et al. 2010, Genakos & Valletti 2012). However, demand for subscription is even less elastic than demand for voice services. For instance, Vodafone et al. (2003) reports price elasticities of mobile subscriptions between -0.08 and -0.54. In addition, when the mobile market is mature, which is the case in the Netherlands, the level of termination rates has no influence on the number of subscriptions any more (see review of empirical findings in Harbord & Hoernig 2012).

Furthermore, in the presence of call externalities (i.e., receivers benefit from being called), higher termination rates may lead to a less intense expansion in the market (Harbord & Hoernig 2012). This effect is particularly relevant in the case of fixed-to-mobile (FTM) call termination. Potential new mobile subscribers would benefit from being called by fixed customers. As a higher FTM termination rate reduces the number of FTM calls, potential mobile users’ would benefit less from the presence of FTM interconnection. This negative effect of call externalities on FTM connectivity is also less relevant in a mature market.

4.3.3 Consumers are never better off with BULRIC plus

The effects of higher termination rates, such that set by BULRIC plus, depend on the retail tariff structure. In the case of linear prices, consumers buy less telephony services at a higher per minute price, thus reducing consumer surplus. Figure 4.8 illustrates this situation: the triangle with thick outline representing consumer surplus in the presence of pure BULRIC shrinks to the small blue triangle. If non-linear tariffs are charged, the fixed fee always extracts the total surplus of consumers. Therefore, there is no change in consumer surplus.

4.3.4 Producer surplus varies for different tariff structures

Producers surplus is created in the wholesale and retail markets. As it is discussed in the chapter on effectiveness, termination rates above incremental costs are accompanied by a positive producer surplus (Figure 3.3). In addition, if a firm has a large proportion of light users, who mainly buy pre-paid packages, these customers remain available for calling parties even at higher retail prices, thus generating almost no termination costs but termination revenues for the called party network. This result holds for both fixed and mobile markets.

The questions are whether producer surplus also increases in the retail market and if not, whether an increase in surplus in the wholesale market outweighs a decrease in surplus in the retail market. The answer again depends on the market in question and the retail price structure.

Whether *mobile network operators* achieve higher or lower producer surplus in the retail market depends again on the retail tariff structure. In the case of *linear prices*, higher termination rates allow for a per minute price that produces positive surplus for the network operator (see green rectangle in Figure 4.8). This surplus is however too high: it is the sum of surplus in the wholesale and retail market. As termination rates based on BULRIC plus partly cover fixed costs, the per minute price contains a double mark-up at the benefit for firms and at the cost of consumers. However, it needs to be noted, that pre-paid packages account only for a small proportion of mobile firms' revenues (see Table 4.3).

If *non-linear prices* are set, network operators transfer total consumer surplus to producer surplus. This value increases by the number of called minutes. Since higher termination rates lead to less called minutes, producer surplus will also be lower. Figure 4.9 illustrates this situation: the triangle with thick green outline representing producer surplus in the presence of pure BULRIC shrinks to the sum of the green triangle and rectangle. This argument holds for mobile and fixed telephony.

The explanation for the waterbed effect lies in the balancing work between termination and retail profits. The waterbed effect is stronger in the case of two-part tariffs and, in particular, influences the fixed fee strongly (Genakos & Valletti 2011). According to the "reversed" waterbed effect, the called party network is willing to ease competition by setting lower fixed fees in order to achieve higher termination revenues. Consequently, producers surplus is created in the call termination market (see also arguments in Chapter 3 about effectiveness). According to theoretical and empirical studies, the overall effect of termination rates on surplus is negligible when non-linear prices are set. This is the so called profit neutrality.

Profit neutrality holds for incumbent firms. However, when a new participant enters the market, asymmetric regulation applies and lower termination rates to be paid to incumbents can benefit this entrant via two channels (Baranes et al. 2011). First, if the entrant pays a lower termination rate and thus can charge a more competitive retail price to increase its market share. In a mature market it happens at the cost of incumbents' market shares. Second, lower termination rates increases entrants' profit, thus strengthening sustainable competition between network owners in the long run. This argument is in line with the goal of the European Commission.

As discussed earlier, *fixed telephony* networks pay FTM termination rates after every minute terminating in mobile networks, and receives FTRs after every mobile call terminating in its network. The difference between these rates is substantial (see Table 4.5). It means that if the calling pattern is balanced – same amount of off-net calls are terminated from fixed-to-mobile and mobile-to-fixed – and there is no termination-based price discrimination, then fixed firms always make a loss after every minute off-net call. The fixed firm can compensate this loss by setting a higher off-net tariff. Based on findings in the literature (e.g. Harbord & Hoernig 2012), the overall effect on total producer surplus is negative.

4.3.5 Total welfare reduces

Deadweight loss, created by less called minutes due to higher termination rates, reduces total welfare. In Table 4.6, the previous results are summarized. It need to be noted, however, that no empirical analysis exists that systematically analyzes the welfare consequences of higher termination rates. In theory, it is also only Harbord & Hoernig (2012) that provides a model with less restrictive assumptions than other studies and thus a more comprehensive study. This article confirms the results in Table 4.6.

As the table shows, mobile network operators benefit from BULRIC plus only for *linear prices*. However this result raises several points for critique. First, this extra surplus is due to double marginalization: network operators set prices above costs in the wholesale market *and* a markup on top of that in the retail market. This conclusion raises the question if market power in the mobile market is reduced effectively to the minimum with the current regulation (see Chapter 3 about the analysis on effectiveness). Second, as Table 4.3 shows, pre-paid packages provide substantially less retail benefits for network operators than post-paid packages (non-linear prices). Therefore, these packages are less relevant for the welfare analysis. For *non-linear prices* BULRIC plus is never beneficial for consumers and mobile network operators. Therefore, for non-linear prices pure BULRIC should be strictly preferred over BULRIC plus and the argument of CBb does not apply. *Fixed firms* and consumers never benefit from BULRIC plus.

Table 4.6 The achieved welfare is higher by pure BULRIC than by BULRIC plus

	Pure BULRIC (P)	relation	BULRIC plus (+)
Mobile market with linear prices			
Consumer surplus (CS)	CS(P)	>	CS(+)
Total producer surplus (PS)	PS(P)	<	PS(+)
Producer surplus (PSr): retail	PSr(P)	<	PSr(+)
Producers surplus (PSw): wholesale	PSw(P)	<	PSw(+)
Deadweight loss (DWL)	DWL(P)	<	DWL(+)
Welfare (W)	W(P)	>	W(+)
Mobile market with non-linear prices			
Consumer surplus (CS)	CS(P)	=	CS(+)
Total producer surplus (PS)	PS(P)	≥	PS(+)
Producer surplus (PSr): retail	PSr(P)	>	PSr(+)
Producers surplus (PSw): wholesale	PSw(P)	<	PSw(+)
Deadweight loss (DWL)	DWL(P)	<	DWL(+)
Welfare (W)	W(P)	>	W(+)
Fixed telephony			
Consumer surplus (CS)	CS(P)	>	CS(+)
Total producer surplus (PS)	PS(P)	>	PS(+)
Producer surplus (PSr): retail	PSr(P)	<	PSr(+)
Producers surplus (PSw): wholesale	PSw(P)	>>	PSw(+)
Deadweight loss (DWL)	DWL(P)	<	DWL(+)
Welfare (W)	W(P)	>>	W(+)

Source: SEO Economic Research based on finding in the literature.

4.3.6 Welfare changes in the Netherlands

For the Netherlands, OPTA derived welfare results to support its recommendation (OPTA 2010; see Table 4.7). The welfare analysis is based on a model. This model accounts for the fixed and mobile telephony market. In the model, the starting point is the price cap on termination rate before 2010 that is 7 eurocent for MTR en 0.69 eurocent for FTR. Changes in welfare are considered when termination rates reduce. According to pure BULRIC, it means 1.2 eurocent for MTR and 0.36 eurocent for FTR; according to BULRIC plus, these rates are 2.4 and 0.37 eurocent, respectively. Demand is assumed to be inelastic (see Section 4.3.2 for details). In addition, the model takes call externalities and network externalities into consideration. The extent of latter externalities is negligible.

The OPTA model considers the similar two type of effect, which is also explained earlier. The termination rate as a cost element is called “cost pass-through” effect and it only affect per minute prices. Consistently to the previous analysis, higher termination rates lead then to higher per minute prices. The termination rate as a revenue element is reflected in the waterbed effect. Similarly to the previous analysis, the “reversed” waterbed effect drives down only the fixed fees as termination rates increase.

Table 4.7 contains the finding of OPTA under the realistic assumption that the waterbed effect is not 100 percent (see e.g. Genakos & Valletti 2011a). These results show that decreasing termination rates – independently of the model – leads to higher welfare (see last row in the column). Furthermore, moving from BULRIC plus to pure BULRIC, total welfare increases further (see columns with italic showing the differences between pure BULRIC and BULRIC plus). This increase in welfare can be seen as a reduction in deadweight loss due to the increased number of called minutes.

The conclusion of the previous welfare analysis validates the results of OPTA concerning the preference of pure BULRIC over BULRIC plus in terms of its impact on the welfare of consumers and producers.

Table 4.7 Comparison of pure BULRIC and BULRIC plus based on OPTA (2010)

Change in <i>million euro per year</i>	Market	Table 14			Table 13		
		Pure BULRIC	BULRIC plus	<i>Difference (pure-plus)</i>	Pure BULRIC	BULRIC plus	<i>Difference (pure-plus)</i>
Consumer surplus	mobile	67	46	21	231	174	57
Producer surplus	mobile	-151	-121	-30	-219	-168	-51
Welfare	mobile	-84	-75	-9	12	6	6
Consumer surplus	fixed	187	148	39	284	221	63
Producer surplus	fixed	70	61	9	34	35	-1
Welfare	fixed	257	209	48	318	256	62
Consumer surplus	total	254	194	60	515	395	120
Producer surplus	total	-81	-60	-21	-185	-133	-52
Welfare	total	173	134	39	330	262	68

Source: OPTA (2010), p. 174-175; sums exclusive BTW; Table 14: cost pass-through 100%, waterbed 50%; Table 13: cost pass-through 50%, waterbed 50%.

4.3.7 Consumers lose substantially more than what firms win

Any change in welfare may put costs in different proportion on consumers and firms. As CBB argues, consumer surplus should not be increased at substantially high costs of firms. Chapter 3 of this report made the proportionality criterion explicit by taking the Hicks compensation criterion in welfare economics as the point of reference. According to this criterion, a strict welfare gain benefiting all consumers and producers is not always feasible and forms a very stringent condition on the regulator.¹⁹ For a Hicks-efficient welfare improvement it suffices to generate a net welfare gain summing up gains and losses for the different parties involved. A proportional regulatory measure satisfies this criterion, as it suggests that it remedies a cost to society – caused by unregulated market power – via a mechanism that does not cause still greater harm to society.

How is welfare change distributed in the mobile and fixed telecommunications market? As Table 4.7 shows, a welfare improvement can be achieved by moving from BULRIC plus to pure

¹⁹ In the economic literature this type of welfare gain would be called Pareto efficient.

BULRIC. All participants in the fixed market benefit from this change. In the mobile market, the results depend on how strong cost pass-through and waterbed effects are. Under a reasonable assumption, according to which the waterbed effect is not full (Genakos & Valletti 2011), consumers always benefit from and firms are always worse off by moving to the pure BULRIC model. The two effects approximately outweigh each other, keeping the change in welfare in mobile market relatively low. However, if the welfare gains are added up in the mobile and fixed market, it shows that welfare is increased. In addition, by a move to pure BULRIC, consumers gain at least twice, sometimes three times more than what firms lose. This result is due to the fact that consumers in both markets benefit from pure BULRIC, while firms in the mobile market make losses.

4.4 Conclusions

In this chapter, the proportionality of regulatory models for determining termination rates is analyzed. Based on theoretical and empirical findings, it can be concluded that BULRIC plus is not a proportional measure. A regulatory method is proportional if it increases welfare, which is the sum of consumer and producer surplus and if the increase in welfare does not influence parties disproportionately (i.e., no large differences appear between consumers and firms in the same market and between fixed and mobile telephony).

In both markets and for both linear and non-linear prices, welfare reduces by a switch from pure BULRIC to BULRIC plus. The underlying reasons are as follows:

- In both markets, *per minute prices increase* due to higher termination costs. *Fixed fees however decrease* due to the “reversed” waterbed effect.
- *Deadweight loss is created.* As a result of higher per minute prices, consumers call less. This reduced volume creates a lost value in the total market, called deadweight loss. In a mature market, such as in the Netherlands, it is unlikely that the number of subscribers increases as a result of lower fixed fees.
- *Consumers are never better off by BULRIC plus.* In the case of linear prices, they benefit less from a higher termination rate: consumer surplus reduces as a result of less called minutes. Furthermore, call externalities cannot be fully utilized in a growing market. In the case of non-linear price, consumer surplus remains zero, thus unchanged.
- *Firms only benefit from BULRIC plus if linear (pre-paid) prices are set. However pre-paid packages account for only small proportion of firms’ revenues.* Total producer surplus is the sum of surplus in the wholesale and retail market. In the mobile market, it increases in the case of linear prices due to increased termination costs and double marginalization. In the case of non-linear prices, the profit neutrality rule holds for incumbent firms: termination rates leave profits of firms unaffected as a result of the waterbed effect. Entrants are worse off by a larger termination rate, such as determined by BULRIC plus. In the fixed telephony, firms are always worse off at a higher MTR as FTRs are substantially lower than MTRs.
- Due to deadweight loss, BULRIC plus leads to *lower welfare*.
- As the analysis of OPTA also shows, the surplus of mobile firms can only increase at a disproportionately larger decrease of consumer surplus.

As a conclusion, BULRIC plus is not a proportional regulatory measure.

Literature

** Empirical papers*

- Analysys Mason (2010a). Fixed network module for OPTA's fixed and mobile LRIC model.
- Analysys Mason (2010b). Mobile network module for OPTA's fixed and mobile LRIC model.
- Analysys Mason (2010c). Mobile termination cost: understanding the move from LRAIC+ to pure LRIC, 25 June 2010, <http://www.analysismason.com>.
- * Andersson, K. & Hansen, B. (2009). Network Competition: Empirical Evidence on Mobile Termination Rates and Profitability. Working Paper No 09/09.
- Armstrong, M. (1998). Network Interconnection in Telecommunications. *The Economic Journal*, 108, 545-564.
- Armstrong, M. (2002). The Theory of Access Pricing and Interconnection. In M. Cave, S. Majumdar and I. Vogelsang (eds.), *Handbook of Telecommunications Economics*, vol. 1, Amsterdam: North-Holland, pp. 297-380.
- Armstrong, M. & Wright, J. (2009). Mobile call termination. *The Economic Journal*, 119, F270-F307.
- * Baranes, E., Benzoni, L. & Hung Vuong, C. (2011). How does European termination rate regulation impact mobile operator performance? *Intereconomics*, 46(6), 346-353.
- BEREC (2011). Work Programme 2012. BEREC Board of Regulators. BoR (11)62, 9 December 2011.
- BEREC (2012a). BEREC Opinion. Phase II investigation pursuant to Article 7a of Directive 2002/21/EC as amended by Directive 2009/140/EC. Case NL/2012/1284 – Call termination on individual public telephone networks provided at a fixed location in the Netherlands. Case NL/2012/1285 – Voice call termination on individual mobile networks in the Netherlands. BoR(12)23, 26 March 2012.
- Berger, U. (2004). Access Charges in the Presence of Call Externalities. *Contributions in Economic Analysis & Policy*, 3(1).
- Bijlsma, M. & van Dijk, M. (2007). Nieuwe generatie netwerken, nieuwe generatie regulering? CPB Document no. 145.
- Carter, M. & Wright, J. (1999). Interconnection in Network Industries. *Review of Industrial Organization*, 14, 1-25.

College van Beroep voor het bedrijfsleven (CBB, 2011), LJN: BR6195, 31 August 2011.

* Cunningham, B.M., Alexander, P.J. & Candeub, A. (2010). Network growth: theory and evidence from the mobile telephone industry. *Information Economics and Policy*, 22, 91–102.

De Bijl, P. & Kocsis V. (2007). Network neutrality and the nature of competition between network operators. *International Economics and Economic Policy*, 4(2): 159-184.

Dessein, W. (2003). Network Competition in Nonlinear Pricing. *The RAND Journal of Economics*, 34(4), 593-611.

* Dewenter, R. & Haucap, J. (2007). Demand Elasticities for Mobile Telecommunications in Austria. *Ruhr Economic Papers* #17.

EC (2009). Commission Recommendation of 7 May 2009 on the Regulatory Treatment of Fixed and Mobile Termination Rates in the EU (2009/396/EC).

ETNO (2008). ETNO Reflection Document on Termination Rates. <http://www.etno.eu/datas/positions-papers/2008/rd281-rpol-termination-rates-etno.pdf>

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

Gans, J. & King, S. (2001). Using ‘bill and keep’ interconnect arrangements to soften network competition. *Economics Letters*, 71, 413-420.

* Genakos, C. & Valletti, T. (2011a). Testing the “Waterbed” Effect in Mobile Telephony. *Journal of the European Economic Association*, 9, 1114–1142.

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

* Genakos, C. & Valletti, T. (2012). Regulating prices in two-sided markets: The waterbed experience in mobile telephony. *Telecommunications Policy*, 36, 360–368.

* Growitsch, C., Marcus, S. & Wernick, C. (2010). The effects of lower Mobile Termination Rates (MTRs) on Retail Price and Demand. *Communications & Strategies*, 80(4), 119-140.

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

Harbord, D. & Hoernig, S. (2012). Welfare analysis of regulating mobile termination rates in the UK with an application to the Orange/T-Mobile merger. Working Paper nr. 571, NOVA School of Business and Economics, October 2012.

- Hoernig, S. & Valletti, T. (2012). Mobile telephony. In Peitz, M. & Waldfogel, J. (Eds). The Oxford Handbook of the Digital Economy. Oxford University Press.
- Jeon, D.S. & Hurkens, S. (2008). A Retail Benchmarking Approach to Efficient Two-Way Access Pricing: No Termination-Based Price Discrimination. *The RAND Journal of Economics*, 39(3), 822-849.
- Jullien, B. & Rey, P. (2008). Notes on the Economics of Termination Charges. IDEI Report nr. 6.
- Jullien, B., Rey, P. & Sand-Zantman, W. (2010). Mobile call termination revisited. Mimeo, 30 August 2010.
- Laffont, J.J., Rey, P. & Tirole, J. (1998). Network Competition I: Overview and Nondiscriminatory Pricing. *The RAND Journal of Economics*, 29, 1-37.
- Laffont, J.J. & Tirole, J. (2000). *Competition in Telecommunications*. Cambridge, MA: MIT Press.
- OPTA (2010). Marktanalyse vaste en mobile gespreksafgifte. FTA-MTA-3. Besluit. OPTA/AM/2010/201951, 7 July 2010.
- OPTA (2012). Marktanalyse vaste en mobile gespreksafgifte. FTA-MTA-3b. Besluit. OPTA/AM/2012/200203, 2 July 2012.
- Ofcom (2007). Mobile call termination. Statement. 27 March 2007.
- Ofcom (2011). Wholesale mobile voice call termination. Statement. 15 March 2011.
- Ofcom (2013). Telecommunications market data tables Q3 2012. Accessible: <http://stakeholders.ofcom.org.uk/binaries/research/cmr/telecoms/Q3-2012.pdf>.
- Schiff, A. (2009). The "Waterbed" Effect and Price Regulation. *Review of Network Economics*, 7(3).
- Valletti, T. & Houppis, G. (2005). Mobile Termination: What is the "Right" Charge? *Journal of Regulatory Economics*, 28(3), 235-258.
- Varian, H.A. (2006). *Intermediate Microeconomics. A Modern Approach*. W.W.Norton & Co.; 7th Edition.
- * Veronese, B. & Pesendorfer, M. (2009). Wholesale termination regime, termination charge levels and mobile industry performance. A study undertaken for Ofcom.

Vodafone (2013). Interim Management Statement for the Quarter Ended 31 December 2012. http://www.vodafone.com/content/index/media/group_press_releases/2012/ims_31december2012.html.

Vodafone, O2, Orange and T-Mobile (2003). Reports on references under section 13 of the Telecommunications Act 1984 on the charges made by Vodafone, O2, Orange and T-Mobile for terminating calls from fixed and mobile networks. http://www.ofcom.org.uk/static/archive/oftel/publications/mobile/ctm_2003/ctm5.pdf



seo economic research

Roetersstraat 29 . 1018 WB Amsterdam . T (+31) 20 525 16 30 . F (+31) 20 525 16 86 . www.seo.nl