APPENDIX B: Ramsey-Boiteux Pricing

The previous Section highlighted that even if the regulator allows the utility to recover the long-run costs attributable to the service, there still remains the problem of recovering the common costs of production. In the absence of the regulator allowing the utility to perfectly price discriminate or charge multi-part tariffs, economics has established that the most efficient way to recover such costs is through allowing prices to be set in accordance with the Ramsey-Boiteux (R-B) pricing principles. This ensures that the common costs can be recovered, while minimising the overall efficiency loss associated with distorting price away from the long-run marginal cost of production. While the rule appears in a number of different guises, in its simplest and most commonly stated form — i.e. where there are no cross-price effects — it involves setting the price of good or service i so that the lower (higher) the own-price elasticity of demand ε_i is, where $\varepsilon_i = -\frac{\partial P_i}{\partial Q_i}\frac{Q_i}{P_i} > 0$,[1] the greater (lower) the proportionate mark-up that is required in price P_i from the marginal cost of production MC_i.[2] Where superscript R denotes the outcomes under the R-B price and the term λ represents what is sometimes referred to as the "Ramsey Number", the textbook R-B (or inverse-elasticity) price is often formally written as satisfying the condition:

$$\frac{P_i^R - MC_i}{P_i^R} = \frac{\lambda}{\varepsilon_i^R}, \quad i = 1, 2\ldots n \text{ and } 0 < \lambda < 1 \tag{B.1}$$

or alternatively, for any two services i and j

$$\frac{\left(P_i^R - MC_i\right)}{P_i^R}\varepsilon_i^R = \frac{\left(P_j^R - MC_j\right)}{P_j^R}\varepsilon_j^R \tag{B.2}$$

This Section provides a detailed analysis of R-B pricing. In particular it examines:

(i) the origins of R-B pricing;

---

1 The negative sign on the own-price elasticity term means that it will be a positive number throughout the analysis in Appendix B.

2 As the analysis here examines the outcome in the long-run, all references to marginal cost should be taken to mean the long-run marginal cost of production. That is, the marginal cost that includes compensation for the opportunity cost of capital.

(ii) the intuition underlying R-B pricing;

(iii) how to derive R-B prices. This examines the case when:

(a) there are no cross-price effects; and

(b) there are cross-price effects;

(iv) how it applies to access pricing regulation;

(v) network externalities and how to derive R-B prices in the presence of a network externality; and

(vi) how to derive R-B prices with linear or constant elasticity demand.

B.1 The Origins of R-B Pricing

R-B pricing derives its origins from the work by Ramsey (1927) and Boiteux (1956),[3] who established similar results from addressing different economic problems. Ramsey's seminal paper on taxation investigated how to minimise the loss in consumer surplus when raising a given amount of tax revenue using distortionary taxes. As he states in the introduction to the paper (p 47):

> The problem I propose to tackle is this: a given revenue is to be raised by proportionate taxes on some or all uses of income, the taxes on different uses being possibly at different rates; how should these rates be adjusted in order that the decrement of utility may be a minimum?

Boiteux meanwhile examined the socially optimal price for a public enterprise monopoly when marginal-cost pricing fails to provide cost recovery. As he noted in the introduction (p 219), his paper was

> *…left with the problem of determining how to amend the marginal cost pricing rule when the firm is subjected to a budgetary condition incompatible with the decision rule*.[4]

Hence the name "Ramsey-Boiteux" pricing in the context of utility pricing acknowledges the work of Ramsey, who established the initial rule (i.e. the "Ramsey Rule" for taxation), and Boiteux, who independently derived the same result in the context of cost recovery for a public utility.

Ramsey illustrated in his paper (at p 54) that in order to raise an infinitesimal amount of tax revenue, the optimal distortionary taxes should be set so as to "*diminish the production of all commodities in the same proportion*" [emphasis in original]. Therefore, where Q_i^0 denotes the quantity of good i demanded in the pristine state when there are no distortionary taxes, the above rule implies that for n goods:

---

3 F.P. Ramsey, "A Contribution to the Theory of Taxation", *Economic Journal* 37, 1927, pp 47-61.

4 M. Boiteux, "Sur la Gestion des Monopoles Publics Astrient á L'Equilibre Budgetaire", *Econometrica* 24, 1956, pp 22-40. As the original article is in French, W. J. Baumol and D.F. Bradford had the paper translated into English, and the citation for this is: M. Boiteux "On the Management of Public Monopolies Subject to Budgetary Constraints", *Journal of Economic Theory* 3, 1971, pp 219-40. The emphasis in the above quote is in the original paper.

$$\frac{dQ_1}{Q_1^0} = \ldots = \frac{dQ_n}{Q_n^0} \tag{B.3}$$

Ramsey found that with linear demand the above relationship would hold for any given level of tax revenue raised by the government, and where the demand for goods was independent, it translated into a requirement that commodities with the least elastic demand should be taxed proportionately the most.

Using an economy-wide model Boiteux established similar results to Ramsey. By also adopting the consumer and producer surplus method and assuming independent demands, Boiteux showed the now familiar result that for each good or service i, the relative divergence between the price and marginal cost should be proportional to the inverse elasticity of demand for good or service i. Curiously, because Boiteux incorrectly defined the elasticity of demand as the proportionate change in price with respect to a proportionate change in quantity (i.e. $\varepsilon_i = -\frac{\partial P_i}{\partial Q_i}\frac{Q_i}{P_i}$, rather than $\varepsilon_i = -\frac{\partial Q_i}{\partial P_i}\frac{P_i}{Q_i}$), he stated (at p 235) that the rule was:

> *The relative divergence between price and marginal cost should be proportionate to the elasticity of the good considered.* [Emphasis in original].

Using the notation developed earlier in this section he therefore concluded that the divergence of price from marginal cost should satisfy the following:

$$\frac{P_i^R - MC_i}{P_i^R} = \lambda\varepsilon_i^R, \; i = 1, 2\ldots n \tag{B.4}$$

However, Boiteux still correctly maintains that the principle is consistent with traditional policy of taxing markets more heavily with less elastic demand.

The Baumol and Bradford (1970) paper explicitly relates the welfare economics literature on the theory of second best with the more specialised analysis on the theory of taxation and public utility regulation.[5] In doing so they are the first to link the results of Ramsey and Boiteux. Further, in summarising the literature, the authors (at p 268) provide four variants on the theorem of the optimal departure from marginal cost-based prices. Whilst variants three and four outlined by Baumol and Bradford are the familiar inverse-elasticity and equi-proportionate reduction in output rules, the authors also illustrate that at the second-best price:

(i) the ratio of the change in profit in market i — where profit is denoted by the term π — resulting from an infinitesimal change in price for any good or service i, must be equal to the ratio of the resulting levels of output (i.e. $\frac{\partial \pi_i}{\partial p_i} \Big/ \frac{\partial \pi_j}{\partial p_j} = Q_i / Q_j$); and

(ii) where there are no cross-price effects, the deviation of price from the long-run marginal cost for any good or service i is proportional to difference between marginal cost and marginal revenue for service i (i.e. $P_i^R - MC_i = (1+\alpha)(MR_i - MC_i)$, i = 1, 2…n and $0 < \alpha < 1$).

The R-B pricing formula has also been extended to incorporate more complex retail pricing problems. For example, Braeutigam (1979) examines the optimal linear price for a natural monopoly given the existence of a competitive fringe,[6] while Brock and Dechert (1985) investigate the optimal R-B price in an inter-temporal setting.[7]

5 R.G. Lipsey and K. Lancaster, "The General Theory of Second Best", *Review of Economic Studies* 24, 1956-7, pp 11-32, coined the phrase "second-best efficiency". Their theory of second best was based around finding an efficient price in a market, given that there was an existing distortion in a related market. Specifically, they examined the optimal price that should be charged for a public utility's output, where it was a substitute for a product provided by a private monopoly. However, more recently, the term second-best efficiency has also been used to describe instances where the best possible outcome for society is achieved given there is some existing distortion yet no related market effects. An example of this more liberal usage of the term is found in the context of public utility pricing where a marginal-cost-based price fails to recover cost. The term second best is often used to describe an R-B price, as it represents the linear pricing regime that maximises efficiency, given that there is some revenue requirement placed upon the firm preventing a first-best outcome.

6 R.R Braeutigam, "Optimal Pricing with Intermodal Competition", *American Economic Review* 69, 1979, pp 38-49.

7 W.A. Brock and W.D. Dechert, "Dynamic Ramsey Pricing", *International Economic Review* 26, 1985, pp 569-91.

Further, Laffont and Tirole (1994)[8] and Armstrong, Doyle and Vickers (1996)[9], have established that the optimal linear access price for a network requiring recovery of its common costs, should be based on R-B pricing principles.

8 J-J. Laffont and J. Tirole, "Access Pricing and Competition", *European Economic Review* 38, 1994, pp 1673-1710.

9 M. Armstrong, C. Doyle and J. Vickers, "The Access Pricing Problem: A Synthesis", *Journal of Industrial Economics* 44, 1996, pp 131-50.

B.2 The Intuition Underlying R-B Pricing

As outlined in Section B.1, in its simplest guise, R-B pricing requires that in order to recover the common costs of production with least impact on social welfare or efficiency, the proportionate mark up in price from the marginal cost must be higher in the relatively more inelastic market. Roughly the intuition underlying this inverse elasticity rule can be outlined with the assistance of the diagram in Figure B.1.

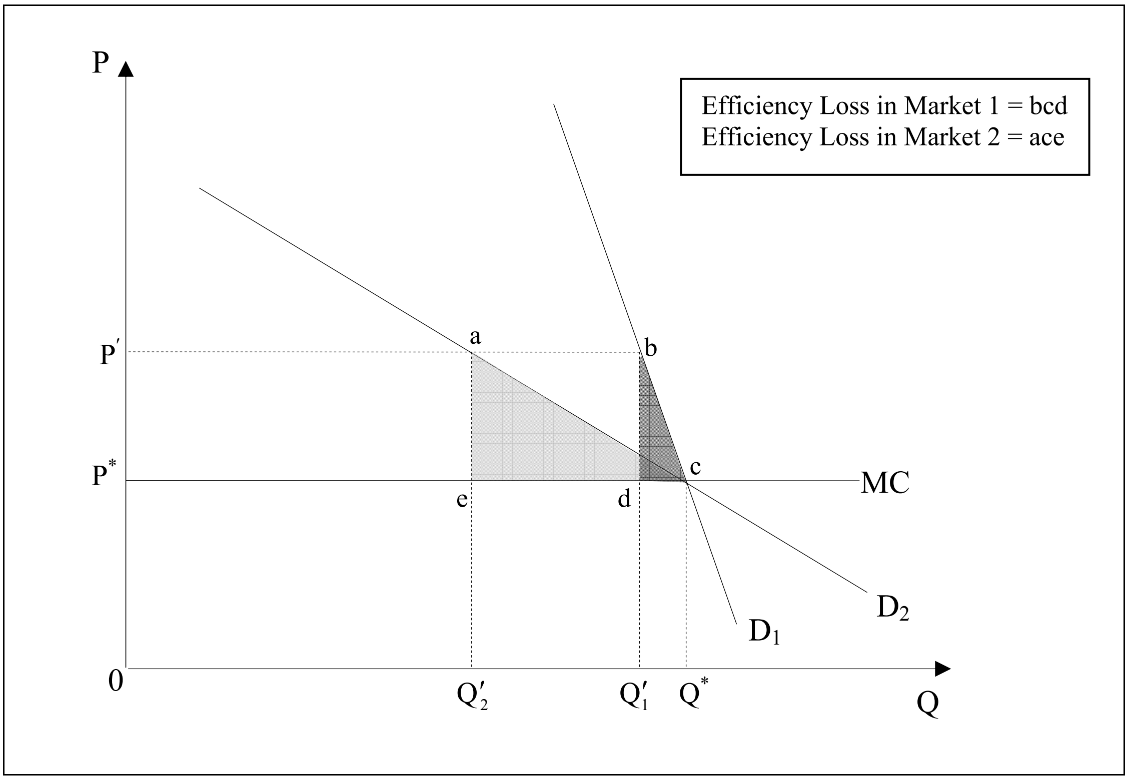
In Figure B.1 it is assumed for simplicity that there are two services Q_1 and Q_2,[10] which are produced under the same constant long-run marginal cost of production MC, and the quantities of each product can be measured along the same horizontal axis.[11] In the two markets there are the respective Hicksian or compensated demand curves D_1 and D_2,[12] and at the initial long-run marginal-cost-based price P^* in each market the same efficient level of output Q^* is consumed. Imagine that there is now an increase in price to P'. From the diagram it is apparent that for this given price increase, the compensated level of demand in market 2 where there is a relatively more elastic demand curve, will decrease by more than the compensated level of demand in market 1 (i.e. $Q_2' < Q_1'$). As the lost value to consumers in market 1 from decreasing output from Q_1' to Q^* is equal to area bcQ^*Q_1', and the additional cost to society of producing these units is only dcQ^*Q_1', there is an overall efficiency loss of area bcd. Meanwhile, in market 2, as the lost value to consumers from decreasing the service from Q_2' to Q^* is acQ^*Q_2', and the additional cost to society of producing these units is ecQ^*Q_2', there is a greater efficiency loss of ace.

10 As the paper addresses telecommunications, it is more appropriate to use the term services throughout, rather than goods.

11 A similar assumption is made in S. Berg and J. Tschirhart, *Natural Monopoly Regulation*, Cambridge University Press, 1988, p 59. Implicitly the assumption is that the products in the two markets are homogenous.

12 Throughout the analysis in this and the following Section it is assumed that the demand curves are Hicksian demand curves that hold utility constant, rather than Marshall-Dupuit demand curves that hold income constant. This assumption is made, because, although Hicksian demands do not generally capture actual quantities consumed in the market — unless the consumer's utility is assumed to be quasi-linear — they provide a true measure of the change in utility to the consumer for any given price change and reflect the consumer's marginal willingness to pay. The Hicksian demand curve is also sometimes referred to as the compensated demand curve.

FIGURE B.1 ILLUSTRATING THE INTUITION UNDERLYING R-B PRICING



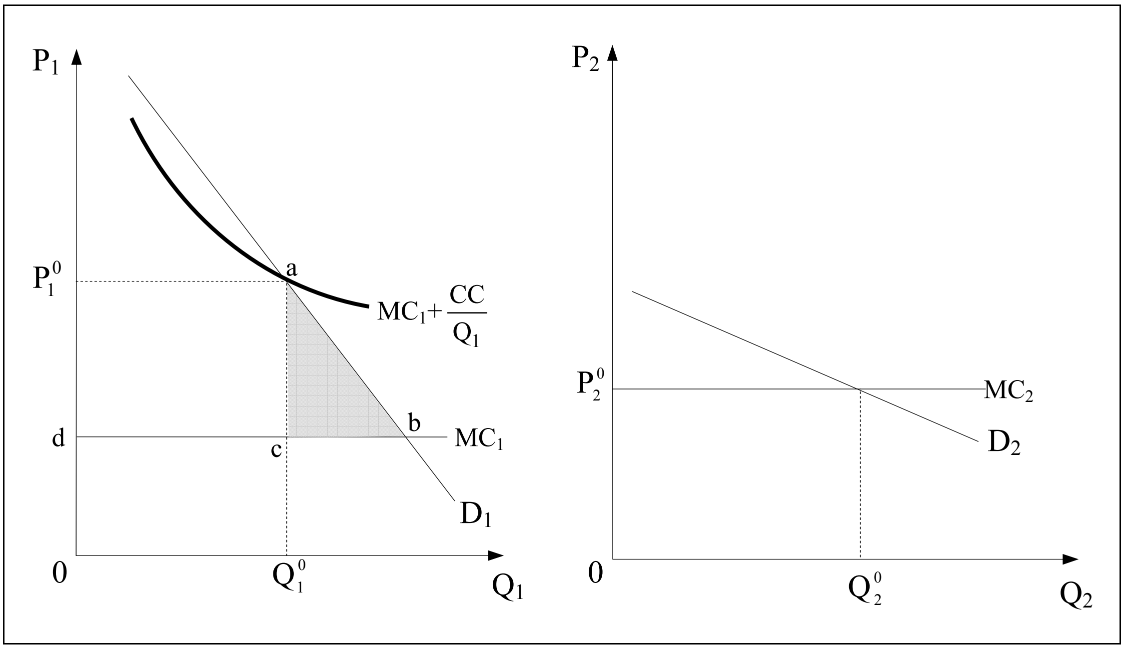
The result highlights that in this particular instance where there is a uniform price increase across both markets, there will be a greater distortion away from the efficient level of consumption Q^* in the market where demand is relatively more elastic, and subsequently a greater efficiency loss. Therefore, a government aiming to maximise welfare whilst raising a given amount of tax revenue — or alternatively a regulator aiming to maximise welfare whilst ensuring that a firm's common costs are recovered — will, rather than charging a uniform price above marginal cost, want to charge a higher price in market 1 than market 2. That is, there should be a higher proportional mark-up in price above the long-run marginal cost in market 1 where demand is relatively more inelastic.[13]

While the previous example roughly illustrates why the proportionate mark up in price above long-run marginal cost should be higher in the market where demand is more inelastic, it does not clearly show why it is not optimal for all common costs to be recovered from the market with the more inelastic demand. The reason why it is optimal to allocate common costs across all markets — i.e. price needs to be above

[13] Of course in this simple example the higher proportionate mark up is equivalent to a higher absolute mark up in price.

the long-run marginal cost in all markets — can be highlighted using an example incorporating the diagrams in Figure B.2.

FIGURE B.2 RECOVERING COMMON COSTS

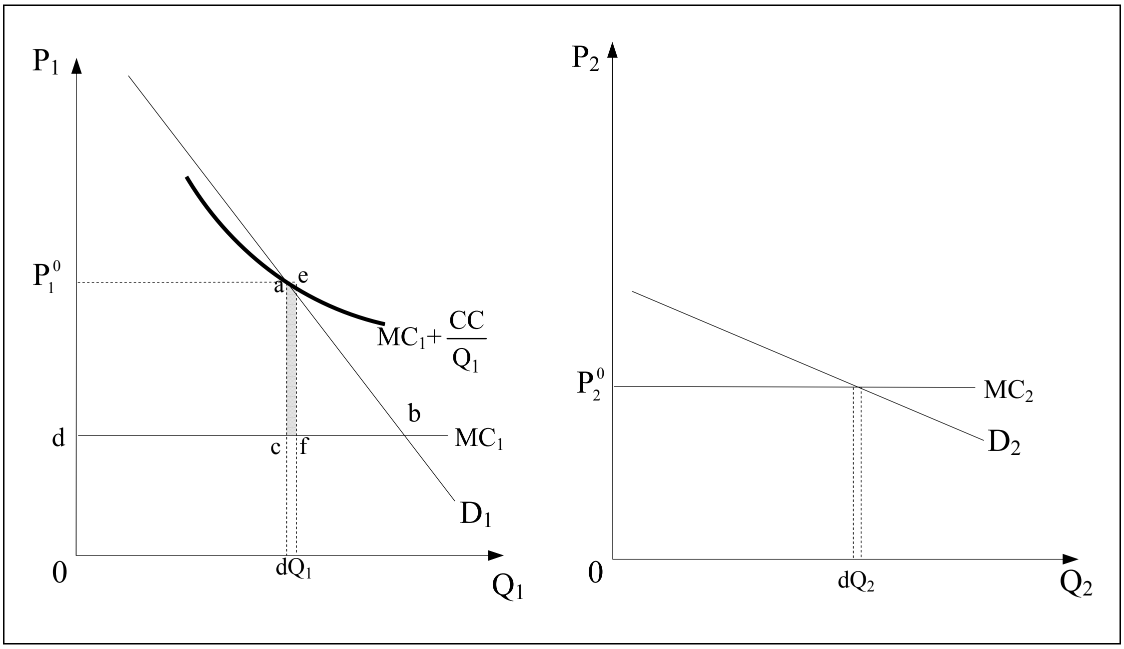


In Figure B.2, the common cost CC — i.e. the cost that cannot be directly attributed to the provision of either good 1 or good 2 — is captured by area P_1^0acd. Significantly, it is assumed throughout the analysis in this paper that if the firm were to charge the unregulated monopoly price in each market — P_1^m and P_2^m — it would over-recover the common cost of production. Further, the assumption made here is that there are no cross-price effects (i.e. $\frac{\partial Q_i}{\partial P_j} = 0$) and that initially the regulator:

- constrains the firm to earning zero profit across both markets; and
- only allows the unattributable cost CC to be recovered in market 1 where demand is relatively inelastic compared to demand in market 2.

The price in market 1 will then be set at P_1^0, where the compensated level of demand is Q_1^0, whilst market 2 yields the long-run marginal cost-based price of P_2^0, and a level of demand Q_2^0. As there is no welfare loss in market 2 the resulting inefficiency from recovering the costs solely from market 1 is equal to area abc.

FIGURE B.3 WELFARE GAIN FROM PRICE REBALANCING



In this setting it can be shown that there will be a welfare gain if the regulator allows the firm to recover the common costs of production by decreasing the price by a small amount in market 1 and increasing the price by a small amount in market 2. The outcome is illustrated in Figure B.3 where for an infinitesimal price decrease in market 1 and a corresponding infinitesimal price increase in market 2, the resulting marginal welfare change is positive. That is, there is a marginal welfare gain in market 1 of area aefc, yet no corresponding marginal deadweight-loss in market 2. This marginal welfare change is formally captured by the following equation:

$$dW = \underbrace{(P_1^0 - MC_1)dQ_1}_{\substack{\text{Marginal Welfare} \\ \text{Gain} > 0}} + \underbrace{(P_2^0 - MC_2)dQ_2}_{\substack{\text{Marginal Welfare} \\ \text{Loss} = 0}} > 0$$

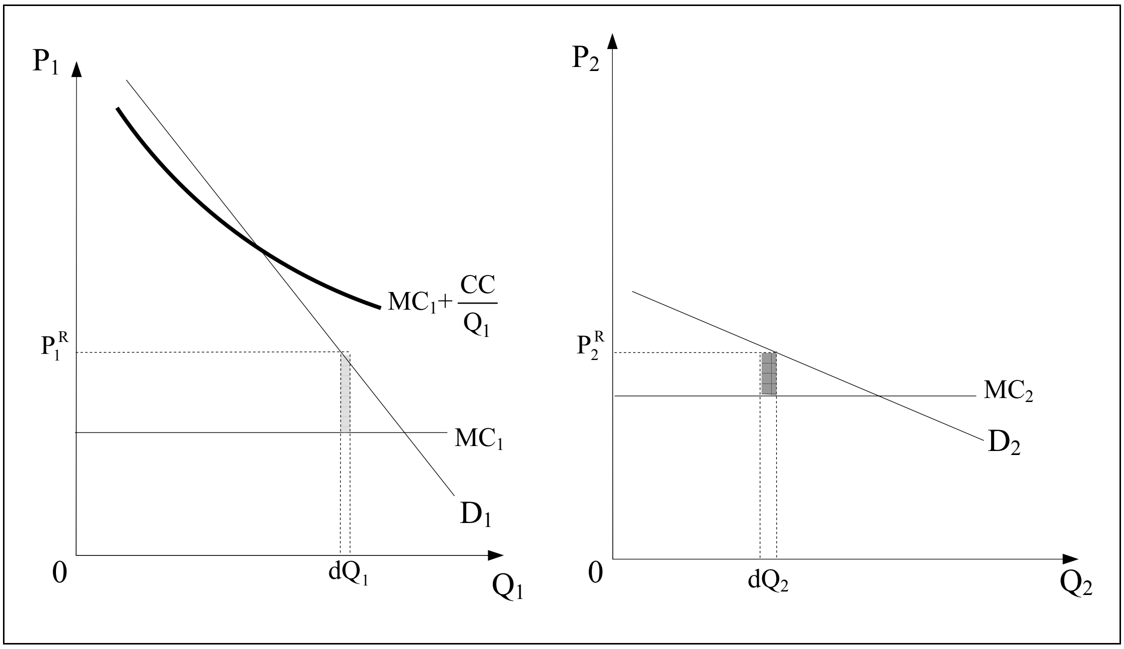
Hence, to maximise welfare across the two markets, prices should rebalance until the marginal welfare gain in market 1 equals the marginal deadweight-loss in market 2.

The above analysis highlights that at the R-B prices (i.e. welfare-maximising linear prices) two conditions must be satisfied. That is:

(i) **The marginal change in profit must be zero i.e. dπ = 0.** This ensures that when price decreases in market 1 and increases in market 2, the firm still earns enough revenue across both markets to cover the common or unattributable cost of production CC. Note, it is important to recognise here that if *the zero profit assumption were not made*, and the firm initially earns a given level of rent — i.e. $\bar{\pi}$, where $\bar{\pi} > 0$ — then the firm would instead be rebalancing prices so as to ensure that its revenues still recovered the common cost of production CC plus the allowed level of profit $\bar{\pi}$; and

(ii) **The marginal change in welfare must be zero i.e. dW = 0.** This maximises the overall welfare gain from price rebalancing, and in this instance involves equating the marginal welfare gain in market 1 with the marginal deadweight-loss in market 2. Such an outcome is illustrated in Figure B.4.

FIGURE B.4 THE OUTCOME AT THE WELFARE-MAXIMISING R-B PRICES



Using conditions (i) and (ii) and the example outlined in Figure B.2, the standard formula for the R-B prices with and without cross-price effects for two services is derived in the following Section.

B.3 Deriving the R-B Price

Where the utility provides n services and there is a common cost of production of CC, the resulting profit is captured by the equation,

$$\pi = \sum_{i=1}^{n} (P_i - MC_i)Q_i - CC \tag{B.5}$$

and the marginal change in profit will be,

$$d\pi = \sum_{i=1}^{n} (P_i - MC_i)dQ_i + Q_i dP_i$$

$$\text{where, } dQ_i = \underbrace{\frac{\partial Q_i}{\partial P_i} dP_i}_{\text{own-price effect}} + \underbrace{\sum_{i \neq j}^{n-1} \frac{\partial Q_i}{\partial P_j} dP_j}_{\text{cross-price effect}} \tag{B.6}$$

While the equation for the marginal change in welfare was outlined in the previous section where it was assumed there were two goods and no cross-price effects, more generally where there are n services, this marginal welfare change is written as,

$$dW = \sum_{i=1}^{n} (P_i - MC_i)dQ_i$$

$$\text{where, } dQ_i = \underbrace{\frac{\partial Q_i}{\partial P_i} dP_i}_{\text{own-price effect}} + \underbrace{\sum_{i \neq j}^{n-1} \frac{\partial Q_i}{\partial P_j} dP_j}_{\text{cross-price effect}} \tag{B.7}$$

Therefore, when there are only two services under consideration, the following relationship must hold at the welfare-maximising R-B prices, P_i^R, i = 1,2:

$$d\pi = (P_1^R - MC_1)dQ_1 + Q_1^R dP_1 + (P_2^R - MC_2)dQ_2 + Q_2^R dP_2 = 0 \tag{B.8a}$$

$$dW = (P_2^R - MC_1)dQ_1 + (P_2^R - MC_2)dQ_2 = 0 \tag{B.8b}$$

$$\text{where, } dQ_i = \frac{\partial Q_i}{\partial P_i} dP_i + \frac{\partial Q_i}{\partial P_j} dP_j \text{, i,j = 1,2 and } i \neq j$$

The above two equations can be used to derive R-B prices when there is unrelated demand (i.e. no cross-price effects) and interrelated demand (i.e. there are cross-price effects) for the products.

B.3.1 The R-B Price with No Cross-Price Effects

When there are no cross-price effects, a change in the price of service j will have no effect on the demand for service i, i.e. $\partial Q_i / \partial P_j = 0$, and $dQ_i = \frac{\partial Q_i}{\partial P_i} dP_i$, i = 1,2. Substituting this outcome into equations (B.8a) and (B.8b) and rearranging yields:

$$\frac{dP_2}{dP_1} = -\frac{(P_1^R - MC_1)\frac{\partial Q_1}{\partial P_1} + Q_1^R}{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2} + Q_2^R} \tag{B.9a}$$

$$\frac{dP_2}{dP_1} = -\frac{(P_1^R - MC_1)\frac{\partial Q_1}{\partial P_1}}{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2}} \tag{B.9b}$$

Equating (B.9a) and (B.9b) and cross-multiplying the terms gives,

$$\frac{Q_1^R}{(P_1^R - MC_1)\frac{\partial Q_1}{\partial P_1}} = \frac{Q_2^R}{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2}}$$

As the own-price elasticity of demand for good i at the R-B prices is $\varepsilon_i^R = -\frac{\partial Q_i}{\partial P_i}\frac{P_i^R}{Q_i^R}$, the above equation can be simplified to give,

$$\frac{(P_1^R - MC_1)}{P_1^R}\varepsilon_1^R = \frac{(P_2^R - MC_2)}{P_2^R}\varepsilon_2^R \tag{B.10}$$

Where there are n goods, the R-B price in equation (B.10) can be written in the more general form outlined in equation (B.2),

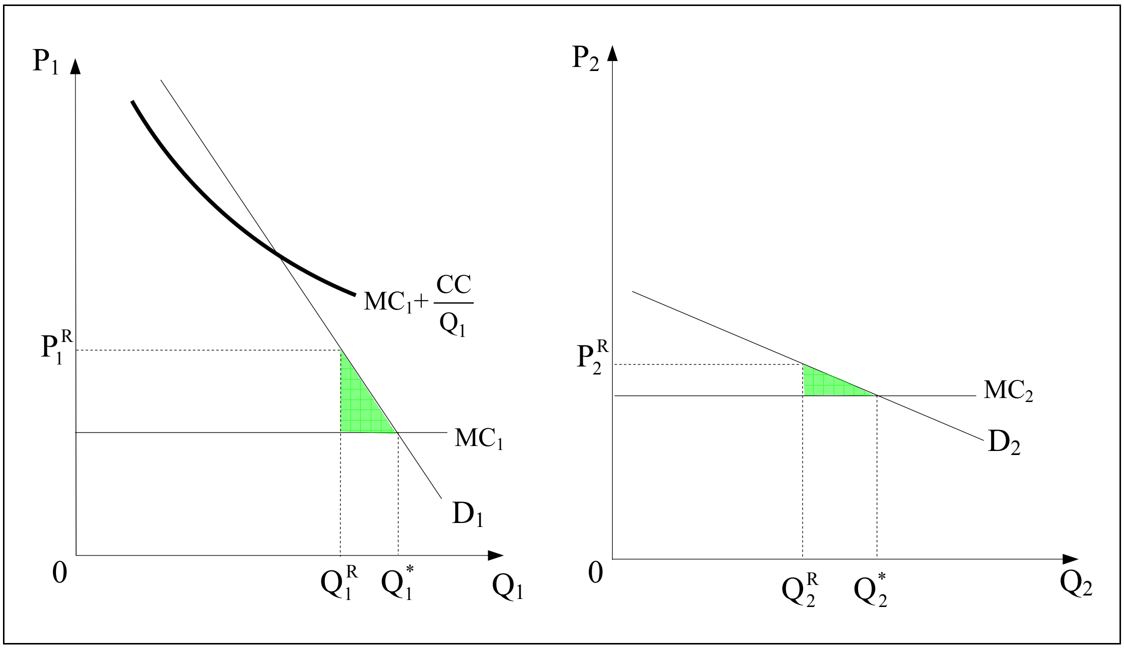
$$\frac{(P_i^R - MC_i)}{P_i^R}\varepsilon_i^R = \frac{(P_j^R - MC_j)}{P_j^R}\varepsilon_j^R$$

The above outcome can then be equated to some constant λ, and the familiar inverse-elasticity formulation for the R-B price outlined in equation (B.1) obtained.

$$\frac{(P_i^R - MC_i)}{P_i^R} = \frac{\lambda}{\varepsilon_i^R}, \quad i = 1,2\ldots n \text{ and } 0 < \lambda < 1$$

The resulting R-B prices outlined in equation (B.10) are depicted in Figure B.5. In the diagrams the shaded green area captures the deadweight-loss from the distorting quantity away from the efficient level of out in each market Q_1^* and Q_2^*.

FIGURE B.5 THE OUTCOME AT THE WELFARE-MAXIMISING R-B PRICES



The intuition for why the value of λ must lie between 0 and 1 is as follows. If λ were equal to zero then the price in each market would just be equal to the long-run marginal cost of production (i.e. $P_i = MC_i$). However, this outcome cannot arise because it requires that there are no common costs of production in the first place (i.e. CC = 0). Meanwhile, if λ were equal to one, then the utility would be charging the unregulated monopoly price for each service i (i.e. $(P_i^m - MC_i)/P_i^m = 1/\varepsilon_i^m$). However, such third-degree monopoly price discrimination is not possible, because in Section B.2 the assumption was made that the regulated firm is subject to a zero profit constraint, and that an unregulated monopoly price leads to the firm over-recovering its common cost of production.

Note that if instead of a zero profit constraint, some positive level of positive profits are allowed (i.e. $\bar{\pi} > 0$), then as outlined by Braeutigam (1980), the same pricing structure arises across the services.[14] The resulting prices, however, will now be

[14] R.R. Braeutigam, "An Analysis of Fully Distributed Cost Pricing in Regulated Industries", *Bell Journal of Economics* 11, 1980, pp 182-96. See p 189, footnote 14, and p 193, footnote 17.

higher, as they are required to not only efficiently allocate the common costs of production, but also the allowed rents across each service. The charges are now effectively the welfare-maximising prices for the given level of common costs and profit that the firm is allowed to earn. Importantly, the higher mark-up in price above marginal cost means that there is also a lower level of welfare or efficiency compared to the instance when zero economic profits are allowed. Further, although in the example used here the R-B price for each service always lies below the unregulated monopoly price, the structure of pricing across markets is identical to that which would arise under third-degree price discrimination of consumers by the monopoly. Joskow (2005) makes this point, stating (at p 80) that,[15]

> …the structure, though not the level, of the Ramsey-Boiteux prices is the same as the prices that would be charged by an unregulated monopoly with an opportunity to engage in third-degree price discrimination.

This highlights that R-B price is a form of price discrimination, and if any firm were to voluntarily set R-B prices it must have some form of existing market power to do so.

To see that the outcome in equation (B.10) is also consistent with the equi-proportionate reduction rule outlined in equation (B.4) for infinitesimal amount of common cost CC, it is necessary to recognise that:

- the infinitesimal price rise associated with recovering an infinitesimal common cost is just $dP_i = (P_i - MC_i)$;

- $dQ_i = \dfrac{\partial Q_i}{\partial P_i} dP_i$;

- the price elasticity of demand is $\varepsilon_i = -\dfrac{\partial Q_i}{\partial P_i}\dfrac{P_i}{Q_i}$; and

[15] P. Joskow, "Regulation of Natural Monopolies", forthcoming in *Handbook of Law and Economics*, A.M. Polinsky and S. Shavell (eds.), Elsvier Science B.V, 2005. Draft available at: http://econ-www.mit.edu/faculty/download_pdf.php?id=1086.

- for an infinitesimal common cost the corresponding price and quantity for good i, will be approximated by the long marginal cost-based price P_1^0 and long-run marginal cost-based quantity Q_1^0.

Substituting the information outlined above, equation (B.10) simplifies to the expression:

$$\frac{dQ_1}{Q_1^0} = \frac{dQ_2}{Q_2^0}$$

or more generally for n goods:

$$\frac{dQ_1}{Q_1^0} = \ldots = \frac{dQ_n}{Q_n^0}$$

B.3.2 The R-B Price with Cross-Price Effects

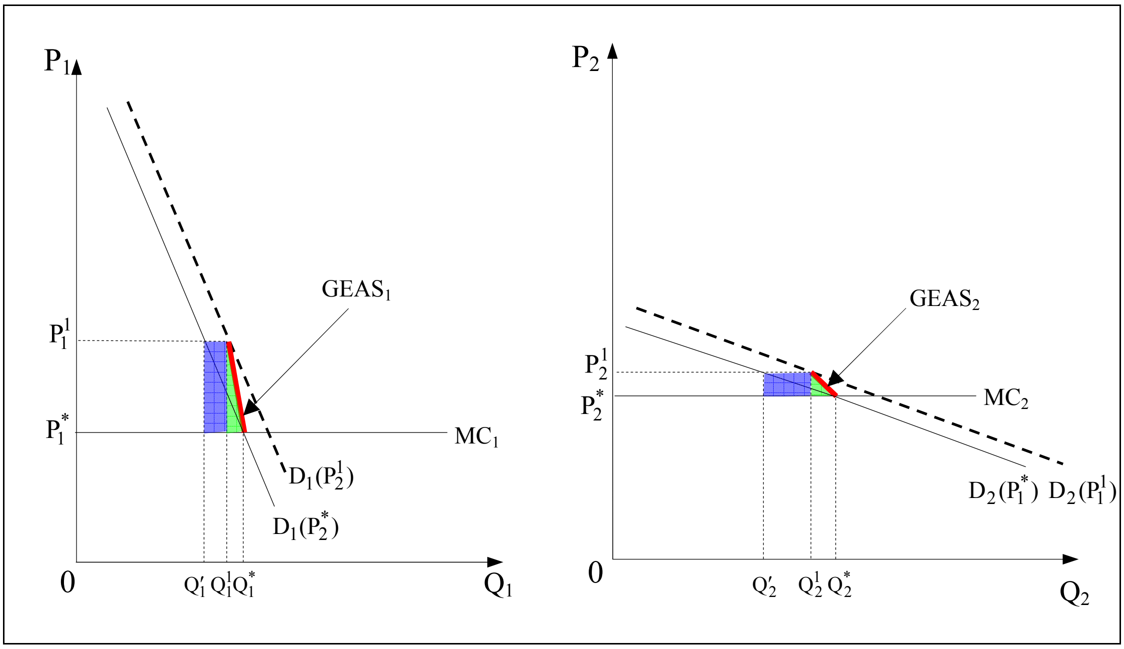
If the demand for goods is interrelated then it implies that a change in the price of some good j will have an impact on the demand for good i, or mathematically that $\partial Q_i / \partial P_j \neq 0$. Therefore, the change in quantity for any good i, dQ_i, will now be equal to,

$$dQ_i = \underbrace{\frac{\partial Q_i}{\partial P_i} dP_i}_{\text{own-price effect}} + \underbrace{\frac{\partial Q_i}{\partial P_j} dP_j}_{\text{cross-price effect}}, \quad i,j = 1, 2 \text{ and } i \neq j$$

Where the two goods are substitutes, an increase in the price of good j will lead to an increase in demand for good i — i.e. $\partial Q_i / \partial P_j > 0$ — while if the two goods are complements, an increase in the price of good j will lead to a decrease in demand for good i — i.e. $\partial Q_i / \partial P_j < 0$. The own-price and cross-price effects of increasing price above marginal cost to recover a common cost are highlighted in Figure B.6 where it is assumed that the two goods — 1 and 2 — are substitutes.

The own-price effect (i.e. $\frac{\partial Q_i}{\partial P_i} dP_i$) implies that there is a movement along the existing demand curve. In Figure B.6, the price increase from the long-run marginal cost-based price of P_i^* to P_i^1, i = 1,2, results in a fall in demand in each market from Q_i^* to Q_i', i = 1,2. **The cross-price effect** (i.e. $dQ_i = \frac{\partial Q_i}{\partial P_j} dP_j$) leads to a shift in the demand curve at the given prices. As the two goods in Figure B.6 are substitutes, there will be an increase in the demand for each good as a result of an increase in the price of the other good. This is reflected by a rightward shift in each demand curve from $D_i(P_j^*)$ to $D_i(P_j^R)$, i,j = 1,2 and $i \neq j$. Therefore, at prices P_i^1, i = 1,2, the cross-price effect results in demand for each good increasing from Q_i' to Q_i^1, i =1,2.

FIGURE B.6 THE OUTCOME WITH CROSS-PRICE EFFECTS



If it is assumed that the terms P_1^1 and P_2^1 in Figure B.6 represent the R-B prices derived in Section B.3.1 where there were no cross-price effects (i.e. $\partial Q_i / \partial P_j = 0$), then it is apparent from the diagram that the existence of positive cross-price effects leads to the firm recovering additional revenues equal to the blue rectangular areas in markets 1 and 2. Therefore, at what were previously the R-B prices there will now be over-recovery of the common costs of production CC. This implies that where the two goods are substitutes, the R-B prices consistent with common cost recovery will be lower than those derived in the previous Section, and where the two goods are complements the R-B prices must be higher than those derived in Section B.3.1.

Rather than examining the movements along and shifts out of the Hicksian demand curve, it also possible to highlight the effects of simultaneously increasing the price above marginal cost, by tracing out the path or loci of the Hicksian demand curves for all the above-cost prices up until price P_i^1, i = 1,2. The resulting curve is a general-equilibrium adjustment schedule (GEAS), which can be used to measure the efficiency change in a general-equilibrium setting.[16] In Figure B.6 the bold red line depicts the GEAS in both markets and the resulting deadweight-loss is captured by the green-shaded triangles underneath the GEAS curves. By again assuming that P_i^1, i = 1,2, denotes the R-B prices derived when there are no cross-price effects, it is evident from the diagram that when the two goods are substitutes (complements), not only is there a greater (smaller) amount of tax revenue at prices P_i^1, i = 1,2, but there will also be a smaller (larger) deadweight-loss.

To derive the condition that the R-B price must satisfy when there are cross-price effects, equation (B.11) for dQ_i is substituted into equations (B.8a) and (B.8b). This is solved in terms of dP_2/dP_1 to give,

$$\frac{dP_2}{dP_1} = -\frac{(P_1^R - MC_1)\frac{\partial Q_1}{\partial P_1} + Q_1^R + (P_2^R - MC_2)\frac{\partial Q_2}{\partial P_1}}{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2} + Q_2^R + (P_1^R - MC_1)\frac{\partial Q_1}{\partial P_2}} \tag{B.12a}$$

$$\frac{dP_2}{dP_1} = -\frac{(P_1^R - MC_1)\frac{\partial Q_1}{\partial P_1} + (P_2^R - MC_2)\frac{\partial Q_2}{\partial P_1}}{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2} + (P_1^R - MC_1)\frac{\partial Q_1}{\partial P_2}} \tag{B.12b}$$

As in the previous sub-section, by equating the outcomes in (B.12a) and (B.12b) and cross-multiplying the terms,

---

16 A.C. Harberger, *Taxation and Welfare*, Little, Brown and Company, Boston, 1974, p 88, effectively describes what is referred to as a GEAS here, stating that the efficiency impact of two or more distortions can be measured using the "loci...of potential equilibrium points, generated in principle by the package of distortions". He shows these loci in diagrams on p 87.

$$\left[(P_1^R - MC_1)\frac{\partial Q_1}{\partial P_1} + (P_2^R - MC_2)\frac{\partial Q_2}{\partial P_1}\right]\frac{1}{Q_1^R} = \left[(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2} + (P_1^R - MC_1)\frac{\partial Q_1}{\partial P_2}\right]\frac{1}{Q_2^R}$$

which can also be expressed as:

$$\begin{aligned} &-\frac{(P_1^R - MC_1)}{P_1^R}\left(-\frac{\partial Q_1}{\partial P_1}\frac{P_1^R}{Q_1^R}\right) - \frac{(P_2^R - MC_2)}{P_2^R}\left(-\frac{\partial Q_2}{\partial P_1}\frac{P_1^R}{Q_2^R}\right)\frac{Q_2^R}{P_1^R}\frac{P_2^R}{Q_1^R} \\ &= -\frac{(P_2^R - MC_2)}{P_2^R}\left(-\frac{\partial Q_2}{\partial P_2}\frac{P_2^R}{Q_2^R}\right) - \frac{(P_1^R - MC_1)}{P_1^R}\left(-\frac{\partial Q_1}{\partial P_2}\frac{P_2^R}{Q_1^R}\right)\frac{Q_1^R}{P_2^R}\frac{P_1^R}{Q_2^R} \end{aligned} \tag{B.13}$$

Recognising here that:

- the own-price elasticity of demand for good i is $\varepsilon_i = -\frac{\partial Q_i}{\partial P_i}\frac{P_i}{Q_i}$;
- the cross-price elasticity of demand is $\varepsilon_{ij} = -\frac{\partial Q_i}{\partial P_j}\frac{P_j}{Q_i}$; and
- the total revenue derived from good i is $TR_i = P_iQ_i$, i = 1,2;

equation (B.13) can be simplified to give:

$$\frac{(P_1^R - MC_1)}{P_1^R}\varepsilon_1^R + \frac{(P_2^R - MC_2)}{P_2^R}\frac{TR_2}{TR_1}\varepsilon_{21}^R = \frac{(P_2^R - MC_2)}{P_2^R}\varepsilon_2^R + \frac{(P_1^R - MC_1)}{P_1^R}\frac{TR_1}{TR_2}\varepsilon_{12}^R$$

As the substitution matrix is symmetric (i.e. $\frac{\partial Q_i}{\partial P_j} = \frac{\partial Q_j}{\partial P_i}$),[17] it follows that $\varepsilon_{ij}^R \frac{TR_j^R}{TR_i^R} = \varepsilon_{ji}^R$, and that,

$$\frac{(P_1^R - MC_1)}{P_1^R}\varepsilon_1^R + \frac{(P_2^R - MC_2)}{P_2^R}\varepsilon_{12}^R = \frac{(P_2^R - MC_2)}{P_2^R}\varepsilon_2^R + \frac{(P_1^R - MC_1)}{P_1^R}\varepsilon_{21}^R \tag{B.14}$$

[17] This is the case for Hicksian demand curves or a Marshall-Dupuit demand where it assumed there is either quasi-linear or homothetic preferences.

This can be rearranged to give,

$$\frac{(P_1^R - MC_1)}{P_1^R}\left(\varepsilon_1^R - \varepsilon_{21}^R\right) = \frac{(P_2^R - MC_2)}{P_2^R}\left(\varepsilon_2^R - \varepsilon_{12}^R\right) \qquad \text{(B.15)}$$

In the above equation the term (ε_i - ε_{ji}), i,j = 1,2 and $i \neq j$; is often referred to as the "superelasticity" of good i, as it captures the full output response as a result of a set of price changes.[18] These superelasticity terms are also the relevant demand elasticities for the respective GEAS 1 and GEAS 2 curves in Figure B.6. Denoting the superelasticity of good i by $\hat{\varepsilon}_i$, it follows that equation (B.15) can be written in terms of the standard inverse-elasticity formulation for R-B prices,

$$\frac{(P_i^R - MC_i)}{P_i^R} = \frac{\lambda}{\hat{\varepsilon}_i^R}, \text{ where } \hat{\varepsilon}_i^R = \varepsilon_i - \varepsilon_{ji}; \ i = 1,2; \ i \neq j; \ \& \ 0 < \lambda < 1 \qquad \text{(B.16)}$$

From this formulation it is apparent that where there is a given amount of common cost that needs to be recovered CC, compared to the case where there are no cross-price effects (i.e. $\partial Q_i/\partial P_j = 0$):

- **When the two goods are substitutes (i.e. $\partial Q_i/\partial P_j > 0$ and $\varepsilon_{ij} < 0$), the R-B price will require a lower proportionate mark-up from the long-run marginal cost of production.** As the GEAS is relatively more inelastic than the Hicksian demand curve when the two goods are substitutes, any given price rise in the market will result in a smaller quantity distortion, leading to higher revenue and a lower deadweight-loss. Therefore, a lower proportionate mark-up in each market is required to recover the common cost of production CC; and

- **When the two goods are complements (i.e. $\partial Q_i/\partial P_j < 0$ and $\varepsilon_{ij} > 0$), the R-B price requires a higher proportionate mark-up from the long-run marginal cost of production.** As the GEAS is relatively more elastic than the

---

[18] For example see S. Berg and J. Tschirhart, *Natural Monopoly Regulation*, Cambridge University Press, 1988, p 66.

Hicksian demand curve when the two goods are complements, any given price rise in the market will result in a greater quantity distortion, leading to lower revenue and a greater deadweight-loss. Therefore, a higher proportionate mark-up in each market is required to recover the common cost of production CC.

As in the previous Section it is possible to write this formulation of the R-B price with cross-price effects in the more general form where there are n goods and there is a relationship between goods i and j. That is,

$$\frac{(P_i^R - MC_i)}{P_i^R} = \frac{\lambda}{\hat{\varepsilon}_i^R}, \text{ where } \hat{\varepsilon}_i^R = \varepsilon_i - \varepsilon_{ji}; \ i = 1, \ldots n; \ i \neq j; \ \& \ 0 < \lambda < 1 \quad \text{(B.17)}$$

B.4 Access Pricing Regulation: Linking the Efficient Component Pricing Rule (ECPR) and R-B Pricing

The analysis so far has only examined R-B pricing and its application to the retail prices of services. The difficulty of translating such pricing principles to wholesale access-pricing regulation is that while the impact upon revenues in the access market can be observed, the welfare outcome can only be measured in the downstream retail market. (This is highlighted by welfare equation (B.10)). Therefore, in order to properly assess the efficiency impact of an access-pricing regime, simplifying assumptions have to be made that link the movement of the access and retail prices, and the quantity of access and output in the downstream retail market. Such assumptions are made in the analysis done by Laffont and Tirole (1994) and Armstrong, Doyle and Vickers (ADV, 1996).

Laffont and Tirole establish that where a vertical-integrated provider (VIP) is subject to some common network cost, the optimal way to recover this is to regulate the VIP so that it charges a Ramsey-Boiteux retail and access price. Using a similar model, ADV illustrate a link between the R-B based access price derived by Laffont and Tirole and the controversial Efficient Component Pricing Rule (ECPR) proposed by Baumol and Willig.

This sub-section briefly examines:

(a) the controversy surrounding the ECPR;

(b) the context in which the ECPR has been considered a second-best efficient access price; and

(c) the optimal access pricing results of ADV and Laffont and Tirole.

B.4.1 The Controversial ECPR

The ECPR — also sometimes referred to as the parity pricing principle or Baumol-Willig (B-W) Rule in recognition of those responsible for its formulation[19] — was promoted as an access-pricing regime that would induce efficient outcomes, whilst still ensuring that a VIP received fair compensation for the use of its essential infrastructure.

Baumol and Sidak (1994a), summarise (at p 178) ECPR setting the access charge so that:[20]

$$\begin{aligned}\text{optimal input price} = {} & \text{the input's direct per unit incrmental cost} + \\ & \text{the opportunity cost to the input supplier of} \\ & \text{the sale of a unit of input}\end{aligned} \tag{B.18}$$

while Pickford (1996) describes the ECPR as an access price "which would leave the incumbent indifferent as to whether it or the rival supplies the unit of final product".[21] It is this property of "leaving the incumbent indifferent" that has been responsible for much of the controversy surrounding the rule, and led to a number of regulators explicitly rejecting it as a method for regulating access prices.

To formally analyse the ECPR, the following simple framework is adopted where it is assumed that:

1. the VIP is subject to Leontief production technology, such that for the incumbent and any entrant a unit of access to the essential facility q is necessary to produce exactly one unit of output in the final retail market Q (i.e. q = Q).

---

19 R.D. Willig, "The Theory of Network Access Pricing", *Issues in Public Utility Regulation*, H. M. Trebing (ed), Michigan State University Public Utility Papers, 1979, and, W.J. Baumol, "Some Subtle Issues in Railroad Regulation", *International Journal of Transport Economics* 10, 1983, pp 341-55.

20 W. J. Baumol and J.G. Sidak, "The Pricing of Inputs Sold to Competitors", *Yale Journal on Regulation* 11, 1994a, pp 171-202.

21 M. Pickford, "Pricing Access to Essential Facilities", *Agenda* 3, 1996, pp 165-76.

2. it is not possible to bypass the incumbent's infrastructure. Access to the facility is an essential input that is required for an entrant to produce the final good sold in the retail market. For each unit of access q supplied, a price of "a" is charged by the VIP;

3. the marginal cost of the VIP supplying the final service is MC_I, which is equal to the sum of the marginal cost of supplying access to itself c_w and the marginal cost of retailing the service c_I (i.e. $MC_I = c_w + c_I$).

4. the entrant faces a private marginal cost of production which is equal to amount it is charged for each unit of access a, and the marginal cost it faces of retailing each unit of the service c_E (i.e. $MC_P = a + c_E$). Depending upon the access price a, this may be greater than or equal to the social marginal cost of production faced by the entrant (i.e. $MC_E = c_w + c_E$);

5. there is a perfectly contestable market to provide the final service, so entry is an "all-or-nothing" affair and the entrant will supply a final product that is a perfect substitute for the incumbent's. Where successful entry does occur, the incumbent will choose to only supply access to its essential infrastructure. The entrant will then service the entire retail market supplying the same quantity as the incumbent Q_I^0, and charging the same fixed price as the incumbent of P_I^0, where $P_I^0 > MC_I$.

Under these assumptions the ECPR-based access price a^{ECPR} outlined in equation (B.18) can mathematically be written as:

$$a^{ECPR} = c_w + \left(P_I^0 - MC_I\right) \tag{B.19}$$

From this equation it is evident that the ECPR links the access and retail price charged by the incumbent,[22] and the difference between the retail price P_I^0 and the access charge a^{ECPR} (i.e. the margin $P_I^0 - a^{ECPR}$) is equal to the incumbent's marginal cost of retailing the service in the contestable downstream activity c_I (as $c_I = MC_I - c_w$). For this reason ADV state (at p 135) that the ECPR is also equivalent to a simple "Margin Rule".

Proponents of the ECPR, such as Baumol and Sidak (1994a, 1994b),[23] emphasise that the access price creates incentives for production efficiency as it encourages the VIP to allow entry by a firm with a lower marginal cost of retailing the service than its own. To see this note that under ECPR the incumbent will be indifferent between providing the final product in the retail market and essential access in the wholesale market, as assuming there are no common costs of production here, the VIPs overall profit will remain unchanged, i.e. $\pi_I = (P_I^0 - MC_I)Q_I^0 = (a^{ECPR} - c_w)Q_I^0$. For a competitor though entry will only occur if at the same price-quantity combination the VIP is able to earn positive profit. In order to achieve positive profit, an entrant must have a lower marginal cost of retailing the service than the incumbent (i.e. $\pi_E = (P_I^0 - MC_p)Q_I^0 = (P_I^0 - a^{ECPR} - c_E)Q_I^0 = (c_I - c_E)Q_I^0 > 0, \text{iff.} c_I > c_E$).

Critics of ECPR have questioned its definition of "opportunity cost". While there is no doubt that compensation is required for the direct marginal cost of providing access to the entrant c_w, including compensation for the marginal loss in profit from the incumbent no longer serving the downstream market ($P_I^0 - MC_I$), seems more difficult to justify. For example, if initially the VIP were an unregulated monopoly, the ECPR would define any lost monopoly rents resulting from deregulation and competition as a marginal cost of production that the access seeker should compensate the access provider for. This creative redefinition of opportunity costs means that the

---

22 J-J Laffont and J. Tirole, *Competition in Telecommunications*, MIT Press, Cambridge, 2000, p 119, refers to ECPR as a "partial (incomplete) regulatory rule that links retail and wholesale prices."

23 W.J. Baumol and J.G. Sidak, *Towards Competition in Local Telephony*, MIT Press, Cambridge, 1994b.

ECPR-based access price cannot guarantee allocatively efficient outcomes for the industry.[24]

B.4.2 The ECPR as a Second-Best Efficient Access Price

ADV illustrate that the ECPR-based access price, as summarised by Baumol and Sidak, is consistent with a second-best efficient outcome when there is initially an irremovable distortion or inefficiency in the retail market served by the incumbent firm. Armstrong (2002) outlines (at p 299) that the price charged by the incumbent P_I may be inefficiently distorted away from the marginal cost MC_I, because of such things as:[25]

(a) having to fund the common costs of providing the service; or

(b) being required to use profits obtained in one market to cross-subsidise another — e.g. geographically uniform retail tariffs.

To highlight the second-best efficiency of the ECPR-based access price, ADV use a more general framework to that outlined by Baumol and Sidak, which formed the basis for the analysis done in Section B.4.1. Thus, to capture the results of ADV, assumptions 1 to 4 are adopted, but assumption number 5 in Section B.4.1 — which relates to perfect contestability — is relaxed here. Instead, this is replaced by the assumption that:

5′. the VIP, which is subject to a common cost of CC, competes in the downstream retail market with a competitive fringe.

The assumption of the competitive fringe, implies that:

24 W.B. Tye, "The Pricing of Inputs Sold to Competitors: A Response", *Yale Journal on Regulation* 19, 1994, pp 203-24.

25 M. Armstrong, "The Theory of Access Pricing and Interconnection", *Handbook of Telecommunications Economics*, Volume 1, M.E. Cave, S.K. Majumdar and I. Vogelsang (eds.), Elsevier Science B.V, 2002.

- now possible for a number of firms to have entered the market;

- entry does not occur on an "all-or-nothing" basis, so the VIP can now simultaneously supply both access q_E to entrants and output Q_I to the retail market;

- entrants are price takers, so each unit of output Q_E will be priced at P_E, which is equal to the private marginal cost of production faced by these firms $MC_P = a + c_E$. The difference between the retail price and the social marginal cost of production is therefore equal to the difference between the access price and the marginal cost of supplying the wholesale access service (i.e. $P_E - MC_E = a - c_w$). Further, as the marginal cost of retailing the service for entrants c_E is constant, it follows that $dP_E = da$, and as q = Q, it implies that $\partial Q_E / \partial P_E = \partial Q_E / \partial a$, $\partial Q_I / \partial P_E = \partial Q_I / \partial a$, and that $dQ_E = dq_E$; and

- the output supplied by entrants Q_E is a substitute for the product supplied by the incumbent Q_I.

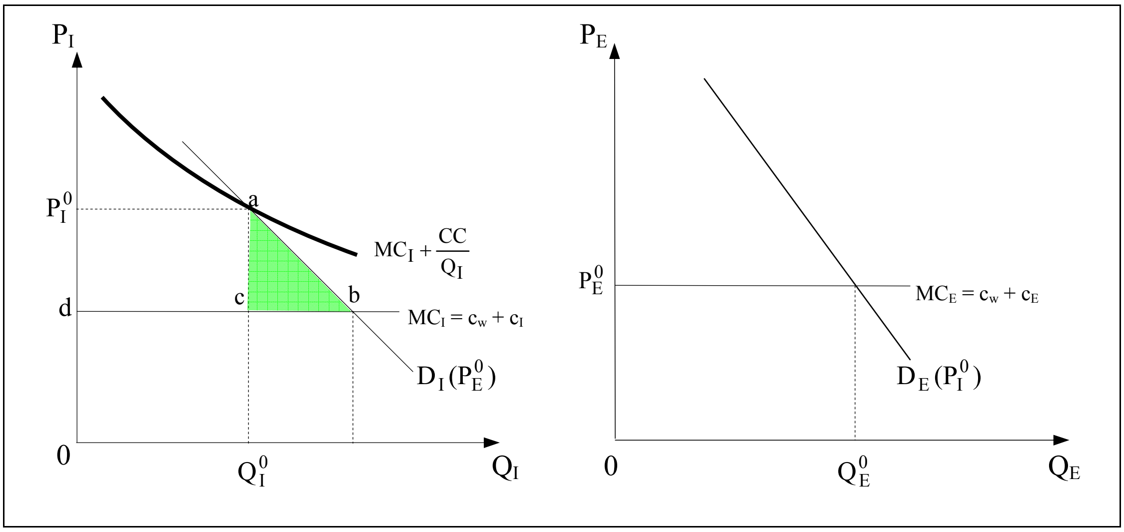
In addition to assumption 5′, a further assumption made for expositional purposes here is that:

6. the VIP is initially able to recover the common costs of its network CC, when it charges a retail market price of P_I^0 that exceeds its long-run marginal cost MC_I, and charges a regulated cost-based access price of $a^0 = c_w$.

Using the assumptions 1 – 4, 5′ and 6, the initial outcome in this market, where there is a marginal cost-based access price, is illustrated in Figure B.8.

The diagram shows that in the retail market served by the VIP, at price P_I^0 the common cost of the network P_I^0acd is recovered and the efficiency loss from distorting price away from the long-run marginal cost is equal to the green-shaded triangle area abc. As initially the regulator requires access to be priced at the marginal cost of using the infrastructure (i.e. a = c_w), the entrant's price P_E^0 is equal to the social marginal cost of production MC_E and there is no production or allocative inefficiency in the competitive fringe. The assumption that a unit of access q is required to produce a unit of output Q, means that the level of output produced Q_E^0 also represents the level of access supplied by the VIP to the competitive fringe.

FIGURE B.8 THE MARGINAL COST-BASED ACCESS PRICE

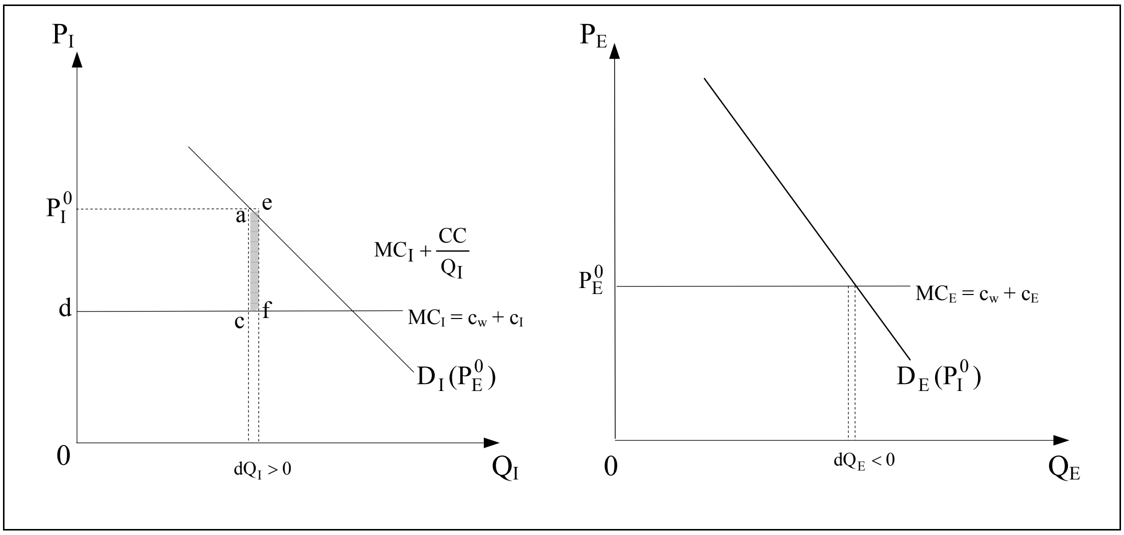


While there is production and allocative efficiency in the competitive fringe when the regulator sets the access price equal c_w, the overall level of welfare is not maximised. This can be highlighted by examining the marginal welfare change resulting from a marginal increase in the regulated access charge a.

A marginal increase in the access charge translates into a marginal increase in the retail price in the competitive fringe (i.e. da = dP_E and $\partial Q_I/\partial P_E = \partial Q_I/\partial a$), and as the VIP's and entrant's products are substitutes, at the given price P_I^0, there will be a marginal increase in the quantity of output demanded Q_I (i.e. $dQ_I > 0$). As the price P_I^0, which the consumer equates its marginal value to, exceeds the marginal cost MC_I, the marginal increase in quantity will generate a marginal welfare gain in the market served by the VIP of $(P_I^0 - MC_I)dQ_I > 0$. Figure B.9 captures this marginal welfare gain by the vertical grey-shaded sliver aefc. As there is no corresponding marginal deadweight-loss in market 2, because the cost-based regulated access price implies that P_E^0 is equal to MC$_E$, there must overall be a positive marginal welfare change. Formally, this result is captured by the following equation:

$$dW\big|_{a=c_w} = \underbrace{\left(P_I^0 - MC_I\right)}_{>0}\underbrace{dQ_I}_{>0} + \underbrace{\left(P_E^0 - MC_E\right)}_{=0}\underbrace{dQ_E}_{>0} > 0,$$
$$\text{where } dQ_I = \frac{\partial Q_I}{\partial a}da \text{ and } dQ_E = \frac{\partial Q_E}{\partial a}da \tag{B.20}$$

FIGURE B.9 THE MARGINAL COST-BASED ACCESS PRICE

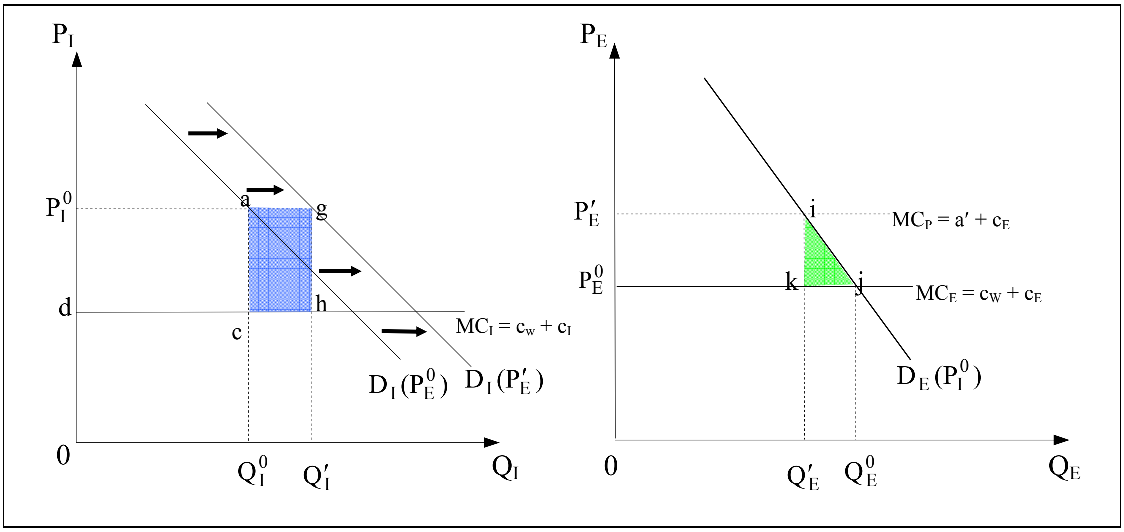


The outcome that a price above marginal cost in the competitive fringe leads to a higher level of welfare is consistent with the rule of second best outlined by Lipsey and Lancaster (1956-57).[26] This states that where there is an existing irremovable distortion in a market, it may not be optimal to set price equal to marginal cost in the related market. In the example outlined here, increasing the regulated access price above the marginal cost of providing access — and subsequently increasing the retail price above the social marginal cost in the competitive fringe — involves trading off a Harberger (1971) rectangle welfare gain in the retail market served by the VIP,[27] with a standard deadweight-loss triangle in the competitive fringe served by entrants. This is highlighted in Figure B.10 for some arbitrary set regulated access price $a' > c_w$.

[26] R.G. Lipsey and K. Lancaster, "The General Theory of Second Best", *Review of Economic Studies* 24, 1956-7, pp 11-32.

[27] A. Harberger, "Three Basic Postulates for Applied Welfare Economics: An Interpretative Essay", *Journal of Economic Literature* 9, 1971, pp 785-97.

FIGURE B.10 THE MARGINAL COST-BASED ACCESS PRICE



The right-hand side diagram illustrates that the access price a′ leads to firms in the competitive fringe facing a private marginal cost and charging a retail price of $P'_E = MC'_P =$ a′ + c_E. This exceeds the social marginal cost of production $MC_E = c_W + c_E$. The result is an allocative inefficiency equal to the green-shaded triangle ijk and a production inefficiency equal to area $P'_E ikP_E^0$. However, as the production inefficiency is simply equal to the increase in profit earned by the VIP from providing access, there is no deadweight-loss associated with it.

The left-hand side diagram highlights that the increase in the regulated access price also has an effect upon the level of demand for the VIP's product. As the output of the competitive fringe and incumbent are substitutes, the decrease in demand for the entrants' product is offset by some increase in the level of demand for the VIP's product at the distorted price P_I^0. This is reflected by the rightward shift in the demand curve from $D_I(P_E^0)$ to $D_I(P'_E)$, which leads to an increase in welfare — or alternatively an increase in profit to the VIP from supplying the retail service — of aghc.

To maximise welfare given the irremovable inefficiency in the VIP's retail market, the regulator must maximise the difference between the Harberger rectangle welfare gain and the standard deadweight-loss triangle in the competitive fringe. This is achieved by setting the access price to equate the marginal welfare gain with the marginal deadweight-loss. Hence, the second-best efficient access price a^* must satisfy the condition:

$$dW = \left(P_I^0 - MC_I\right)dQ_I + \left(P_E^* - MC_E\right)dQ_E = 0 \text{, where } P_E^* = a^* + c_E \qquad (B.21)$$

Substituting in for P_E^*, MC_E, dQ_I and dQ_E in the above equation, gives the expression,

$$dW = \left(P_I^0 - MC_I\right)\frac{\partial Q_I}{\partial a}da + \left(a^* - c_w\right)\frac{\partial Q_E}{\partial a}da = 0$$

Solving this in terms of a^* yields,

$$a^* = c_w + \sigma_a\left(P_I^0 - MC_I\right), \text{ where } \sigma_a = -\frac{\partial Q_I/\partial a}{\partial Q_E/\partial a} \qquad (B.22)$$

and it follows that that retail price for services supplied by the competitive fringe will be,

$$P_E^* = c_E + c_w + \sigma_a\left(P_I^0 - MC_I\right), \text{ where } \sigma_a = -\frac{\partial Q_I/\partial a}{\partial Q_E/\partial a} \qquad (B.23)$$

The second-best access price in equation (B.22) appears consistent with the condition stated by Baumol and Sidak in their summary of the ECPR. That is,

- the term c_w is the direct (marginal) cost of supplying access; and
- the term $\sigma_a\left(P_I^0 - MC_I\right)$ represents the (marginal) opportunity cost of the VIP supplying a unit of access and no longer serving some portion of the retail market.

ADV refer (at p 139) to σ_a in equation (B.22) as the "*displacement ratio*", as it captures the rate of substitution by consumers between the service supplied by the VIP and entrant for a small change in the access price.[28] Where the products are

---

28 ADV derive a similar outcome to that given by equation (B.22), when they relax the assumption of no bypass. The major difference is that the displacement ratio σ_{IE} involves a far more complex expression, as it captures not only the possibility of demand-side substitution, but also the possibility of supply-side substitution.

imperfect substitutes the value of σ_a lies between 0 and 1 (i.e. $0 < \sigma_a < 1$). In the special cases:

- where there are **no cross-price effects** between the products (i.e. $\partial Q_I/\partial a = 0$), σ_a is zero, and the second-best access price a^* is just set equal to the direct (marginal) cost of supplying access c_w. Here, if the access price were raised above c_w, the retail price would lie above the marginal cost in the competitive fringe and create a deadweight-loss triangle. The increase in price however has no impact on the level of demand for the incumbent's product at price P_I^0, so there is now no Harberger rectangle welfare gain offsetting the deadweight-loss. Consequently, pricing above the direct marginal cost of providing access in these circumstances is not second best as it generates an unambiguous welfare loss; and

- where the two products are **perfect substitutes** (i.e. $-\partial Q_E/\partial a = \partial Q_I/\partial a$), σ_a is equal to one, and the second-best access price a^* simplifies to the Margin Rule often associated with ECPR, which was outlined in equation (B.19).

The problem with the second-best efficiency of the ECPR is that precisely the same result is achieved if the initial rents to the incumbent are labelled an irremovable distortion. The second-best efficiency of the ECPR does not by itself guarantee that the incumbent no longer earns monopoly profits. The outcome relies on the regulator being able to accurately identify such things as all the relevant common costs of the VIP, and remove all unnecessary rents. This is something that ADV and Armstrong implicitly assume away as a problem in their analysis. Thus, the criticism of the ECPR initially highlighted by Tye still holds true, and where there are common costs, the optimal access price must not only be second best, but should also eradicate any rents accruing to the incumbent prior to entry taking place. Such a regime ensures that the access provider earns a normal rate of return on the investment, because it correctly treats all above-normal profits as a removable inefficiency.

One problem the regulator obviously encounters in accurately identifying such costs is that it has limited information and is faced with an incumbent that has incentives to overstate its costs. For instance, where the VIP is forced to fund some deficit that arises from another service, there is an incentive for the access-provider to overstate its losses so that it can increase the access price it charges and maintain its monopoly rent. Pelcovits (1999) highlights such a problem when he outlines a debate that occurred in the US in 1993,[29] over the appropriate level of the universal service obligation (USO). In that case, the incumbent argued that it required US 20 billion dollars per year to cover the cost of the USO, while entrants argued that the cost was only US 3.7 billion dollars per year. Although not of the same magnitude, there have been similar ongoing debates in Australia between Telstra and the regulator, the Australian Competition and Consumer Commission (ACCC), over the appropriate mark-up in access price that is required to meet the access deficit contribution (ADC).

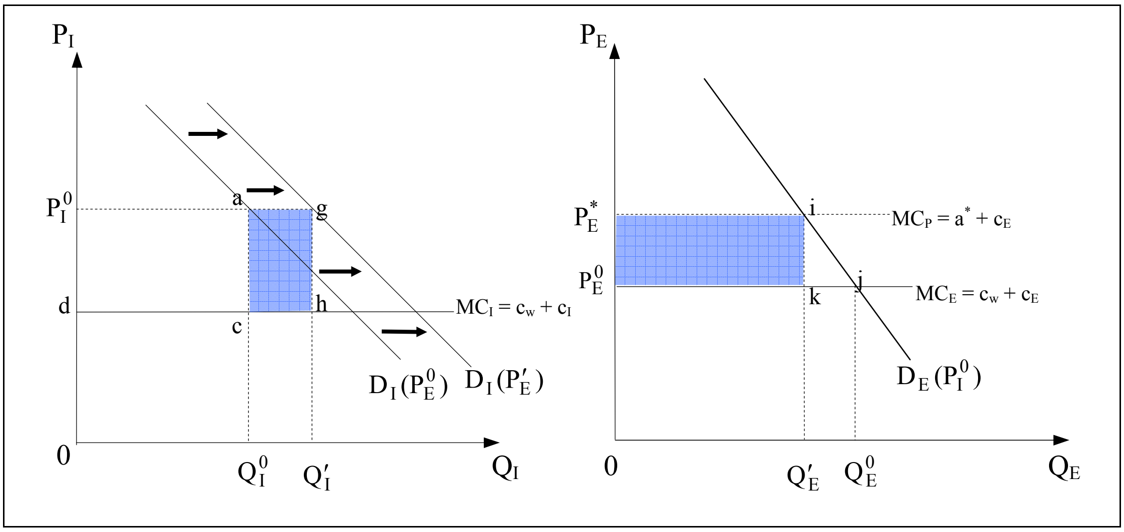
To see the problems created by inappropriately identifying the irremovable distortion, imagine that in the example outlined in this section, some portion of the initial distortion of the VIP's price P_I^0 away from marginal cost MC_I, consists of economic rents to the firm. In this framework, given a displacement ratio of $0 < \sigma_a < 1$, the incorrect classification of economic rents as part of an irremovable distortion, means that when the access price a is set above c_w, the height of the Harberger rectangle is greater than it should otherwise be. The increased height of the rectangle means that compared to the case where there is legitimate common cost recovery and no rents, there is a greater marginal welfare gain, which justifies a larger second-best access price a^* and marginal deadweight-loss in the competitive fringe. While the access price here maximises the relevant trade-offs, it should not be considered optimal, because it has failed to remove the unnecessary inefficiencies existing in the VIP's market.

---

29 M.D. Pelcovits, "Application of Real Options Theory to TELRIC Models: Real Trouble or Red Herring", *The New Investment Theory of Real Options and its Implications for Telecommunications Economics*, Kluwer Academic Publishers, Boston, 1999.

Another crucial point to recognise about the regulated second-best access price a^* derived in equation (B.22), was that the VIP was not subject to any profit constraint. Therefore, although overall welfare increased as the access price was distorted away from the marginal cost of supplying access, it was also the case that the VIP was able to increase its level of profit. The result is highlighted in Figure B.11, where the increase in the access price from a^0 to a^* generates an increase in profit for the VIP of area $P_E^* ik P_E^0$ in the access market and area aghc in the retail market. As assumption 6 established that at the initial retail market price P_I^0 and cost-based access price a^0 the VIP earned zero profit, these areas must also represent the super-normal profit derived from charging the above-cost access price a^*.

FIGURE B.11 THE INCREASE IN PROFIT TO THE VIP



While welfare and profit increased in the example outlined in this Section, the regulator would have achieved superior welfare outcomes, if it had been able to constrain the VIP to still earning zero profit. That is, a higher overall level of welfare is obtained from rebalancing the access and retail prices, until prices are reached that simultaneously maximise welfare (i.e. $dW = 0$) and ensure profit remains unchanged (i.e. $d\pi = 0$).[30] As the reader will recognise from Section B.2, these two conditions

30 The model outlined in Section B.4.2 is similar to that often employed by Telstra to promote its economic arguments. That is, Telstra often maintains that there are allocative efficiency gains from increasing an access charge, yet consistently fails to rebalance any of the existing prices to take into account that it was initially already earning a normal profit. By neglecting the issue of price rebalancing, Telstra implicitly treats existing prices as if they are an irremovable inefficiency distortion, and is requesting an access charge that allows it to earn an above-normal profit. Examples of where Telstra has used such arguments can be found in the context of its undertakings it has provided for the LSS.

are identical to those that must be satisfied to achieve the optimal R-B retail prices. Placing the analysis done earlier in the context of the model used here, it is possible to derive the optimal access and retail prices highlight the link established by ADV between ECPR and R-B pricing.

B.4.3 Linking the ECPR with R-B Pricing

In the framework outlined in Section B.4, the profit earned by the VIP is equal to the profits that it obtains from providing the downstream retail market and the upstream wholesale access service. Hence, the expression for total profit is,

$$\pi = (P_I - MC_I)Q_I + (a - c_w)q_E - CC$$

As it was assumed earlier that there is:

- Leontief production technology, so that a unit of access is required to produce a unit of output $Q_E = q_E$; and

- a competitive fringe that sets price equal to the private marginal cost of production, (i.e. $P_E = a + c_E$, $MC_E = c_w + c_E$, and $P_E - MC_E = a - c_w$);

the above expression for profit can be rewritten as,

$$\pi = (P_I - MC_I)Q_I + (P_E - MC_E)Q_E - CC$$

Thus, the optimal prices in the retail market served by the VIP and the competitive fringe — P_I^* and P_E^* — must satisfy the following conditions:

$$d\pi = (P_I^* - MC_I)dQ_I + Q_I^* dP_I + (P_E^* - MC_E)dQ_E + Q_E^* dP_E = 0 \quad \text{(B.24a)}$$

$$dW = (P_I^* - MC_I)dQ_I + (P_E^* - MC_E)dQ_E = 0 \quad \text{(B.24b)}$$

$$\text{where, } dQ_i = \frac{\partial Q_i}{\partial P_i}dP_i + \frac{\partial Q_i}{\partial P_j}dP_j, \text{ i,j = I,E and } i \neq j$$

It is apparent that by replacing the subscripts 1 and 2 with I and E, and superscript R with *, the above two conditions are identical to those outlined in equations (B.8a) and (B.8b) in Section B.3. Consequently the optimal prices charged by the VIP and the competitive fringe will be R-B prices, and must satisfy the expression in equation (B.14), such that,

$$\frac{(P_I^R - MC_I)}{P_I^R}\varepsilon_I^R + \frac{(P_E^R - MC_E)}{P_E^R}\varepsilon_{IE}^R = \frac{(P_E^R - MC_E)}{P_E^R}\varepsilon_E^R + \frac{(P_I^R - MC_I)}{P_I^R}\varepsilon_{EI}^R \quad \text{(B.25)}$$

$$\text{where, } \varepsilon_i = -\frac{\partial Q_i}{\partial P_i}\frac{P_i}{Q_i},\ i = I,E \text{ and } \varepsilon_{ij} = -\frac{\partial Q_i}{\partial P_j}\frac{P_j}{Q_i},\ i,j = I,E\ \&\ i \neq j$$

Equating the outcome in (B.25) to the constant θ, where $0 < \theta < 1$, substituting in $a^R - c_w$ for $P_E^R - MC_E$, it is possible to derive expressions for the optimal access and retail price found by ADV. That is, equation (B.25) can be broken up into the following two conditions,

$$\frac{(a^R - c_w)}{P_E^R}\varepsilon_E^R + \frac{(P_I^R - MC_I)}{P_I^R}\varepsilon_{EI}^R = \theta \quad \text{(B.26a)}$$

$$\frac{(P_I^R - MC_I)}{P_I^R}\varepsilon_I^R + \frac{(a^R - c_w)}{P_E^R}\varepsilon_{IE}^R = \theta \quad \text{(B.26b)}$$

These equations can also be expressed in the form,

$$(a^R - c_w)\left(-\frac{\partial Q_E}{\partial P_E}\frac{1}{Q_E^R}\right) + \left(P_I^R - MC_I\right)\left(-\frac{\partial Q_E}{\partial P_I}\frac{1}{Q_E^R}\right) = \theta \quad \text{(B.27a)}$$

$$\left(P_I^R - MC_I\right)\left(-\frac{\partial Q_I}{\partial P_I}\frac{1}{Q_I^R}\right) + (a^R - c_w)\left(-\frac{\partial Q_I}{\partial P_E}\frac{1}{Q_I^R}\right) = \theta \quad \text{(B.27b)}$$

As $\partial Q_E / \partial P_I = \partial Q_I / \partial P_E$, the above two equations can be expressed in terms of the access and retail price as,

$$a^R = c_w - \left(P_I^R - MC_I\right)\left(\frac{\partial Q_I / \partial P_E}{\partial Q_E / \partial P_E}\right) + \theta P_E^R \frac{Q_E^R}{P_E^R}\left(-\frac{\partial Q_E}{\partial P_E}\right)^{-1}$$

$$P_I^R = MC_I - (a^R - c_w)\left(\frac{\partial Q_E / \partial P_I}{\partial Q_I / \partial P_I}\right) + \theta P_I^R \frac{Q_I^R}{P_I^R}\left(-\frac{\partial Q_I}{\partial P_I}\right)^{-1}$$

Rearranging and simplifying the above expressions, gives the R-B access and retail charges derived by ADV at pp 139-41 and Armstrong at p 322,

$$a^R = c_w + \sigma_{P_E}\left(P_I^R - MC_I\right) + \frac{\theta P_E^R}{\varepsilon_E^R} \qquad \text{(B.28a)}$$

$$P_I^R = MC_I + \sigma_{P_I}(a^R - c_w) + \frac{\theta P_I^R}{\varepsilon_I^R} \qquad \text{(B.28b)}$$

$$\text{where } \sigma_{P_i} = -\frac{\partial Q_j / \partial P_i}{\partial Q_i / \partial P_i},\ i = I,E \ \&\ i \neq j$$

It follows from (B.28a) that the R-B retail price charged by the competitive fringe will then be,

$$P_E^R = c_E + c_w + \sigma_{P_E}\left(P_I^R - MC_I\right) + \frac{\theta P_E^R}{\varepsilon_E^R}$$

As $\partial Q_I/\partial P_E = \partial Q_I/\partial a$ and $\partial Q_E/\partial P_E = \partial Q_E/\partial a$, σ_{P_E} will be equal to σ_a, and the outcome in equation (B.28a) can be rewritten as,

$$a^R = \underbrace{c_w + \sigma_a\left(P_I^R - MC_I\right)}_{\text{ECPR access charge}} + \underbrace{\frac{\theta P_E^R}{\varepsilon_E^R}}_{\text{R-B mark-up}}$$

ADV and Armstrong outline that the optimal access charge is now a combination of the ECPR level plus a R-B mark-up.[31] The reason that the R-B access charge is above the level specified by the ECPR — which applied for a given price P_I^0 — is that a higher access price raises more net revenue to cover the common cost of production, which allows the retail price charged by the VIP to be lowered.

ADV (p 141), Armstrong (pp 322-3) and Laffont and Tirole (2000, p 103), outline that the access price in equations (B.28a) and (B.28b) are equivalent to the optimal prices for the access and retail price derived by Laffont and Tirole (1994). To see this

[31] ADV and Armstrong, both refer to this as a Ramsey mark-up, rather than a Ramsey-Boiteux mark-up.

for the access charge, substitute the expression for $(P_I^R - MC_I)$ in (B.28b) into equation (B.28a) to give,

$$a^R = c_w + \sigma_{P_E}\left(\sigma_{P_I}\left(a^R - c_w\right) + \frac{\theta P_I^R}{\varepsilon_I^R}\right) + \frac{\theta P_E^R}{\varepsilon_E^R}$$

Collecting terms in the above equation so that,

$$\left(a^R - c_w\right)\left(1 - \sigma_{P_E}\sigma_{P_I}\right) = \theta P_E^R\left(\frac{\left(\frac{\sigma_{P_E} P_I^R}{P_E^R}\right)}{\varepsilon_I^R} + \frac{1}{\varepsilon_E^R}\right)$$

and recognising that $\sigma_{P_E}\sigma_{P_I} = \frac{\varepsilon_{EI}^R \varepsilon_{IE}^R}{\varepsilon_I^R \varepsilon_E^R}$ and $\frac{\sigma_{P_E} P_E^R}{P_I^R} = -\frac{\varepsilon_{EI}^R}{\varepsilon_E^R}$,

$$\left(a^R - c_w\right)\left(\frac{\varepsilon_I^R \varepsilon_E^R - \varepsilon_{EI}^R \varepsilon_{IE}^R}{\varepsilon_I^R \varepsilon_E^R}\right) = \theta P_E^R\left(\frac{\varepsilon_I^R - \varepsilon_{EI}^R}{\varepsilon_I^R \varepsilon_E^R}\right)$$

Rearranging the above equation yields the optimal access price,

$$a^R = c_w + \theta P_E^R \frac{1}{\varepsilon_E^R}\left[\frac{\varepsilon_E^R \varepsilon_I^R - \varepsilon_E^R \varepsilon_{EI}^R}{\varepsilon_I^R \varepsilon_E^R - \varepsilon_{EI}^R \varepsilon_{IE}^R}\right]$$

and using identical methodology the optimal retail price can be solved for,

$$P_I^R = MC_I + \theta P_I^R \frac{1}{\varepsilon_I^R}\left[\frac{\varepsilon_I^R \varepsilon_E^R - \varepsilon_I^R \varepsilon_{IE}^R}{\varepsilon_I^R \varepsilon_E^R - \varepsilon_{IE}^R \varepsilon_{EI}^R}\right]$$

By setting $\hat{\eta}_i^R = \varepsilon_i \frac{\varepsilon_i \varepsilon_j - \varepsilon_{ij}\varepsilon_{ji}}{\varepsilon_i \varepsilon_j - \varepsilon_i \varepsilon_{ij}}$, i,j = I,E & $i \neq j$; Laffont and Tirole (at p 1678) provide the following simplified expressions for the R-B access and retail charges,[32]

$$a^R = c_w + \frac{\theta P_E^R}{\hat{\eta}_E^R} \qquad \text{(B.29a)}$$

$$P_I^R = MC_I + \frac{\theta P_I^R}{\hat{\eta}_I^R} \qquad \text{(B.29b)}$$

32 The difference between the term $\hat{\eta}_i$ used here, and that used by Laffont and Tirole at p 1678, arises due to the different expression for the cross-price elasticity. Laffont and Tirole define the cross-price elasticity as $\varepsilon_{ij} = (\partial Q_i/\partial P_j)(P_j/Q_i)$, while through the analysis here the cross-price elasticity has been set so that $\varepsilon_{ij} = -(\partial Q_i/\partial P_j)(P_j/Q_i)$.

Laffont and Tirole (at p 1678) describe $\hat{\eta}_i^R = \varepsilon_i \dfrac{\varepsilon_i\varepsilon_j - \varepsilon_{ij}\varepsilon_{ji}}{\varepsilon_i\varepsilon_j - \varepsilon_i\varepsilon_{ij}}$ as a "superelasticty" term. The reason that it differs from the superelasticity term outlined in equation (B.17) is that the constant θ differs from the constant λ used in equation (B.17). By instead using the term λ from equation (B.17), the R-B access and retail price charged by the VIP will be,

$$a^R = c_w + \frac{\lambda P_E^R}{\hat{\varepsilon}_E^R} \tag{B.30a}$$

$$P_I^R = MC_I + \frac{\lambda P_I^R}{\hat{\varepsilon}_I^R} \tag{B.30b}$$

where $\hat{\varepsilon}_i^R = \varepsilon_i - \varepsilon_{ji}$; $i = 1,\ldots n$ & $i \neq j$

Before proceeding with a worked example, it is important to recognise that the ability to link the access price derived here with the R-B retail price in Section B.3, relied upon assumptions that allowed the changes in the access price and quantity of access demanded to be directly translated into changes in the retail price and quantity of the final product demanded, and subsequently, changes in overall welfare. In particular, the two crucial assumptions made in this analysis — which were also made by Laffont and Tirole and ADV — were that:

- the vertically-integrated provider (VIP) competes with and supplies access to a competitive fringe in the downstream retail market. The competitive fringe prices on the basis of the private marginal cost it faces, which consists of the access price charged by the VIP and a given marginal cost of retailing the service (i.e. $P_E = MC_P = a + c_E$), so any given change in the access price must be equal to a change retail price. That is, there is one hundred per cent cost pass-through of any changes in the access price charged by the VIP to the retail price charged by the competitive fringe; and

- the industry is subject to Leontief production technology, so a unit of demand for access by the competitive fringe in the wholesale market q_E is exactly equal to a unit of output demanded in the final product market Q_E.

B.5 Deriving the R-B Pricing Relationship with Network Externalities

In examining the efficient pricing of telecommunications services, economists — such as Squire (1973), Rohlfs (1979), Willig (1979), Brown and Sibley (1986), Mitchell and Vogelsang (1991), and Armstrong (2002)[33] — have identified the potential for two types of externalities to arise, the network externality and the call externality. If these externalities cannot be properly internalised, then both will have an impact upon the efficient price that should be charged for an access and call service.

B.5.1 Network Externalities

Liebowitz and Margolis (2002)[34] outline on p 76 that network externalities, or *network effects* as they are more appropriately described,[35] arise because:

---

[33] For example see: L. Squire, "Some Aspects of Optimal Pricing for Telecommunications", Bell Journal of Economics 4, pp 515-25; J. Rohlfs, "Economically-Efficient Bell-System Pricing", Bell Laboratory Discussion Papers No. 138, January 1979, pp 5-9; R. D. Willig, "The Theory of Network Access Pricing", in H. Trebing (ed.), *Issues in Public Utilty Regulation*, Michigan State University, East Lansing, 1979, pp 109-52; S.J. Brown and D.S. Sibley, *The Theory of Public Utility Pricing*, Cambridge University Press, 1986, pp 197-9; B.M. Mitchell and I. Vogelsang, *Telecommunications Pricing*, Cambridge University Press, Cambridge, 1991, pp 55-61; M. Armstrong, "The Theory of Access Pricing and Interconnection", *Handbook of Telecommunications Economics*, Volume 1, M.E. Cave, S.K. Majumdar and I. Vogelsang (eds.), Elsevier Science B.V., Amsterdam, 2002, pp 337-45.

[34] S.J. Liebowitz and S.E. Margolis, "Network Effects", *Handbook of Telecommunications Economics*, Volume 1, M.E. Cave, S.K. Majumdar and I. Vogelsang (eds.), Elsevier Science B.V., Amsterdam, 2002.

[35] For example, see Liebowitz and Margolis (2002) at p 76-8, and S.J. Liebowitz and S.E. Margolis, "Network Externality: An Uncommon Tragedy", *Journal of Economic Perspectives* 8, 1994, pp 133-50. Both papers draw a distinction between network externalities and network effects. The authors maintain that while the term network externality is often used to describe how an existing consumers' value is affected by a change in the size of the network, it is more appropriate to refer this as a *network effect*, as the term network externality should be reserved for only describing those instances where there is a market failure, or a network effect that is not being internalised in the market. Liebowitz and Margolis (2002) state on pp 77-8 that:

> The use of the term 'externality' to mean something different in this literature than it does in the rest of economics is likely to create confusion. Unfortunately the term externality has sometimes been used carelessly in the networks literature.

They note that although network participants are unlikely to internalise the impact that their joining the network has on existing subscribers, network owners may well internalise such effects.

> As the number of users of a product or network increases, the value of the product or network to the other users changes.

In the context of telecommunications, it means that each new subscriber to a telephone service increases the number of communication opportunities that are available to existing users of the network. Rohlfs (1979) states at p 5 that it is "a classic case of an externality", as existing users "benefit from actions taken and paid for by new users."

To illustrate how the existing subscribers' values may be affected by the network size, the following simple example is used. Imagine there are initially two subscribers A and B to a telecommunications network. In this instance there will only be two possible call services that can be offered — A to B and B to A. However, if a new subscriber C joins the network, there is the opportunity for six call services to be supplied — A to B, A to C, B to A, B to C, C to A, and C to B. More generally, where there are N subscribers, there will be N×(N-1) possible call services that the network is able to supply, and each new subscriber creates the opportunity for an additional 2N call services.[36]

As outlined in the above example, network externalities have an impact upon all existing subscribers to the network. Due to this widespread impact it is often argued that network externalities are very difficult to internalise without some form of corrective pricing of the subscription service (i.e. the access service). This is highlighted by Rohlfs (1979), who notes on p 5 that, because it involves many people, it "probably cannot be fully internalised without corrective pricing", and that the "effect of such externalities is that the economically efficient price of access to the network is below marginal cost of access".

While emphasising the importance of network externalities and prescribing the need for a lower access price in their presence, Rohlfs (1979) recognises that they may be partially internalised in such instances where:

---

36 This example is an illustration of what has become known as "Metcalfe's Law". For a discussion of Metcalfe's Law see C. Shaprio and H.R. Varian, *Information Rules*, Harvard Business School Press, Boston, 1999, pp 183-4.

- a business subscribes to a telephone service to accommodate its customers, and the customer of the business benefits and pays for the service; and

- an individual party subscribes to accommodate communication with friends and relatives, or a group of people agree to subscribe jointly to communicate with one another.

In addition to the potential for this type of internalisation, a number of economists have also noted that at higher levels of subscription for fixed-line telephony voice in the US, the marginal network externality is likely to have diminished. For example:

- Mitchell (1978) outlines on p 518 that he chose to ignore network externalities in modelling the optimal US fixed-line telephony rates on the basis that, "they should have only a limited marginal effect on demand in a system with a high saturation of subscribers.";[37]

- Kahn and Shew (1987) states on p 242 that it may be the case that once subscription exceeds per cent, "marginal subscribers have come to consist disproportionately of people relatively isolated from society generally, to whose hypothetical addition to the network existing subscribers would impute progressively smaller values.";[38] and

- Sidak and Spulber (1997) also highlight on p 548 that the significance of network externalities "become less important as more and more subscribers are connected to the network" and that in relation to the fixed-line voice telephony, "once subscription rises to more than 95 percent of all households,

---

37 B.M. Mitchell, "Optimal Pricing of Local Telephone Service", *American Economic Review* 68, 1978, pp 517-37.

38 A. E. Kahn and W. B. Shew, "Current Issues in Telecommunications Regulation: Pricing", *Yale Journal on Regulation* 4, 1987, pp 191-256.

the remaining positive externalities that may be achieved on the margin surely become quite small."[39]

Econometric analysis by Iimi (2005) illustrates that there has been a similar pattern in relation to the significance of network externalities in the cellular phone market in Japan.[40] By comparing his findings on the Japanese mobile market with those of Okada and Hatta (1999),[41] Iimi concludes that the higher level of market saturation and increased product differentiation, has led to conventional network effects no longer being a crucial factor in choosing a mobile phone carrier.

B.5.2 A Method for Capturing Network Externalities — Rohlfs (1979)

To examine the socially-optimal price for the US Bell fixed-line telephony system, Rohlfs (1979) defined a term that was designed to simply capture the impact of network externalities on the demand for access to the local telephony service. The work that follows summarises the formal analysis done by Rohlfs to establish the externality factor, and examines how the value of this externality factor is determined.[42]

Rohlfs outlines on p 3 that the marginal private value for access is equal to,

$$\mathrm{mpv}_j = \frac{\hat{u}_j - u_j}{\partial u_j / \partial y_j}$$

where, mpv_j = marginal private value of access to individual j;

[39] J.G. Sidak and D.F. Spulber, *Deregulatory Takings and the Regulatory Contract*, Cambridge University Press, Cambridge, 1997.

[40] A. Iimi, "Estimating Demand for Cellular Phone Services in Japan", *Telecommunications Policy* 29, 2005, pp 3-23.

[41] Y. Okada and K. Hatta, "The Interdependency of Telecommunications Demand and Efficient Price Structure", *Journal of the Japanese and International Economies* 13, 1999, pp 311-35.

[42] The formal analysis from Rohlfs (1979) is also summarised in S.J. Brown and D.S. Sibley, *The Theory of Public Utility Pricing*, Cambridge University Press, 1986, pp 197-9.

$\hat{u}_j$ = utility of individual j if he/she joins the network;

u_j = utility of individual j if he/she does not join the network; and

y_j = individual j's income

while the marginal social value of access at p 5 is equal to

$$msv_j = \sum_k \frac{\hat{u}_k - u_k}{\partial u_k / \partial y_k}$$

where, msv_j = marginal social value of individual j's joining the network

$\hat{u}_k$ = individual k's utility if j joins the network;

u_k = individual k's utility if j does not joins the network; and

y_k = k's income

Rohlfs then defines the "externality factor" for access at p 6 as being equal to

$$e = \frac{1}{n} \sum_{\substack{\text{marginal} \\ \text{subscribers} \\ j}} \frac{msv_j}{mpv_j}$$

where, n = the number of marginal subscribers to the network

That is, the externality factor e is the average ratio of the marginal social value to the marginal private value for marginal subscribers. Rohlfs outlines at p 6 in equation (1-4), that the marginal condition for an unconstrained Pareto optimum now requires that,

$$e \times P_{access} = \text{Marginal Cost of Access} \qquad \text{(B.31)}$$

or alternatively that the price for the access service or subscription is set equal to the $(\text{Marginal Cost of Access})/e$.

The problem of estimating network externalities is well documented. For example:

- Rohlfs (1979) states at p 6 that, "measuring externalities to determine e is extraordinarily difficult";

- Brown and Sibley state at p 198 that: "Clearly, it is difficult to measure e"; and

- Crandall and Sidak (2004) state on p 299 that: "The difficult question is how to value the network externality".[43]

In the absence of any empirical or econometric estimate for the network externality, Rohlfs (1979) outlines on p 6, that "the appropriate value of e must be determined by judgement". For example, by arguing on p 7 that it is "not unreasonable" to assume the value of the communication link is equal between the existing and new subscriber, he establishes that if there is no way to internalise the externality, the marginal social benefit will be twice the marginal private benefit to that user and e = 2. Accounting though for the possibility that the network externality can be internalised, Rohlfs concludes on p 7 that,

> …the appropriate value of e may be considerably less than 2. However, it would exceed 1 to the extent that the externality cannot be completely internalised.

This leads Rohlfs to analyse the economically efficient price for the Bell System's fixed-line telephony operations employing the values for e of 1, 1.5 and 2, which he claims, "covers a reasonable range of plausible values of e".

Griffin (1982) used the externality factor developed by Rohlfs, along with the same values for the term "e", in his later analysis on the efficient pricing for the US fixed-

---

[43] R.W. Crandall and J.G. Sidak, "Should Regulators Set Rates to Terminate Calls on Mobile Networks", *Yale Journal on Regulation* 21, 2004, pp 264-319.

line telephony system.[44] Like Rohlfs, Griffin found that the economically efficient prices were not very sensitive to the choice of e.

In recognition of Griffin's work representing the first time that the Rohlfs' externality factor appeared in a published Journal article, the term "e" is now referred to as the "Rohlfs-Griffin" factor (or R-G factor), and based upon the initial values chosen by Rohlfs, it appears that it is now accepted that the R-G factor has a lower bound of 1 and an upper bound of 2. While the R-G factor was initially used to assess optimal pricing in the fixed-line telephony system, the R-G factor has also recently been employed to assess optimal pricing in the mobile network in the UK. For example, in the UK the regulator used a value for e of between 1.3 and 1.7, in order to estimate the efficient price for the mobile termination access service.[45]

B.5.3 Depicting the R-G Factor and the Impact of the Network Externality

The constant value of the R-G externality factor e, where $1 < e \leq 2$, implies that wherever a subscriber derives a positive marginal private benefit from subscription, there must exist a positive marginal network externality. Further, the constant value of the ratio implies that, as suggested would occur in the analysis in Section 5.1, the marginal network externality becomes less significant as the level of subscription rises. The impact of the R-G factor and the resulting estimated deadweight-loss in competitive market framework are illustrated in the price-subscription space diagram in Figure B.12. A similar diagram appears in Bomsel, Cave, Le Blanc and Neumann (BCLN, 2003) at p 22,[46] which they note was also used in the UK by OFTEL and the Competition Commission.

---

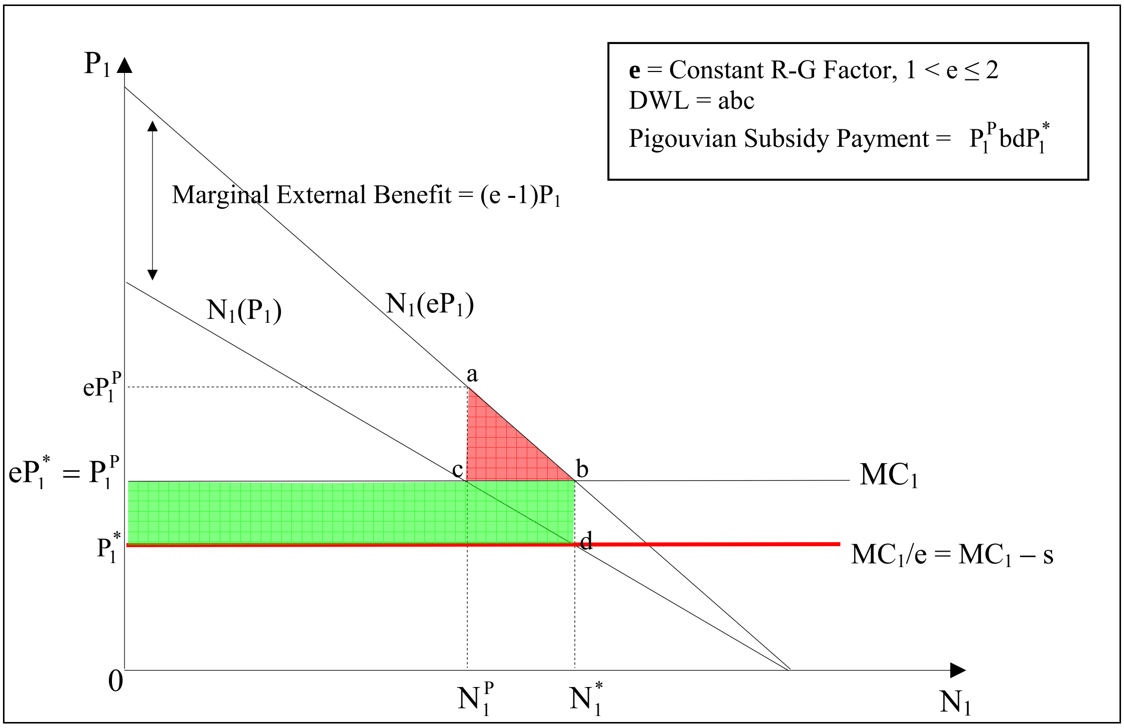
44 J.M.Griffin, "The Welfare Implications of Externalities and Price Elasticities for Telecommunications Pricing", *The Review of Economics and Statistics* 64, 1982, pp 59-66.

45 The UK regulator used a value for the R-G factor of between 1.3 and 1.7, in order to estimate the efficient price for the mobile termination access service. See Oftel, *Review of the Charge Control on Calls to Mobiles*, 26 September 2001, p 72, paragraph A4.45, available at: http://www.ofcom.org.uk/static/archive/oftel/publications/mobile/ctm0901.pdf.

46 O. Bomsel, M. Cave, G. Le Blanc and K-H Neumann, "How Mobile Termination Charges Shape the Dynamics of the Telecom Sector", Final Report, CERNA, wik Consult, University of Warwick, 9 July 2003, available at http://www.cerna.ensmp.fr/Documents/OB-GLB-F2M-FinalReport.pdf.

The diagram illustrates that a competitive market for subscribers leads to a marginal cost-based price for access of P_1^P being charged to each of the N_1^P subscribers. As the competitive provider fails to take into account the positive network externality, which is estimated using a given value of the R-G factor e, where $1 < e \leq 2$, it supplies below the estimated welfare-maximising level of subscription services N_1^*. The under-supply of subscription, equal to $(N_1^* - N_1^P)$, results in a DWL of area abc. That is, because the pricing regime does not take into account the network externality estimated using the R-G factor, a socially sub-optimal network size is reached.

FIGURE B.12 DEPICTING THE NETWORK EXTERNALITY USING THE R-G FACTOR



Where it is assumed there are no marginal DWLs associated with raising tax revenues,[47] the Government can achieve the first-best level of subscription N_1^* that removes the DWL, by providing a Pigouvian subsidy on each unit of access of $s = P_1^*(e - 1)$.[48] This subsidy decreases the private marginal cost of supply to MC_1/e.

---

[47] H.F. Campbell and K.A. Bond, "The Cost of Public Funds in Australia", *Economic Record* 73, 1997, pp 22-34, estimate that in Australia the marginal cost to society for a $1 of government expenditure is approximately $1.24.

[48] Many authors use the spelling "Pigovian" in describing this type of subsidy. While this spelling appears to be widely accepted by many academic journals and economists, particularly in the US, strictly speaking the correct spelling is Pigouvian. The reason is that only this spelling gives the appropriate recognition to the original exponent of this type of tax or subsidy, the eminent economist A.C. Pigou.

In the above diagram, as $N_1(P_1)$ represents the inverse demand curve for subscription, the welfare gain from taking into account the network externality — captured by area abc — can be estimated using the equation,

$$\int_{MC_1}^{eMC_1} N_1(eP_1)dP_1 - N_1(MC_1)MC_1(e-1)$$

while the total subsidy payment S — captured in the diagram by area $P_1^P bdP_1^*$ — will be equal to,

$$S = sN_1^* = P_1^* N_1^* (e-1) = TR_1^* (e-1)$$

BCLN (2003) note at pp 22-3, that while the total subsidy payment outlined above yields the optimal outcome, it should be regarded as a maximum total subsidy payment S_{max}. The authors illustrate that if the subsidy could be individually tailored to each of the marginal subscribers from N_1^P to N_1^*, then a minimum subsidy payment of S_{min} could be made, where $S_{min} = \frac{e}{2}\left(N_1^* - N_1^P\right)$,[49] and is captured by the area cbd in Figure B.12. In assessing the feasibility of such a minimum subsidy payment, BCLN state at p 23 that while "perfect price discrimination is unrealistic in the real world, it is also not appropriate to assume that no opportunities for price differentiation exist."

B.5.4 R-B Pricing with the R-G Factor

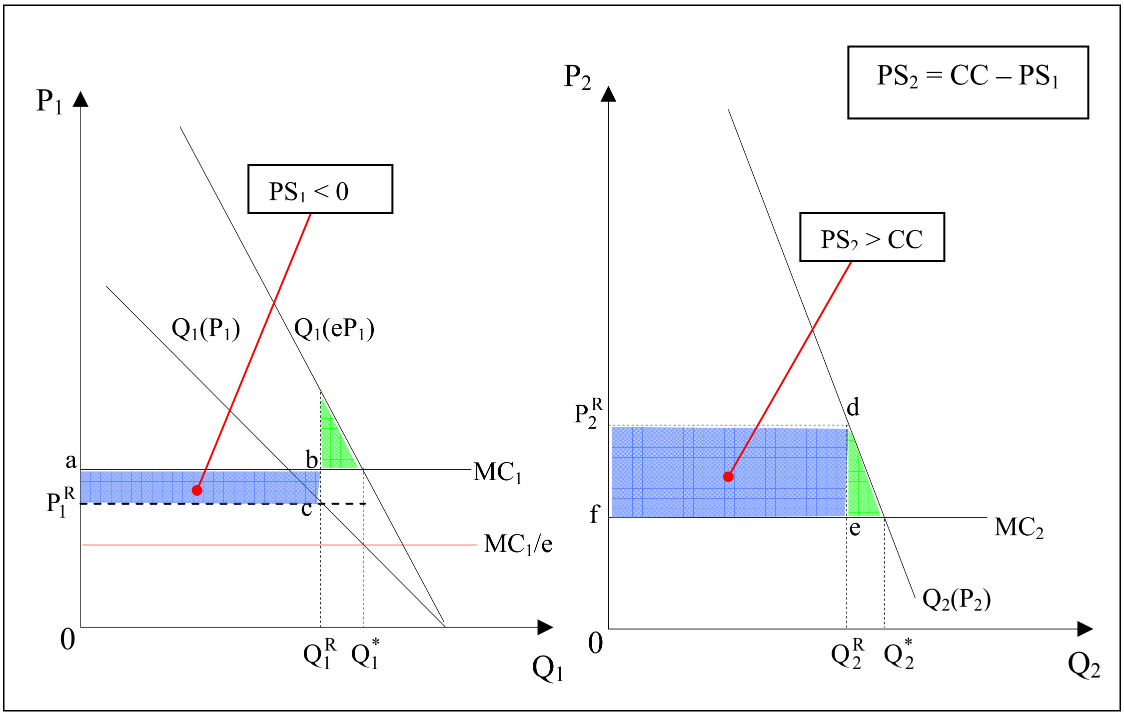
The previous section illustrates that where the R-G factor is used to estimate the network externality for the access or subscription service, the efficient outcome requires that the price for subscription is set equal to MC/e rather than the standard marginal cost of supply MC. Consequently, where there are two services, common network costs, and an externality estimated by the R-G factor e on the access service, there will be a change in the standard R-B pricing outcomes. Now the price for the access or subscription service needs to be marked up above MC/e, while the price for the other service will have the standard mark up above marginal cost. Compared to

[49] BCLN appear to get the equation for S_{min} slightly incorrect, as they do not multiply the expression by ½.

the instance where there are no externalities, there will now be a lower R-B price for subscription, which means that a higher R-B price is required for the other service in order to still recover the common network cost CC. As it is possible for the marked-up R-B price of the access service to still be below the marginal cost of production, the other service may even be required to contribute an amount that exceeds the common cost CC. Such an outcome is illustrated in Figure B.13, where it is assumed that:

- there is a common network cost of CC;
- there are two services 1 and 2;
- service 1 is an access service subject to a network externality e. Here, instead of using the notation N, the term Q is employed to denote the number of subscribers or the quantity of demand for the access service; and
- there are no cross-price effects (i.e. $\frac{\partial Q_i}{\partial P_j} = 0$, i,j = 1,2 & i ≠ j).

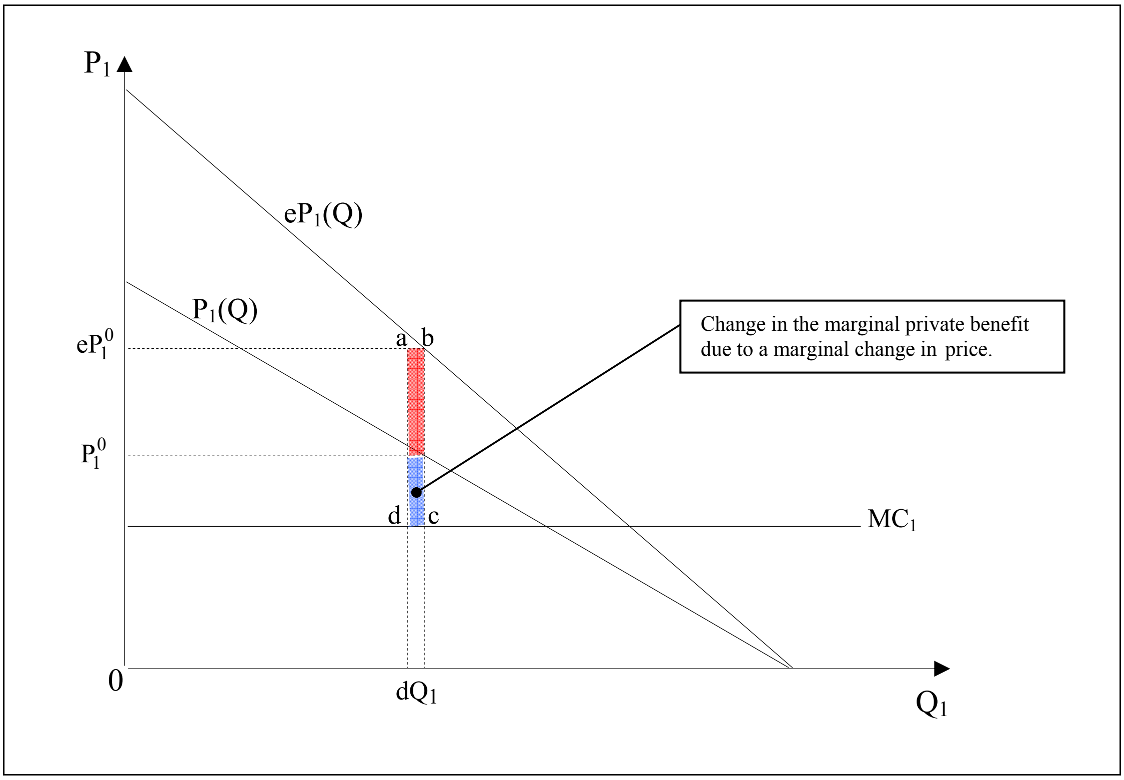
FIGURE B.13 R-B PRICING WITH AN ACCESS EXTERNALITY



In the diagram the R-B price for service 1 leads to a price of P_1^R that is set above MC_1/e, but lies below the marginal cost of providing service 1, MC_1. The result is a loss or negative producer surplus in market 1 (PS_1) equal to the blue-shaded rectangular area $abcP_1^R$, and a DWL equal to the green-shaded triangle area. In market 2, the R-B for the service is such that it yields a producer surplus (PS_2) equal to area $P_2^R def$, which is greater than the common cost of production CC. This blue-shaded area is in fact equal to the common cost plus an amount to cover the negative producer surplus obtained in the access market, i.e. $P_2^R def = CC + abcP_1^R$.

Adopting the assumptions outlined above, it is possible to derive formally the adjustment that is required to the standard R-B pricing formula in equation (B.10), to take into account the access externality. To assist with this derivation the diagram in Figure B.14 is also used.

FIGURE B.14 THE MARGINAL WELFARE CHANGE WITH A NETWORK EXTERNALITY



The diagram indicates that at price P_1^0, a marginal change in price will lead to a marginal change in output of dQ_1 and a marginal welfare change equal to area abcd

— the combination of the blue and red slivered areas — which can mathematically be written as, $dW_1 = (eP_1^R - MC_1)dQ_1$. While the change in profit is still captured by the expression in equation (B.8a), the change in welfare outcomes given by equation (B.8a) must now take into account the impact of the network externality that is estimated by the R-G factor.

Thus, the change in profit and welfare will be given by,

$$d\pi = (P_1^R - MC_1)dQ_1 + Q_1^R dP_1 + (P_2^R - MC_2)dQ_2 + Q_2^R dP_2 = 0 \quad \text{(B.32a)}$$

$$dW = (eP_1^R - MC_1)dQ_1 + (P_2^R - MC_2)dQ_2 = 0 \quad \text{(B.32b)}$$

$$\text{where, } dQ_i = \frac{\partial Q_i}{\partial P_i} dP_i \text{, i,j = 1,2}$$

Substituting dQ_i into equations (B.32a) and (B.32b) and rearranging, yields:

$$\frac{dP_2}{dP_1} = -\frac{(P_1^R - MC_1)\frac{\partial Q_1}{\partial P_1} + Q_1^R}{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2} + Q_2^R} \quad \text{(B.33a)}$$

$$\frac{dP_2}{dP_1} = -\frac{(eP_1^R - MC_1)\frac{\partial Q_1}{\partial P_1}}{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2}} \quad \text{(B.33b)}$$

Equating (B.33a) and (B.33b) and cross-multiplying terms,

$$\frac{(P_1^R - MC_1)\frac{\partial Q_1}{\partial P_1} + Q_1^R}{(eP_1^R - MC_1)\frac{\partial Q_1}{\partial P_1}} = \frac{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2} + Q_2^R}{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2}}$$

Substituting $(P_1^R - MC_1)\frac{\partial Q_1}{\partial P_1} = (eP_1^R - MC_1)\frac{\partial Q_1}{\partial P_1} + P_1^R(1-e)\frac{\partial Q_1}{\partial P_1}$ into the above, yields

$$\frac{(eP_1^R - MC_1)\frac{\partial Q_1}{\partial P_1} + P_1^R(1-e)\frac{\partial Q_1}{\partial P_1} + Q_1^R}{(eP_1^R - MC_1)\frac{\partial Q_1}{\partial P_1}} = \frac{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2} + Q_2^R}{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2}}$$

which can be simplified to,

$$\frac{P_1^R(1-e)\frac{\partial Q_1}{\partial P_1} + Q_1^R}{(eP_1^R - MC_1)\frac{\partial Q_1}{\partial P_1}} = \frac{Q_2^R}{(P_2^R - MC_2)\frac{\partial Q_2}{\partial P_2}}$$

This can be rewritten as

$$\frac{Q_1^R\left((1-e)\frac{\partial Q_1}{\partial P_1}\frac{P_1^R}{Q_1^R} + 1\right)}{\frac{\left(P_1^R - \frac{MC_1}{e}\right)}{P_1^R} e\frac{\partial Q_1}{\partial P_1} P_1^R} = \frac{Q_2^R}{\frac{(P_2^R - MC_2)}{P_2^R}\frac{\partial Q_2}{\partial P_2} P_2^R}$$

$$\frac{\left(-(1-e)\left(-\frac{\partial Q_1}{\partial P_1}\frac{P_1^R}{Q_1^R}\right) + 1\right)}{-\frac{\left(P_1^R - \frac{MC_1}{e}\right)}{P_1^R} e\left(-\frac{\partial Q_1}{\partial P_1}\frac{P_1^R}{Q_1^R}\right)} = \frac{1}{-\frac{(P_2^R - MC_2)}{P_2^R}\left(-\frac{\partial Q_2}{\partial P_2}\frac{P_2^R}{Q_2^R}\right)}$$

As the own-price elasticity of demand for good i is $\varepsilon_i = -\frac{\partial Q_i}{\partial P_i}\frac{P_i}{Q_i}$, the above equality can be rearranged to give,

$$-\frac{\left(P_1^R - \frac{MC_1}{e}\right)}{P_1^R}\left[\frac{e\varepsilon_1^R}{1-(1-e)\varepsilon_1^R}\right] = -\frac{\left(P_2^R - MC_2\right)}{P_2^R}\varepsilon_2^R$$

Multiplying this through by –1 gives an expression for the R-B prices adjusted for the R-G factor that is assumed to capture the network externality on the access service 1,

$$\boxed{\frac{\left(P_1^R - \frac{MC_1}{e}\right)}{P_1^R}\left[\frac{e\varepsilon_1^R}{1-(1-e)\varepsilon_1^R}\right] = \frac{\left(P_2^R - MC_2\right)}{P_2^R}\varepsilon_2^R} \qquad \text{(B.34)}^{50}$$

While Rohlfs first proposed that an adjustment was required to the efficient R-B price in the presence of an externality for the access service, it was Griffin in equation (16) at p 65, who first provided the type of expression outlined in equation (B.34).

Finally, the expression actually provided by Griffin in equation (16) does differ slightly from that outlined in equation (B.34). Griffin found that,

$$\boxed{\frac{\left(P_1^R - \frac{MC_1}{e}\right)}{P_1^R}\left[\frac{e\varepsilon_1^R}{1+(1-e)\varepsilon_1^R}\right] = \frac{\left(P_2^R - MC_2\right)}{P_2^R}\varepsilon_2^R} \qquad \text{(B.35)}$$

This difference between the two outcomes is that instead of defining the own-price elasticity as being $\varepsilon_i = -\frac{\partial Q_i}{\partial P_i}\frac{P_i}{Q_i}$, Griffin instead used $\varepsilon_i = \frac{\partial Q_i}{\partial P_i}\frac{P_i}{Q_i}$ in his analysis.[51] The outcome in the above equation is also presented in the Appendix of Brown and Sibley (1986) at p 199, although it appears that Brown and Sibley should actually be using the expression in equation (B.34) rather than that in equation (B.35), as they assume earlier at p 195 that the own-price elasticity is captured by the expression $\varepsilon_i = -\frac{\partial Q_i}{\partial P_i}\frac{P_i}{Q_i}$.

---

50 Where there are cross-price effects a similar formula is derived, but the own-price elasticity of demand terms in equation (B.34) are replaced by superelasticity terms. That is,

$$\frac{\left(P_1^R - \frac{MC_1}{e}\right)}{P_1^R}\left[\frac{e\hat{\eta}_1^R}{1-(1-e)\hat{\eta}_1^R}\right] = \frac{\left(P_2^R - MC_2\right)}{P_2^R}\hat{\eta}_2^R, \quad \hat{\eta}_i^R = \varepsilon_i\frac{\varepsilon_i\varepsilon_j - \varepsilon_{ij}\varepsilon_{ji}}{\varepsilon_i\varepsilon_j - \varepsilon_i\varepsilon_{ij}}, \ i,j = 1,2 \ \& \ i \neq j$$

As the derivation is slightly more complex than is the case for the example with no cross-price effects, which is outlined above, it is not provided in this Appendix.

51 While Griffin does not explicitly outline the formula used for the own-price elasticity, the paper consistently presents the numerical estimates as negative numbers.

B.5.5 The Call Externality

The call externality arises because a telephone call provides benefits to both the originator and the recipient of the call, but only one party pays for the service — i.e. normally the calling party. Mitchell and Vogelsang define the call externality at p 55 as, "the benefit of a call to the party that does not have to pay for the call", and Brown and Sibley note at p 197 that where the benefit to the recipient and the originator is the same, "the social benefit from the call is twice the marginal cost of the call".

Squire (1973) was one of the first to formally analyse the impact of the call externality on efficient pricing.[52] He established that in the presence of a call externality, pricing according to marginal cost was no longer efficient. Instead Squire states on p 524 that

> In general, the price of calls should be less than their marginal costs by a margin representing the external benefit received by the "callee."

A similar outcome is derived by Armstrong (2002) using a model that is designed to capture the impact of call externalities on the price of mobile termination. He finds on p 344 that:[53]

> Therefore, if mobile subscribers derive a benefit from incoming calls, then the regulator should set the termination charge *below* cost in order to encourage calls from the fixed sector.

Economists such as Littlechild (1975),[54] Rohlfs (1979), Willig (1979) and Brown and Sibley (1986), however, suggest that in practice call externalities do not provide a strong case for a call price reduction. They note that unlike the marginal network externality, which affects all subscribers to the network, the call externality only

---

52 L. Squire, "Some Aspects of Optimal Pricing for Telecommunications", *Bell Journal of Economics* 4, 1973, pp 515-25.

53 The outcome is mathematically shown by Armstrong in equation (46) on p 343.

54 S.C. Littlechild, "Two-part Tariffs and Consumption Externalities", *Bell Journal of Economics* 6, 1975, pp 661-70.

impacts upon two callers, so it can probably be efficiently internalised.[55] This is highlighted by Willig, who states on p 133 that:

> The expression in equation (55) makes it clear that the marginal network externality effects that are relevant for pricing are potentially spread over all consumers with network access. In contrast, the effects of potential uninternalized values of incoming flows discussed earlier were concentrated on one consumer. For this reason, it was argued that such values were indeed likely to be privately internalized.

Further, Rohlfs (1979) provides the example on p 5 that where two frequent callers arrange to call each other half the time, the call externality will be roughly internalised. Rohlfs (2002) notes that in the UK Oftel reached a similar conclusion in assessing the call externality for mobile phone services. It considered that even if corrective pricing could not be implemented, users would still be able to internalise a large portion of their call externalities.[56]

The summary of the literature by Hahn (2003) seems to suggest that another possible explanation for call externalities being ignored is that, such externalities are very difficult to measure.[57] He notes on p 950 that "there has been no market mechanism by which consumers can express their preference for incoming calls". Similarly, Mitchell and Vogelsang (1991) outlines on p 60 that a difficulty when assessing call externalities is that the demand for incoming calls is not readily observable, and cannot easily be related to other variables such as, outgoing calls, other purchases, or the number of subscribers. Mitchell and Vogelsang on p 60, and Hahn on p 950, also question the ease with which call externalities can actually be internalised by parties. Citing the work of Acton and Vogelsang (1990),[58] they note that the Coase theorem is hard to apply in such situations, because the negotiations between a caller and a receiver that would lead to the internalisation of the call externalities, themselves,

---

[55] Brown and Sibley (1986) also outline on p 197 that there are likely to be telephone calls that people will be annoyed to receive, and these will impose a negative call externality on the call recipient.

[56] J.H. Rohlfs, *Annex A: Network Externalities and Their Internalization with Respect to the UK Mobile Market Network*, 19 April 2002, p 3, available at http://www.ofcom.org.uk/static/archive/oftel/publications/mobile/ctm_2002/annex_a.pdf.

[57] J-H Hahn, "Nonlinear Pricing of Telecommunications with Call and Network Externalities", *International Journal of Industrial Organization* 21, pp 949-67.

[58] J.P. Acton and I. Vogelsang, "Telephone Demand Over the Atlantic — Evidence from Country-Pair Data", Technical Report R-3715-NSF/MF, RNAD, Santa Monica, 1990.

require costly telephone calls to be made. Both authors also summarise the results from Einhorn (1990),[59] and Mitchell and Vogelsang states on p 61 that:

> The main result that can be derived in this context is that the importance of the call externality relative to the network externality increases with subscriber penetration.

Therefore, in relation to the optimal R-B prices, Mitchell and Vogelsang conclude that, "the price/marginal-cost markup for calls, relative to that for access, should decrease as penetration rate increases."

59 M. A. Einhorn, "Regulatory Biases in Network Pricing with Access and Usage Externalities", mimeo, 1990.

B.6 R-B Pricing — Some Worked Examples

The analysis in this Section provides a number of worked examples that shows how the R-B prices can be derived when it is assumed there is once again a common cost of production to recover and there is:

- linear demand and no cross-price effects; and
- constant elasticity demand and no cross-price effect.

B.6.1 R-B Prices with Linear Demand and No Cross-Price Effects

It is assumed in Section B.5.1 that the inverse linear demand curve for service i is,

$$P\,(Q_i) = A_i - B_iQ_i, \text{ where } A_i > 0,\ B_i > 0,\ \&\ i = 1, 2, \ldots n \quad \text{(B.36)}$$

As the demand curve is linear, the elasticity of demand at any given price P_i and quantity Q_i will now be equal to $\varepsilon_i = P_i(Q_i)/B_iQ_i$. As B_i is constant, the elasticity of demand will vary at different prices or different points along the demand curve. Therefore, where estimated elasticities and initial prices are being used to calculate the linear demand curve, it should be expect that the resulting elasticities at the R-B price will not be the same as those at the initial prices.

B.6.1.1 Consumer and Producer Surplus, Deadweight-loss, and the Overall Welfare under Linear Demand

The inverse demand curve given by equation (B.36) is illustrated in Figure B.15. Assuming there is some arbitrary price for each service i, P_i^0 — where P_i^0 > MC_i — quantity Q_i^0 will be consumed where,

$$Q_i^0 = \left(A_i - P_i^0\right)/2B_i \quad \text{(B.37a)}$$

Further, there will be the following expressions for the consumer surplus, producer surplus and deadweight-loss in markets i = 1,…n.

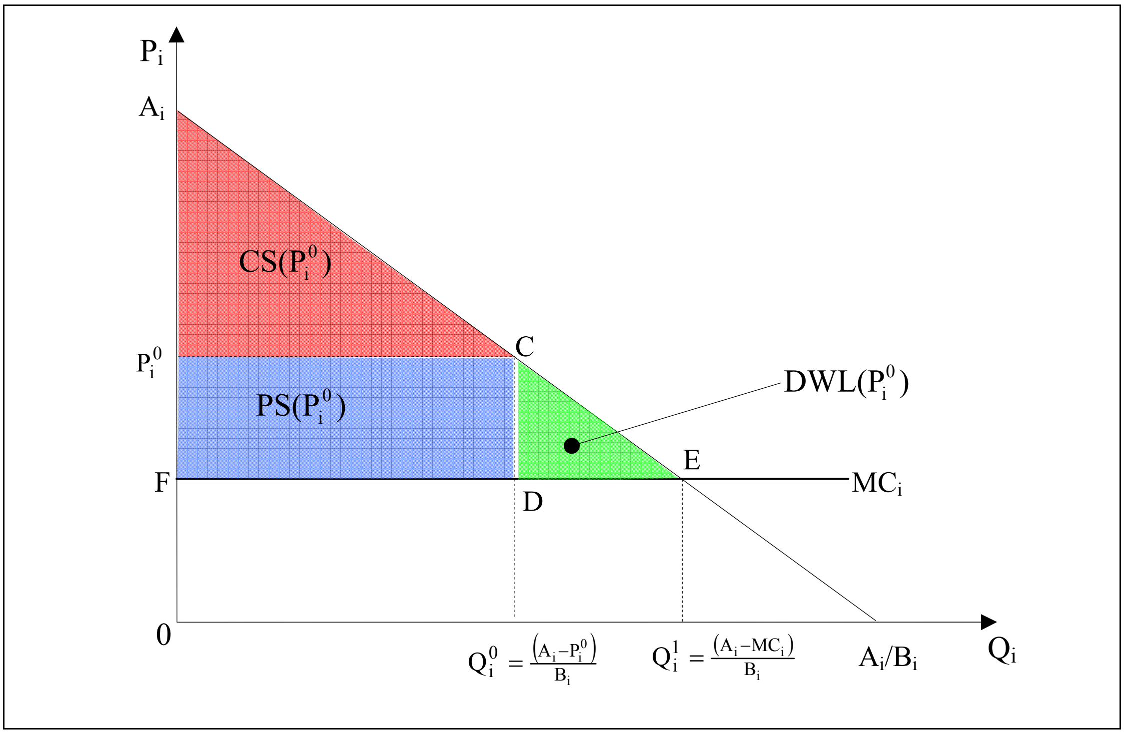
$$CS(P_i^0) = \frac{(A_i - P_i^0)^2}{2B_i} \tag{B.37b}$$

$$PS(P_i^0) = \frac{(P_i^0 - MC_i)(A_i - P_i^0)}{B_i} \tag{B.37c}$$

$$DWL(P_i^0) = \frac{(P_i^0 - MC_i)^2}{2B_i} \tag{B.37d}$$

The areas capturing the outcomes in equations (B.37b) to (B.37d) are illustrated in Figure B.15. The red-shaded triangle area $A_i CP_i^0$ captures the consumer surplus, the blue-shaded rectangle area $P_i^0 CDF$ captures the producer surplus, and the green-shaded triangle area CED captures the deadweight-loss.

FIGURE B.15 CONSUMER AND PRODUCER SURPLUS AND THE DEADWEIGHT-LOSS



As there is a common cost of the network providing these services of CC, the overall level of profit (i.e. π), welfare (i.e. S) or inefficiency at price P_i^0, i = 1,…n, can formally be written,

$$\pi = \sum_{i=1}^{n} PS(P_i^0) - CC \tag{B.38a}$$

$$S = \sum_{i=1}^{n} \left(CS(P_i^0) + PS(P_i^0) \right) - CC = \sum_{i=1}^{n} CS(P_i^0) + \pi \tag{B.38b}$$

$$\text{Total Inefficiency } = \sum_{i=1}^{n} DWL(P_i^0) \tag{B.38c}$$

It is important to note here that these expressions for the overall level of profit, welfare and inefficiency at the given price P_i^0 are not exclusive to the linear demand curve, and will also hold in the next section for the analysis using the constant elasticity demand curve.

B.6.1.2 Solving for the R-B Price

With linear inverse demand curve in equation (B.36) the expression in equation (B.10) generalised for n services, simplifies to

$$\frac{P_i^R - MC_i}{A_i - P_i^R} = \frac{P_j^R - MC_j}{A_j - P_j^R}, \ \forall \ i \neq j \ \& \ i,j = 1,...n \tag{B.39}$$

Solving for the price of service j as a function of the R-B price of service i yields

$$P_j^R = \frac{P_i^R \left(A_j - MC_j \right) - A_j MC_i + A_i MC_j}{A_i - MC_i}, \ \forall \ i \neq j \ \& \ i,j = 1,..n \tag{B.40}$$

As the prices are set so as to recover the common network costs CC and the allowed level of profit $\overline{\pi}$ = 0, the above prices must also satisfy the constraint,

$$\sum_{j=1}^{n} \left(P_j^R - MC_j \right) Q_j^R = CC + \overline{\pi}, \ j = 1,...n \tag{B.41}$$

While R-B prices are normally derived for outcomes where $\overline{\pi}$ = 0, the above equation also allows for the derivation of prices when there is some level of positive profit (i.e. $\overline{\pi} > 0$). As outlined in Section B.3, when there are positive profits allowed, the resulting prices will have the same structure as R-B prices, but will be set a higher level to efficiently distribute the common costs and the allowed rents across services i = 1,…n. The prices in this instance — where $\overline{\pi} > 0$ — maximise welfare for the given level of common cost and the allowed level of profit.[60]

Substituting in an expression for Q_j^R using the equation for the linear demand curve (i.e. $Q_j^R = \frac{A_j - P_j^R}{B_j}$), and substituting equation (B.40) into equation (B.41), yields a quadratic expression in P_i^R,

$$\left(P_i^R\right)^2 - P_i^R\left(A_i + MC_i\right) + \left(A_i MC_i + \frac{CC+\overline{\pi}}{\sum_{j=0}^{n} \frac{1}{B_j}\left(\frac{A_j - MC_j}{A_i - MC_i}\right)^2} \right) = 0 \qquad \text{(B.42)}$$

Applying the quadratic formula the expression for the R-B price for each service i can be derived.

$$P_i^R = \frac{\left(A_i + MC_i\right) - \sqrt{\left(A_i - MC_i\right)^2 - 4\left(\frac{CC+\overline{\pi}}{\sum_{j=0}^{n} \frac{1}{B_j}\left(\frac{A_j - MC_j}{A_i - MC_i}\right)^2}\right)}}{2}, \; i = 1,\ldots n \qquad \text{(B.43)}$$ [61]

The outcome in equation (B.43) can be substituted into equations (B.37a) to (B.37d), to derive expressions for the quantity demanded, the consumer surplus, the producer surplus and the deadweight-loss at the R-B price for the respective services 1 through to n. Similarly using these outcomes the solutions for the overall profit, overall welfare and total inefficiency can be found using equations (B.38a) through to (B38.c).

As the welfare-maximising competitive market price P_i^* in each of these markets is just $P_i^* = MC_i$, and the unregulated monopoly price P_i^m under linear demand is

---

60 R.R. Braeutigam, "An Analysis of Fully Distributed Cost Pricing in Regulated Industries", *Bell Journal of Economics* 11, 1980, pp 182-96, outlines that the R-B prices can be derived for greater than zero profit on p 189 footnote 14, and on p 193 footnote 17.

61 There is also a maximum value from solving the quadratic, however that solution can be ignored as the price exceeds the monopoly price that the firm can charge in market i.

simply $P_i^m = \dfrac{A_i + MC_i}{2B_i}$, equations (B.37a) through to (B.37d) and equations (B.38a) through to (B.38c), can also be used to solve for all the competitive and unregulated monopoly market outcomes. These can be compared with the R-B outcomes to highlight the welfare superiority of R-B pricing over the unregulated monopoly outcome, but welfare inferiority of R-B pricing over the first-best competitive market outcome.

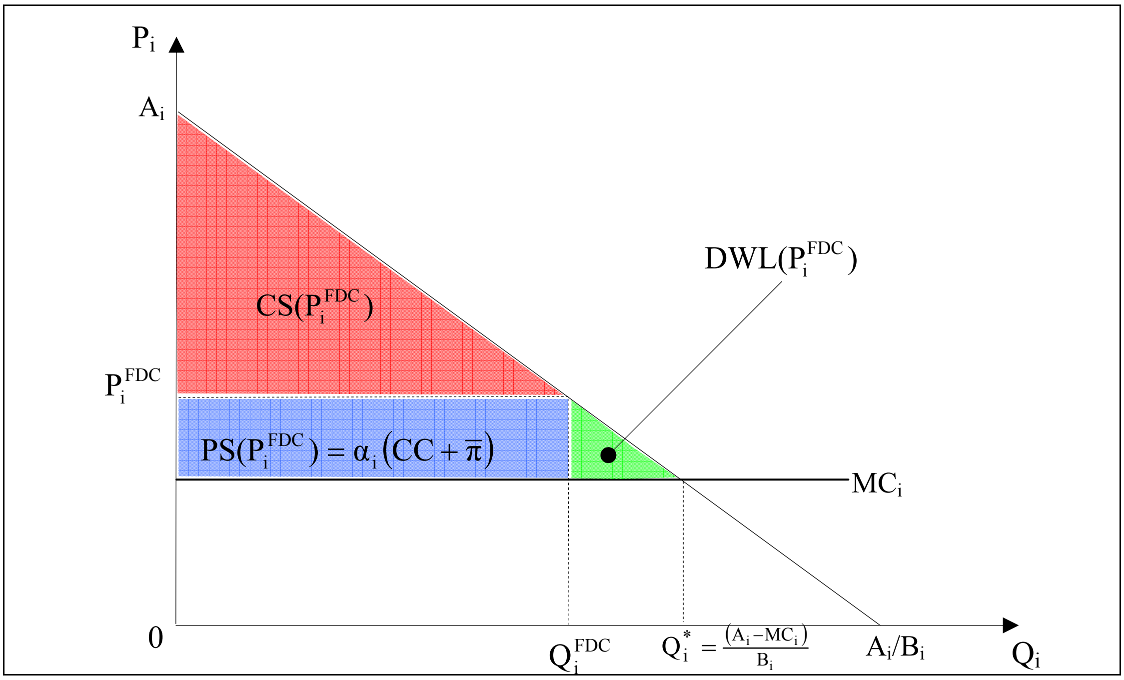
B.6.1.3 Comparing R-B Pricing with Fully-Distributed Cost (FDC) Pricing

Under FDC pricing for each service i (i.e. P_i^{FDC}) allows the firm to achieve common cost recovery and its allowed level of profit $\bar{\pi}$. This is done by letting the firm obtain a producer surplus in each market that is equal to some proportion α_I of its common cost and allowed profit. That is,

$$\left(P_i^{FDC} - MC_i\right)Q_i^{FDC} = \alpha_i\left(CC + \bar{\pi}\right),\ \ i = 1, \ldots n\ \&\ \sum_{i=0}^{n} \alpha_i = 1 \qquad \text{(B.44)}$$

A diagram capturing the outcome in equation (B.44) is illustrated in Figure B.16.

FIGURE B.16 FULLY-DISTRIBUTED COST PRICING



In equation (B.44) the term α_I can be arbitrarily chosen, or as Braeutigam (1980) outlines,[62] set through three specific methods based upon:

(1) Relative Output (i.e. $\alpha_i = \frac{Q_i}{\sum_{i=1}^{n} Q_i}$);

(2) Attributable Cost i.e. ($\alpha_i = \frac{C_i}{\sum_{i=1}^{n} C_i}$); or

(3) Gross Revenue i.e. ($\alpha_i = \frac{TR_i}{\sum_{i=1}^{n} TR_i}$).

In order to derive an expression for P_i^{FDC}, $Q_i^{FDC} = \frac{(A_i - P_i^{FDC})}{B_i}$ is substituted into equation (B.44), resulting in the following quadratic expression for P_i^{FDC},

$$\left(P_i^{FDC}\right)^2 - \left(A_i + MC_i\right)P_i^{FDC} + A_i MC_i + \alpha_i B_i F = 0\text{, } i = 1,\ldots n \qquad \text{(B.45)}$$

Using the quadratic formula the solution for P_i^{FDC} is,

$$P_i^{FDC} = \frac{(A_i + MC_i) - \sqrt{(A_i - MC_i)^2 - 4\alpha_i B_i CC}}{2}\text{, } i = 1,\ldots n \qquad \text{(B.46)}$$[63]

Using methodology identical to that done under R-B pricing, expressions for output, consumer surplus, producer surplus, and the DWL in each market i can be derived,

62 R.R. Braeutigam, "An Analysis of Fully Distributed Cost Pricing in Regulated Industries", *Bell Journal of Economics* 11, 1980, pp 182-96.

63 There are of course actually two prices from solving the quadratic. However, the maximum price can be rejected, as this price exceeds the price that would be charged by an unregulated monopoly in market i.

along with the overall levels of profit, welfare and inefficiency. The outcomes under FDC pricing can then be compared with those under R-B pricing, and used to highlight the welfare superiority of the cost allocation mechanism under R-B pricing.

B.6.2 R-B Prices with Constant Elasticity Demand and No Cross-Price Effects

As the title suggests, unlike the linear demand curve, the price elasticity of demand is the same or constant along all points of the constant elasticity demand curve, and where the term $\varepsilon_i > 0$ denotes the price elasticity of demand, the constant elasticity demand curve for service i will take the form,

$$Q_i = A_i P_i^{-\varepsilon_i}\text{, where } A_i > 0,\ i = 1, 2, \ldots n \qquad \text{(B.47)}$$

Therefore, the resulting R-B prices will have the same elasticity as any other price that could be set in market i.

B.6.2.1 Consumer and Producer Surplus, Deadweight-loss, and the Overall Welfare under Constant Elasticity Demand

The demand curve given by equation (B.47) is illustrated in Figure B.17. Assuming there is some arbitrary price for each service i, P_i^0 — where $P_i^0 > MC_i$ —quantity Q_i^0 will be consumed where,

$$Q_i^0 = A_i \left(P_i^0\right)^{-\varepsilon_i} \qquad \text{(B.48a)}$$

Further, there will be the following expressions for the consumer surplus, producer surplus and deadweight-loss in markets i = 1,…n.

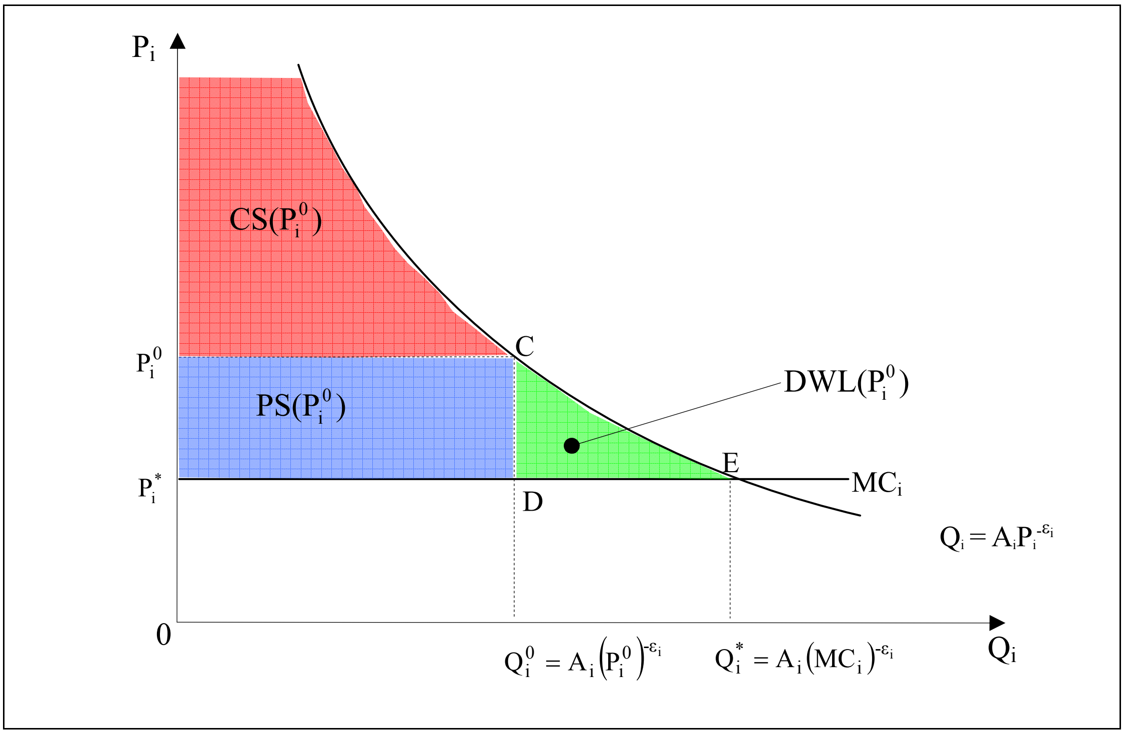
$$CS(P_i^0) = \lim_{\hat{P}_i \to \infty} \int_{P_i^0}^{\hat{P}_i} Q_i dP_i = \begin{cases} \dfrac{-A_i P_i^{0(1-\varepsilon_i)}}{1-\varepsilon_i}, & \text{if } \varepsilon_i > 1 \\ \infty & \text{, if } \varepsilon_i \le 1 \end{cases} \qquad \text{(B.48b)}^{64}$$

$$PS(P_i^0) = \left(P_i^0 - MC_i\right) A_i \left(P_i^0\right)^{-\varepsilon_i} \qquad \text{(B.48c)}$$

$$DWL(P_i^0) = A_i \left[\left(\frac{(P_i^0)^{-\varepsilon_i} - MC_i^{-\varepsilon_i}}{1-\varepsilon_i} \right) - \left(P_i^0 - MC_i\right)(P_i^0)^{-\varepsilon_i} \right] \qquad \text{(B.48d)}$$

The areas capturing the outcomes in equations (B.48b) to (B.48d) are illustrated in Figure B.17. The red-shaded area captures the consumer surplus, the blue-shaded rectangle area P_i^0CDF captures the producer surplus, and the green-shaded triangle area CED captures the deadweight-loss.

FIGURE B.17 CONSUMER AND PRODUCER SURPLUS AND THE DEADWEIGHT-LOSS



64 To find the solution for the consumer surplus with constant elasticity of demand at price i.e. $CS(P_i^0)$, the indefinite integral is solved to give: $CS(P_i^0) = \lim_{\hat{P}_i \to \infty} \int_{P_i^0}^{\hat{P}_i} \left(A_i P_i^{1-\varepsilon_i}\right) dP_i = \lim_{\hat{P}_i \to \infty} \left[\frac{A_i P_i^{1-\varepsilon_i}}{1-\varepsilon_i} \right]_{P_i^0}^{\hat{P}_i} = \lim_{\hat{P}_i \to \infty} \left[\frac{A_i \hat{P}_i^{1-\varepsilon_i}}{1-\varepsilon_i} \right] - \left[\frac{A_i P_i^{0\,1-\varepsilon_i}}{1-\varepsilon_i} \right]$.

Both Equation (B.48b) and the diagram indicate that unlike linear demand, with constant elasticity demand, the consumer surplus will be undefined. While an approximation can be made, solutions for the indefinite integral will only exist when $\varepsilon_i > 1$. Therefore, while equations (B.38a) through to (B.38c) can once again be used to solve for the overall level of profit (i.e. Π), welfare (i.e. S), or inefficiency at price P_i^0, the solution for overall welfare will also be undefined when $\varepsilon_i > 1$. Consequently, under constant elasticity of demand, a more meaningful welfare comparison involves comparing the total level of inefficiency under the different pricing regimes.

B.6.2.2 Solving for the R-B Price

It was established earlier that where λ represents a constant — sometimes referred to as the "Ramsey Number" —the R-B price for n services must satisfy,

$$\frac{P_i^R - MC_i}{P_i^R}\varepsilon_i^R = \frac{P_j^R - MC_j}{P_j^R}\varepsilon_j^R = \lambda,\ \forall\ i \neq j\ \&\ i,j = 1,...n$$

Rearranging the above expression, the R-B price for service i is,

$$P_i^R = \frac{MC_i \varepsilon_i^R}{\varepsilon_i^R - \lambda},\ i = 1,\ldots n \qquad \text{(B.49)}$$

Dropping the superscript R on the elasticity term for the remainder of this Section, as in the analysis here it remains unchanged, the quantity demanded at the R-B price can be expressed as,

$$Q_i^R = A_i\left(\frac{MC_i \varepsilon_i}{\varepsilon_i - \lambda}\right)^{-\varepsilon_i},\ \ i = 1,\ldots n \qquad \text{(B.50)}$$

As the R-B price and quantity must also satisfy the constraint outlined in equation (B.41) — i.e. $\sum_{i=1}^{n}\left(P_i^R - MC_i\right)Q_i^R = CC + \bar{\pi}$, i = 1,...n — the outcomes in equations (B.49) and (B.50) can be substituted into equation (B.41) to give,

$$\sum_{i=1}^{N}\left(\frac{\lambda A_i MC_i}{\varepsilon_i-\lambda}\right)\left(\frac{MC_i\varepsilon_i}{\varepsilon_i-\lambda}\right)^{-\varepsilon_i}-CC-\bar{\pi}=0 \qquad \text{(B.51)}^{65}$$

While there is no analytical solution for λ in equation (B.51), given values for the other parameters, and by recognising that λ lies between the value of 0 and 1, numerical techniques can be used to iterate towards a numerical solution. For example, in Microsoft Excel the tool Goalseek employs such a process, and it can subsequently be used to solve for the Ramsey number.

By substituting the value of λ into equation (B.49), the R-B price can be derived, and the solution used in equations (B.48a) through to (B.48d), and equations (B.38a) through to (B.38c), to derive the relevant outcomes under the R-B price with constant elasticity demand. The result can then be compared with the competitive market outcome where price is $P_i^* = MC_i$ and the unregulated monopoly outcome where price is $P_i^m = \frac{MC_i\varepsilon_i}{\varepsilon_i - 1}$.[66]

In order to compare the R-B outcome with FDC pricing, the expression for the constant elasticity demand curve in equation (B.48a) is substituted into equation (B.44) to give,

$$A_i\left(P_i^{FDC}\right)^{1-\varepsilon_i} - A_i MC_i\left(P_i^{FDC}\right)^{-\varepsilon_i} - \alpha_i\left(CC+\bar{\pi}\right)=0,\ i = 1,\ldots n\ \&\ \sum_{i=0}^{n}\alpha_i = 1 \quad \text{(B.52)}$$

As with the R-B price, numerical techniques can once more be employed to solve for the FDC price.

65 Brown and Sibley derive the same outcome on p 41.

66 There will only be a solution for the unregulated monopoly outcome with a constant elasticity demand curve where the value of the elasticity of demand is greater than one.