TOPICS IN PRICE CAP REGULATION

Thesis submitted to the Social Studies Faculty Board
in partial fulfilment of the requirements for the Degree of D.Phil.

Simon Cowan

Worcester College
Oxford

April 1995

Abstract

TOPICS IN PRICE CAP REGULATION

Simon Cowan
Worcester College
Oxford

D.Phil. Thesis
Hilary Term 1995

This thesis examines the theoretical properties of different price cap schemes that have been applied in the UK and the USA. The objective is to assess the consequences for price structures and welfare of different ways of defining the regulated price index. Chapter 2 surveys the literature on regulation under asymmetric information that is related to price caps. Chapter 3 presents a general analysis of five main types of price cap when the regulated firm sets linear tariffs. The schemes are the Tariff Basket (TB) scheme, the Fixed Weights (F) scheme, the Average Revenue (AR) scheme, the Average Revenue (Lagged) scheme, and the Paasche Price Index (PPI) scheme. The TB and PPI schemes generate efficient price structures in the long run, whereas prices are inefficient under the other schemes. In Chapter 4 the consequences of allowing freedom to set different prices, relative to the case of uniform prices, are analyzed. The conditions for price freedom to be desirable are derived for the case where the price level is not regulated. When the price level is regulated it is shown that AR regulation can cause welfare to be below the level that obtains without any regulation. Chapter 5 contains an analysis of the five price caps examined in Chapter 3 for the case where the firm sets a two-part tariff. The AR and PPI schemes are dominated, and the conditions under which TB, F and ARL are optimal are established. Chapter 6 explores some issues in the regulation of nonlinear tariffs by AR and TB price caps. Chapter 7 considers some extensions of the analysis. It is shown that when quality is a choice variable, the regulator is concerned about income distribution and there is demand growth the TB scheme can be adapted and retains its desirable properties. Chapter 8 contains conclusions and suggestions for future work.

Acknowledgments

I would like to thank my supervisor and colleague John Vickers for his encouragement, suggestions and patience. In addition I have benefitted greatly from discussions with Mark Armstrong. Participants of several seminars at the Institute of Economics and Statistics at Oxford made very helpful suggestions. I am indebted to Barrie Wigmore, who funds the Wigmore Fellowship in Economics at Worcester College, for the payment of my University fees and computing expenses. Worcester College and the Institute of Economics and Statistics provided excellent and supportive environments in which to pursue this research. I am also grateful to the Office of Fair Trading and the Economic and Social Research Council who funded a related project on the Regulation of Firms with Market Power that I was a member of in 1991 and 1992. Finally I would like to thank my wife, Meriel Raine, for her support throughout the period that I have worked on the thesis.

TOPICS IN PRICE CAP REGULATION

Contents

List of Figures

CHAPTER 1

INTRODUCTION

The objective of this thesis is to assess theoretically some schemes that have been used to regulate the prices of natural monopolies. Price regulation of natural monopolies has been an important area of research over the last decade or so. There are two main reasons for this. First, privatization, in the UK and elsewhere, of firms that retained significant monopoly positions focused the attention of policymakers on the need to regulate prices to prevent the abuse of market power. The Littlechild report on the regulation of British Telecom's profitability (Littlechild, 1983) is an important landmark in the development of policy towards price caps. Second, theorists recognized that that regulators typically have an informational disadvantage relative to the firms that they regulate. This observation led to the construction of models that explicitly allow for information asymmetries and the consequent need to give a regulated firm *incentives* to behave efficiently.

A price cap is a contract between the regulator and the regulated firm that specifies the maximum level of price(s) for a specified length of time. The purest form of price cap is one where the level of the cap is never revised, so the firm faces an infinite "regulatory lag". Since regulators typically have less information than firms about production techniques, factor prices and the scope for cost reduction, the use of price caps offers an attractive way to provide incentives for efficient input choices. These incentives are perfect if the price cap is pure (as long as the firm is willing to participate). In practice reviews occur at regular, fixed intervals and the regulator takes account of new information to ensure that prices do not move too far out of line with costs. The longer the interval between price reviews the greater the incentive to be productively efficient.

The effect that a price cap, whether pure or not, has on input choices is now reasonably well-understood – see Armstrong, Cowan and Vickers (1994, Chapters 2 and 3) for an assessment. But *productive efficiency* is not the only concern of regulators.

Two further possible objectives of regulators are *allocative efficiency* and *the minimization of informational rents.* Laffont and Tirole (1993) discuss the trade-offs between these three objectives. A price cap provides strong incentives for productive efficiency but the need to ensure the firm's participation means that the level of the cap has to be "high", yielding a price that is generally above cost (allocative inefficiency), which in turn generates large rents for the firm. In general Laffont and Tirole argue for price controls that are intermediate between price caps and cost-of-service regulation, although in their more practical work, e.g. Laffont and Tirole (1994) on interconnection pricing, they assume that their proposals are implemented through price caps.

In a multiproduct setting there is a further complication in the design of a price cap.[1] Unless each price is capped separately, which is not the case in practice, the regulator must specify an *index* of prices. It turns out that the choice of index is of crucial importance to the overall performance of a price cap. In this thesis we assess the implications of different price cap schemes for allocative efficiency. In order to focus on this issue it is assumed that price caps are pure. This implies that we can ignore both productive efficiency issues and the objective of the minimization of informational rents. The assumption that price caps are pure also allows straightforward characterization of the properties of price cap schemes that are dynamic in the sense that the firm's constrained choice of prices in one period affects the set of prices that are available in subsequent periods. We shall show that some price cap schemes can generate price structures that are allocatively inefficient and have poor welfare properties, whereas other schemes align the regulator's and the firm's objectives and offer the prospect of both efficient pricing and good welfare performance.

There are two questions that are relevant for policymakers to which the analysis of different price caps can be applied. First, to what extent should firms be allowed to disaggregate tariffs that currently do not reflect relative costs of supply? Many utilities have been required to offer identical terms to customers irrespective of differences in the

[1]See, for example, Braeutigam and Panzar (1993).

costs of serving different customer types, such as urban and rural consumers. There is now strong pressure from regulated firms and some customers for "de-averaged" tariffs that reflect these cost differences, and careful consideration must be given to the form of the price cap scheme that is used to regulate these disaggregated charges. Second, when firms practise second-degree price discrimination, e.g. through two-part tariffs or other forms of nonlinear pricing, to what extent should they be allowed to "rebalance" different parts of the tariff? Typically regulated firms argue for the ability to raise fixed charges in two-part tariffs and lower prices of marginal purchases.

The structure of the thesis is as follows. In Chapter 2 some of the literature on regulation is surveyed and assessed. There are two main parts. Baron and Myerson (1982) and Laffont and Tirole (1986) use the techniques of information economics to develop optimizing models, where the regulator knows everything *except* one parameter of the firm's cost function. In addition the regulator is able to make transfers to the firm. Although these models provide much insight they are of limited practical use. Following Loeb and Magat (1979) and Vogelsang and Finsinger (1979) a parallel literature has developed that examines the consequences of various *ad hoc* regulatory schemes. To implement these schemes the regulator needs only limited information on sales, prices and costs. Such schemes are closely related to some of the price cap schemes that are analyzed in later chapters.

In Chapter 3 the analysis of price caps begins. It is assumed that the firm sells in multiple markets, or has many products, and that tariffs are linear. The properties of five main price cap schemes are examined. These schemes are either the same as, or closely related to, actual price cap schemes that are used in the UK and the USA. A series of questions is asked. How do the schemes rank as far as consumers, the firm and the regulator are concerned? What are the consequences of each scheme for price structures? Some preliminary answers to these questions are given, based on a general model.

Chapter 4 focuses on the implications of disaggregating uniform tariffs, although much of the analysis can be given a more general interpretation. In the first part of the chapter it is assumed that price levels are not regulated, and the welfare consequences of a move from a uniform to a de-averaged tariff are assessed using a linear model. This allows comparison with the existing literature on third-degree price discrimination. The analysis is then extended to the case where the price level is controlled using a capped price index. It is shown that the method of price regulation based on capping average revenue can have some perverse consequences for price structures, whereas other price cap schemes perform better in general.

In Chapters 5 and 6 the analysis of price caps is extended to the case of second-degree price discrimination. In Chapter 5 the firm sets a two-part tariff and a simple model is used that allows complete characterization of each scheme. It turns out that there is no scheme that dominates all the others. Different schemes can be optimal, depending on the discount factor and on the slope of the marginal cost function (relative to the demand curve). In Chapter 6 a simple model of fully nonlinear pricing is used to explore the properties of some of the price caps.

Chapter 7 contains some extensions of the earlier analysis. Regulated firms usually have flexibility in the choice of the quality of service that they offer, and the interaction of the quality and pricing decisions is assessed. Regulators are concerned about the consequences for income distribution of changes to price structures, and it is shown that such concerns can be explicitly integrated into a price cap. In some utilities there are consumption or network externalities and the implications for these for price caps is discussed. Finally we relax of one of the main assumptions of the earlier analysis, that demand conditions are stationary, and show how price caps can be adjusted to take account of demand growth.

Chapter 8 contains the conclusions to be drawn from the analysis. Suggestions for future research are also made.

CHAPTER 2

MODELS OF REGULATION WITH ASYMMETRIC INFORMATION

2.1 Introduction

The recent literature on regulation of monopolies stresses the rôle of asymmetric information. The firm is likely to have better information than the regulator about technologies, factor prices and demand conditions. In this chapter we survey some recent models of regulation with asymmetric information. The purpose of the chapter is to provide some background for the analysis of price caps in the following chapters. The focus is on the analysis of regulatory mechanisms that are designed to achieve regulatory objectives while taking into account the regulator's lack of information. In some circumstances the regulator's lack of information is of a particular kind. She knows the forms of the cost and/or demand functions but is ignorant of one parameter in one of the functions. If she knows the distribution from which this parameter is drawn then the regulatory problem is one of mechanism design. The regulator and the firm play a Bayesian game. The application of mechanism design techniques to regulation, using modelling tools first used by Mirrlees (1971), was begun by Baron and Myerson (1982), and was developed by Laffont and Tirole (1986, 1993) and others. Following Loeb and Magat (1979) and Vogelsang and Finsinger (1979), a parallel literature has developed that assumes that the regulator's information problem is much more severe than that assumed in the Bayesian literature. In these models the regulator does not know the functional form of the cost function, and might not know the demand function. Bayesian modelling techniques are not used – the regulator is not assumed to optimize – and instead the positive and normative implications of various *ad hoc* regulatory schemes are analyzed. These schemes involve either the provision of a subsidy or the imposition of a constraint on the firm's pricing. The emphasis of this chapter is on these "non-Bayesian" regulatory schemes and on their dynamic properties. Some of the price-cap schemes analyzed in following chapters are closely related to the "non-Bayesian" regulatory schemes considered in this chapter.

Although the literature can be divided into Bayesian and "non-Bayesian" parts, for purposes of exposition it is more natural to separate the models according to the assumed source of information asymmetry. We start by considering the benchmark case of full information in 2.2. In section 2.3 we consider models where the regulator knows the demand function but does not know all the parameters of the cost function. Models where the regulator knows the cost function but not the demand function are discussed in Section 2.4. In section 2.5 models where the regulator is ignorant of the parameters of demand and of technology are presented. Some concluding remarks are made in 2.6. Other surveys of parts of the literature are Train (1991), Doyle (1993), the bibliographic notes of Laffont and Tirole (1993) and Armstrong, Cowan and Vickers (1994, Chapters 2 and 3, mainly written by Mark Armstrong).

2.2 The Full-Information Benchmark: Marginal Cost and Ramsey Pricing

When the regulator has complete information about cost and demand conditions, and enough instruments, she can ensure that the firm charges optimal prices. If she can make lump sum transfers to the firm then, provided there are no other distortions in the economy, the first-best scheme involves pricing each product at marginal cost. The transfer is positive if the cost function exhibits decreasing (ray) average costs as marginal cost pricing would not allow all costs to be covered. If, however, for some reason she is unable to make transfers then the constraint that the firm must at least break even using revenue derived from consumers is added to the regulator's problem. This is the classic problem of Ramsey pricing, based on Ramsey's (1927) analysis of second-best commodity taxation. Boiteux (1956, 1971) was the first to extend Ramsey's analysis to public utility pricing. The revival of interest among Anglo-Saxon economists in Ramsey pricing started with the papers by Diamond and Mirrlees (1971) and Baumol and Bradford (1970) and the latter contains an interesting historical discussion of Ramsey pricing. The influence of Ramsey pricing on regulatory agencies is discussed by Faulhaber and Baumol (1988). Before presenting the Ramsey model the assumptions and results on demand (that will be used throughout the thesis) will be presented.

Consumers have quasi-linear utility functions $U(\mathbf{q}) + m$ where $\mathbf{q}$ is an n x 1 vector of quantities of the regulated firm's products, U(.) is a concave direct utility function and m is consumption of the numéraire good. The budget constraint is $\mathbf{p} \cdot \mathbf{q} + m \leq y$, where y is the consumer's income, $\mathbf{p}$ is the price vector for the n goods and $\mathbf{p} \cdot \mathbf{q} \equiv \sum_{i=1}^{n} p^i q^i$. The price of the numéraire is normalized to unity. The consumer's indirect utility function is $V(\mathbf{p}) + y$. If a fixed charge, A, must be paid in order to purchase any of the n goods then the indirect utility function is $V(\mathbf{p}) - A + y$ as long as some of the n goods are bought. Demand for good i, where i = 1, . . , n, is $q^i(\mathbf{p}) = - \partial V(\mathbf{p})/\partial p^i$ by a simple application of Roy's Identity. A unit increase in p^i increases the cost of buying the original bundle by $q^i(\mathbf{p})$ and, since the marginal utility of income is constant and equal to unity, the loss of utility is $q^i(\mathbf{p})$. The function $V(\mathbf{p})$ is decreasing and convex in prices. In the single product case this is the same as saying that the demand curve slopes down. For the multiproduct case Varian (1992, pp 102-103) shows that the indirect utility function is a quasi-convex function of prices and the proof that it is convex in the case of quasi-linear utility is given by Varian (1985, p 871). The expenditure function is the utility level minus $V(\mathbf{p})$, and $V(\mathbf{p})$ must be convex since an expenditure function is concave in prices. A more intuitive argument based on Vogelsang (1983) is instructive. Change the price vector $\mathbf{p_1}$ to $\mathbf{p_2}$ and give extra income of $\sum_{i=1}^{n} q_1^i [p_2^i - p_1^i]$ to the consumer, where $q_1^i = q^i(\mathbf{p_1})$.[1] This is exactly the compensation that is required to enable him to buy the original bundle $\mathbf{q_1}$ and is the Slutsky compensation amount (Varian, 1992, p 135). If he continued to purchase $\mathbf{q_1}$ in spite of the price changes he would derive the same utility $V(\mathbf{p_1})$ as before. In general, however, he will choose a different bundle, so $\mathbf{q_2} \neq \mathbf{q_1}$ and, by revealed preference,

$$V(\mathbf{p_2}) + \sum_{i=1}^{n} q_1^i [p_2^i - p_1^i] \geq V(\mathbf{p_1}). \tag{2.1}$$

Since $q^i(\mathbf{p_1}) = - \partial V(\mathbf{p_1})/\partial p^i$, (2.1) can be rearranged to give

[1]The notation established here will be in use in most subsequent chapters. p_t^i denotes the price of product i sold in discrete time period t.

$$V(\mathbf{p_2}) - V(\mathbf{p_1}) \geq \sum_{i=1}^{n} \frac{\partial V(\mathbf{p_1})}{\partial p^i}[p_2^i - p_1^i]$$

implying that $V(\mathbf{p})$ is convex.

The assumption of quasi-linear utility justifies the use of partial equilibrium techniques and is standard in the literature. It should, however, be treated with some caution. Varian (1992, p 165) argues that it might be sensible to model the demand for products such as pencils with a quasi-linear utility function because income effects are unlikely to be important for such products. But income effects might be important in the demand for services such as gas supply or telecommunications. Furthermore many privatized utilities in Britain have shareholders who also consume the firm's output and the firm might best serve its shareholder's interests by not maximizing profits. Having noted these caveats we shall, however, follow the existing literature in ignoring them.

We now return to Ramsey pricing. The firm's rent or income is

$$I = A + \Pi(\mathbf{p}) = A + \sum_{i=1}^{n} p^i q^i(\mathbf{p}) - C[q^1(\mathbf{p}), \ldots, q^n(\mathbf{p})] \tag{2.2}$$

where the cost function C[.] is general – it does not necessarily depend on the different products in an additive form. The fixed fee received, if any, is denoted by A, and $\Pi(\mathbf{p})$ denotes the firm's operating profit. Because the firm is assumed to be a natural monopoly the cost function is taken to be subadditive – see Panzar (1989) for a discussion of multiproduct cost concepts – although the analysis is completely general. The regulator's problem is to maximize the weighted sum of consumer surplus and rent subject to the latter being equal to some given non-negative number $\bar{I}$.[2] Often $\bar{I}$ is taken to be zero. The Lagrangian is:

$$\mathcal{L} = V(\mathbf{p}) - A + \alpha[A + \Pi(\mathbf{p})] + \mu\left\{A + \Pi(\mathbf{p}) - \bar{I}\right\} \tag{2.3}$$

[2]Many authors treat the problem as that of maximizing consumer surplus alone subject to profits being equal to $\bar{I}$. The problems are equivalent, but it is more convenient for our purposes to include profit in the objective function as it facilitates comparison with the Laffont-Tirole framework discussed below.

where α is the weight on the firm's income in the objective function, satisfying $0 < \alpha \leq 1$, and μ is the Lagrange multiplier. We assume initially that $\mathrm{A} = 0$. The first-order condition for the kth price is

$$\frac{\partial \mathcal{L}}{\partial \mathrm{p}^k} = -\mathrm{q}^k + (\alpha + \mu)\frac{\partial \Pi}{\partial \mathrm{p}^k} = 0. \tag{2.4}$$

It follows from (2.4) that, for any pair of products k and l, $(\partial\Pi/\partial \mathrm{p}^k)/(\partial\Pi/\partial \mathrm{p}^l) = \mathrm{q}^k/\mathrm{q}^l$. The iso-surplus curve (the combinations of p^k and p^l that give the consumer the same indirect utility) is tangential to the iso-rent curve with Ramsey prices. Ramsey prices are illustrated in Figure 2.1. The firm's rent increases with prices, but consumer surplus falls as prices rise. Ramsey prices are at the point of tangency of the lowest iso-surplus curve with the iso-rent (iso-profit) constraint. The slope of the iso-surplus locus is $-\mathrm{q}^1/\mathrm{q}^2$.

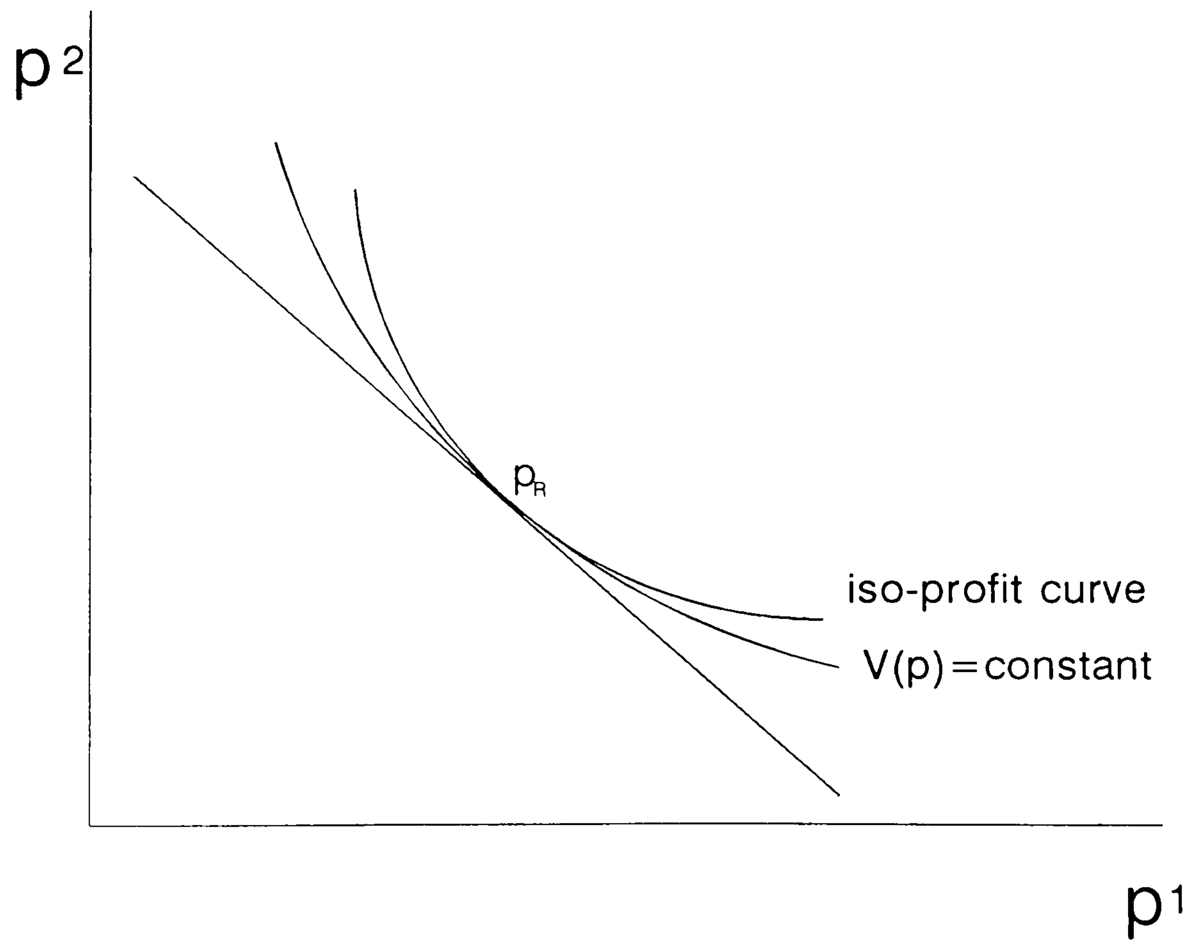


Figure 2.1 Ramsey Prices

The first-order conditions for the Ramsey problem (2.4) are necessary but not sufficient for a local maximum for welfare – see Atkinson and Stiglitz (1980, p 374) or Dixit and Norman (1980, p 174). The second-order condition requires that the iso-rent curve is more convex than the iso-surplus curve, as in Figure 2.1. Often, however, the interesting question about a regulatory scheme is whether the prices it induces meet the necessary conditions for Ramsey pricing and for this question the second-order conditions can be ignored.

When the regulator has complete information about demand and costs she can delegate the pricing decision to the firm by imposing a particular constraint on the firm's prices. Let the firm maximize rent subject to $\sum_{i=1}^{n} [p^i_R - p^i]q^i(\mathbf{p}_R) \geq 0$, where $\mathbf{p}_R \equiv (p^1_R, \ldots, p^n_R)$ is the Ramsey price vector. This constraint is the straight line in Figure 2.1. It goes through $\mathbf{p}_R$ and has a gradient of $-q^1(\mathbf{p}_R)/q^2(\mathbf{p}_R)$. It is clear from the diagram that the firm maximizes profits subject to this price constraint by choosing the Ramsey prices. A simpler form of regulation would, of course, be to order the firm to set the known Ramsey prices. In practice it is unlikely that the regulator will know what the Ramsey prices and quantities are. Her problem then is to choose the level and slope of the price constraint. In this and the following chapters we shall discuss various constraints that can and have been applied when the regulator does not know the Ramsey prices.

Another way to express the first order condition (2.4) uses the symmetry of the substitution terms:

$$\sum_{i=1}^{n} [p^i - MC^i]\frac{\partial q^k}{\partial p^i} = -\frac{(\alpha + \mu - 1)}{\alpha + \mu}\, q^k. \tag{2.5}$$

where $MC^i \equiv \partial C/\partial q^i$. If the rent target is close to the rent level with marginal cost pricing then the left-hand side of (2.5) can be interpreted as the change in the demand for good k that results from a deviation of prices from marginal costs. Equation (2.5) says that the proportional reduction in demand should be the same for each good since the term $(\alpha + \mu - 1)/(\alpha + \mu)$, known as the Ramsey number, is independent of k. One well-known special case is that of independent demands. In this case $\partial q^k/\partial p^i = 0$ for $i \neq k$ and (2.4) and (2.5) give

$$\frac{p^k - MC^k}{p^k} = \frac{(\alpha + \mu - 1)}{\alpha + \mu} \cdot \frac{1}{\eta^k} \tag{2.6}$$

where $\eta^k \equiv -(\partial q^k/\partial p^k)p^k/q^k$ is the own price-elasticity of demand. This condition says that the percentage deviation of price from marginal cost (the Lerner index) should be higher in those markets with less elastic demand and is the famous inverse-elasticity rule.

Under what conditions are Ramsey prices equal to marginal costs? There are two main ways this could hold. First, if the weight on rents is the same as on consumer surplus ($\alpha = 1$), and there are constant returns to scale on all product lines, marginal cost pricing is both desirable and feasible. Second, if the firm could set a fixed fee this would substitute for the unavailable lump-sum transfer and Ramsey prices would be marginal cost prices. In the Lagrangian (2.3) the first-order condition for A yields $\mu = 1 - \alpha$. Substituting this into (2.5) gives the marginal cost pricing result. The level of A is chosen to cover the deficit from marginal cost pricing. For this to hold it must be the case that no consumers are driven from the market by the existence of a fixed charge. In many of the models considered in the remainder of this chapter it is assumed that the regulator can make transfers to the firm or that she can set a fixed fee to be paid by consumers that does not alter demand. Note that the optimal fixed fee might be negative. If the cost function is U-shaped and average cost is increasing when price equals marginal cost, then operating profits generated by marginal cost pricing might be more than sufficient to meet the profit constraint, and it is better to return money to the consumers if $\alpha < 1$.[3]

This characterization of the Ramsey problem is the conventional one and it

[3]This can still be a natural monopoly provided demand is not too large — see Waterson (1988, Chapter 2) for an example with quadratic costs.

produces a Ramsey number that depends on the firm-specific parameters of both tastes and technology. Laffont and Tirole (1990, 1993) discuss the Ramsey problem in a different way. They note that the assumption that the regulator does not make transfers is not consistent with her benevolence – see Laffont and Tirole (1993, p 32). Transfers to the firm are costly because of the inevitable distortions created when ideal lump-sum taxes are not available – indeed Ramsey (1927) starts from this premise – but transfers are not *infinitely* costly. Laffont and Tirole, whose models will be discussed more fully in the next section, adopt the accounting convention that the firm pays its operating profits $\Pi(\mathbf{p})$ to the regulator, who in return makes a *net* transfer τ to the firm.[4] The key difference between the Laffont-Tirole welfare function and that used above is that government revenue has a higher weight than either consumer surplus or the firm's rent (which are weighted equally). This is because revenue in the hands of the government enables distortionary taxation to be reduced. The cost of raising £1 of government revenue is £$(1+\gamma)$ where $\gamma > 0$ represents the shadow cost of public funds. In terms of our notation the regulator's objective function is:

$$\mathrm{W} = \mathrm{V}(\mathbf{p}) + \tau - (1+\gamma)[\tau - \Pi(\mathbf{p})].$$

Welfare is consumer surplus plus the firm's rent less the total monetary transfer the regulator makes, $\tau - \Pi(\mathbf{p})$, times the total cost of each pound of that transfer. This is equivalent to

$$\mathrm{W} = \mathrm{V}(\mathbf{p}) + (1+\gamma)\Pi(\mathbf{p}) - \gamma\tau$$

which shows that leaving the firm with a net rent, i.e. a positive τ, is costly. With complete information the regulator sets $\tau = 0$ and prices are given by the usual Ramsey formula (2.5) except that $\alpha + \mu$ is replaced by $1 + \gamma$.[5] The Ramsey number thus does

[4]The *net* transfer τ should be distinguished from the gross transfer T used in other models. Laffont and Tirole (1993) do not interpret τ as the fixed fee, A, in a two-part tariff because they assume that consumers are heterogeneous enough that some are excluded from the market when a two-part tariff is set.

not depend on the firm's demand or cost conditions but is exogenous to the firm (although γ can be determined endogenously in a general equilibrium model).

The Ramsey model can be extended. In later chapters we shall characterize Ramsey prices when the firm sets two-part and nonlinear tariffs, when quality is a choice variable, when there are network externalities, and when there is regulatory preference for some customers over others.

2.3 Asymmetries of Information about Costs

In this section we consider models where the regulator does not know at least one characteristic of the cost function, but does know the demand function. This case was first analyzed by Loeb and Magat (1979). Their contribution was to show that this information asymmetry does not preclude the implementation of first-best full-information pricing, but their model produces the extreme result that consumers derive no net surplus and the firm's rent is large. The assumption that leaving rent to the firm is costly led to the application of Bayesian techniques by Baron and Myerson (1982) and Laffont and Tirole (1986).

2.3.1 Surplus Subsidy Schemes

When the regulator knows all the characteristics of the demand function but does not know any of the characteristics of the cost function then she can still induce efficient pricing. The key assumption is that lump-sum transfers or a fixed fee are feasible. Throughout this section it will be assumed for ease of notation that the firm only produces one good though all of the results can be generalized to the multiproduct case.

Loeb and Magat (1979) presented a model in which the firm is induced to price efficiently in spite of the asymmetry of information on the cost side. The regulator's objective is to maximize the unweighted sum of consumer surplus, producer surplus and

[5] Note that the profit constraint is now $\tau \geq 0$.

government revenue: $W \equiv V(p) + \Pi(p)$. Note that in the welfare function the transfer cancels out. In the absence of differential welfare weights it is simply a transfer to the firm, whether it is provided by the general taxpayer or by consumers as part of a two-part tariff. The Loeb and Magat scheme (L-M) is very simple. It delegates the pricing decision to the firm (Baron, 1989, pp 1359-1361) and gives the firm a transfer payment equal to $V(p)$, i.e. consumer surplus. This means that the firm's objective function is the same as social welfare. It follows that price is equal to marginal cost whatever the shape of the cost function (as long as $W(p)$ is non-negative at this price).[6]

There is an obvious problem with L-M.[7] If the transfer is paid by the consumers then the distributional consequences are extreme since all the social surplus derived from the market is appropriated by the firm – it is as if the firm was practising first-degree price discrimination. Loeb and Magat suggest two alternative ways to reduce the extremity of this outcome. First, if there are many identical firms that could operate in the market then the regulator could auction a franchise for the sole right to be in the market. If the auction is competitive then the successful firm would offer the whole of the social surplus gained with marginal cost pricing for the right and the regulator would then appropriate the gains (which could be returned to consumers in lump-sum fashion). Franchising is unlikely to be feasible in utility industries where there are substantial sunk costs, although there are examples of franchise schemes which do appear to work such as those for providing municipal water supplies in France. The second way to alleviate the distributional problem is to set an upper limit, $\bar{p}$, to the allowed price, and then set the transfer equal to the increase in consumer surplus derived from changing price from $\bar{p}$ to p, i.e. $T = V(p) - V(\bar{p})$. Since $V(\bar{p})$ is a given number it does not affect the firm's pricing decision, so marginal cost pricing will still prevail. But the regulator has to choose $\bar{p}$. If she sets $\bar{p}$ too low then the firm might not be able to make a positive profit and would shut down. Some knowledge of the cost

[6]The L-M scheme ensures that the firm produces if and only if social welfare with marginal cost pricing is non-negative.

[7]See also Sharkey (1979) for a discussion of some of the practical problems with the L-M scheme.

function would be necessary in order to reduce the danger that the firm's participation constraint is breached.

The scheme designed by Sappington and Sibley (1988) builds on the Loeb and Magat scheme and substantially alleviates its defects. Their scheme, known as the Incremental Surplus Subsidy scheme (ISS), can be thought of as a dynamic discrete-time version of L-M. Each period the firm is keeps its operating profit defined by $\Pi_t \equiv p_t q(p_t) - C(q(p_t))$ and receives a transfer from the regulator equal to the increase in consumer surplus from period $t-1$ to period t less the operating profit achieved in period $t-1$. The firm's income in period t is then

$$I_t \equiv V(p_t) - V(p_{t-1}) + \Pi_t - \Pi_{t-1} \tag{2.7}$$

which is the change in total welfare, $W_t - W_{t-1}$. The regulator must have slightly more information to implement ISS rather than L-M because lagged costs must be observable, although it is only total cost levels that must be observable – knowledge of the functional form or parameters of the cost function is not necessary. A further trivial assumption is that the regulator should be able to observe and remember last period's revenue. The firm's objective is to choose prices in each period to maximize the present value of income. Its problem is

$$\underset{\{p_t\}}{\text{Max}}\ Z = \sum_{t=1}^{\infty} \beta^{t-1}\{W_t - W_{t-1}\} = -W_0 + (1-\beta)\sum_{t=1}^{\infty}\beta^{t-1}W_t \tag{2.8}$$

The discount factor, β, satisfies $0 \leq \beta < 1$ and the initial price, p_0, is given.[8] The first-order conditions give

[8] If p_0 is chosen by the firm, and the firm knows that the ISS scheme will be imposed from period 1, then it has an incentive to raise p_0 above the unconstrained profit maximizing level, because this increases the rent earned in period 1. Since rents are zero from period 2 onwards its objective function is $\Pi(p_0) + \beta[V(p_1) - V(p_0) + \Pi(p_1) - \Pi(p_0)]$. Its choice of p_1 is independent of p_0. The first-order condition for p_0 is $(1-\beta)\Pi'(p_0) + \beta q(p_0) = 0$, which implies $\Pi'(p_0) < 0$. Anticipation of regulation will usually encourage such strategic behaviour – see Sappington (1980), Law (1993) and the discussion in 2.5.

$$(1-\beta)\Big(p_t - MC_t\Big)q'(p_t) \;=\; 0 \qquad\qquad t = 1, 2, \ldots$$

where $q'(p_t) \equiv dq(p_t)/dp$. There is marginal cost pricing every period. Since maximizing $(1-\beta)\beta^{t-1}W_t$ is the same as maximizing W_t, it is optimal to price at marginal cost each period.

Figure 2.2 shows that by choosing p_1 the firm earns a rent in period 1 equal to the area DFGH, which is the change in welfare. The firm does better to price at p_* immediately, yielding a rent of DJH. Any delay in reaching p_*, for example setting p_1 and then $p_2 = p_*$, is undesirable because this delays the date at which FGJ is earned, and since $\beta < 1$, the present value of profits from this strategy is below that with immediate marginal cost pricing.

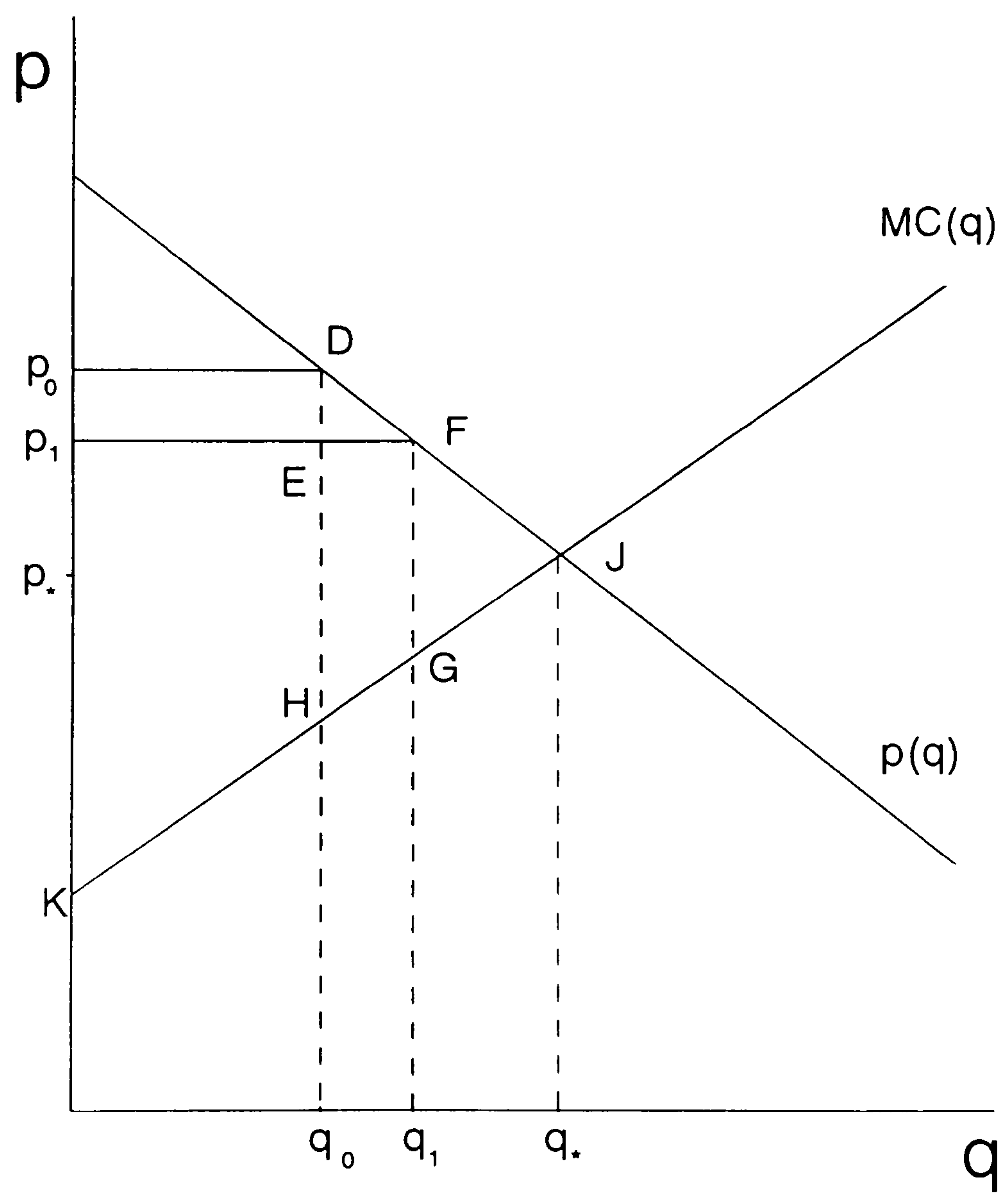


Figure 2.2 The Incremental Surplus Subsidy Scheme

Under some dynamic regulatory schemes that require knowledge of lagged costs there can be an incentive to indulge in purely wasteful expenditures in order to manipulate future regulatory constraints (Sappington, 1980) but under ISS the firm minimizes production costs each period. An extra pound spent on waste this period costs £1 now and will raise the subsidy in the next period by £1, since the previous period's costs are reimbursed. This reduces the present value of the firm's income stream by $(1-\beta)$. Thus the optimal level of waste is zero. Similarly if there are intertemporal effects in the cost function because, for example, capital expenditure reduces current and future operating costs the firm has the socially correct investment incentives (Sappington and Sibley, 1988, p 304).

The ISS scheme improves on L-M as it reduces the firm's rent. Income is positive only in the first period. Thereafter the firm will set a constant price and its income is zero.[9] Sappington and Sibley (1988, p 303) show that it is possible to tax away the first period rent if the regulator knows β. From period 2 onwards the regulator subtracts $(1/\beta - \epsilon)I_{t-1}$ from I_t where $\epsilon > 0$ and is arbitrarily small. This does not alter the pricing decision and $I_t = 0$ from $t = 2$ onwards anyway. The present value of the firm at date 1 is thus $Z = I_1 - \beta(1/\beta - \epsilon)I_1 = \beta\epsilon I_1$ which can be made arbitrarily small by setting ϵ very low. Stefos (1990) and Sappington and Sibley (1990) debate the potential problems for the ISS scheme including taxation when there are exogenous increases in costs that could force the firm to make a loss in the period in which costs rise. But even in cases where the cost function is stationary there is a problem with the ISS taxation scheme. After date 1 the firm has a negative present value since it must pay a tax of $(1/\beta - \epsilon)I_1$ in period 2 and its income stream is zero thereafter. Thus it will want to quit the relationship after period 1 – any firm with a cash flow which becomes negative and will not turn positive will want to quit before the first negative cash flow.

Two further comments on the L-M and ISS schemes should be made. First,

[9] If there are scale economies operating profits will be negative. From period 2 onwards the firm receives a positive transfer equal to minus the previous period's operating profits (which were negative) and this transfer will be just sufficient to cover its operating losses in each period.

since the schemes ensure that social welfare is maximized in all periods the regulator will want to maintain the schemes, even if she cannot make a commitment to do so. In other words there is no time-consistency problem for the regulator. Second, even in cases where the demand function is not common knowledge the schemes can be implemented using optional two-part tariffs that induce the firm to reveal correctly the demand function, as long as consumers are identical or first-degree price discrimination is feasible. Graham and Vernon (1991) and Sibley (1989) show this for the L-M and ISS schemes respectively.

2.3.2 Bayesian Models

In the models considered in the previous subsection the regulator did not need to know anything about the *form* of the cost function, although in the ISS scheme she had to be able to observe lagged total costs. In this subsection we discuss models of regulation under asymmetric information when the regulator knows the functional form of the cost function and knows the distribution of a parameter that characterizes this function but does not know its actual value. The problem is one of mechanism design (Fudenberg and Tirole, 1991, Chapter 7). These models have been the subject of several surveys including Baron (1989), Caillaud et al. (1988), Laffont and Tirole (1993) and Armstrong, Cowan and Vickers (1994, Chapters 2 and 3) so the treatment here will be brief.

Baron and Myerson (1982) (hereafter B-M) presented the first Bayesian model of regulation, which highlights the rôle of adverse selection. The description of B-M here is influenced by Baron (1989). The crucial difference between B-M and Loeb and Magat (1979) is that B-M assume that the regulator's objective function is a weighted sum of consumer surplus and the firm's income with a weight on the firm's income that is strictly below unity. This might be because consumers are located in one area whereas many of the firm's shareholders live elsewhere. The effect of giving consumer surplus a higher weight than profits is to make transfers socially costly, and thus to generate a trade-off between rent extraction and allocative efficiency.

In the B-M model the regulator knows that the cost function is of the form $C = \theta q + K$ where the fixed cost K is common knowledge, demand is given by $q = q(p)$ and θ is the marginal cost parameter.[10] The results can be extended to more general cost functions. The distribution function of θ, $F(\theta)$, has density $f(\theta)$, and both are common knowledge. The firm has private knowledge of the realization of θ. The support of $F(\theta)$ is $[\underline{\theta}, \overline{\theta}]$ where $\underline{\theta} > 0$. We assume that it is always desirable for the firm to operate. The regulator offers a menu of price and transfer pairs $\{p(\hat{\theta}), T(\hat{\theta})\}$ that determines the allowed price and transfer for each report $\hat{\theta}$ that the firm makes about its true cost parameter θ. By the Revelation Principle attention can be restricted to mechanisms that induce truth-telling by the firm ($\hat{\theta} = \theta$). The regulator's problem is to maximize expected welfare subject to the incentive-compatibility constraints (which require that truth-telling is at least as profitable as lying) and the individual rationality constraints (which require that each type of firm is willing to participate). For a given θ welfare is

$$\begin{aligned} W(\theta) &= V(p(\theta)) - T(\theta) + \alpha I(\theta) \\ &= V(p(\theta)) + [p(\theta) - \theta]q(p(\theta)) - K - (1-\alpha)I(\theta). \end{aligned}$$

The second expression follows by substituting in for $T(\theta) = I(\theta) - [p(\theta) - \theta]q(p(\theta)) + K$. This shows that positive rents are costly when $\alpha < 1$.

The incentive compatibility (IC) constraints are represented by $dI(\theta)/d\theta = -q(p(\theta))$. Intuitively if the firm's marginal cost falls by $d\theta$ to $\theta - d\theta$ then it can earn the same price and transfer by lying and announcing that its parameter remains at θ. Its rent, I, will increase by $q(p(\theta))d\theta$ which is the reduction in the cost of producing the given output. In order to give the firm an incentive not to lie the regulator must thus increase the firm's rent by this amount. The individual rationality (IR) constraints reduce to the requirement that the least efficient type must be willing to

[10]In this chapter θ will always stand for the parameter about which the firm has hidden knowledge. This is usually, but not always, a parameter of the cost function.

participate. Since more efficient (lower θ) types can always mimic the least efficient type their IR constraints are automatically satisfied if $I(\overline{\theta}) \geq 0$. Integrating the IC constraint gives the following expression for the firm's rent

$$I(\theta) = \int_{\theta}^{\overline{\theta}} q(p(\tilde{\theta}))d\tilde{\theta} + I(\overline{\theta}). \tag{2.9}$$

Since the regulator prefers not to leave the firm with rents she sets $I(\overline{\theta}) = 0$. It can also be shown that another implication of incentive compatibility is that the price should be non-decreasing in θ. Treating $I(\theta)$ as a state variable and $p(\theta)$ as the control variable the regulator's Hamiltonian is

$$H = \{V(p) + [p - \theta]q(p) - K - (1 - \alpha)I\}f(\theta) - \phi(\theta)q(p)$$

where $\phi(\theta)$ is the Pontryagin multiplier on the IC constraint. Among the necessary conditions for optimality are:

$$\frac{\partial H}{\partial p} = [p - \theta]q'(p)f(\theta) - \phi(\theta)q'(p) = 0 \tag{2.10}$$

$$\frac{\partial H}{\partial I} = -(1 - \alpha)f(\theta) = -\frac{d\phi}{d\theta}. \tag{2.11}$$

Integrating (2.11) and using the transversality condition, $\phi(\underline{\theta}) = 0$, implies

$$\phi(\theta) = (1 - \alpha)F(\theta).$$

Substituting this into (2.10) yields the expression for the optimal price as a function of θ:

$$p(\theta) = \theta + (1 - \alpha)\frac{F(\theta)}{f(\theta)} \tag{2.12}$$

while the firm's rent is given by (2.9). The condition that price is non-decreasing in θ holds if $F(\theta)/f(\theta)$ is non-decreasing, which holds for many common distributions.

This confirms that when $\alpha = 1$ the B-M mechanism induces marginal cost pricing whatever the realization of θ, as in the L-M scheme. When $\alpha < 1$ price is always above marginal cost except when the firm has the lowest feasible cost $\underline{\theta}$ since $F(\underline{\theta}) = 0$.[11] The reason for the distortion of price above marginal cost is that transfers are costly. The firm receives an "information rent" of $\int_{\theta}^{\overline{\theta}} q(p(\tilde{\theta}))d\tilde{\theta}$ which depends positively on quantity. This rent can be reduced by raising $p(\theta)$, so lowering quantity demanded and the firm's rent. The regulator trades off the benefit from the reduction of information rents against the cost in terms of lower consumer surplus when p is raised above θ.

The B-M model and the models based on it are adverse selection models in which the agent (the firm) has private information about an exogenous variable. Another agency-theoretic model is the moral hazard model in which the agent has private information about both an exogenous productivity parameter and an endogenous effort variable. The most famous moral hazard models of regulation are those of Laffont and Tirole (1986, 1990, 1991 and 1993). For ease of exposition the version presented here will be for the single-product case and we shall assume that marginal cost is constant – see Laffont and Tirole (1990 and 1993, Chapter 3) for more general models. Marginal cost is $\theta - e$ where e is the effort level of the firm that is not observed by the regulator and θ is an exogenous productivity parameter whose value is known by the firm but not the regulator. The regulator observes marginal cost $(\theta - e)$ and output q and, since fixed costs are common knowledge, she knows total costs $C = (\theta - e)q + K$ but she cannot tell if a high marginal cost is due to low effort or an unfortunate realization of θ. This means that she faces a moral hazard problem in addition to the adverse selection problem. The disutility of effort for the firm (or its manager) is given

[11] This is an "absence of distortion at the top" result which is typical in models of mechanism design. The lowest type $\underline{\theta}$ induces efficient pricing. Mirrlees (1971) has a model of optimal income taxation in which the individual with the highest ability has a marginal tax rate of zero. Similarly in models of nonlinear pricing (see Tirole, 1988, Chapter 3 or Wilson (1993)) the highest consumer usually faces a marginal price equal to marginal cost, at least when regulation is absent.

by the increasing and convex function $\psi(e)$. This is not observed by the regulator – if it was then the effort level could be inferred by inverting the function. As mentioned in section 2.2 Laffont and Tirole (L-T) adopt the accounting convention that the firm's revenue is paid to the regulator and the firm receives reimbursement for its production costs, C. In addition the regulator makes a net transfer payment τ to the firm. The firm's net rent is then $I \equiv \tau - \psi(e)$. An important assumption is that paying £1 to the firm costs the government £$(1+\gamma)$. The regulator's welfare for a given θ is

$$\begin{aligned} W &= V(p) - (1+\gamma)[\tau + (\theta - e)q(p) + K - pq(p)] + I \\ &= V(p) + (1+\gamma)pq(p) - (1+\gamma)[\psi(e) + (\theta - e)q(p) + K] - \gamma I \end{aligned}$$

where the second expression follows from the definition of the firm's rent I. If the regulator had full information then the following first-order conditions characterize the optimum:

$$\frac{p-(\theta-e)}{p} = \frac{\gamma}{(1+\gamma)}\cdot\frac{1}{\eta}$$

$$\psi'(e) = q(p)$$

$$I = 0.$$

The first equation is a Ramsey pricing result – remember η is the price elasticity of demand. If government revenue is costless then $\gamma = 0$ and the firm should price at marginal cost. The equation for effort says that the firm should invest in effort until the marginal cost of extra effort equals the marginal benefit in terms of lower operating costs. Finally, since the regulator does not suffer from an information asymmetry there is no need to give the firm any costly rent.

With asymmetric information the regulator's problem is to choose p, e and τ (as functions of θ) to maximize expected welfare subject to the incentive compatibility and individual rationality constraints. As in the B-M model the relevant individual rationality constraint is $I(\overline{\theta}) \geq 0$ and at the optimum this is satisfied with equality. The incentive compatibility constraints are represented by $dI/d\theta = -\psi'(e)$ and the firm's information rents are thus $\int_{\theta}^{\overline{\theta}} \psi'(e(\tilde{\theta}))d\tilde{\theta}$. If the value of θ drops by $d\theta > 0$ then the firm can make extra rents of $\psi'(e)d\theta$ by claiming that the productivity parameter remains at θ. Another implication of incentive compatibility is that marginal cost is non-decreasing in θ. The Hamiltonian for the firm is

$$H = \{V(p) + (1+\gamma)pq(p) - (1+\gamma)[\psi(e) + (\theta - e)q(p) + K] - \gamma I\}f(\theta) - \phi(\theta)\psi'(e)$$

where $\phi(\theta)$ is the Pontryagin multiplier. The regulator chooses $p(\theta)$ and $e(\theta)$, and $I(\theta)$ is the state variable. The first order condition for the choice of p yields

$$\frac{p - (\theta - e)}{p} = \frac{\gamma}{(1+\gamma)} \cdot \frac{1}{\eta}. \tag{2.13}$$

The necessary condition for the choice of e is

$$-(1+\gamma)[\psi'(e) - q(p)]f(\theta) - \phi(\theta)\psi''(e) = 0$$

and the maximum principle gives $d\phi/d\theta = \gamma f(\theta)$. Integrating this and using the transversality condition yields $\phi(\theta) = \gamma F(\theta)$. Substituting this into the first order condition for e gives

$$\psi'(e) = q(p) - \frac{\gamma}{1+\gamma} \frac{F(\theta)}{f(\theta)} \psi''(e). \tag{2.14}$$

Equation (2.13) shows that there is Ramsey pricing even under asymmetric information. There is no need to distort prices to minimize the firm's information rents. This differs from the B-M result where the full-information benchmark was marginal cost pricing but with asymmetric information price is generally above marginal cost. In B-M the firm's marginal cost is unobservable, so the only way to reduce information rents is to raise p to limit output. Under L-T marginal cost is observable and so can be reimbursed, and under a condition the cost-reimbursement rule is sufficient to limit rents without requiring any price distortion. The necessary and sufficient general condition for this result (which is known as the "incentive-pricing dichotomy") is that there is a function ζ such that $C = C(\zeta(\theta, e), q)$. The cost function assumed above clearly satisfies this. See Laffont and Tirole (1990 or 1993, Chapter 3.6) for a full discussion. Intuitively if the cost function satisfies this property then changes in quantity do not affect the ability of the firm to convert productivity increases into rent by reducing effort.

The solution to the trade-off between encouraging effort and extracting rent from the firm is given in (2.14). The following intuition is based on Laffont and Tirole (1993, pp 65-66). A small increase in effort, de, for types in $[\theta, \theta + d\theta]$ who number $f(\theta)d\theta$ lowers marginal cost by de and raises the cost of effort by $\psi'(e)de$. The social value of this increase in productive efficiency is $(1+\gamma)[1-\psi'(e)]de.f(\theta)d\theta$. This raises the rent of all those who are more efficient than θ, and there are $F(\theta)$ in the interval $[\underline{\theta}, \theta]$. Their rent increase by $\psi''(e)de.d\theta$, and the social cost of this extra rent is $\gamma\psi''(e)de.d\theta.F(\theta)$. Trading the benefit of extra productive efficiency against the cost of extra rent gives (2.14). Note that the optimal trade-off gives an outcome that is intermediate between a fixed-price contract (a price-cap) and a cost-plus contract (rate-of-return regulation). The full-information level of effort is only achieved by the most efficient firm. It is feasible to give the firm an incentive to put in optimal effort (characterized by $\psi'(e) = q(p)$), but such a price-cap is not desirable because it leaves the firm with excessive rents. Conversely rate-of-return regulation which leaves rents very low would induce insufficient effort.

Baron and Myerson and Laffont and Tirole apply the techniques of information economics to regulation. Nevertheless their approach has limited utility as a guide for policy. The information asymmetry concerns the value of one parameter of one function. Everything else is common knowledge. In particular the regulator knows the functional form of the cost function. In practice it is very unlikely that the regulator will know the relevant functional form, especially as most natural monopolies produce multiple products with scope economies. Even if the functional form is known it is likely to be characterized by more than one unknown parameter. Both marginal cost and fixed cost, for example, might be private information and might not be perfectly correlated. Modelling this case is difficult and produces messy results because the incentive compatibility constraints are partial differential equations. Baron (1989, p 1376) states that "the difficulty in extending ... the theory ... to multiple information parameters constitutes a significant limitation to the application of the theory of mechanism design with asymmetric information." See McAffee and McMillan (1988) for some general results in this area and Laffont and Tirole (1993, Chapter 3.7) for some special cases in the regulation context where multidimensionality is straightforward to model because the cost function is additively separable. It is also unrealistic to assume that the regulator has full knowledge of the demand function.[12]

2.4 Asymmetries of Information about Demand

Most of the literature on regulation under asymmetric information assumes that the regulator knows the demand function but does not know the realization of one parameter of the cost function. The models of Lewis and Sappington (1988a) and Riordan (1984) examine the opposite case where the regulator knows the parameters of the cost function but does not know the realization of a demand parameter. They show that a simple scheme to induce efficient prices without leaving a rent to the firm is available under these conditions as long as marginal costs are non-decreasing.

[12]One criticism of the Laffont and Tirole models has been their reliance on transfers, which are very rarely observed in the regulation of natural monopolies and are often banned. Laffont and Tirole (1993, Chapter 2) argue, however, that the lack of transfers does not alter their key insights. When the firm sets a two-part tariff that excludes some customers, optimal pricing still involves Ramsey-type principles and the trade-off between rent extraction and incentives has a similar form to (2.14). The main difference is that the shadow price of rents is type-contingent, rather than being determined by economy-wide data, since it is now consumers who provide the rent to the firm necessary to elicit effort.

We focus on the model of Lewis and Sappington (1988a), which is more general than that of Riordan. The demand function is $q(p, \theta)$ where θ is now a parameter that shifts demand. We assume that $\partial q/\partial\theta > 0$. Indirect utility is $V(p, \theta)$. Welfare, as in B-M, is a weighted average of consumer surplus plus profits, i.e. $W = V(p, \theta) - T + \alpha[T + pq(p, \theta) - C(q(p, \theta)]$. We know from the discussion of Ramsey pricing with transfers in 2.2 that the full-information price is equal to marginal cost, and that the firm's profits should be taxed, leaving it with no rent (as long as $\alpha < 1$). The optimal full-information price and transfer are thus

$$p^*(\theta) = C'(q(p^*(\theta), \theta) \tag{2.15}$$

$$T^*(\theta) = C(q(p^*(\theta), \theta) - p^*(\theta)q(p^*(\theta), \theta). \tag{2.16}$$

When there is asymmetric information about the demand parameter θ and $C'' \geq 0$ the regulator asks the firm to report θ and sets the price and transfer given in (2.15) and (2.16). To prove this all that needs to be shown is that the firm has no incentive to lie about θ, and is willing to participate. If the firm exaggerates the value of θ then it will earn a higher price, but will lower the output on which it earns the higher price and will reduce the transfer. Its rent from announcing $\hat{\theta} > \theta$ is

$$\begin{aligned} I(\hat{\theta}) &= T^*(\hat{\theta}) + p^*(\hat{\theta})q((p^*(\hat{\theta}), \theta) - C(q(p^*(\hat{\theta}), \theta)) \\ &= C(q(p^*(\hat{\theta}), \hat{\theta})) - C(q(p^*(\hat{\theta}), \theta)) + p^*(\hat{\theta})\Big[q((p^*(\hat{\theta}), \theta) - q(p^*(\hat{\theta}), \hat{\theta})\Big] \\ &= \int_{\theta}^{\hat{\theta}} \frac{\partial q((p^*(\hat{\theta}), x)}{\partial\theta}\Big[C'(q(p^*(\hat{\theta}), x)) - p^*(\hat{\theta})\Big]dx \end{aligned}$$

To sign this expression we must sign the term in square brackets. When $\theta = \hat{\theta}$, (2.15) implies $p^*(\hat{\theta}) = C'(q(p^*(\hat{\theta}), \hat{\theta}))$. For values of θ below $\hat{\theta}$ output is lower than at $\hat{\theta}$ and since marginal cost is non-decreasing in output this implies $C'(q(p^*(\hat{\theta}), \theta) \leq p^*(\hat{\theta})$. This implies that $I(\hat{\theta}) \leq 0$ and that exaggeration of θ is unprofitable. Similarly it can be shown that if the firm understates θ it will lose more in operating profits than it gains from the increase in T. Participation is assured because the firm makes a rent of zero when it correctly announces θ. Thus the first-best is implementable even if there is asymmetric information about demand as long as marginal cost is non-decreasing.

This result is not surprising. The simplest case is where the cost function is affine: $C = cq(p, \theta) + K$ where c is the constant marginal cost and K is the fixed cost. Since the regulator knows c and K, all she needs to do is to tell the firm to price at marginal cost and give the firm a transfer to cover its fixed costs, i.e. $p(\theta) = c$ and $T(\theta) = K$. The value of θ is irrelevant to efficient pricing when marginal cost is constant. Riordan (1984) extends this intuition to the case where the firm has a capacity constraint and there is a peak-load pricing problem.

When, however, marginal cost declines with output and the "single-crossing property" holds, the regulator cannot implement the first-best price and transfer. The "single-crossing property" implies that the greater θ is the happier the firm is to forego transfer payments in order to get a higher price. This implies that that p rises with θ in any incentive-compatible mechanism (Lewis and Sappington, 1988a, Lemma 2). But with declining marginal costs the first-best price *falls* as θ rises. The best the regulator can do is to set a fixed price and transfer that do not depend on the announced value of θ. The fixed price will depend on the prior beliefs about θ given by the density $f(\theta)$, the known cost function and the weights in the welfare function.

These models show that the regulator's optimal policy depends on whether her ignorance is about costs or about demand. When she does not know costs (and transfers are costly), as in the B-M and L-T models, the first-best outcome is only obtained when the cost parameter θ attains its lowest value. When she is ignorant of a demand parameter, however, the first-best outcome is often feasible (even if transfers are costly) and its desirability depends not on the realization of the demand parameter θ but on whether marginal costs are decreasing or not.

2.5 Asymmetries of Information about Costs and Demand

The next step is to analyze regulation when the firm has better information about both costs and demand. This section will be divided into three subsections. In subsection 2.5.1 the Bayesian model of Lewis and Sappington (1988b) will be discussed. They assume the regulator knows the functional forms of the demand and cost functions but does not know the parameters which characterize them. The dynamic mechanism of Vogelsang and Finsinger (1979) in which the regulator does not know the functional forms of demand and costs is presented in 2.5.2. This mechanism, under some restrictive assumptions, induces the firm to set Ramsey prices in the limit. A class of related mechanisms due originally to Scott (1952, 1978) but since revived by Finsinger and Vogelsang (1982) and Hagerman (1990) is discussed in 2.5.3. These include a rôle for transfers and thus generate marginal cost pricing in the limit. In addition they allow some of the restrictive assumptions of the Vogelsang-Finsinger mechanism to be relaxed.

2.5.1 The Bayesian Model of Lewis and Sappington

Lewis and Sappington (1988b) combine the B-M model with the Lewis and Sappington (1988a) model. The regulator knows the demand and cost functions but does not know the value of shift parameters in the cost and demand functions. They claim that there are a number of similarities between the optimal policies with one and two sources of uncertainty. First, there are situations in which price always exceeds expected marginal cost, in order to limit the information rents from exaggerating costs. Second, the optimal price rises with both the cost parameter and the demand parameter. They argue that there are also features that do not arise in either of the cases where there is only one source of uncertainty. The main difference is that prices may optimally be below actual marginal cost in order to prevent the firm from understating demand.

The main reason for being sceptical about the value of a Bayesian approach to asymmetries about both costs and demand should be clear. The uncertainties that Lewis and Sappington model are about the values of parameters that enter demand and cost functions that are common knowledge. But it is most unlikely that the regulator will know these functional forms. She will also find it difficult to form Bayesian priors, and if she could they would be of very limited value if the functional forms were unknown. The next two subsections show that mechanisms do exist which, under some assumptions, can induce efficient pricing even when the regulator does not know the functional forms of demand and costs.

2.5.2 The Vogelsang-Finsinger Mechanism and Related Schemes

The mechanism proposed by Vogelsang and Finsinger (1979) (V-F) is an ingenious but ultimately flawed device to induce a naturally monopolistic firm to set Ramsey prices in the long run. The regulator is unable to make transfers to the firm, but prices in period t are constrained to satisfy

$$\sum_{i=1}^{n} p_t^i q_{t-1}^i \le E_{t-1} \tag{2.17}$$

where E_{t-1} denotes the firm's observed expenditure in period $t-1$. The value of last period's quantities at the proposed prices must be at most equal to last period's expenditure. When there is only one good the constraint says that p_t cannot exceed average cost in period $t-1$. To implement this scheme the regulator does not need any knowledge of the demand function or of the current value of demand. The only demand information she needs is last period's quantities. Similarly she does not have to know anything about the cost function, but must be able to measure last period's expenditure. Since she knows last period's expenditure, quantities and prices she also knows last period's profits, $\Pi_{t-1} = \sum_{i=1}^{n} p_{t-1}^i q_{t-1}^i - E_{t-1}$. Subtracting $\sum_{i=1}^{n} p_{t-1}^i q_{t-1}^i$ from both sides of (2.17) gives another way of writing the constraint:

$$\sum_{i=1}^{n} (p_{t-1}^i - p_t^i) q_{t-1}^i \ge \Pi_{t-1} \tag{2.18}$$

A third way to write the constraint is

$$1 - \frac{\Pi_{t-1}}{\mathbf{p}_{t-1} \cdot \mathbf{q}_{t-1}} \geq \frac{\mathbf{p}_t \cdot \mathbf{q}_{t-1}}{\mathbf{p}_{t-1} \cdot \mathbf{q}_{t-1}}$$

where we have used the inner product notation. The right-hand side of this inequality is the Laspeyres price index and the left-hand side is at most 1 if $\Pi_{t-1} \geq 0$. As long as profits are always non-negative the constraint implies a monotonically decreasing Laspeyres chain index.

In order to analyze further properties of the mechanism we assume, following V-F, that demand and cost conditions are stationary so there are no future exogenous shocks that shift costs or demand. We can now say something about the change over time of consumer welfare. Using the result (2.1) on the convexity of $V(\mathbf{p})$ in (2.18) gives

$$V(p_t) - V(p_{t-1}) \geq \sum_{i=1}^{n} (p_{t-1}^i - p_t^i) q_{t-1}^i \geq \Pi_{t-1}. \tag{2.18$'$}$$

The *change* in consumer surplus is at least as large as the *level* of profits in $t-1$. The firm returns its previous rents to consumers. In particular if $\Pi_{t-1} \geq 0$ the change in consumer surplus is non-negative. Equation (2.18′) also implies that the change in welfare (defined as consumer surplus plus profits) is at least as large as current profits. Subtracting Π_{t-1} from both sides of (2.18′) and adding Π_t gives

$$W_t - W_{t-1} \equiv V(p_t) - V(p_{t-1}) + \Pi_t - \Pi_{t-1} \geq \Pi_t.$$

Thus if $\Pi_t \geq 0$ then both consumer surplus and welfare increase. The last point to be made about (2.18) is that if prices are constant then profits cannot be positive. In any steady state profits are zero.

Under what conditions will profits be non-negative, thus guaranteeing that consumer surplus and welfare (weakly) increase each period? Vogelsang and Finsinger make two assumptions. First, the firm is myopic. It maximizes current profits without caring about the fact that this period's choice of prices affects the next period's constraint. One immediate implication of myopia is that the firm will never spend money on wasteful expenditure to manipulate future price constraints. Second, the cost function has decreasing ray average costs. This means that if the output vector $\mathbf{q}$ is multiplied by some scalar $\alpha \geq 1$ then $C(\alpha\mathbf{q}) \leq \alpha C(\mathbf{q})$ and is a generalization of single-product scale economies. It is sufficient but not necessary for the cost function to be subadditive, i.e. for the industry to be a natural monopoly. This assumption allows the firm to make a non-negative profit if profits were non-negative in the previous period. This is easiest to explain in the single-product case. Setting p_t equal to the maximum allowed by (2.17) gives $p_t = C_{t-1}/q_{t-1} = AC_{t-1}$ where AC denotes average cost. This implies that $\Pi_t = AC_{t-1}q_t - C_t$. The sign of Π_t is thus the same as the sign of $AC_{t-1} - AC_t$, which is positive since $p_t \leq p_{t-1}$ and average cost falls with output.

Thus under myopia and decreasing ray average costs, and given an initial price vector $\mathbf{p_0}$ that generates $\Pi_0 > 0$, the V-F mechanism causes consumer surplus and welfare to increase (weakly) each period, profits are non-negative because the firm maximizes profits each period, and profits to tend to zero (if a steady state is reached). We now show that the steady-state price vector satisfies the necessary condition for Ramsey prices. The treatment here follows Vogelsang (1989a). The original V-F assumptions will be used but we allow the firm to be a present value maximizer. The firm's problem is to choose the sequence of price vectors to maximize the present value of profits subject to the V-F constraint (2.18). The Lagrangian is:

$$\mathcal{L} = \sum_{t=1}^{\infty} \beta^{t-1} \left\{ \Pi(\mathbf{p_t}) + \lambda_t \left(\sum_{i=1}^{n} [p^i_{t-1} - p^i_t] q^i_{t-1} - \Pi(\mathbf{p_{t-1}}) \right) \right\} \tag{2.19}$$

where λ_t is the multiplier for the constraint in period t. The first-order condition for the choice of p^k_t gives

$$\frac{\partial \Pi_t}{\partial p_t^k} - \lambda_t q_{t-1}^k + \beta\lambda_{t+1}\left(q_t^k + \sum_{i=1}^{n} [p_t^i - p_{t+1}^i]\frac{\partial q_t^i}{\partial p_t^k} - \frac{\partial \Pi_t}{\partial p_t^k}\right) = 0. \tag{2.20}$$

A marginal increase in p_t^k affects current profits. It tightens the period t constraint by q_{t-1}^k, which costs the firm $\lambda_t q_{t-1}^k$ in period t and affects the period t+1 constraint. If the firm reaches a steady state in which all prices and quantities are constant then all the time subscripts can be dropped from (2.20) yielding

$$(1-\beta\lambda)\frac{\partial \Pi}{\partial p^k} = \lambda(1-\beta)q^k \quad \text{for } k = 1, \ldots, n. \tag{2.21}$$

We can put a bound on the value of λ. Since profits are zero in a steady-state the constraint binds, $\lambda > 0$, and the right-hand side of (2.21) is positive. On the left-hand side we must have either $(1-\beta\lambda) < 0$ and $\partial\Pi/\partial p^k < 0$ or $(1-\beta\lambda) > 0$ and $\partial\Pi/\partial p^k > 0$. But if we had $\partial\Pi/\partial p^k < 0$ then the constraint would not bind since (2.21) holds for all n. Thus $\partial\Pi/\partial p^k > 0$ and $\lambda < 1/\beta$. It follows from (2.21) that

$$\frac{\partial\Pi/\partial p^k}{\partial\Pi/\partial p^l} = \frac{q^k}{q^l},$$

i.e. the necessary condition for Ramsey pricing holds. We also know that the firm's rent in steady state is zero. Note that it is possible that the second-order condition for Ramsey pricing is not satisfied at prices that satisfy (2.21). Vogelsang and Finsinger (1979) argue that such a case is unlikely.

The V-F mechanism is neat but it relies on restrictive assumptions. In an important critique Sappington (1980) showed that if the firm anticipates the imposition of V-F regulation in the following periods (or if V-F regulation is already underway) and $\beta > 0$ then it can have an incentive to indulge in wasteful expenditures. These are directly unproductive but are privately profitable because they allow higher prices in succeeding periods as the level of the allowed price vector rises with E_{t-1}. In future periods the waste can be reduced and higher profits can be reaped. Sappington presents a finite horizon example where the strategic behaviour induced by V-F regulation can lower welfare below the level achieved with no regulation. In steady state there will not be any waste. An extra unit of waste costs £1 currently but relaxes next period's constraint by £1. This relaxation has value λ next period and so has present value $\beta\lambda$. But we showed above that $\beta\lambda < 1$, so waste is unprofitable in steady-state. Other general results about waste in the infinite horizon V-F model are not available, but positive waste along the path to the steady state cannot be ruled out, and in the special cases that Sappington considers will certainly occur. Thus Sappington argues (p 369) that "the most attractive feature of the V-F regulatory mechanism may, in fact, prove to be its tragic flaw". The attractive feature is the lack of detailed cost and demand information that the regulator needs, but the presence of lagged expenditure in the constraint allows the firm to act strategically, as long as it values future payoffs.

The V-F model has been combined by Logan, Masson and Reynolds (1989) with the model of stochastic regulatory review of Bawa and Sibley (1980) – see also Lipman (1985) and Sibley (1985). If the regulator chooses to review the firm's prices in period t then prices must satisfy the constraint:

$$\sum_{i=1}^{n} p_t^i q_{t-1}^i \leq wL_{t-1} + sK_{t-1}$$

where L_{t-1} is the amount of labour hired in $t-1$, w is the wage rate, K_{t-1} is the capital stock and s is the target rate of return which exceeds the true (but unknown) cost of capital r. The profits that the regulator targets are $(s-r)K_{t-1}$, which is unobservable, and actual profits are $\sum p_{t-1}^i q_{t-1}^i - wL_{t-1} - rK_{t-1}$, so the difference between actual and target profits at time $t-1$ is $\Delta_{t-1} \equiv \sum p_{t-1}^i q_{t-1}^i - wL_{t-1} - sK_{t-1}$, which is observable. The firm faces a probability of review $\Phi(\Delta_{t-1})$ each period, which rises as Δ_{t-1} diverges from 0 in either direction and satisfies $\Phi(0) = 0$.[13] The regulator is assumed to react more frequently to profit overruns than to profit shortfalls. Using dynamic

[13] The regulator's behaviour is fully characterized by the probability of review function $\Phi(\Delta_t)$ and is not derived from optimizing behaviour.

programming techniques Logan *et al.* show that this scheme converges to a steady state characterized by both productive efficiency and Ramsey pricing. Productive efficiency occurs because in a steady state actual profits equal target profits and the probability of review is zero. Since the probability of review is zero the regulatory lag is effectively infinite and there is no benefit from Averch-Johnson-type overcapitalization even though $s > r$. The Ramsey pricing result occurs for similar reasons to those in the V-F model. This model captures a feature of US-style cost-of-service regulation in which the review period is endogenous and the likelihood of review depends on the firm's excess profits.

2.5.3 Approximate Surplus Subsidy Schemes

A closely related scheme to V-F was proposed by Finsinger and Vogelsang (F-V) (1981, 1982).[14] It is also related to the price caps considered in Chapter 5. It was originally proposed by Scott (1952, 1978) as a method for evaluating the performance of nationalized industries but was independently rediscovered by Finsinger and Vogelsang. Another very similar version has been proposed by Hagerman (1990). The F-V scheme is also related to the ISS scheme examined in 2.3.1. Unlike ISS the regulator does not know the demand function, although cost and demand functions are known to be stationary. Unlike V-F the regulator is able to grant the firm transfers. As usual the transfer can be thought of as a fixed fee in a two-part tariff. Under the F-V scheme the firm receives a transfer equal to the *linear approximation* to the change in consumer surplus less the operating profit earned in $t-1$. The transfer is

$$T_t = [p_{t-1} - p_t]q(p_{t-1}) - \Pi_{t-1} = C_{t-1} - p_t q(p_{t-1}) \tag{2.23}$$

where for ease of notation a one-good model has been assumed.[15] The transfer depends only on easily observable accounting data such as lagged costs and lagged demand and

[14]See also Gravelle (1983).

[15]The results can be generalized to the multiproduct case.

does not depend on any further knowledge of demand or costs. The first part of the transfer, $[p_{t-1} - p_t]q(p_{t-1})$, is the linear approximation to the actual change in consumer surplus, $V(p_t) - V(p_{t-1})$. Since $V(p)$ is convex the actual change in consumer surplus exceeds $[p_{t-1} - p_t]q(p_{t-1})$. The firm's income in period t is

$$I_t = \Pi_t + T_t = p_t[q(p_t) - q(p_{t-1})] - C(q(p_t)) + C(q(p_{t-1})).$$

This is an approximation to the change in welfare, whereas under ISS income is exactly the change in welfare. As long as $I_t \geq 0$ welfare will increase (weakly) each period under F-V. By the convexity of $V(p)$ we have

$$W_t - W_{t-1} \equiv V(p_t) - V(p_{t-1}) + \Pi_t - \Pi_{t-1} \geq q_{t-1}[p_{t-1} - p_t] + \Pi_t - \Pi_{t-1} \equiv I_t$$

so if $I_t \geq 0$, $W_t - W_{t-1} \geq 0$. The firm's income is bounded above by the change in welfare. This is illustrated in Figure 2.2. If the firm sets p_1 it receives compensation for the price cut of p_0DEp_1, which is $[p_0 - p_1]q_0$, and must pay back $\Pi_0 = p_0DHK$.[16] Its income in period 1 is then EFGH. This is lower than its income would be under ISS for the same p_1 by area DFE, which is the net increase in consumer surplus.

The F-V scheme can be thought of as a special case of the V-F mechanism when one product is completely inelastic in demand. This could be because the firm sets a two-part tariff that does not affect the number of consumers. Let $q^1 \equiv 1$ be the fixed number of consumers. The V-F constraint (2.17) is then, if it binds, $p_t^1 + p_t^2 q_{t-1}^2 = E_{t-1}$. This implies a value of the fixed charge p_t^1 exactly equal to the F-V transfer in (2.23).

The firm's problem is to choose the sequence of prices to maximize the present value of its income stream:

$$Z = \sum_{t=1}^{\infty} \beta^{t-1} \Big\{ p_t[q(p_t) - q(p_{t-1})] - C(q(p_t)) + C(q(p_{t-1})) \Big\}.$$

[16] We normalize fixed costs to zero.

One desirable property of the scheme is clear. The firm has no incentive to spend on waste because the net present value of a pound spent on waste is $-1+\beta<0$. Expenditure on waste is exactly returned in the next period via a higher transfer but because of discounting this is not profitable. The first-order condition for p_t gives

$$q(p_t) - q(p_{t-1}) + [p_t - MC_t - \beta(p_{t+1} - MC_t)]q'(p_t) = 0. \tag{2.24}$$

At the margin a small increase in p_t raises current operating profits by $q_t + [p_t - MC_t]q'(p_t)$ and reduces the current transfer by q_{t-1}. It also affects next period's transfer. Since $T_{t+1} = C_t - p_{t+1}q_t$, a unit rise in p_t raises next period's transfer by $(MC_t - p_{t+1})q'(p_t)$. These effects on the t+1 transfer are discounted by β. This is the trade-off represented in (2.24).

What happens in a steady state? Setting prices constant in (2.24) yields $q(p_t) = q(p_{t-1})$ and so $p_t = p_{t+1} = MC_t$. This should not be surprising when the relationship to the V-F mechanism is noted.[17] The V-F mechanism generates Ramsey pricing in the limit, but when a transfer is possible Ramsey prices are just marginal cost prices. We know that the F-V scheme can be thought of as a special case of V-F where $q^1 \equiv 1$. Finsinger and Vogelsang (1982) provide a formal proof of convergence to the steady state.

F-V has several advantages over V-F. It relies on more general assumptions and the welfare implications are more favourable. The cost function can be completely general and need not exhibit decreasing ray average costs, the firm need not be assumed to be myopic and the steady-state price is equal to marginal cost. The firm will not want to waste whereas under the V-F mechanism waste cannot be ruled out. The superiority of F-V is not surprising since it is effectively the V-F mechanism with the addition of lump-sum transfers.

[17] Hagerman (1990) proposes a version of the V-F mechanism that is effectively the same as assuming that the regulator can make transfers of the F-V type, subject to an upper bound on the transfer.

Suppose that the regulator knows the demand curve and could implement either ISS or F-V. The ISS scheme is preferable from a welfare point of view if social welfare is the unweighted sum of consumer and producer surpluses. The firm also prefers the ISS scheme since the convexity of consumer surplus guarantees that income with ISS, $V(p_t) - V(p_{t-1}) + \Pi_t - \Pi_{t-1}$, exceeds income under F-V, $[p_{t-1} - p_t]q_{t-1} + \Pi_t - \Pi_{t-1}$. But the preference of consumers depends on their discount factor, which we denote by β_c. Under ISS welfare is maximized immediately, but the increment to welfare does not accrue to *consumers* until period 2. But under F-V consumers capture some of each change in welfare. In particular their surplus increases in the first period, unlike in the ISS case. This implies that when $\beta_c = 0$ the F-V scheme is preferred by customers to ISS. By continuity F-V is also preferred for some positive values of β_c.

One interesting problem with F-V arises when costs depend on effort.[18] Let production costs be given by $C(q(p_t), e_t)$ where e_t is the effort level. Assume $\partial C/\partial e < 0$ and $\partial^2 C/\partial e^2 \geq 0$. Note that there are still no intertemporal effects in the cost function – e_t does not directly affect C_{t+1}. Effort might be thought of as R&D expenditure that is not observed by the regulator, is not included in her calculation of total expenditure and either depreciates fully each period or is current spending anyway. As in the Laffont-Tirole models the cost of effort is given by the increasing, convex function $\psi(e_t)$, where $\psi(0) = 0$. The firm's rent is now

$$I_t = p_t[q(p_t) - q(p_{t-1})] - C(q(p_t), e_t) + C(q(p_{t-1}), e_{t-1}) - \psi(e_t)$$

Note that the lagged cost of effort, $\psi(e_{t-1})$, does not appear in the transfer because it is unobservable. A marginal increase in effort lowers current production costs and so raises current operating profits by $-\partial C_t/\partial e_t$. It costs the firm $\psi'(e_t)$ in period t. But it also lowers the next period's transfer by $\partial C_t/\partial e_t$, which is discounted by β. The Kuhn-Tucker condition for effort is thus

[18] A similar problem was examined by Vogelsang (1983).

$$\left(-1+\beta\right)\frac{\partial C_t}{\partial e_t} - \psi'(e_t) \leq 0. \tag{2.25}$$

The socially optimal effort level for a given value of q_t is given by $-\partial C_t/\partial e_t = \psi'(e_t)$ and (2.25) implies an inefficient effort level along the path to steady state, except when $\beta = 0$. In steady state effort is zero whatever the value of β, since the firm receives no net income to balance against the cost of effort. Exactly the same problem would occur if the firm is subject to the ISS scheme but the cost of effort is unobservable. The problem arises because the firm pays back its previous operating profits, $p_{t-1}q_{t-1} - C_{t-1}$, but these are larger than its true profits by $\psi(e_{t-1})$. Vogelsang (1983) suggests an alternative transfer that generates long run efficiency even if $\psi(e)$ is unobservable. The new transfer is

$$T_t = T_{t-1} + q_{t-1}[p_{t-1} - p_t] \tag{2.26}$$

which differs from that in (2.23) in two ways.[19] First, the firm does not have Π_{t-1} taxed away, implying that costs do not enter into the transfer at all. Second, the firm is allowed to retain the previous transfer T_{t-1}. Vogelsang shows that this produces optimal effort (given the price chosen) each period.

Cox and Isaac (1987) describe the results of laboratory experiments of the Loeb and Magat (L-M) (1979) scheme and the Finsinger and Vogelsang subsidy process based on (2.26). The L-M scheme performed well but the scheme described by (2.26) did not. The problem was that the autoregressive characteristic of (2.26) makes it very unforgiving of errors. If price is too low by mistake in one period and is raised to compensate in the next period then the firm faces a heavy tax. Cox and Isaac observed that bankruptcy frequently occurred.

The ISS scheme uses the true change in consumer surplus as part of the transfer and ensures that $p = MC$ from the first period. The F-V scheme sets a transfer that

[19]The scheme is also the same as the subsidy adjustment process proposed by Finsinger and Vogelsang (1981).

depends on a linear approximation to the change in consumer surplus and produces a gradual approach to marginal cost pricing. Vogelsang (1988) asks whether a better approximation to the change in consumer surplus produces a scheme that in some sense is in between F-V and ISS. The second-order approximation for $V(p_t) - V(p_{t-1})$ is

$$\begin{aligned} V(p_t) - V(p_{t-1}) &\simeq q(p_{t-1})[p_{t-1} - p_t] + \tfrac{1}{2} q'(p_{t-1})[p_{t-1} - p_t]^2. \\ &\simeq \tfrac{1}{2}[q_{t-1} + q_t][p_{t-1} - p_t] \end{aligned} \tag{2.27}$$

The regulator does not know the slope of the demand function $q'(p_{t-1})$ but she can approximate it by $[q_t - q_{t-1}]/[p_{t-1} - p_t]$, giving the second expression in (2.27). Vogelsang shows that a scheme which includes this approximation in the transfer might not reach a steady state where $p = MC$, and can result in cycles. The difference between the intermediate scheme and both ISS and F-V is that under both ISS and F-V *at most* the exact change in $V(p)$ is transferred, whereas under the intermediate scheme the approximation can sometimes lead to more than the increment in $V(p)$ being transferred (for example when the demand function is convex). It is this over-payment that leads to the possibility of cycles with the intermediate scheme.

The assumption underlying the F-V mechanism is that welfare is the unweighted sum of consumer surplus and profits, and thus the full-information benchmark is marginal cost pricing. In the Laffont and Tirole models, however, the benchmark is Ramsey pricing – see (2.13). Laffont and Tirole show (1990, 1993, Chapter 2) that their regulatory scheme can be implemented by delegating pricing decisions to the firm. In particular they suggest that the firm receives a transfer of

$$T_t = \frac{[p_{t-1} - p_t]q_{t-1}}{(1+\gamma)(1-\beta)}$$

where both β, the discount factor, and γ, the shadow cost of public funds are assumed to be known. Like F-V the L-T transfer is based on the linear approximation to the change in $V(p)$, but unlike F-V lagged profits are not taxed away. The firm chooses its price and effort level subject to the above transfer. It is allowed to keep a constant fraction, $b \leq 1$, of its profits including the transfer. Its rent in period t is thus $b[p_t q_t - C_t + T_t] - \psi(e_t)$. The parameter b reflects the trade-off between incentives and rent extraction given by (2.14). The cost of effort, which is unobservable, is borne fully by the firm. Laffont and Tirole (1993, pp. 142-143) show that the firm will in steady state set the optimal (Ramsey) price, and each period its effort level is optimal given the price that is set.

2.6 Conclusions

The models surveyed in this chapter are differentiated by their assumptions about the source and extent of the informational asymmetry, and by their modelling techniques. One characteristic that is common to all the models, except the Vogelsang-Finsinger mechanism discussed in 2.5.2, is the use of transfers. Another typical characteristic of the non-Bayesian schemes is the limitation of the firm's rents by taxation of lagged profits. This occurs in the ISS, V-F and F-V schemes and requires the regulator to be able to observe the previous period's costs. If even lagged cost observation is not feasible then the regulatory scheme cannot be conditioned on costs and the only feasible alternative is to set a price cap. In the following chapters we shall examine the properties of non-Bayesian mechanisms that do not tax profits and (usually) do not rely on transfers, i.e. price caps.

CHAPTER 3

PRICE CAPS WITH LINEAR TARIFFS

3.1 Introduction

If the regulator is unable to observe costs, has limited demand information and cannot make transfers, then the only regulatory option is to control prices. Price caps have become the subject of much academic interest since their application to regulated utilities in Britain and the United States in the 1980s. In this chapter we analyze and compare the properties of five types of price cap that can be applied to a multi-product monopolist. Products might be differentiated by location (natural gas supplied to a customer located near the beachhead costs less to transport than gas supplied to the other side of the country), by time of purchase (electricity supplied at peak times costs more because of the need for extra capacity) or by intrinsic characteristics (a long-distance telephone call is different from a local call). Usually the firm is allowed some freedom to alter relative prices subject to a cap on average prices. In addition it can set prices below the cap, but such behaviour is only rational if the constraint does not bind or if a regulator contemplating a formal tightening of the cap might be appeased by a reduction in short-run profits below the maximum feasible level. In Britain the cap is fixed for a known period of four or five years within which the regulator commits more or less not to intervene. At the review date a new cap is set on the basis of expected future productivity gains and demand growth rather than on the basis of past profits. For more details on price caps see Armstrong, Cowan and Vickers (1994, Chapter 6), Rees and Vickers (1995), and the symposium on price-cap regulation in the *Rand Journal of Economics* Autumn 1989, especially the introduction by Acton and Vogelsang and the paper by Beesley and Littlechild.

In this chapter we assume that the firm is restricted to setting a linear tariff. The consumer's payment for each product is strictly proportional to the quantity purchased. There are no fixed fees or quantity discounts and purchases of one product are not tied to purchases of another. Special cases of linear tariffs are (i) third-degree price discrimination, where the markets are separable and total cost depends only on total output, and (ii) peak-load pricing, where the firm sells the same product at different times of the day or year with costs that depend on the load factor. We must distinguish *linear* tariffs from *uniform* tariffs. If there are n products then revenue with a linear tariff is $\sum_{i=1}^{n} p^i q^i$, where prices for each product can differ. A uniform tariff is a special case of a linear tariff that has the same price, p, for all products, so revenue is $p\sum_{i=1}^{n} q^i$. When comparing the properties of the different price caps it is often useful to assume that prices were uniform before the price cap was imposed. This could be justified by both its analytical convenience and the fact that under public ownership many utility services were priced uniformly.

The first type of price cap that we analyze is known as the *tariff basket* mechanism (TB). This is used in the UK to regulate British Telecom and a version applies to the water companies. The average price that is capped is a Laspeyres price index, and this method of regulation is used for firms that sell products that cannot sensibly be added up, such as local and long-distance telephone calls. The second type uses *fixed weights* in the price index and will be denoted by F. The third type is *average revenue* regulation (AR). Predicted revenue from all the firm's regulated activities is divided by predicted total output and the resulting average revenue is constrained not to exceed a given number. This requires commensurability of the firm's outputs, i.e. it must be possible to add them to find a measure of total output. In Britain average revenue is the basis of regulation of British Gas, the twelve regional electricity companies, British Airports Authority and Manchester Airport. An alternative type of regulation that requires commensurability of the products is *average revenue (lagged)* regulation (ARL). This puts a cap on the value of the previous period's quantities at this period's prices divided by lagged total output. It has been used in the United States to regulate AT&T.[1] The final type of price cap is new, and we shall call it the *Paasche Price Index* (PPI). Like AR it requires forecasts of demand,

[1] In AT&T's case there are three separate caps that apply to distinct baskets of services. Services within a particular basket are commensurable. See Sappington and Sibley (1992).

but unlike AR it does not require commensurability of outputs. It is included in the chapter for the purposes of comparison.

General assumptions that are made throughout this chapter are: (i) demand and cost functions are stationary; (ii) the regulator does not have current or past cost information; (iii) the regulator might know current demand for any proposed prices and certainly knows previous demands; (iv) the regulator cannot make transfers to the firm; (v) the firm cannot chooses its quality level; (vi) all customers are final consumers, with quasi-linear utility functions; (vii) income is optimally distributed; (viii) there are no externalities in consumers' utility functions; (ix) the definitions of markets or products are given (e.g. the cut-off point between peak and off-peak periods cannot be chosen by the firm); and (x) the regulator commits to the price cap and does not revise it.

Assumption (i) is standard in the literature, and facilitates welfare comparisons across periods. It is relaxed in Chapter 7 where we allow for shifts in demand. Assumption (ii) means that the regulator cannot condition prices on cost information, so she cannot use, for example, the Vogelsang-Finsinger scheme (see Chapter 2.5.2). Assumption (iii) allows the regulator to control prices as a function of current demand in the AR and PPI cases. The no-transfers assumption (iv) means that surplus subsidy schemes are not available. Assumptions (v), (vii), and (viii) are relaxed in Chapter 7, where we consider the rôles of quality regulation, income distribution considerations and network externalities. The force of assumption (x) is to make the price cap "pure". With no cost information the regulator cannot observe profits, even with a lag, and she thus cannot revise the price cap to force the firm to return profits to consumers. In reality price caps are revised at fixed intervals, but the no-revision case is a useful benchmark, and allows comparison with the non-Bayesian mechanisms such as those of Vogelsang and Finsinger (1979) and Finsinger and Vogelsang (1982) discussed in Chapter 2.

The focus is on the study of implementable price caps that are related to actual schemes that are used in Britain and the United States, although we do not allow for revisions of the price caps. The schemes are all non-Bayesian. We argued in Chapter 2 that the Bayesian analysis of regulation, while providing many insights, is not especially helpful for policymakers. But non-Bayesian mechanisms are inevitably somewhat *ad hoc.* Rather than designing an optimal scheme from first principles we characterize existing and potential price caps and ask a series of questions to determine the properties of each scheme. What information does the regulator need in order to implement the scheme? What is the rationale, if any, for the particular form of price cap? What happens to profits, consumer surplus and total welfare when the cap is imposed? Are regulated price structures likely to be efficient? Which type of cap do the regulator, consumers and the firm prefer?

All price caps schemes that satisfy the above assumptions will induce the firm to minimize costs. Unlike a firm facing the Vogelsang-Finsinger (1979) mechanism, a firm with a price cap has no incentive to waste. Costs play no rôle in the level of the price cap and waste will simply lower profits. When costs depend on current effort the firm will always exert the effort level that is optimal given the prices that are set. This is because it keeps any incremental profits. Indeed the provision of incentives for cost reduction was one of the main arguments for price cap regulation when it was first introduced in Britain (Littlechild, 1983). To ease the notational burden, however, and because the focus is on prices, we shall assume that effort is not a choice variable. This is without loss of generality because effort levels will always be optimal *for given prices.*

We now summarize briefly the existing literature. Bradley and Price (1988) show that AR typically generates prices that do not have the Ramsey property using a linear example. Vickers and Yarrow (1988, pp 101-107 and 271-273) show that when markets have different elasticities of demand and marginal cost is common, AR leads to inefficient price structures. Armstrong, Cowan and Vickers (1994, Chapter 3) show that prices under AR do not have the Ramsey property in a model with common marginal cost and inelastic demands. These results are generalized in 3.2.3. Bradley and Price (1988) also argue that TB regulation is preferable to AR, but they interpret TB regulation as a form of Vogelsang-Finsinger (1979) regulation. The V-F mechanism requires decreasing ray average costs and, more importantly, the constraint is progressively tightened since the firm has to return to consumers the previous period's profit. It does not seem reasonable to compare AR where the level of the cap is constant through time to V-F where the constraint becomes progressively tighter. Our form of TB regulation is defined without reference to profits or costs. We shall find in Chapters 4 and 5 that there are well-defined cases where AR generates an outcome that is preferable in welfare terms to TB. Armstrong and Vickers (1991) consider whether a firm should be allowed to depart from uniform pricing subject to a cap on average prices. They focus on the case of AR regulation, and do not consider dynamic issues. They also analyze the fixed weights case. Vogelsang (1989b) and Brennan (1991) show that TB generates efficient pricing. Sappington and Sibley (1992) analyze ARL assuming that demand is linear, that the firm can set a fixed fee that acts as a transfer and that the horizon is finite. Our analysis of ARL considers the infinite horizon, and allows for many products with a general demand function, but we delay consideration of two-part tariffs until Chapter 5. All the results on ARL in this chapter are thus new.

The chapter is organized as follows. In Section 3.2 we present the main results on the tariff basket, fixed weight, average revenue, average revenue (lagged) and Paasche Price Index schemes. Two further schemes that are closely related to the five main ones are discussed in Section 3.3. Conclusions are in 3.4.

3.2 Five Price Caps

The order in which we analyze each price cap is determined by both history and analytical convenience. The tariff basket scheme was applied to British Telecom in 1984, average revenue regulation was first applied in 1986, and average revenue (lagged) was applied to AT&T in 1989. PPI has not been applied anywhere, for reasons that will become apparent, and likewise fixed weights regulation has not been used. More importantly the order chosen facilitates comparison between them. The diagrammatic representation of average revenue regulation draws on the tariff basket mechanism. Both the average revenue (lagged) scheme and the PPI scheme can be thought of as close relations to TB and AR. The fixed weights scheme is similar to TB. In each case we derive general results that follow simply from the application of the price constraint under the assumptions described in 3.1 and the assumption that the firm maximizes the present value of its profit stream. We also characterize the firm's constrained optimization problem, with the emphasis on steady-state properties. Often there is little that can be said about the stability of, and path to, a steady state when the firm sets linear tariffs. More can be said when the firm sets a two-part tariff, and we postpone discussion of these important issues until Chapter 5. We also compare each scheme with the preceding ones, and ask which scheme consumers, the firm and the regulator prefer.

3.2.1 The Tariff Basket Mechanism

This scheme requires that the cost to the representative consumer of buying the bundle purchased in period $t-1$ at period t prices should be no greater than at $t-1$ prices. In symbols,

$$\sum_{i=1}^{n} [p^i_{t-1} - p^i_t] q^i(p^i_{t-1}) \geq 0. \tag{3.1}$$

This can be compared with equation (2.18) which defines the Vogelsang-Finsinger mechanism. The difference is that lagged profits do not appear on the right hand side of (3.1). Rearranging gives

$$\frac{\sum_{i=1}^{n} p^i_t q^i_{t-1}}{\sum_{i=1}^{n} p^i_{t-1} q^i_{t-1}} \leq 1$$

where the left-hand side is a Laspeyres price index. A third way to express (3.1) is:

$$\sum_{i=1}^{n} \frac{(p^i_t - p^i_{t-1})}{p^i_{t-1}} \cdot \frac{p^i_{t-1} q^i_{t-1}}{\sum_{i=1}^{n} p^i_{t-1} q^i_{t-1}} \leq 0, \tag{3.1'}$$

which requires the weighted average growth rate of prices to be at most zero. The weight on the growth rate of p^i is the share of product i in total revenue in period $t-1$. The price control formulae for British Telecom (BT) and the water companies differ slightly from (3.1′). Instead of having 0 on the right-hand side they have $RPI - X$, where RPI is the percentage rate of growth of the Retail Prices Index and X is an exogenous number. The X factor gives the required fall each year in average real prices – see Armstrong, Cowan and Vickers (1994, Chapter 6). Since we are assuming stationarity in the cost and demand functions, and that the regulator never revises the cap, without essential loss of generality we take X to be zero. Similarly we assume that general consumer price inflation is zero.[2] In order to implement the TB mechanism the regulator needs only information on lagged quantities and prices and on the proposed prices.

What is the rationale for the TB mechanism? First, it allows for the averaging of prices of incommensurable products. A Laspeyres price index does not require the calculation of total output. Second, it guarantees that consumer surplus increases (weakly) each period, given stationarity of demand. This follows from the convexity of the indirect utility (or consumer surplus) function $V(\mathbf{p})$. We have

$$V(\mathbf{p}_t) - V(\mathbf{p}_{t-1}) \geq \sum_{i=1}^{n} [p^i_{t-1} - p^i_t] q^i(p^i_{t-1}) \geq 0 \tag{3.2}$$

where we have used (3.1). Intuitively in period t the representative consumer is able to purchase the $t-1$ period bundle at cost no greater than in period $t-1$. If he continues to purchase the $t-1$ bundle, and the constraint holds with equality, he is as well off as before. In general he will purchase a different bundle and will be strictly better off.[3] The only case where the same bundle is purchased, assuming $\mathbf{p}_t \neq \mathbf{p}_{t-1}$, is when all

[2] In practice even perfectly anticipated inflation can affect the decisions of firms because of imperfect indexation, for example in the corporate tax system.

[3] We can also use the Laspeyres price index interpretation to show this. When this price index is below the total expenditure index (which is 1 by the assumption of stationarity) the consumer is better off. See Varian (1993, pp 132-133).

products are perfect complements. We assume throughout this chapter that this is not the case, i.e. that the Slutsky substitution matrix is negative definite and not just negative semi-definite. In geometrical terms the iso-consumer surplus curve in price space is assumed to be strictly convex.

From the firm's point of view the TB scheme always allows it to choose the previous period's prices. This is important if the regulator is worried about the cap being too tight and the firm earning sub-normal profits. If profits in period 0, before the scheme starts, are known to be non-negative then the regulator can be confident that TB will not drive the firm to bankruptcy. The fact that the previous period's prices are available under TB also means that the firm is better off with TB than it would be if it was required simply to keep prices fixed at $\mathbf{p}_0$. The firm's profits need not fall from one period to the next. In fact if the firm is myopic the sequence of profits will be non-decreasing. Since consumer surplus is non-decreasing also the TB mechanism generates a non-decreasing sequence of welfare levels when the discount factor is zero. This result is independent of the weights on consumer surplus and profits in social welfare. In the more general case where future profits are not fully discounted we cannot be certain that profits will be non-decreasing but we do know that because the previous period's prices are available the present value of profits is above the value when prices remain at $\mathbf{p}_0$, i.e. $\sum_{t=1}^{\infty} \beta^{t-1}\mathrm{H}(\mathbf{p}_t) \geq \Pi(\mathbf{p}_0)/(1-\beta)$, where β is the discount factor. As long as the social discount factor is the same as that of the firm the imposition of TB will raise welfare above that obtained with keeping prices at $\mathbf{p}_0$.

In choosing prices the firm has an optimal control problem. It maximizes the present value of profits subject to the sequence of constraints, taking account of the fact that this period's prices affect next period's constraint. We shall characterize the price path and the steady state (if it exists) without being fully rigorous. The Lagrangian for this problem is[4]

[4]We could alternatively have written the problem as a dynamic programme. The Bellman equation is $\mathrm{W}_t(\mathbf{p}_{t-1}) = \mathrm{Max}\ \Pi(\mathbf{p}_t) + \beta\mathrm{W}_{t+1}(\mathbf{p}_t)$ subject to (3.1), where W(.) is the value function. If W(.) is differentiable and the solution is interior we can use the first-order and envelope conditions to characterize the sequence of prices. The Lagrangian approach involves much less notation. See Chow (1993) for a general argument for using Lagrangians rather than the Bellman equation.

$$\mathcal{L} = \sum_{t=1}^{\infty} \beta^{t-1}\left\{\Pi(\mathbf{p}_t) + \lambda_t \sum_{i=1}^{n} [p^i_{t-1} - p^i_t] q^i(\mathbf{p}_{t-1})\right\}.$$

We define $\Pi(\mathbf{p}_t) \equiv \sum_{i=1}^{n} p^i_t q^i(\mathbf{p}_t) - C[q^1(\mathbf{p}_t), \ldots, q^n(\mathbf{p}_t)]$ and assume that $\Pi(\mathbf{p})$ is concave. The first order conditions give

$$\frac{\partial \Pi_t}{\partial p^k_t} - \lambda_t q^k_{t-1} + \beta\lambda_{t+1}\left\{\sum_{i=1}^{n} [p^i_t - p^i_{t+1}]\frac{\partial q^i_t}{\partial p^k_t} + q^k_t\right\} = 0 \quad \text{for all k and t} \tag{3.3a}$$

$$\sum_{i=1}^{n} [p^i_{t-1} - p^i_t] q^i(\mathbf{p}_{t-1}) = 0 \quad \text{for all t} \tag{3.3b}$$

assuming that the constraint always binds. The interpretation of the pricing equation (3.3a) is as follows. An infinitesimal increase in p^k_t changes current profits by $\partial\Pi_t/\partial p^k_t$. Pushing up p^k_t forces the firm to cut other prices in period t, which has current cost $\lambda_t q^k_{t-1}$. It also affects the period t+1 constraint, by the term in curly brackets, and each pound of this effect has present value in period t of $\beta\lambda_{t+1}$.

What happens if the firm reaches a steady state in which prices are constant? Dropping the time subscripts in (3.3a) and (3.3b) and setting $p_{t-1} = p_t = p_{t+1}$ gives:

$$\frac{\partial \Pi}{\partial p^k} = \lambda(1-\beta)q^k \quad \text{for } k = 1, \ldots, n. \tag{3.4}$$

We can divide equation (3.4) for product k by the same equation for product l to get:

$$\frac{\partial\Pi/\partial p^k}{\partial\Pi/\partial p^l} = \frac{q^k}{q^l}$$

which is the necessary condition for Ramsey pricing – see Chapter 2.2. As usual, however, we cannot guarantee in general that (3.3) is sufficient for Ramsey pricing.[5] Since profits are not returned to consumers via the price constraint we expect profits to

[5] Brennan (1991) proves convergence to Ramsey prices using a more restricted model. He assumes that there are no demand interdependencies, that the firm is myopic ($\beta = 0$) and that the price vector that maximizes $V(\mathbf{p})$ subject to $\Pi(\mathbf{p}) = \bar{\Pi}$ for all $\bar{\Pi}$ is unique.

be strictly positive in steady state. This Ramsey pricing result has been proved by Vogelsang (1989b) and Brennan (1991).

The TB mechanism is illustrated in Figure 3.1. The initial prices are $\mathbf{p}_0$, where $\Pi(\mathbf{p}_0) > 0$. The period 1 constraint is the straight line TB_1, which has slope $-q_0^1/q_0^2$ and goes through $\mathbf{p}_0$. Any price vector that satisfies the constraint puts the representative consumer on a lower iso-surplus curve than V_0. If the firm was myopic it would choose $\mathbf{p}_M$ on TB_1, where profits are maximized. The non-myopic firm, however, prefers not to maximize current profits subject to the constraint, as the first-order condition (3.3a) demonstrates. Instead it chooses prices such as $\mathbf{p}_1$, yielding a superior constraint, TB_2, in the next period than could be obtained by pricing at $\mathbf{p}_M$. The change in relative prices induces a change in relative demand, so TB_2 is flatter than TB_1.

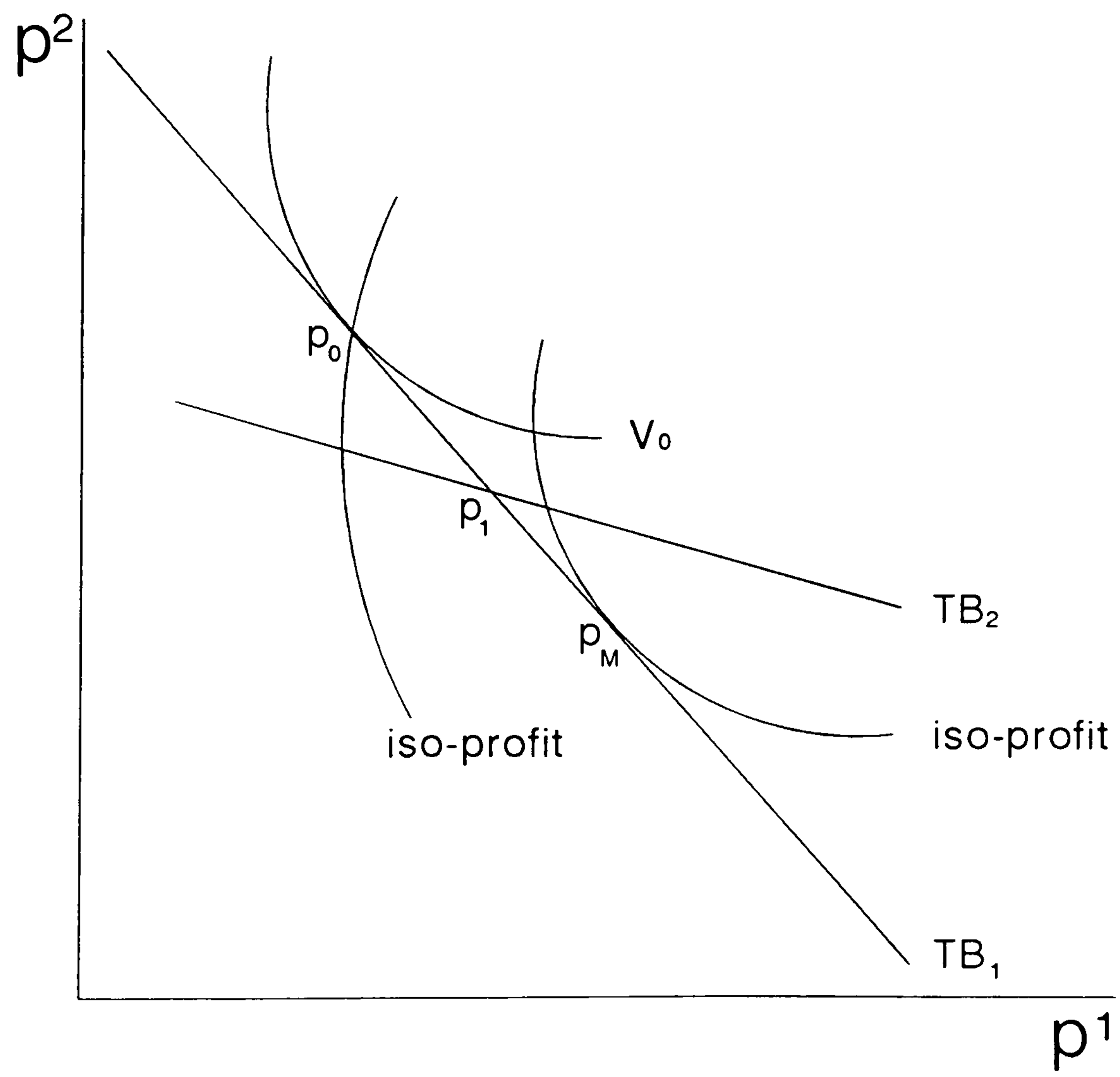


Figure 3.1 The Tariff Basket Scheme

The intuition for the Ramsey pricing result can also be seen in Figure 3.1. Suppose that the steady-state prices are $\mathbf{p}_0$, where the iso-surplus curve is *not* tangential to the iso-profit curve. It is clear that there are price vectors, such as $\mathbf{p}_1$ and $\mathbf{p}_M$, that are both feasible under TB and preferable to remaining at $\mathbf{p}_0$. Thus $\mathbf{p}_0$ cannot be a steady-state point. Only when the iso-profit locus and the iso-surplus locus are tangential will there be no feasible and desirable moves.

Collecting our results gives the following proposition.

Proposition 3.1 *With an infinite sequence of tariff basket constraints prices will be such that:*

(i) *Consumer surplus increases (weakly) each period;*

(ii) *For $\beta > 0$ the present value of the profit stream is at least equal to* $\Pi(\mathbf{p}_0)/(1-\beta)$;

(iii) *For $\beta > 0$, and if the social discount factor is the same as the firm's, welfare is higher than when prices are kept at $\mathbf{p}_0$, whatever the weights in the social welfare function;*

(iv) *For $\beta = 0$ the profit sequence is non-decreasing;*

(v) *For $\beta = 0$ the welfare sequence is non-decreasing;*

(vi) *In steady state prices satisfy the necessary conditions for Ramsey pricing (Vogelsang, 1989b, and Brennan, 1991).*

The tariff basket mechanism has many desirable properties. The imposition of TB raises increases consumer surplus each period and allows the firm to increase profits. We did not need to know anything about the firm's pricing behaviour to prove these results. In steady state, if this exists, prices have the Ramsey property. We shall see below that the welfare effects of other price caps are more uncertain.

3.2.2 Fixed Weights Regulation

The fixed weight scheme (F) is a close relation to the TB scheme. Many of the results in this section, although not those on the comparison of F to TB, are based on Armstrong and Vickers (1991). The fixed weight constraint is:

$$\sum_{i=1}^{n} [p_0^i - p_t^i]w^i \geq 0.$$

where $\{w^i\}$ are the weights. The constraint applies for all periods. The natural weights to use are the base-period quantities: $\{w^i\} = \{q^i(\mathbf{p}_0)\}$. The constraint in period 1 is exactly the same as the TB constraint for that period, but thereafter the weights and the reference price vector, $\mathbf{p}_0$, are not updated. The firm will set the same price vector in all periods after the first. Consumers will be at least as well off with F as with the initial price vector since the constraint under F is the same as the first TB constraint. The firm will earn at least as much profit as with the original prices since the latter are available.

We can also show that the firm prefers TB to F. Under TB the firm can choose prices in period 1 that maximize profits subject to the price constraint, and can then retain the same prices for ever. Thus it can obtain the same profits as under F. But in general it can do better, for example by setting the same prices as under F in period 1 and thereafter taking advantage of the fact that the updating of the weights in the TB constraint offers more profitable price vectors, as shown in Figure 3.1. The preference of consumers between TB and F is more difficult to determine. Under F consumer surplus rise between periods 0 and 1 and thereafter is constant, whereas under TB it increases each period. But, as Figure 3.1 suggests, consumer surplus under F at prices $\mathbf{p}_M$ will be typically be higher than period 1 surplus under TB, as the firm acts strategically to manipulate future constraints. Clearly the discount factor will be important in determining which scheme gives a higher present value of consumer surplus. When the firm is myopic TB leads to the same outcome as F in period 1, but the updating of subsequent constraints and profit maximization ensure that both consumer surplus and profits rise each period. As long as the regulator has a positive discount factor TB will be preferable in this case. But we cannot say which of the two schemes generates higher welfare in general because of the uncertainty about the relative effect on consumers.

Prices under F will only satisfy the Ramsey property in a special case. The first-order condition for the kth price is

$$\frac{\partial \Pi}{\partial \mathrm{p}_1^k} = \lambda \mathrm{q}^k(\mathbf{p}_0).$$

Dividing this by the first-order condition for the price of product l gives:

$$\frac{\partial \Pi / \partial \mathrm{p}_1^k}{\partial \Pi / \partial \mathrm{p}_1^l} = \frac{\mathrm{q}^k(\mathbf{p}_0)}{\mathrm{q}^l(\mathbf{p}_0)}.$$

Only if the firm chooses to set $\mathbf{p}_1 = \mathbf{p}_0$, i.e. if the initial prices are Ramsey prices, will pricing be efficient under F.

In the general case where $\mathbf{p}_0$ is not the Ramsey price vector both profits and consumer surplus would be higher in period 2 if the regulator updated the constraint so that $\mathbf{p}_2$ is constrained by $\sum_{i=1}^{n} [\mathrm{p}_1^i - \mathrm{p}_2^i]\mathrm{q}_1^i \geq 0$. Because both sides would be better off with updating, the regulator might find it hard to commit to the fixed weight scheme although, as we shall see in Chapter 5, it can yield higher welfare than TB.

We summarize the properties of the fixed weight scheme in the following proposition.

Proposition 3.2 *Under fixed weights regulation prices will be such that:*

(i) Consumer surplus is higher in period 1 than in period 0, and thereafter is constant.

(ii) Profits are higher in period 1 than in period 0, and thereafter constant. The

firm prefers TB to F.

(iii) Welfare is higher in period 1 than period 0, and thereafter constant.

(iv) If $\beta = 0$ but consumers and the regulator have positive discount factors then TB yields higher present values of welfare and consumer surplus than F.

(iv) Prices will only satisfy the necessary conditions for Ramsey pricing if the period 0 prices themselves have the Ramsey property.

3.2.3 Average Revenue Regulation

Average Revenue (AR) regulation can be applied when the firm's products can be added together to derive a meaningful measure of total output. Examples are therms of gas, kilowatt hours of electricity and numbers of airport passengers. Average revenue in period t, defined as $\sum_{i=1}^{n} p_t^i q_t^i / \sum_{i=1}^{n} q_t^i$, is constrained to be at most some predetermined level, $\bar{p}_t$. In the British cases this level is given by the previous period's average revenue, although there are also provisions to carry forward any unused price increases. The *rate of growth* of average revenue is capped by RPI − X, in the same way that under TB the average rate of growth of prices is constrained by (3.1′). We continue to assume RPI and X are both zero. Since we assume that the constraint always binds we also take the level of the cap to be fixed each period at $\bar{p}$, which is defined by $\bar{p} \equiv \sum_{i=1}^{n} p_0^i q_0^i / \sum_{i=1}^{n} q_0^i$. Under average revenue regulation, therefore, the firm faces a static problem and we shall often drop time subscripts. The AR constraint is

$$\sum_{i=1}^{n} (\bar{p} - p^i) q^i(p) \geq 0. \tag{3.5}$$

In order to implement this the regulator needs to know what demands will be at particular prices. Unlike TB and F, which only required lagged demand information, AR regulation characterized by (3.5) implies that the regulator can forecast demand.[6] If the regulator has this amount of demand information then it would be better to use

[6] In practice neither the regulator nor the firm can forecast demand perfectly. British firms regulated by AR face penalties for incorrect estimation of demand (and other variables such as RPI).

the Loeb and Magat scheme (see 2.3.1), assuming that profits have the same weight in the welfare function as consumer surplus. This would induce immediate marginal cost pricing of each product even though the regulator does not know the cost function. But we have assumed that the regulator cannot make transfers, which precludes the use of the Loeb-Magat scheme.

The rationale for AR appears to be its simplicity and ease of comprehension. There is also a natural interpretation as a constraint on the profit margin. If variable costs depend linearly on total output, i.e. $C = c\sum_i q^i$, then variable profits under AR are $(\bar{p} - c)\sum_i q^i$. Profit per unit of output is fixed at $\bar{p} - c$. As long as $\bar{p} > c$ the firm has an incentive to maximize output. When some costs are exogenous and stochastic, average revenue regulation allows for a relatively simple form of cost passthrough, as $\bar{p}$ can be adjusted for exogenous changes in marginal cost. Suppose marginal cost is $c_1 + c_2$, where c_1 is the cost that the firm controls and c_2 is an exogenous input cost over which the firm has little or no control. If c_1 is known only by the firm but c_2 is common knowledge then the regulator might want to allow the firm to pass through changes in c_2, so that average revenue minus c_1 remains fixed. The constraint would then effectively be a cap on value added per unit of output. For more on cost passthrough see Armstrong, Cowan and Vickers (1994, Chapters 2 and 6).

Another possible rationale for AR comes from index number theory. If the Paasche price index, P, (which uses current quantity weights) exceeds 1 while income is constant then consumer surplus is lower (Varian, 1993, pp 132-133). A necessary condition for consumer surplus to rise is thus $P \leq 1$. If, as in Armstrong and Vickers (1991), the initial price vector is uniform, so all prices equal $\bar{p}$, then the Paasche price index is $P = \sum_i p^i q^i(\mathbf{p}) / \sum_i \bar{p} q^i(\mathbf{p})$. If $P \leq 1$, which is true if and only if (3.5) holds, then the necessary condition for consumer surplus to rise is satisfied. This condition is, however, not sufficient, and we will show that when $P = 1$ consumer surplus typically falls.

What can we say about the effect of AR regulation on consumer surplus when initial prices are not necessarily equal? We have the following result.

Result 1. Ambiguity of the effect on consumer surplus of Average Revenue regulation.

(i) *Consumer surplus will be at most equal to* $V(\mathbf{p}_0)$ *if* $\sum_{i=1}^{n}[\bar{p} - p_0^i]q_1^i \geq 0$.

(ii) *Consumer surplus will at least equal to* $V(\mathbf{p}_0)$ *if* $\sum_{i=1}^{n}[\bar{p} - p_1^i]q_0^i \geq 0$.

Proof.

(i) We revert to the subscript notation. Define $\bar{\mathbf{p}} \equiv (\bar{p}, \ldots, \bar{p})$ as the uniform price vector. The initial price vector is $\mathbf{p}_0$ (which is not necessarily the same as $\bar{\mathbf{p}}$) and the price vector under AR is $\mathbf{p}_1$. By convexity of $V(\mathbf{p})$,

$$V(\mathbf{p}_0) - V(\mathbf{p}_1) \geq \sum_{i=1}^{n}[p_1^i - p_0^i]q_1^i. \tag{3.6}$$

If (3.5) holds with equality then $\sum_{i=1}^{n} p_1^i q_1^i = \sum_{i=1}^{n} \bar{p} q_1^i$. Substituting into (3.6) gives

$$V(\mathbf{p}_0) - V(\mathbf{p}_1) \geq \sum_{i=1}^{n}[\bar{p} - p_0^i]q_1^i. \tag{3.6$'$}$$

If the right-hand side of (3.6′) is non-negative then the consumer is at least as well-off under the initial prices, $\mathbf{p}_0$, as under AR regulation.

(ii) To show this we write out (3.2) in our current notation,

$$V(\mathbf{p}_1) - V(\mathbf{p}_0) \geq \sum_{i=1}^{n}[p_0^i - p_1^i]q_0^i. \tag{3.7}$$

By definition $\sum_i p_0^i q_0^i \equiv \bar{p}\sum_i q_0^i$ and the right-hand side of (3.7) is $\sum_i[\bar{p} - p_1^i]q_0^i$. If this is non-negative then $V(\mathbf{p}_1) \geq V(\mathbf{p}_0)$. □

Checking the conditions in Result 1 is straightforward, requiring only simple price and demand information. The sign of the change in consumer surplus depends on the firm's pricing decision, whereas under TB and F it is always positive, whatever prices are set.

Corollary 1. *If the initial prices are uniform then AR regulation makes consumers worse off (Armstrong and Vickers, 1991). If the initial prices are not uniform then consumers can be better off under AR regulation, and will be better off if the firm chooses prices subject to the average revenue constraint that are uniform.*

The first part of the corollary, proved by Armstrong and Vickers (1991), holds because $\mathbf{p}_0 = \bar{\mathbf{p}}$ implies that condition (i) in Result 1 holds. Any movement along the AR constraint away from $\bar{\mathbf{p}}$ lowers V. The second part of the corollary is just the reverse of this. If initial prices are $\mathbf{p}_0 \neq \bar{\mathbf{p}}$ then setting $\mathbf{p}_1 = \bar{\mathbf{p}}$ raises V. Another corollary is:

Corollary 2. *If initial prices are uniform then consumers prefer tariff basket or fixed weight regulation to average revenue regulation.*

Under TB and F consumers always benefit, but when $\mathbf{p}_0 = \bar{\mathbf{p}}$ they are worse-off with AR.

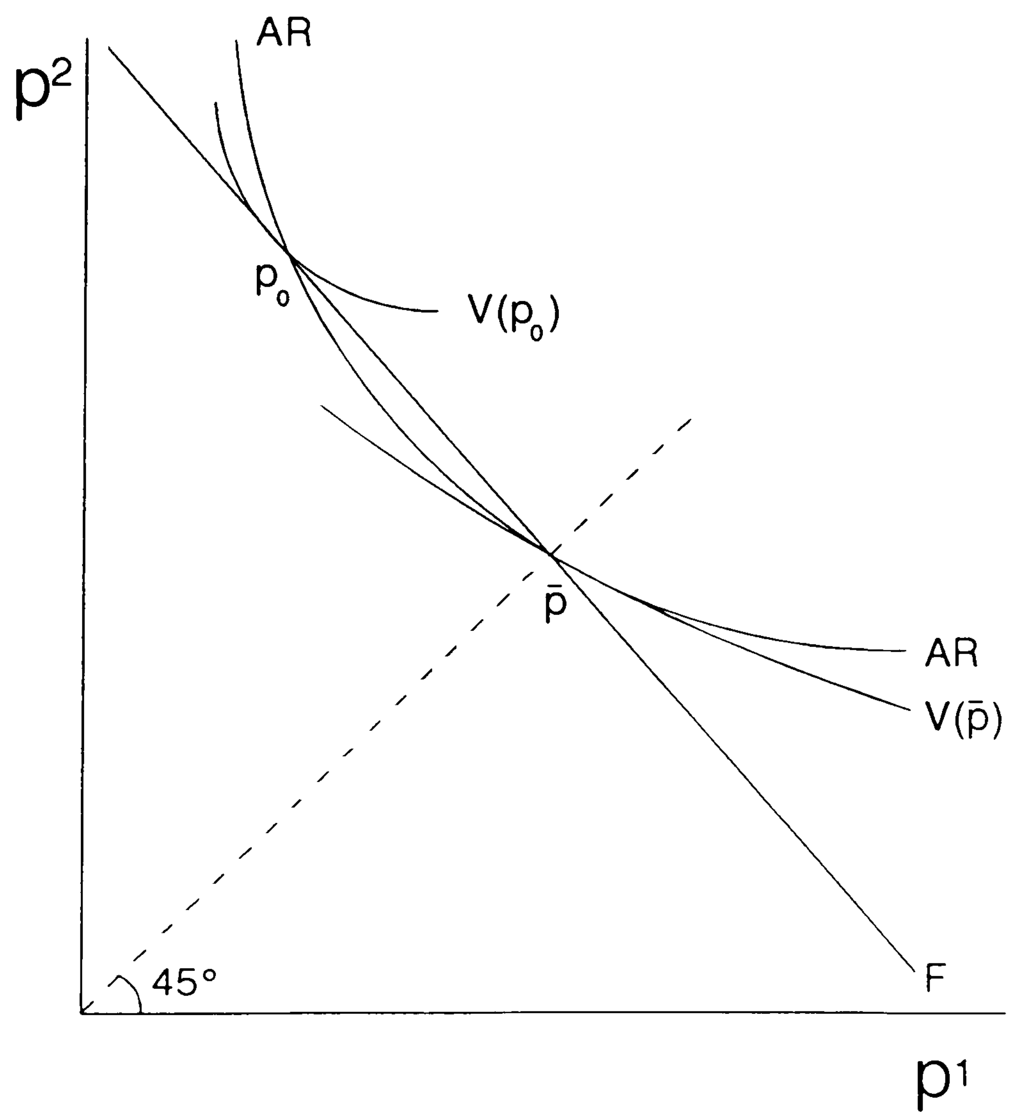


Figure 3.2 The Average Revenue Constraint

Result 1 is illustrated in Figure 3.2. With initial prices $\mathbf{p}_0$ the fixed weights constraint would be the straight line F (which is also the first TB constraint). We use this to determine average revenue earned with prices $\mathbf{p}_0$, which equals the price cap, $\bar{p}$. The line F is defined by $\sum_i p_0^i q_0^i = \sum_i p_1^i q_0^i$. When $p_1^1 = p_1^2$ along F we have, by construction, $p_1^1 = p_1^2 = \bar{p} = \sum_i p_0^i q_0^i / \sum_i q_0^i$. Average revenue in period 0 is thus given by the intersection of F with the 45° line. If $\mathbf{p}_1$ is on the AR curve in between $\mathbf{p}_0$ and $\bar{\mathbf{p}}$ then the consumer is on a lower iso-surplus curve than $V(\mathbf{p}_0)$, ensuring $V(\mathbf{p}_1) > V(\mathbf{p}_0)$. The fixed weights constraint is satisfied in this region, which is sufficient for consumer surplus to rise. Conversely if $\mathbf{p}_1$ is along AR to the left of $\mathbf{p}_0$ consumer surplus falls. There are several points to note about the AR constraint. First, if the products were perfect complements then the AR curve and the iso-surplus curve would both coincide with the F constraint. Second, the AR constraint lies above the $V(\bar{\mathbf{p}})$ curve, with tangency at $\bar{\mathbf{p}}$. At $\bar{\mathbf{p}}$ the AR constraint is more convex than $V(\bar{\mathbf{p}})$, and Corollary 1 shows that consumer surplus is maximized along the AR curve when all prices equal $\bar{p}$. Third, the general convex shape of the AR constraint means that it is more likely that there will be problems meeting the second-order conditions for the firm's problem than under TB or F where the constraints are linear.

What happens to profits when AR regulation is imposed? Since the initial prices, $\mathbf{p}_0$, are available, the firm does at least as well under AR as it did in period 0. In addition we are interested in the preferences of the firm between AR and TB and between AR and F. We first prove a lemma that is also useful in the analysis of the ARL scheme.

Lemma. *Average revenue under TB when initial prices are uniform is at most equal to* $\bar{p}$, *and is strictly below* $\bar{p}$ *when the chosen prices are not all equal to* $\bar{p}$.

Proof. We know from Proposition 3.1 that the TB mechanism generates a non-decreasing sequence of consumer surplus levels, i.e., $V(\mathbf{p}_t) \geq V(\mathbf{p}_{t-1}) \geq \ldots \geq V(\mathbf{p}_0)$. In particular when $\mathbf{p}_0 = \bar{\mathbf{p}}$, $V(\mathbf{p}_t) \geq V(\bar{\mathbf{p}})$. By convexity of $V(\mathbf{p})$,

$$V(\bar{\mathbf{p}}) - V(\mathbf{p}_t) \geq \sum_{i=1}^{n} [p_t^i - \bar{p}] q_t^i.$$

Suppose that average revenue, $\sum_{i=1}^{n} p_t^i q_t^i / \sum_{i=1}^{n} q_t^i$, *under the TB scheme* strictly exceeds $\bar{p}$, i.e. $\sum_{i=1}^{n} [p_t^i - \bar{p}] q_t^i > 0$. Using the convexity inequality this implies $V(\bar{\mathbf{p}}) > V(\mathbf{p}_t)$, which contradicts the fact that $V(\mathbf{p}_t) \geq V(\bar{\mathbf{p}})$ under TB. Thus

$$\sum_{i=1}^{n} [p_t^i - \bar{p}] q_t^i \leq 0$$

and $\mathbf{p}_t$, the TB price vector, generates average revenue at most equal to $\bar{p}$. To prove that the inequality is strict when $\mathbf{p}_t \neq \bar{\mathbf{p}}$, suppose not, so $\sum_i [\bar{p} - p_t^i] q_t^i = 0$. By the convexity inequality $V(\bar{\mathbf{p}}) \geq V(\mathbf{p}_t)$. But by the assumption that products are not perfect complements $V(\mathbf{p}_t) > V(\bar{\mathbf{p}})$ for $\mathbf{p}_t \neq \bar{\mathbf{p}}$, which implies a contradiction. □

We use Lemma in the next subsection, and to prove the following result.

Result 2. The firm's ranking of Average Revenue and Tariff Basket regulation. *When initial prices are uniform the firm prefers to be regulated by AR rather than TB.*

Proof. The lemma implies that under AR the firm can choose the sequence of prices that the TB-regulated firm chooses. In general it can do better. If the firm under TB chooses $\mathbf{p}_t \neq \bar{\mathbf{p}}$ then $\sum_i [\bar{p} - p_t^i] q_t^i > 0$, so the AR-constrained firm has a feasible set of prices that lies strictly outside the sequence of constraint sets under TB, except at $\bar{\mathbf{p}}$ where the constraint sets are tangential. □

Corollary 3. *When initial prices are uniform the firm prefers Average Revenue regulation to Fixed Weights regulation.*

This follows because TB is always preferred by the firm to F, whether initial prices are uniform or not.

Result 2 can be seen in Figure 3.2. If initial prices are uniform at $\bar{\mathbf{p}}$ the first TB constraint (not shown) is tangential to $V(\bar{\mathbf{p}})$ and the AR constraint at $\bar{\mathbf{p}}$, and all subsequent TB constraints cannot cross the iso-surplus curve through $\bar{\mathbf{p}}$, $V(\bar{\mathbf{p}})$. Since the AR constraint lies above $V(\bar{\mathbf{p}})$ except when prices are uniform, the price sequence under TB never crosses the AR constraint. The same argument shows that consumers prefer TB to AR when initial prices are uniform (Corollary 2 above). We should also note that the effect of regulation on welfare when $\mathbf{p}_0 = \bar{\mathbf{p}}$ is ambiguous. Armstrong and Vickers (1991) show this by example. Profits rise with regulation but consumer surplus falls. We shall discuss this case more fully in the next chapter.

When initial prices are not uniform the comparisons of AR with TB and F in terms of the preferences of consumers and the firm are not clear cut. In Figure 3.2 we know that if $\mathbf{p}_1$ is along AR between $\mathbf{p}_0$ and $\bar{\mathbf{p}}$ then consumers will be better off than at $\mathbf{p}_0$. Thus welfare will rise if $\mathbf{p}_1$ is in that region. But if $\mathbf{p}_1$ is on the part of AR to the left of $\mathbf{p}_0$ then consumer surplus falls. In the former case consumers *might* prefer AR to TB or F, but in the latter case they *certainly* prefer TB or F. The effect thus depends on the firm's choice of prices. Similarly we cannot say which scheme the firm prefers without knowing more about price choices. The AR constraint has two sections that are outside TB_1 and one section that is "tighter" than TB_1.

TB generates Ramsey prices in steady state. Are equilibrium prices under AR efficient? The answer is "no", except in the uninteresting case of complete symmetry. The firm's Lagrangian is

$$\mathcal{L} = \Pi(\mathbf{p}) + \lambda \sum_{i=1}^{n} [\bar{p} - p^i] q^i(\mathbf{p}). \tag{3.8}$$

The first-order condition for p^k, $k = 1, \ldots, n$ is

$$\frac{\partial \mathcal{L}}{\partial p^k} = \frac{\partial \Pi}{\partial p^k} - \lambda q^k + \lambda \sum_{i=1}^{n} [\bar{p} - p^i] \frac{\partial q^i}{\partial p^k} = 0. \tag{3.9}$$

The second-order conditions require that the iso-profit locus is more convex than the AR constraint. We have the following result.

Result 3. Inefficiency of pricing under Average Revenue regulation. *Regulated prices satisfy the necessary conditions for Ramsey efficiency if and only if the firm optimizes by setting the same price in each market.*

Proof. Ramsey pricing requires $\partial\Pi/\partial p^k = \omega q^k$, for all k, for some ω that is independent of k. If $p^i = \bar{p}$ then (3.9) gives $\partial\Pi/\partial p^k = \lambda q^k$, for all k so the necessary conditions for Ramsey pricing are satisfied with $\lambda = \omega$. To show that setting all prices equal to $\bar{p}$ is necessary for Ramsey pricing suppose that prices have the Ramsey property. It must then be the case that, for all k,

$$\lambda q^k + \lambda \sum_{i=1}^{n} [p^i - \bar{p}] \frac{\partial q^i}{\partial p^k} = \omega q^k$$

which implies

$$\sum_{i=1}^{n} [p^i - \bar{p}] \frac{\partial q^i}{\partial p^k} = \frac{1}{\lambda}(\omega - \lambda) q^k.$$

Multiplying by $[p^k - \bar{p}]$ and summing gives

$$\sum_{k=1}^{n} \sum_{i=1}^{n} [p^i - \bar{p}] \frac{\partial q^i}{\partial p^k} [p^k - \bar{p}] = \frac{1}{\lambda}(\omega - \lambda) \sum_{k=1}^{n} [p^k - \bar{p}] q^k. \tag{3.10}$$

Since the constraint binds $\sum_{k=1}^{n} [p^k - \bar{p}] q^k = 0$ and the right hand side of (3.10) equals 0. By the negative definiteness of the Slutsky substitution matrix the left-hand side is zero only if $\bar{p} = p^k$ for all k. □

Thus average revenue-constrained prices *can* be efficient, in the sense that consumer surplus is maximized for the given profit level, but this only occurs when demand and cost conditions are such that the firm wants to choose uniform prices. To see whether pricing is efficient all the regulator needs to check is whether prices are equal or not. Figure 3.2 can be used to illustrate Result 3. Suppose that the firm chooses prices $\mathbf{p_0}$ where the highest iso-profit curve (not drawn) is tangential to AR. Since $\mathbf{p_0} \neq \bar{\mathbf{p}}$ the iso-profit curve at $\mathbf{p_0}$ is not tangential to the iso-surplus curve $V(\mathbf{p_0})$.

We can now say a little more about the welfare comparison of AR with both TB and F. One case where both TB and F are socially preferred to AR is where the imposition of AR does not change any prices and initial prices are not uniform. Formally we assume that when $\mathbf{p}_0 \neq \bar{\mathbf{p}}$ the prices $\mathbf{p}$ that maximize $\Pi(\mathbf{p})$ subject to (3.5) are just $\mathbf{p}_0$. Neither consumer surplus nor profit changes. Under TB, however, the firm will change prices and we know from Proposition 3.1 that the present value of profits increases and that consumer surplus rises each period. Similarly under F both consumer surplus and profits rise in the first period. But in general the welfare comparison is ambiguous, as will be shown in the next two chapters.

The existing literature on AR regulation provides special cases of Result 3. Bradley and Price (1988) use a two-product model with linear costs and linear, independent demand functions. With identical costs but different demands in the two markets the Ramsey price structure requires a higher price in the high-demand market. The AR-regulated firm, however, exaggerates the price difference by setting a higher price in the high-demand market than the Ramsey price. With similar demands but different costs again the price differentials are exaggerated. The Ramsey pricing structure has identical proportional mark-ups of prices over costs but under average revenue regulation mark-ups will be higher in high-cost markets. Vickers and Yarrow (1988, pp 272-273) obtain a similar result for an n-product model with identical isoelastic demand functions and constant but different marginal costs. Armstrong, Cowan and Vickers (1994, Chapter 3) take the case where marginal costs are identical across products, demands are independent and price elasticities of demand are below unity. They show that prices are higher in markets with *more* elastic demands, which is the opposite result to the Ramsey rule. In the next chapter we show that the distortion in relative prices under AR can be very serious, and that there are some cases where AR regulation generates a welfare outcome that is inferior to the case of pure monopoly with no regulation.

We now summarize the main results on AR regulation.

Proposition 3.3 *Under AR regulation:*

(i) *When initial prices are uniform, consumer surplus falls and the welfare effect is ambiguous (Armstrong and Vickers, 1991). Consumers prefer TB and F to AR and the firm prefers AR to both TB and F.*

(ii) *When initial prices are not uniform, consumer surplus can rise or fall.*

(iii) *Profits are higher than in the initial period:* $\Pi(\mathbf{p}) \geq \Pi(\mathbf{p}_0)$.

(iv) *The welfare comparisons between AR and TB, and between AR and F, are ambiguous.*

(v) *Prices satisfy the Ramsey property if and only if they are chosen to be equal to* $\bar{\mathrm{p}}$.

Although we cannot be sure whether TB, F or AR will generate higher welfare, in general there are good reasons for preferring one of TB or F. If consumer surplus has a higher weight in the regulator's welfare function than profits, TB or F might be preferable because they guarantee an increase in consumer welfare, unlike AR which often lowers consumer surplus. Also both the TB and the F schemes ensure that total welfare will increase, but AR does not. We shall see in the next chapter that AR can lead to serious pricing distortions. There are also some practical reasons in favour of TB or F. They can be imposed whether products are commensurable or not, whereas AR requires commensurability. AR requires either that the regulator knows current demands, or can impose a correction factor which penalizes errors in the firm's forecasts and this complicates the operation of the formula. Second-order conditions are less likely to be satisfied in the case of AR.

3.2.4 Average Revenue (Lagged) Regulation

Under Average Revenue (Lagged) (ARL) regulation, the regulator constructs an average revenue measure using period t prices and period $t-1$ quantities, and requires this to be less than a fixed number, $\bar{p}$. The firm faces the constraint

$$\frac{\sum_{i=1}^{n} p_t^i q^i(\mathbf{p}_{t-1})}{\sum_{i=1}^{n} q_{t-1}^i} \leq \bar{p}$$

which can be written as

$$\sum_{i=1}^{n} [\bar{p} - p_t^i] q^i(\mathbf{p}_{t-1}) \geq 0 \tag{3.11}$$

and as

$$\frac{\sum_{i=1}^{n} p_t^i q_{t-1}^i}{\sum_{i=1}^{n} p_{t-1}^i q_{t-1}^i} \cdot \frac{\sum_{i=1}^{n} p_{t-1}^i q_{t-1}^i}{\bar{p} \sum_{i=1}^{n} q_{t-1}^i} \leq 1. \tag{3.11$'$}$$

As in the AR scheme we take $\bar{p} \equiv \sum_i p_0^i q_0^i / \sum_i q_0^i$. Implementation of ARL does not require current demand information, but the products must be commensurable.

What is the rationale for ARL? First, the fact that current demand information is not necessary makes it readily implementable, like TB and F. Second, in a one-period context the scheme is the same as both TB and F. This can be seen in (3.11$'$), since $\sum_{i=1}^{n} p_{t-1}^i q_{t-1}^i / \sum_{i=1}^{n} q_{t-1}^i = \bar{p}$ when $t = 1$. We should note that from period 2 onwards the ARL constraints are not the same as the TB constraints (or the F constraint), since the cap, $\bar{p}$, remains fixed. If instead the cap was defined as actual average revenue in the previous period then ARL would be exactly the same scheme as TB. This can be seen in (3.11$'$). We analyze the fixed price cap case since the actual scheme used for AT&T has a fixed cap (apart from the usual indexing and efficiency factors).

The ARL scheme is a hybrid of the TB and AR schemes. What can we say about the effect of ARL on consumer surplus simply from a study of the properties of (3.11)? We already know that in the first period the ARL constraint is the same as the TB constraint, so consumer surplus in period 1 is no less than that in period 0. By inspection of (3.11) we know that in steady state *actual* average revenue $\sum_i p^i q^i / \sum_i q^i$, equals $\bar{p}$. We have the following implications for consumer surplus.

Result 4. Effect of ARL on consumer surplus. *Under ARL:*

(i) *Consumer surplus is higher in period 1 than in period* 0: $V(\mathbf{p}_1) \geq V(\mathbf{p}_0)$.

(ii) *In a steady state with prices* $\mathbf{p}$, $V(\mathbf{p}) \leq V(\mathbf{p}_0)$ *if* $\sum_i [\bar{p} - p_0^i] q^i(\mathbf{p}) \geq 0$.

(iii) *In a steady state* $V(\mathbf{p}) \geq V(\mathbf{p}_0)$ *if* $\sum_i [\bar{p} - p^i] q_0^i \geq 0$.

(iv) *In a steady state* $V(\mathbf{p}) \leq V(\mathbf{p}_0)$ *if* $\mathbf{p}_0 = \bar{\mathbf{p}}$.

(v) *If average revenue in period* $t-1$ *exceeds* $\bar{p}$ *then* $V(\mathbf{p}_t) > V(\mathbf{p}_{t-1})$.

Proof. Part (i) follows from the fact that the first ARL constraint is the same as the first TB constraint. Parts (ii) and (iii) follow from the fact that, in steady state, actual average revenue equals $\bar{p} = \sum_i p_0^i q_0^i / \sum_i q_0^i$, and Result 1 in 3.2.3. Part (iv) is a special case of part (ii). Part (v) follows from inspection of (3.11′). If $\sum_i p_{t-1}^i q_{t-1}^i / \sum_i q_{t-1}^i > \bar{p}$ then $\sum_i p_t^i q_{t-1}^i / \sum_i p_{t-1}^i q_{t-1}^i < 1$, implying that surplus is higher in period t. □

We should note that although something can be said about consumer surplus in period 1 and in steady state relative to the level in period 0, we can say little about the path of consumer surplus after period 1 without knowing how the firm behaves, although we do have part (v) of Result 4. When $\mathbf{p}_0 = \bar{\mathbf{p}}$ we would expect the total effect on discounted consumer surplus to depend on the consumers' discount factor, since consumer surplus rises at first, but falls by the time the steady state is reached. A higher consumer discount factor will increase the likelihood that consumers lose overall.

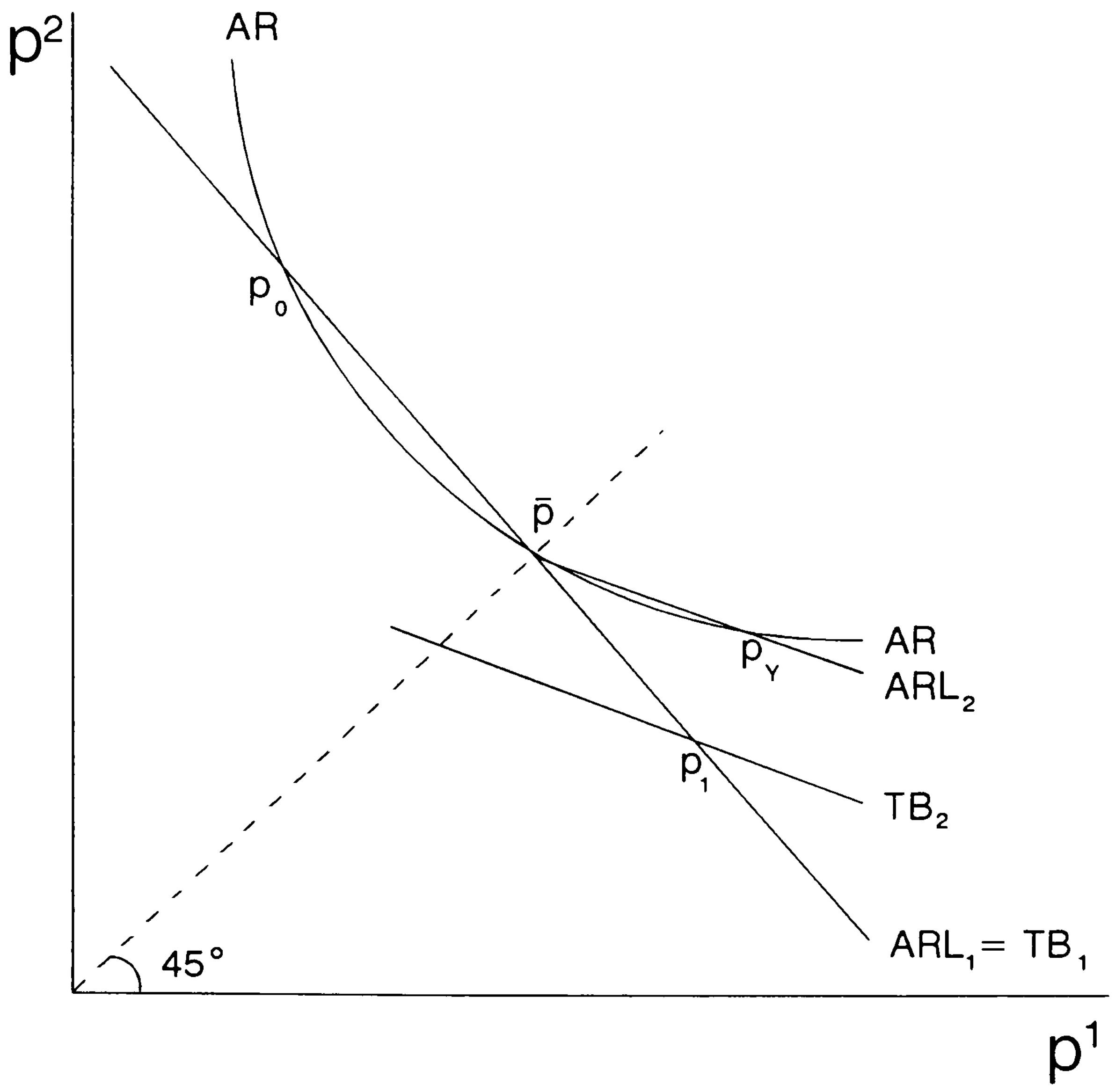


Figure 3.3 Average Revenue (Lagged) Constraints

In Figure 3.3 we illustrate the first two ARL constraints. They go through $\bar{\mathbf{p}}$ and are linear with slope of $-q^1_{t\text{-}1}/q^2_{t\text{-}1}$. The period 1 constraint is $\text{ARL}_1 = \text{TB}_1$. If the firm chooses prices $\mathbf{p}_1$ on ARL_1 then this will lower demand for product 1 and raise it for product 2, relative to demands at $\mathbf{p}_0$. If the TB scheme were in operation the new constraint would be TB_2. Since the ARL constraint goes through $\bar{\mathbf{p}}$ the new constraint is ARL_2. This has the same gradient as TB_2, but goes through $\bar{\mathbf{p}}$. If prices in period 2 are chosen along ARL_2 to the right of $\mathbf{p}_Y$ then relative demand for product 1 will fall further, since at $\mathbf{p}_Y$ there is an iso-surplus curve that is tangential to ARL_2 with slope given by $-q^1_1/q^2_1$, and the ARL constraint for period 3 will be flatter than ARL_2. The average revenue locus (AR) describes the set of feasible steady state prices.

We have seen that the effect on consumers of ARL depends on their discount factor. What about the firm? As usual the firm does better with ARL than by pricing at $\mathbf{p}_0$ for ever. This follows simply from revealed preference. The firm could always keep prices at $\mathbf{p}_0$, since this is on ARL_1 and AR by construction, and the fact that it can choose not to do so implies that the present value of the profit stream is at least $\Pi(\mathbf{p}_0)/(1-\beta)$. There is, however, no guarantee that profits will increase each period if the firm is myopic, as there was with TB, since the firm is not necessarily able to maintain the previous period's prices. In Figure 3.3 if the firm chose prices along ARL_2 in between $\bar{\mathbf{p}}$ and $\mathbf{p}_Y$ then it will not be able to sustain those prices in the next period.

Before characterizing the firm's constrained maximization problem we see how the ARL scheme compares with TB and F. We shall compare ARL with AR after we have examined the pricing decision.

Result 5. The firm's ranking of Average Revenue (Lagged), Tariff Basket and Fixed Weight regulation. *When initial prices are uniform, the firm prefers ARL to TB, and TB to F.*

Proof. We show that any sequence of prices that satisfies the relevant TB constraints can be chosen under ARL, and that in general the ARL-constrained firm can do better. We already know that the TB constraint for period 1 is the same as the ARL constraint. From period 2 onwards the TB constraint is

$$\sum_{i=1}^{n} p_{t-1}^{i} q_{t-1}^{i} \geq \sum_{i=1}^{n} p_{t}^{i} q_{t-1}^{i} \text{ for } t = 2, 3, \ldots$$

which implies

$$\frac{\sum_{i=1}^{n} p_{t-1}^{i} q_{t-1}^{i}}{\sum_{i=1}^{n} q_{t-1}^{i}} \geq \frac{\sum_{i=1}^{n} p_{t}^{i} q_{t-1}^{i}}{\sum_{i=1}^{n} q_{t-1}^{i}}.$$

We know from the Lemma in 3.2.3 that actual average revenue under TB is at most equal to $\bar{p}$ when initial prices equal $\bar{p}$. So

$$\bar{p} \geq \frac{\sum_{i=1}^{n} p_{t-1}^{i} q_{t-1}^{i}}{\sum_{i=1}^{n} q_{t-1}^{i}} \geq \frac{\sum_{i=1}^{n} p_{t}^{i} q_{t-1}^{i}}{\sum_{i=1}^{n} q_{t-1}^{i}} \text{ for } t \geq 2.$$

Thus TB prices satisfy the ARL constraint. Since under TB actual average revenue is strictly below $\bar{p}$, except when all prices equal $\bar{p}$, the ARL-constrained firm can usually do better than under TB. ARL is preferred to F by Proposition 3.2. □

One way for the firm regulated by ARL to behave is to follow TB prices until the steady state is reached. Prices in steady state under TB are Ramsey prices. As we shall see the firm under ARL can always choose a more profitable sequence of prices than Ramsey prices, as long as $\beta > 0$. Thus the firm earns the same profit stream as under TB by this strategy until the steady state TB prices are reached, and thereafter does better.

The preferences of consumers between ARL, TB and F are less clear. If both the firm and the consumers have zero discount factors then the schemes are equivalent as consumers will only care about the first period outcome. But in the more general case where $\beta > 0$ we cannot say more without precise knowledge of the prices that are chosen under each scheme. We would expect, however, that ARL is worse for consumers than both TB and F since ARL is more "generous" to the firm when initial prices are uniform. Since there is uncertainty about which scheme consumers prefer we cannot at this stage say which scheme generates higher welfare.

The firm chooses a sequence of feasible prices to maximize the present value of profits, taking account of the dynamics in the price constraint. Its Lagrangian is:

$$\mathcal{L} = \sum_{t=1}^{\infty} \beta^{t-1} \left\{ \Pi(\mathbf{p}_t) + \lambda_t \sum_{i=1}^{n} [\bar{p} - p_t^i] q^i(\mathbf{p}_{t-1}) \right\}.$$

The first order conditions give:

$$\frac{\partial \Pi_t}{\partial p_t^k} - \lambda_t q_{t-1}^k + \beta \lambda_{t+1} \sum_{i=1}^{n} [\bar{p} - p_{t+1}^i] \frac{\partial q_t^i}{\partial p_t^k} = 0 \quad \text{for all k and t} \tag{3.12}$$

$$\sum_{i=1}^{n}[\bar{p} - p_t^i]q^i(\mathbf{p}_{t-1}) = 0 \quad \text{for all t}$$

We focus first on the steady state, assuming one exists. Dropping time subscripts to denote the steady state in (3.12) yields:

$$\frac{\partial \Pi}{\partial p^k} = \lambda\left[q^k + \beta\sum_{i=1}^{n}[p^i - \bar{p}]\frac{\partial q^i}{\partial p^k}\right] \quad \text{for all k}$$
$$\sum_{i=1}^{n}[\bar{p} - p^i]q^i(\mathbf{p}) = 0. \tag{3.12'}$$

The pricing equations in (3.12′) can be compared with the first-order conditions in the AR problem, (3.9). The difference is that the term multiplying the sum is β rather than 1.

Result 6. Inefficiency of pricing under Average Revenue (Lagged) regulation. *Under ARL, prices in steady state satisfy the Ramsey property if and only if either the firm is myopic or all prices are equal to* $\bar{p}$.

Proof. The proof is very similar to the proof of inefficiency in 3.2.3 and details are omitted. If $\beta = 0$, or if $p^i = \bar{p}$ for all i, then $\partial\Pi/\partial p^k = \lambda q^k$ for all k. If steady-state prices have the Ramsey property, then by the argument in 3.2.3 we have, analogously to (3.10),

$$\beta\sum_{k=1}^{n}\sum_{i=1}^{n}[p^i - \bar{p}]\frac{\partial q^i}{\partial p^k}[p^k - \bar{p}] = \frac{1}{\lambda}(\omega - \lambda)\sum_{k=1}^{n}[p^k - \bar{p}]q^k.$$

Since in steady state the right hand side is zero by (3.12′), either $\beta = 0$, or, if $\beta > 0$, $p^i = \bar{p}$ for all i. □

If the firm is myopic it chooses p_t^k so that $\partial\Pi/\partial p_t^k = \lambda_t q_{t-1}^k$, so in steady state there are Ramsey prices. The ARL price cap, like the Vogelsang-Finsinger mechanism, has this desirable property when the firm is myopic.[7] The reason uniform prices imply Ramsey efficiency is as follows. We know that at $\bar{\mathbf{p}}$ the AR curve is tangential to both iso-surplus curve. If the highest feasible iso-profit curve is also tangential to the steady-state ARL constraint at $\bar{\mathbf{p}}$ then the iso-profit curve is tangential to the iso-surplus curve.

In Figure 3.4 we provide a loose argument for the general case where pricing is inefficient. Suppose that prices $\mathbf{p} \neq \bar{\mathbf{p}}$ have the Ramsey property in steady state, and that $\beta > 0$. The firm could sacrifice some current profits by pricing at $\mathbf{p_a}$ instead. This would shift the next period's constraint to $\mathrm{ARL_b}$, allowing the firm to make higher profits than at $\mathbf{p}$, for example at $\mathbf{p_b}$. As long as $\beta > 0$ the firm is likely to find some sequence $\{\mathbf{p}, \mathbf{p_a}, \mathbf{p_b}\}$ such that the present value of profits from this cycle of prices is higher than pricing at $\mathbf{p}$. Only when $\beta = 0$ will the firm prefer to remain at $\mathbf{p}$. This argument is made more rigorous in Chapter 5.

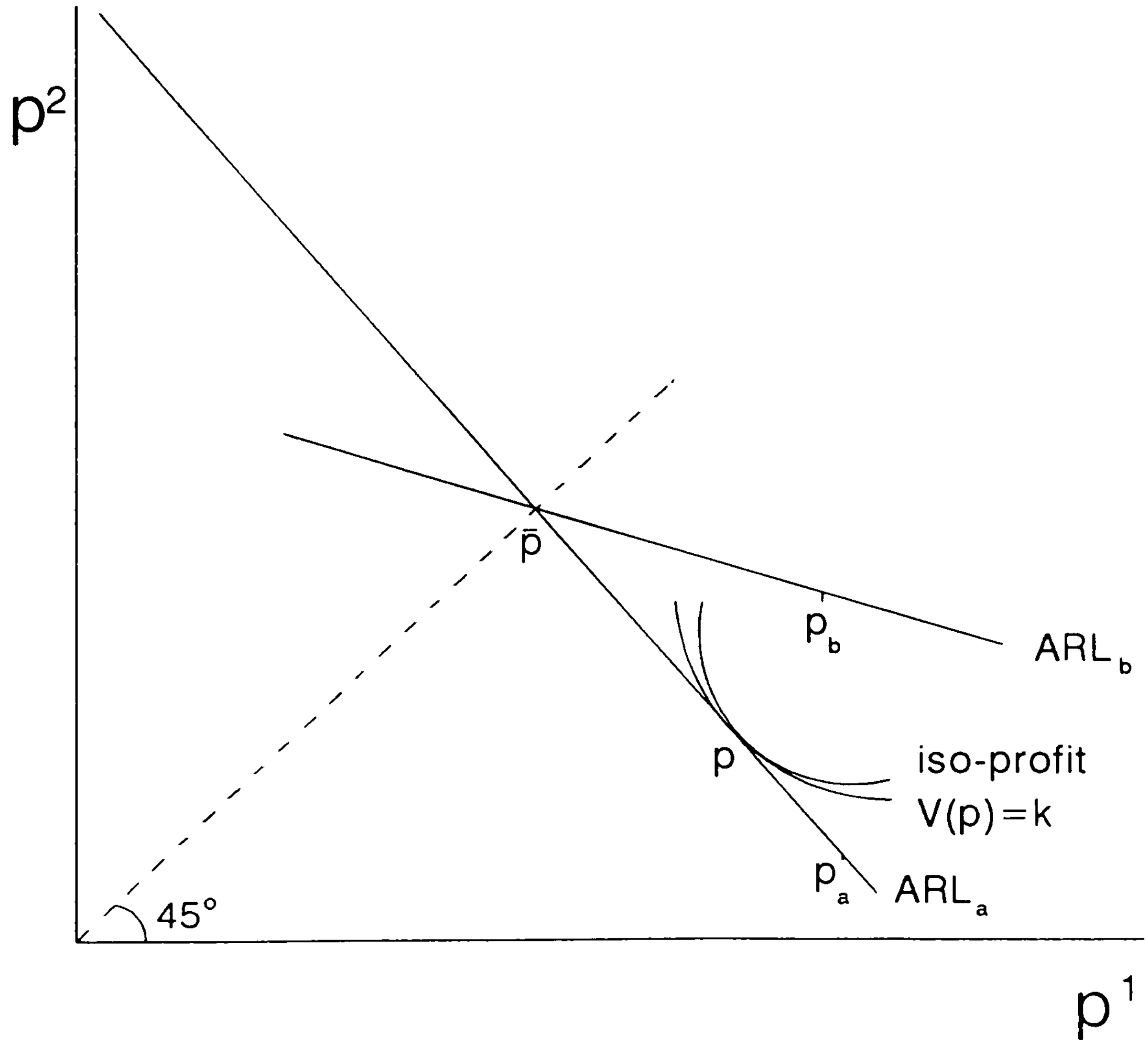


Figure 3.4 Inefficient Prices under Average Revenue (Lagged) Regulation

[7]Another scheme that generates allocative efficiency only when $\beta = 0$ is the price tax scheme devised by Tam (1981). Tam suggests imposing a tax of $p_t q_{t-1}$. The firm's rent is then $p_t(q_t - q_{t-1}) - C(q(p_t))$. A firm selling a single product will have a steady state price of $MC/(1-\beta)$. See Finsinger and Vogelsang (1985) and Gravelle (1985) for criticisms of Tam's scheme

We can now compare the effects of ARL and AR on the consumer and the firm. We could not tell in general which scheme either the firm or the consumer prefers simply from examining the two constraints as there are typically prices that are available under one scheme that are not available under the other. If both the firm and consumers have zero discount factors, and initial prices are uniform, then ARL is preferred by consumers to AR, while the firm prefers AR. These findings follow from Proposition 3.3(i), as ARL is the same as F under these conditions. In the more general case where $\beta > 0$ we cannot prove that these preferences are the same, but there is a presumption that they are.

In Figure 3.5 we show steady states for ARL when $\beta = 0$ (at $\mathbf{p}_1$), ARL when $\beta > 0$ (at $\mathbf{p}_2$) and AR (with prices $\mathbf{p}_3$). We know from (3.12′) that ARL steady-state prices are on the AR locus. The slopes of the three downward sloping straight lines through $\bar{\mathbf{p}}$ are the slopes of the iso-surplus loci at the three steady-state price vectors, and we can interpret these lines as ARL constraints. Result 6 showed that when $\beta = 0$ the necessary conditions for Ramsey prices are satisfied in steady state. Thus the iso-profit locus Π_1 is tangential to the ARL constraint, ARL_1, at $\mathbf{p}_1$. It immediately follows that Π_1 is not tangential to AR, and that there are more profitable price vectors available under AR regulation. Profits are maximized subject to AR at $\mathbf{p}_3$. The intermediate price vector $\mathbf{p}_2$ illustrates a possible steady state for ARL with $\beta > 0$. We know from Result 6 that the iso-profit curve Π_2 at $\mathbf{p}_2$ is not tangential to ARL_2 when $\beta > 0$ and from the first-order condition (3.12′) we know Π_2 is not tangential to AR. Although prices such as $\mathbf{p}_n$ are feasible under ARL and yield higher profits initially, they are not sustainable because the ARL constraint in the following period will be steeper than ARL_2 as quantities demanded change relative to those at $\mathbf{p}_2$. The diagram suggests that steady-state prices under ARL with a non-myopic firm will be intermediate between the steady state achieved under myopia and the prices chosen under AR.

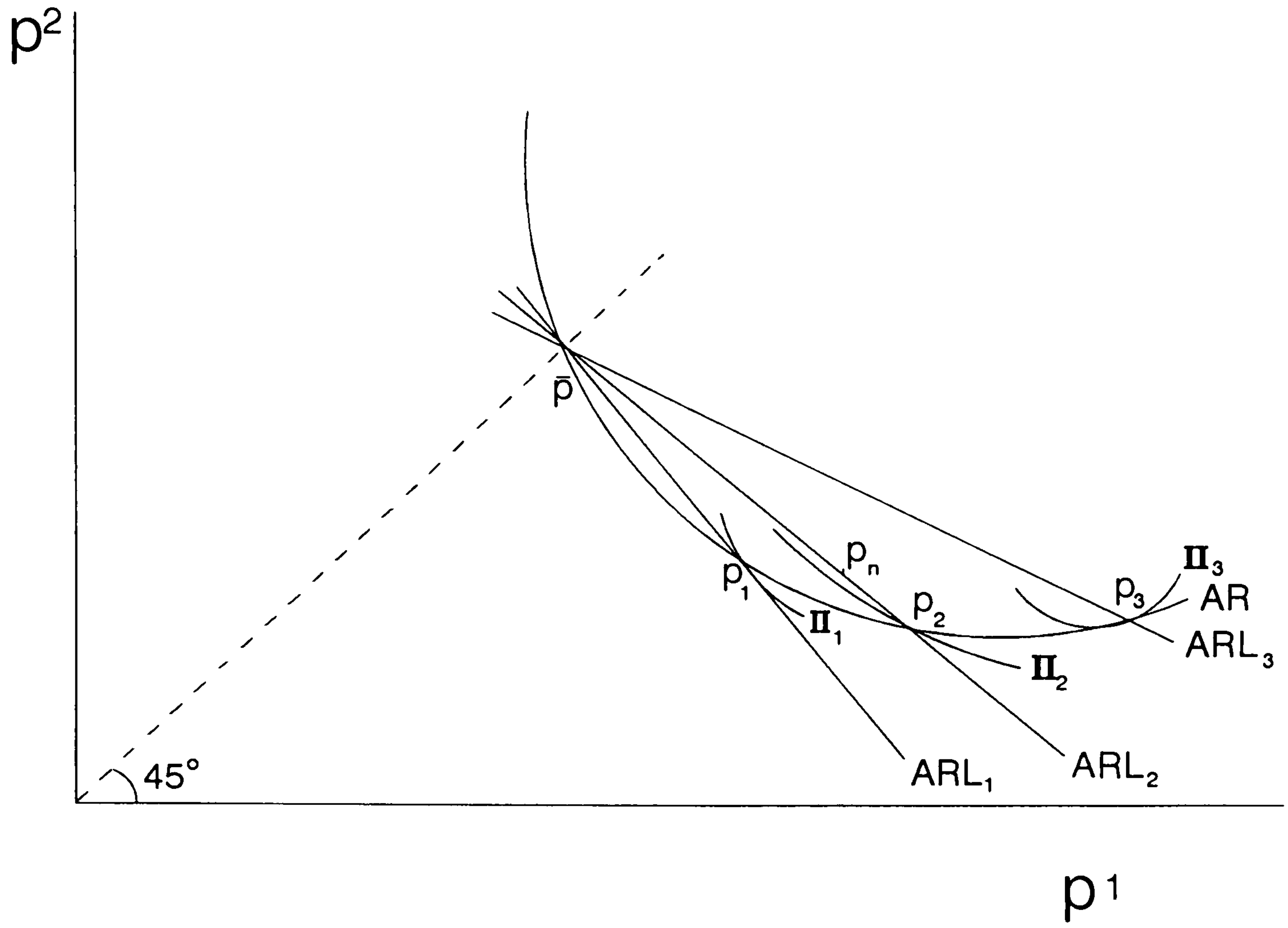


Figure 3.5 Steady-State Prices under ARL and AR

In the two-product example of Figure 3.5 consumers are better off the closer prices on AR are to $\bar{\mathbf{p}}$.[8] Thus in steady state they prefer ARL with myopia to ARL when $\beta > 0$, which in turn is preferable to AR. When the firm acts strategically (i.e. $\beta > 0$) the steady state gives a lower level for consumer surplus than when the firm is myopic. Similarly the firm makes higher steady state profits when $\beta > 0$, but these are not as high as under AR. From these observations we make the following conjecture.

Conjecture. Comparison of ARL and AR.

If initial prices are uniform then consumers prefer ARL to AR, while the firm prefers AR to ARL.

Under ARL the firm must choose prices in period 1 that raise consumer surplus above

[8]The intuition for this is as follows. Since the iso-surplus curve at, say, $\mathbf{p}_3$ is tangential to ARL_3 the latter is the same as a tariff basket or fixed weights constraint with initial prices $\mathbf{p}_3$. Since $\mathbf{p}_2$ lies below this line consumers are better off at $\mathbf{p}_2$.

that obtained at $\bar{\mathbf{p}}$. In steady state, by Result 4 part (iv), consumer surplus is lower than at $\bar{\mathbf{p}}$, but is above that under AR. Thus we expect that the present value of consumer surplus is higher under ARL than under AR. Similarly since the period 1 ARL constraint is not as good for the firm as AR, and in steady state the firm makes higher profits under AR, we expect the present value of profits to be higher under AR than under ARL. We should also note that it is likely that ARL yields higher welfare than AR because the steady state under ARL is closer to $\bar{\mathbf{p}}$, thus reducing the distortions that can occur under AR. In Chapter 4 we consider examples where the conjecture holds, although it is assumed that $\beta = 0$. In Chapter 5 we prove the conjecture when the firm sets a two-part tariff, and we find that welfare is always higher under ARL than under AR.

We summarize our results in the following proposition.

Proposition 3.4 *With ARL regulation:*

(i) *Consumer surplus is higher in period 1 than in period 0, and is lower in steady state than at* $\bar{\mathbf{p}}$. *If* $\mathbf{p}_0 \neq \bar{\mathbf{p}}$, *then consumer surplus in steady state can be above or below* $V(\mathbf{p}_0)$.

(ii) *The firm prefers ARL to TB or F. For consumers ARL is equivalent to TB and F in period 1 if* $\beta = 0$, *and so they also prefer ARL to AR if initial prices are uniform., while the firm prefers AR to ARL.*

(iii) *The present value of profits exceeds* $\Pi(\mathbf{p}_0)/(1-\beta)$.

(iv) *Steady state prices satisfy the Ramsey necessary conditions if and only if either* $\beta = 0$ *or the firm chooses* $\mathbf{p} = \bar{\mathbf{p}}$.

3.2.5 The Paasche Price Index Scheme

This is a scheme that we include in this chapter for purposes of comparison. The price constraint is:

$$\sum_{i=1}^{n}[p_{t-1}^{i} - p_t^{i}]q^{i}(\mathbf{p}_t) \geq 0. \tag{3.13}$$

This is similar to (3.1), but the quantities that multiply the price changes are current rather than lagged. It also resembles the AR constraint (3.5), except that the quantity terms multiply individual price changes rather than the difference between initial average revenue and the current price. Rearranging (3.13) shows that the PPI requires that the Paasche Price Index, $P = \sum_{i=1}^{n} p_t^i q_t^i / \sum_{i=1}^{n} p_{t-1}^i q_t^i \leq 1$.

The rationale for PPI is simple. The convexity result (3.6) can be written as

$$V(\mathbf{p}_{t-1}) - V(\mathbf{p}_t) \geq \sum_{i=1}^{n} [p_t^i - p_{t-1}^i]q_t^i. \tag{3.14}$$

If the right-hand side is positive then consumer surplus is lower in period $t-1$ than in period t. Equivalently consumer surplus falls if $P > 1$. A necessary condition for consumer surplus to *rise* is then that $P \leq 1$. We should note that (3.13) satisfies only the necessary condition for consumer surplus to increase, whereas the TB scheme *guarantees* that consumer surplus increases. Another point is that PPI does not require products to be commensurable. It is thus a scheme that, like TB, can be applied generally.

The effect on consumer welfare of PPI is, however, the opposite of TB. In particular, assuming the constraint always binds, consumer surplus *falls* each period. The proof is simple. If the right-hand side of (3.14) is zero, then $V(\mathbf{p}_{t-1}) \geq V(\mathbf{p}_t)$. Clearly consumers prefer TB and F to PPI. As far as the firm is concerned the present value of its profits exceeds $\Pi(\mathbf{p}_0)/(1-\beta)$ by the usual revealed preference argument. It is also straightforward to show that the firm prefers PPI to TB. By convexity we have

$$\sum_{i=1}^{n} [p_{t-1}^i - p_t^i]q_t^i \geq V(\mathbf{p}_t) - V(\mathbf{p}_{t-1}) \geq \sum_{i=1}^{n} [p_{t-1}^i - p_t^i]q_{t-1}^i. \tag{3.15}$$

If the TB constraint holds then $\sum_i [p_{t-1}^i - p_t^i]q_t^i \geq \sum_i [p_{t-1}^i - p_t^i]q_{t-1}^i \geq 0$ and the PPI constraint is also satisfied. The firm could mimic the TB prices, and do better in general. As usual since the firm prefers PPI to TB it also prefers PPI to F.

We can also compare PPI with AR when initial prices are uniform. The period 1 PPI constraint is exactly the same as the AR constraint under this condition. The period 2 PPI constraint allows the firm to keep the prices it chose in the first period, and to choose some others as well. The firm prefers PPI to AR, since it can always choose the AR prices in period 1 and keep them constant thereafter, but in general it can do better. When $\beta = 0$ and $\mathbf{p}_0 = \bar{\mathbf{p}}$ the firm will choose the same prices under PPI as under AR in period 1, but thereafter will choose prices that drive consumer surplus even lower. If consumers have a positive discount factor then they would prefer AR to PPI. The preference of consumers between PPI and AR when the firm's discount factor $\beta > 0$ is uncertain as it depends on the path of prices under PPI. The effect on welfare of PPI is also unclear. The present value of profits is higher than with $\mathbf{p}_0$, but consumer surplus falls each period.

To compare PPI with ARL we use the conjecture made in 3.2.4 and the result just shown. For $\mathbf{p}_0 = \bar{\mathbf{p}}$ the firm prefers PPI to AR, and we have guessed that the firm prefers AR to ARL, so PPI is likely to be preferred to ARL. Certainly if $\beta = 0$ then the firm prefers PPI to AR to ARL, and consumers definitely prefer ARL to AR to PPI, since ARL is equivalent to F in this case.

We now turn to the firm's optimization problem under PPI. We shall, as usual, take a "first-order condition" approach. The time path of prices and the steady state are characterized without attempting to prove that the firm will in fact reach a steady state. In Chapter 5 we shall present a more rigorous argument. In particular we find the precise conditions under which prices do converge to the steady state. In the present case the Lagrangian is:

$$\mathcal{L} = \sum_{t=1}^{\infty} \beta^{t-1} \left\{ \Pi(\mathbf{p}_t) + \lambda_t \sum_{i=1}^{n} [p^i_{t-1} - p^i_t] q^i(\mathbf{p}_t) \right\}$$

and the first-order conditions are:

$$\frac{\partial \Pi_t}{\partial p_t^k} - \lambda_t q_t^k + \lambda_t \sum_{i=1}^{n} [p_{t-1}^i - p_t^i] \frac{\partial q_t^i}{\partial p_t^k} + \beta \lambda_{t+1} q_{t+1}^k = 0 \quad \text{for all k and t} \tag{3.16a}$$

$$\sum_{i=1}^{n} [p_{t-1}^i - p_t^i] q^i(\mathbf{p}_t) = 0 \quad \text{for all t} \tag{3.16b}$$

assuming that the constraint always binds. In steady state, if it exists, we have:

$$\frac{\partial \Pi}{\partial p^k} = \lambda(1-\beta) q^k \quad \text{for } k = 1, \ldots, n$$

This resembles the steady state equation (3.4) for the TB scheme. Thus in steady state prices satisfy the necessary conditions for Ramsey efficiency.

In spite of some unattractive features of PPI, for example the decline in consumer surplus, it has the Ramsey property in the long run, which at first glance might seem surprising. General results about PPI are summarized in:

Proposition 3.4 *With PPI regulation:*

(i) Consumer surplus falls (weakly) each period that the constraint binds.

(ii) The present value of profits is at least $\Pi(\mathbf{p}_0)/(1-\beta)$.

(iii) Consumers prefer TB and F to PPI, and the firm prefers PPI to TB and F, whatever the initial prices.

(iv) Consumers prefer ARL to AR and AR to PPI when $\beta = 0$ and initial prices are uniform.

(v) The firm prefers PPI to AR when initial prices are uniform.

(vi) The firm prefers PPI to ARL when initial prices are uniform and $\beta = 0$.

(vii) In steady state prices satisfy the necessary conditions for Ramsey pricing.

3.3 Some Related Price Cap Schemes

In this section we discuss two further feasible price cap schemes. The first is a generalization of TB and PPI, while the second encompasses both AR and ARL. The main reason for studying them is to shed some light on the schemes analyzed in section 3.2.

We have seen that both TB and PPI have steady state prices that satisfy the Ramsey property, assuming a steady state exists. Examination of the first-order conditions in (3.3) and (3.16) shows that the reason for this is that in the constraints the price *changes*, $[p^i_{t-1} - p^i_t]$, are multiplied either by lagged or by current demands. In steady state price changes are by definition zero and the first-order conditions reduce to the Ramsey form. An obvious generalization of both TB and PPI is to multiply price changes by a weighted average of lagged and current demands. This constraint is

$$\sum_{i=1}^{n}[p^i_{t-1} - p^i_t]\{\epsilon q^i(\mathbf{p}_{t-1}) + (1-\epsilon)q^i(\mathbf{p}_t)\} \geq 0 \tag{3.17}$$

where ϵ is a weight chosen by the regulator. The TB scheme has $\epsilon = 1$ and PPI has $\epsilon = 0$.

The choice of $\epsilon = 1/2$ has a nice rationale. We know from (3.15) that the change in consumer surplus, ΔV, satisfies $\sum_i[p^i_{t-1} - p^i_t]q^i_t \geq \Delta V \geq \sum_i[p^i_{t-1} - p^i_t]q^i_{t-1}$, and that these are both linear approximations of ΔV. A simple average of these would give a "better" approximation to ΔV. This constraint is

$$\frac{1}{2}\sum_{i=1}^{n}[p^i_{t-1} - p^i_t](q^i_{t-1} + q^i_t) \geq 0. \tag{3.18}$$

Deaton and Muellbauer (1980, p 188) note that Hicks (1946) proposed the left-hand side of (3.18) as a measure of the change in consumer welfare. The left-hand side of (3.18) is exactly equal to ΔV when demand is linear. We prove this in Appendix 3.1. The constraint (3.18) then requires the firm to choose prices that at least keep consumer surplus constant at $V(p_{t-1})$. There will be immediate Ramsey pricing, since the price constraint is the iso-surplus curve. With linear demand the constraint (3.18) is a natural intermediate case between the tariff basket, which ensures $\Delta V \geq 0$, and the PPI scheme, which implies $\Delta V \leq 0$.[9] In general, however, although steady-state prices are efficient there is no guarantee that these prices will be reached. Vogelsang (1988) shows that his surplus subsidy scheme based on (2.27), which is closely related to (3.18), will not generate convergence when demand is nonlinear and the discount factor is high enough. Convergence is only guaranteed when demand is linear, in which case it is immediate.

The AR and ARL constraints can also be encompassed. Multiplying the differences between $\bar{p}$ and current price by a weighted average of lagged and current quantities yields:

$$\sum_{i=1}^{n} [\bar{p} - p_t^i](\epsilon q_{t-1}^i + (1-\epsilon)q_t^i) \geq 0.$$

When $\epsilon = 0$ the AR scheme operates and when $\epsilon = 1$ we have ARL. There are a few results on this constraint. First, as in the cases of ARL and AR, in steady state average revenue will be equal to $\bar{p}$. Second, Ramsey pricing is unlikely. The first-order condition in steady state for the firm's pricing problem under this constraint is

$$\frac{\partial \Pi}{\partial p^k} = \lambda q^k + \lambda[1 - \epsilon(1-\beta)] \sum_{i=1}^{n} [p^i - \bar{p}] \frac{\partial q^i}{\partial p^k}.$$

By a similar proof to that used for Result 3 there will be Ramsey pricing if and only if either all prices are uniform or $\epsilon = 1/(1-\beta)$. We shall come across this case again in

[9] The assumption of linearity generally simplifies the welfare analysis of third-degree price discrimination. Robinson (1933) shows that with linear demands a price-discriminating profit-maximizing monopolist will set the same output as when it cannot discriminate (assuming that no new markets are opened up) and that welfare falls with discrimination. Ireland (1992) shows that linear restrictions on the relationship between prices and unit costs (including common price-cost differences and common percentage mark-ups as special cases) raise welfare if demand is linear. We discuss the linear case more fully in Chapter 4.

Chapter 5. Under pure ARL we have $\epsilon = 1$ and we already know from Result 6 that if $\beta = 0$ under ARL the necessary conditions for Ramsey pricing are satisfied. The AR scheme operates when $\epsilon = 0$, so it cannot be the case that $\epsilon = 1/(1-\beta)$, and only when prices are uniform will they satisfy the Ramsey conditions.

3.4 Conclusions

We now summarize briefly the main points of this chapter. First, the firm has the following rankings of the schemes, where the symbol "$\succeq$" indicates preference: PPI $\succeq$ TB $\succeq$ F whatever the initial prices; while when initial prices are uniform we have PPI $\succeq$ AR $\succeq$ TB $\succeq$ F, ARL $\succeq$ TB $\succeq$ F. With uniform initial prices and myopia AR $\succeq$ ARL. Second, consumers have the following rankings: TB $\succeq$ PPI and F $\succeq$ PPI, whatever the initial prices; TB $\succeq$ AR and F $\succeq$ AR when initial prices are uniform; ARL $\succeq$ AR $\succeq$ PPI when initial prices are uniform and the firm is myopic. Third, under two of the schemes (TB and PPI) the necessary conditions for Ramsey pricing are satisfied in the limit, while Ramsey pricing only occurs under very special circumstances with F, AR and ARL. Fourth, both TB and F guarantee an increase in welfare relative to keeping prices constant (whatever the relative weights on consumer surplus and profits in welfare), whereas the other three schemes have ambiguous effects on welfare. Fifth, three of the constraints (TB, F and ARL) have given weights in the current period and are thus linear. We conjecture that it is unlikely that AR or PPI will be the best scheme. In the next chapter we consider in more detail the comparison between the schemes when demands are linear, and in Chapter 5 the case of a simple two-part tariffs is analyzed, where the focus is on the dynamic properties of each scheme. We shall find much more can be said about the properties and desirability of each scheme.

Appendix 3.1

Proof that the approximation to $V(p_t) - V(p_{t-1})$ given by equation (3.18) is exact when demand is linear.

Taking $\mathbf{p}_{t-1}$ as the reference price vector gives:

$$\Delta V = \mathbf{DV}(\mathbf{p}_{t-1})'[\mathbf{p}_t - \mathbf{p}_{t-1}] + \tfrac{1}{2}[\mathbf{p}_t - \mathbf{p}_{t-1}]'\mathbf{D}^2\mathrm{V}.[\mathbf{p}_t - \mathbf{p}_{t-1}] \tag{A.1}$$

where $\mathbf{DV}(\mathbf{p}_{t-1})$ is the gradient of V(.) evaluated at $\mathbf{p}_{t-1}$, and $\mathbf{D}^2\mathrm{V}$ is -1 times the Slutsky matrix. Similarly taking the Taylor series about $\mathbf{p}_t$ gives:

$$-\Delta V = \mathbf{DV}(\mathbf{p}_t)'[\mathbf{p}_{t-1} - \mathbf{p}_t] + \tfrac{1}{2}[\mathbf{p}_{t-1} - \mathbf{p}_t]'\mathbf{D}^2\mathrm{V}.[\mathbf{p}_{t-1} - \mathbf{p}_t] \tag{A.2}$$

Note that $\mathbf{D}^2\mathrm{V}$ does not depend on prices by the assumption of linearity and thus the quadratic forms in (A.1) and (A.2) are equal. Subtracting (A.2) from (A.1) gives

$$2\Delta V = [\mathbf{p}_{t-1} - \mathbf{p}_t]'\{-\mathbf{DV}(\mathbf{p}_{t-1}) - \mathbf{DV}(\mathbf{p}_t)\}. \tag{A.3}$$

Roy's identity implies $-\mathbf{DV}(\mathbf{p}_t) = (q^1(\mathbf{p}_t), \ldots, q^n(\mathbf{p}_t))$, and similarly for $\mathbf{DV}(\mathbf{p}_{t-1})$, so (A.3) shows that the approximation (3.18) is exact in the linear case. □

CHAPTER 4

PRICE FLEXIBILITY AND PRICE CAPS

4.1 Introduction

In this chapter we ask whether a firm that currently sets uniform prices in different markets should be allowed to set prices that differ. The analysis is motivated by the fact that there are growing pressures in Britain and other countries to allow utilities that have set uniform tariffs to differentiate their prices. British Gas, for example, has had prices that did not differentiate according to location, payment method, time of year or size of customer in spite of the cross-subsidization that this implied (see Armstrong, Cowan and Vickers, 1994, Chapter 8). The introduction of competition into its industrial and domestic supply markets and moves towards partial vertical separation have induced British Gas to reconsider its uniform tariff policy. There have also been moves to "de-average" tariffs in the water, electricity and telecommunications industries, and the question of uniform pricing is important in the economics of postal services. The growth of privatization and liberalization has weakened the concept of a utility offering a guaranteed service to all customers at uniform prices because costs of serving customers can differ according to location, load factor, volume and payment method.

We first consider the welfare effects of allowing a multimarket monopolist that previously set uniform prices the freedom to set its relative prices when the price level is not regulated. The model is then extended to the case where average prices are regulated by price caps. A standard reason for allowing prices to differ is that there is variation in costs, but this can lead to some paradoxical effects on welfare when the price level is regulated. Some general results are obtained, but much of the analysis is based on a simple model with linear demands and costs. The linear model is very tractable and allows us to find simple sufficient and necessary conditions for the paradoxical welfare effects of price-level regulation to occur.

The welfare effects of third-degree price discrimination with no regulation of the price level were first explored by Robinson (1933). She showed (pp. 188-195) that with linear demand and costs there is a net welfare loss from allowing discrimination, assuming that it does not open up new markets. Schmalensee (1981a) showed that a necessary condition for third-degree price discrimination to raise welfare when marginal cost is constant is that output increases. The reason price discrimination lowers welfare in the linear case is that output remains constant. The given output level is then inefficiently distributed because marginal utilities differ across consumers. Varian (1985) shows that Schmalensee's necessary condition can be generalized to the case of interdependent demands and non-decreasing marginal costs. Schwartz (1990) proves that the necessary condition also holds with decreasing marginal costs.

These authors do not allow for marginal costs that differ across markets. We generalize their results to allow for variation in costs as well as in demand functions. Ireland (1992) considers general linear restrictions on the relationship between prices and marginal costs and shows that these raise welfare above the unrestricted level in the linear model. Of course the linear model is a special case. Robinson (1933) attempts a more general analysis, and Greenhut and Ohta (1976) discuss her results. More recently Nahata, Ostaszewski and Sahoo (1990) show that when demand functions are nonlinear price discrimination can raise or lower all prices. Malueg (1993) also examines third-degree price discrimination with nonlinear demand.

Another literature examines price discrimination in locational models. Relevant analyses include Greenhut and Ohta (1972), Greenhut and Greenhut (1975), Holahan (1975) and Phlips (1983, Chapter 4). Typically consumers are located uniformly along a line with identical, linear demand functions. The welfare consequences of allowing price freedom relative to the case of charging in full for transport costs are analyzed and the main result is that price freedom raises welfare because it opens up more distant markets. Our model differs because consumer demands can vary with location, no new markets are opened up when price freedom is allowed, and we treat as the starting point the case that is relevant for many utilities of a uniform *delivered* price rather than a uniform *mill* price. Our model can be given a peak-load pricing interpretation in addition to the locational one.

Armstrong and Vickers (1991) consider the welfare effects of allowing third-degree price discrimination subject to a cap on average prices.[1] Our model in section 3 generalizes their linear model to allow for the important case where marginal costs vary across markets. The welfare analysis of average revenue regulation turns out to depend crucially on whether marginal costs can be different. When there is variation in costs there are non-pathological examples where average revenue regulation can lead to (i) some prices being above their unregulated levels, (ii) welfare below the level with no regulation, and (iii) situations where *relaxation* of the price cap raises welfare and even generates a Pareto improvement. Fixed weights regulation, however, is not subject to any of these perverse effects, and certainly raise welfare above the level with uniform pricing.

The organization of this chapter is as follows. In Section 4.2 we consider the case where the price level is not regulated but different degrees of price flexibility are allowed. In 4.3 price level regulation is analyzed. In 4.4 we consider the cases of tariff basket, average revenue (lagged) regulation and a mixed constraint. Concluding remarks are in 4.5.

4.2 Price differentiation with no regulation of price levels.

Varian (1985) derives some useful general results on price discrimination and welfare. Although he initially allows costs to differ across markets, in the statements of his main results he assumes marginal costs are the same in each market. We consider the general case where costs can vary across markets, but retain the assumption that marginal cost within a market is constant. There are n markets. The reference prices are denoted by the vector $\mathbf{p^0} \equiv (p_1^0, p_2^0, \ldots, p_n^0)$ where subscripts now denote the market and superscripts denote the price vector. Letting $\mathbf{p} \equiv (p_1, p_2, \ldots, p_n)$ denote

[1]Bradley and Price (1988) discuss a simple linear model with variation in costs, but they do not provide a welfare analysis.

the new price vector, convexity of consumer surplus V(.) implies that

$$\Delta V \equiv V(\mathbf{p}) - V(\mathbf{p^0}) \geq \sum_{i=1}^{n}(p_i^0 - p_i)q_i(\mathbf{p^0}).$$

Adding the change in profits gives a lower bound to the change in welfare:

$$\Delta W \equiv W(\mathbf{p}) - W(\mathbf{p^0}) \equiv \Delta V + \Delta\Pi \geq \sum_{i=1}^{n}(p_i - c_i)\Delta q_i$$

where $\Delta q_i \equiv q_i(\mathbf{p}) - q_i(\mathbf{p^0})$. An upper bound for the change in welfare can be derived in similar fashion, giving:

$$\sum_{i=1}^{n}(p_i^0 - c_i)\Delta q_i \geq \Delta W \geq \sum_{i=1}^{n}(p_i - c_i)\Delta q_i. \tag{4.1}$$

When costs are uniform we have $c_i = c$ for all i, and if the reference price vector is also uniform we can take $p^0 - c$ out of the sum on the left-hand side of (4.1) to obtain $(p^0 - c)\sum_{i=1}^{n}\Delta q_i \geq \Delta W$. Noting that $p^0 > c$ this yields Schmalensee's (1981a) result that a necessary condition for price discrimination to raise welfare is that total output increases. When costs vary, however, the necessary condition for the new prices to be preferable is that the vector of mark-ups with the reference prices multiplied by the changes in output is positive. Ireland (1992) notes that with costs that differ jointly sufficient conditions for $\mathbf{p}$ to be preferable to $\mathbf{p^0}$ are that the new margins are equal and output does not fall. In this case the second part of (4.1) can be written as $\Delta W \geq m\sum_{i=1}^{n}\Delta q_i$ where $m \equiv p_i - c_i$ for all i. If output does not fall then the change in welfare is non-negative. The bounds in (4.1) are useful for prove several subsequent welfare results.

We now consider what happens when the firm is required to set a uniform delivered price, but is able to choose the level of this price. It is assumed that all markets are served at the uniform price. We retain the assumption that marginal costs are constant within a market and allow general demand functions. Profits with uniform pricing are

$$\Pi(\text{Uniform}) = \sum_{i=1}^{n} (p - c_i) q_i.$$

Adding and subtracting the mean marginal cost $E(c) \equiv \sum_i c_i/n$ within the brackets yields another useful expression:

$$\Pi(\text{Uniform}) = [p - E(c)]Q(\text{Uniform}) - n\text{Cov}(c_i,\ q_i) \tag{4.2}$$

where $Q(\text{Uniform}) \equiv \sum_i q_i$ and $\text{Cov}(c_i,\ q_i)$ is the covariance of costs with sales, defined by $\text{Cov}(c_i,\ q_i) \equiv \sum_i c_i q_i/n - E(c)E(q)$, where $E(q) \equiv \sum_i q_i/n$. The first term in (4.2) reflects the standard monopolist's problem of choosing p to trade off the average margin against total sales. The covariance term captures the effect that differing costs have on profits. This term equals zero if costs are identical or if demand functions are the same. But if costs are positively correlated with sales, for example in a peak-load context, then there is a loss to the firm from the requirement that prices be uniform.

We now turn to the linear model. Demand functions are linear and independent, $q_i = a_i - p_i$. It is assumed that there is a natural way of measuring sales in each market, e.g. therms or kilowatt hours, that implies that it is meaningful to talk about total output even though the costs of supplying markets may differ. This is a natural assumption to make for utilities that supply a homogeneous product. The firm's problem is to choose the common value of p to maximize profits of

$$\Pi = \sum_{i=1}^{n} \left(p - c_i\right)\left(a_i - p\right),$$

yielding the following price:

$$p = \frac{E(a) + E(c)}{2}, \tag{4.3}$$

where $E(a) \equiv \sum_{i=1}^{n} a_i/n$ is the mean demand intercept. We assume that all markets are served under uniform pricing, i.e. $a_i > [E(a) + E(c)]/2$ for all i. Total output is

$$Q(\text{Uniform}) = \sum_{i=1}^{n} (a_i - p) = \sum_{i=1}^{n} a_i - np = \tfrac{n}{2}[E(a) - E(c)]. \tag{4.4}$$

This is half of the output level with marginal cost pricing.

Welfare is the sum of consumer and producer surplus, i.e.

$$W = \tfrac{1}{2}\sum_{i=1}^{n} (a_i - p)^2 + \sum_{i=1}^{n} (p - c_i)(a_i - p),$$

which can be written as

$$\tfrac{1}{2}\sum_{i=1}^{n} [a_i - E(a) + E(a) - p]^2 + \sum_{i=1}^{n} [p - E(c) + E(c) - c_i](a_i - E(a) + E(a) - p)$$

$$= \tfrac{n}{2}[\sigma_a^2 + (E(a) - p)^2] - n\sigma_{ac} + n[p - E(c)](E(a) - p)$$

where $\sigma_a^2 \equiv \frac{1}{n}\sum_i a_i^2 - [E(a)]^2$ is the variance of a and $\sigma_{ac} \equiv \frac{1}{n}\sum_i a_i c_i - E(a)E(c)$ is the covariance of a and c. Substituting for p from (4.3) gives

$$W(\text{Uniform}) = \tfrac{3n}{8}[E(a) - E(c)]^2 + \tfrac{4n}{8}[\sigma_a^2 - 2\sigma_{ac}]. \tag{4.5}$$

When markets are identical output in each market is $(a - c)/2$ and the absolute mark-up is $p - c = (a - c)/2$, yielding profits of $(a - c)^2/4$. With linear demand and costs consumer surplus is half of monopoly profits, so welfare in each market is $3(a - c)^2/8$, and when multiplied by the number of markets we obtain the first term in (4.5). The remaining term captures the effect on welfare of variation in demand and the covariance of demand with costs. Welfare falls the more strongly costs are correlated with demand, i.e. the higher σ_{ac} is, assuming that it is positive. If costs differ, but not demand, welfare is equal to the first term in (4.5) and is unaffected by the extent of cost variation.

There is another way to interpret (4.5). What is the best way to distribute total output given by (4.4)? It is straightforward to show that any given output level is best distributed when marginal social surplus is equalized across markets, i.e. the absolute mark-ups $p_i - c_i$ should be equal. This is a general result as long as the weights on consumer surplus and profits in welfare are equal and does not depend on the specific demand and cost assumptions made here. The problem of choosing prices to maximize welfare subject to total output being equal to that in (4.4) yields

$$p_i = c_i + \frac{E(a) - E(c)}{2} \quad \text{for } i = 1, \ldots, n.$$

The welfare level with equal mark-ups is

$$W(\text{Equal Mark-Ups}) = \tfrac{3n}{8}[E(a) - E(c)]^2 + \tfrac{4n}{8}[\sigma_a^2 - 2\sigma_{ac} + \sigma_c^2]. \tag{4.6}$$

The difference between W(Uniform) and W(Equal Mark-Ups) is

$$W(\text{Equal Mark-Up}) - W(\text{Uniform}) = \tfrac{n}{2}\sigma_c^2. \tag{4.7}$$

When costs do not vary the best way of distributing the output is to have uniform pricing, because this entails equal marginal social surpluses. But when $\sigma_c^2 > 0$ uniform pricing entails a loss of welfare relative to the case where mark-ups are equal.

What happens when the firm has freedom to set different prices? It is assumed that such freedom does not induce the firm to supply new markets.[2] Again we start with some general comments. For any demand functions, and constant marginal costs, we can write

[2]The fact that price freedom can open up new markets is an important argument in favour of price freedom (see Hausman and Mackie-Mason (1988) and Varian (1985)). Layson (1994a) considers the conditions under which a new market is likely to be opened up. Layson (1994b), however, shows by example that price freedom can shut out a market that is served under uniform pricing.

$$\Pi(NC) = \sum_{i=1}^{n} (p_i - c_i)q_i = [E(p) - E(c)]Q(\mathbf{p}) + n[Cov(p_i, q_i(\mathbf{p})) - Cov(c_i, q_i(\mathbf{p})]$$

where NC indicates "No Constraint". The standard trade-off between the average mark-up, $E(p) - E(c)$, and total output, $Q(\mathbf{p})$, is captured by the first term. The last two terms capture the main relative price effects. The firm wants to have prices in markets with above average sales to raise $Cov(p_i, q_i)$. But it also wants to have high prices in markets where costs are high, because this lowers $Cov(c_i, q_i)$. In a peak-load pricing context, where high demand is associated with high marginal cost, we would expect the two covariance terms to be approximately equal and thus to offset each other, and if mark-ups are chosen to be equal they are exactly equal.

In the linear model the firm's problem with price flexibility is

$$\underset{\{p_i\}}{\text{Max}} \ \Pi = \sum_{i=1}^{n} (p_i - c_i)(a_i - p_i),$$

and the first-order conditions yielding the standard monopoly pricing result

$$p_i = \frac{a_i + c_i}{2} \qquad \text{for } i = 1, \ldots, n. \tag{4.8}$$

It is straightforward to check that total output under NC, Q(NC) is the same as that under uniform pricing, given in (4.4). This implies that whether price freedom or uniform pricing generates higher welfare depends only on which scheme induces better distribution of the given level of output. In general the mark-ups implied by (4.8) will not be equal, so output is inefficiently distributed with price freedom.

Welfare under price freedom is

$$W(NC) = \frac{3}{8} \sum_{i=1}^{n} (a_i - c_i)^2.$$

Adding and subtracting E(a) and E(c) within the bracket gives

$$W(NC) = \frac{3}{8}\sum_{i=1}^{n}(a_i - E(a) + E(a) - E(c) + E(c) - c_i)^2$$

and so

$$W(NC) = \frac{3n}{8}[E(a) - E(c)]^2 + \frac{3n}{8}[\sigma_a^2 - 2\sigma_{ac} + \sigma_c^2] \tag{4.9}$$

where $\sigma_c^2 = \frac{1}{n}\sum_i c_i^2 - [E(c)]^2$ is the variance of c. Taking the difference between (4.6) and (4.9) gives:

$$W(\text{Equal Mark-Ups}) - W(NC) = \frac{n}{8}[\sigma_a^2 - 2\sigma_{ac} + \sigma_c^2] = \frac{n}{8}\mathrm{Var}(a - c) \geq 0.$$

The one case where price freedom generates equal mark-ups is when $\mathrm{Var}(a-c) = 0$, which implies that a and c are perfectly correlated with $\sigma_c^2 = \sigma_a^2 = \sigma_{ac}$. Except in this very special case there is a distributional loss from price freedom.

The difference between welfare with no constraint and with uniform pricing is

$$W(NC) - W(\text{Uniform}) = \frac{n}{8}\Big[3\sigma_c^2 + 2\sigma_{ac} - \sigma_a^2\Big]. \tag{4.10}$$

Robinson's result that welfare is higher with uniform pricing when costs do not vary is a special case of (4.10). Uniform costs imply that $\sigma_c^2 = \sigma_{ac} = 0$, so (4.10) reduces to $-n\sigma_a^2/8 \leq 0$. This is equal to zero only in the uninteresting case of no demand differences. Otherwise the effect of allowing price discrimination on welfare is negative when costs are uniform because total output stays constant and the output is inefficiently distributed.

If costs vary then the clear result of the Robinson case is not as obvious. In the special case of perfect correlation and equal variances we know that W(NC) = W(Equal

Mark-Ups) > W(Uniform). More generally there is a presumption that allowing some price flexibility might raise welfare, since the firm would want to push up prices in markets with higher costs and lower them in markets with lower costs. A sufficient condition to reverse the Robinson result is $\sigma_c^2 \geq \sigma_a^2 > 0$. This follows because the bracketed term in (4.10) is $\sigma_c^2 + 2\sigma_{ac} + \sigma_a^2 + 2(\sigma_c^2 - \sigma_a^2) = \mathrm{Var}(a + c) + 2(\sigma_c^2 - \sigma_a^2)$, and $\mathrm{Var}(a + c) > 0$. This is illustrated in Figure 4.1 for the case where $\sigma_a^2 = 0$. The profit-maximizing uniform price is $p_U = [a + E(c)]/2$. Price freedom leads to a rise in the price of the higher cost market to p_2 and a fall in the price in the lower cost market to p_1. Linearity implies $p_2 - p_U = p_U - p_1$, and since the demand curve slopes down (consumer surplus is convex in prices) total consumer surplus is increased. Price freedom also benefits the firm as it can always choose to price uniformly and other options are now available. Thus there is a welfare gain.

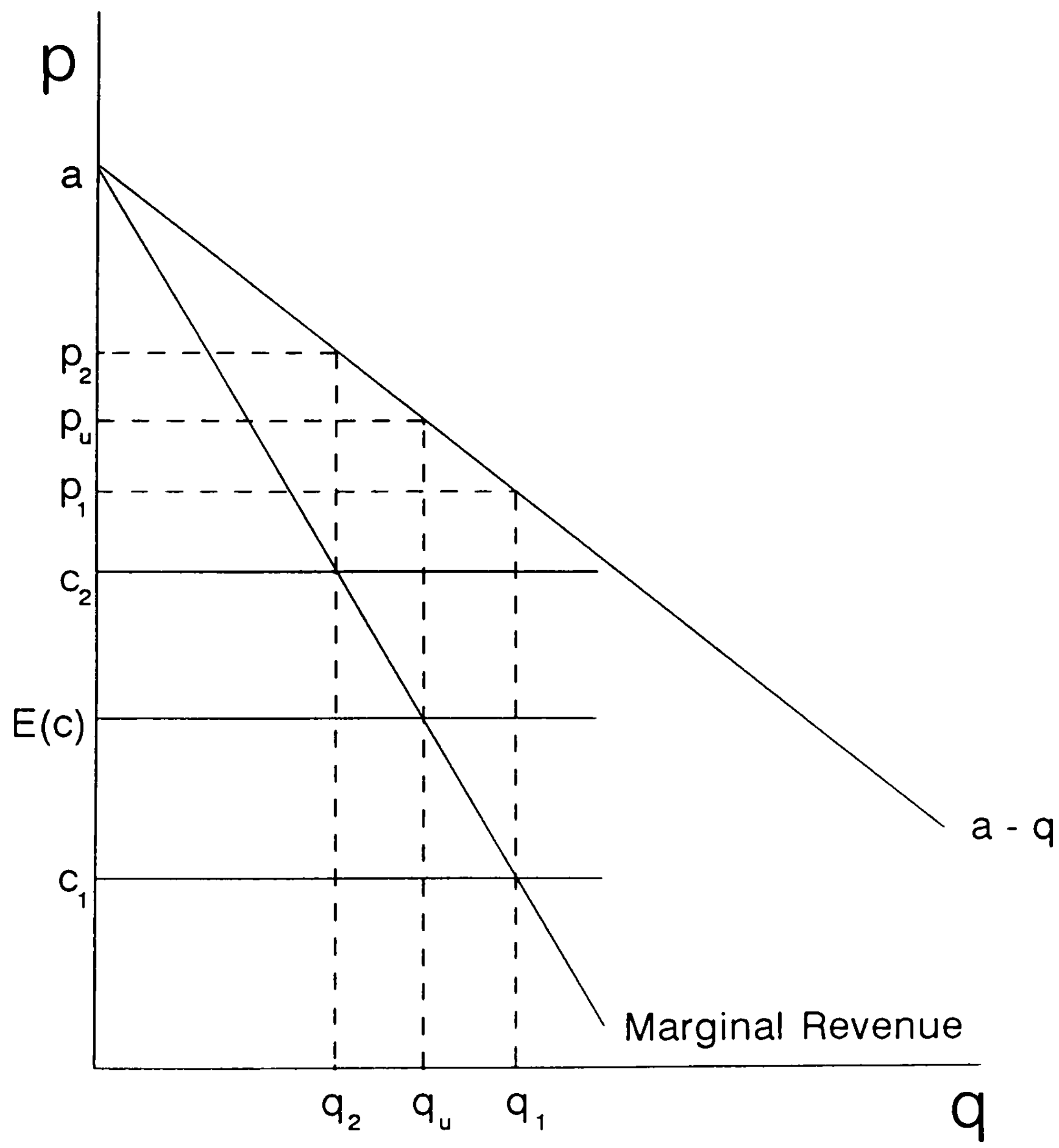


Figure 4.1 Uniform and Unrestricted Pricing when Costs Vary

Inspection of (4.10) also indicates that if costs and demand are positively correlated (without perfect correlation) then the greater this correlation the more likely it will be that price freedom raise welfare. We also know that if $\sigma_{\mathrm{ac}} \geq 0$ then $\sigma_{\mathrm{c}}^2 > \sigma_{\mathrm{a}}^2/3$ is sufficient for W(NC) > W(Uniform). Positive correlation is natural in a peak-load pricing context. The markets would be defined according to the time of the day or year, and the need for extra capacity to service peak periods will imply that marginal costs are higher then.

We can use the welfare bounds (4.1) to establish conditions under which price freedom is desirable or not. The reference price vector $\mathbf{p^0}$ is the vector of uniform prices given in (4.3). Substituting the relevant prices and quantities into (4.1) gives

$$\tfrac{\mathrm{n}}{2}[\sigma_{\mathrm{ac}} + \sigma_{\mathrm{c}}^2] \geq \mathrm{W(NC)} - \mathrm{W(Uniform)} \geq \tfrac{\mathrm{n}}{4}[\sigma_{\mathrm{c}}^2 - \sigma_{\mathrm{a}}^2].$$

This confirms that when there is no cost variation, i.e. $\sigma_{\mathrm{c}}^2 = \sigma_{\mathrm{ac}} = 0$, it is better to have uniform pricing. A necessary condition for welfare to be higher with no constraint is that $\sigma_{\mathrm{ac}} + \sigma_{\mathrm{c}}^2 > 0$, while a sufficient condition for welfare to rise with price freedom is that there is greater variation in costs than in demand.

We now summarize the main results of this section. With linear demands and costs total output is the same whether the firm is restricted to uniform pricing or is allowed complete freedom to set prices, as long as no new markets are opened up. The welfare comparison between the two depends only the distribution of this given output level. The best way to distribute a given output is to equalize marginal social surplus across markets by setting uniform absolute mark-ups. Uniform pricing is better when costs are uniform, while price freedom is preferable when costs and demand are perfectly correlated and have the same variance, because in both cases the firm sets uniform mark-ups. More generally price freedom is preferable if there is greater variation in costs than in demand, and is more likely to be preferable the greater the positive correlation between demand and costs.

4.3 Price Caps

In this section we extend the analysis of the previous section to the case where there is regulation of average prices as well as of price structure. Two static price caps, average revenue (AR) and fixed weights (F) regulation, are assessed using the linear model. The welfare effects of allowing price freedom are very different in the two cases.

4.3.1 Average Revenue Regulation

The firm faces the constraint that average revenue cannot exceed a fixed number $\bar{p}$, which is equal to the initial uniform price. We already know from Result 3 in Chapter 3 that relative prices are inefficient under AR in the sense that consumer surplus is not maximized subject to the given level of profits. We shall see in this section that this inefficiency can be very serious. Before exploring the linear model we consider the general problem. The only assumption that is made is that marginal costs are constant. Using the same manipulation used to derive (4.2) profits under AR are:

$$\Pi(AR) = \sum_{i=1}^{n} (\bar{p} - c_i)q_i^{AR} = [\bar{p} - E(c)]Q(AR) - n\text{Cov}(c_i,\ q_i^{AR})$$

where $Q(AR) \equiv \sum_i q_i^{AR}$ is total output under AR. In general there are two effects of imposing AR. First, as long as $\bar{p} > E(c)$, the firm wants to maximize total output. Second, for any given total output, the firm chooses prices to minimize the covariance of costs with sales. It pushes up prices in markets with high costs to reduce sales in below the average sale level, and cuts prices in markets with low costs. When costs are uniform we have $\bar{p} > c$ and $\text{Cov}(c_i,\ q_i) = 0$, so the firm simply wants to maximize total output, but with varying costs the firm's problem is more complex.

What happens if the constraint is "tight", so $\bar{p} = E(c)$? In this case the firm's objective is to minimize $\text{Cov}(c_i,\ q_i)$, i.e. to make it as negative as possible. The problem with this behaviour from a welfare point of view is that it takes no account of the strength of demand. Take, for example, the standard peak-load pricing problem. In this case the regulator wants sales to be *above* average in periods where costs are above average, but the imposition of AR with $\bar{p} = E(c)$ gives the firm the opposite incentive. This suggests that relative pricing might be very inefficient in such circumstances.

Another way of interpreting the behaviour of the firm uses the Lagrangian for the firm's problem, where λ_{AR} is the multiplier on the AR constraint:

$$\mathcal{L} = \sum_{i=1}^{n} (p_i - c_i)q_i + \lambda_{AR} \sum_{i=1}^{n} (\bar{p} - p_i)q_i.$$

Assuming that $0 < \lambda_{AR} < 1$ this can be written as:

$$\mathcal{L} = (1 - \lambda_{AR}) \sum_{i=1}^{n} \left\{ p_i - \left[c_i - \frac{\lambda_{AR}}{1 - \lambda_{AR}} (\bar{p} - c_i) \right] \right\} q_i.$$

Thus it is as if the firm is unregulated but marginal cost in market i is now $c_i - \lambda_{AR}(\bar{p} - c_i)/(1 - \lambda_{AR})$. In markets where $\bar{p} > c_i$, the effective marginal cost is below c_i, but in markets where $\bar{p} < c_i$ effective marginal cost is above c_i, so the firm wants to lower sales. The tighter the constraint is the greater $\lambda_{AR}/(1 - \lambda_{AR})$ becomes and thus the greater the incentive to reduce sales in markets where $\bar{p} < c_i$.

We shall show shortly that the inefficiency in pricing that AR can induce can be so great that welfare is *below* the level with no regulation at all. This possibility will be explored in some detail, firstly in general and then using the linear model, before we compare welfare under AR, W(AR), with welfare under uniform pricing at $\bar{p}$. We first use the welfare bounds in (4.1). Treating the unconstrained prices as the reference price vector and the prices under AR as the new price vector the upper bound in (4.1) is

$$\sum_{i=1}^{n} (p_i^{NC} - c_i)[q_i^{AR} - q_i^{NC}] \geq W(AR) - W(NC). \tag{4.1'}$$

A sufficient condition for $W(AR) \leq W(NC)$ is that the left-hand side of (4.1′) is non-positive. In general little can be said about the conditions under which this condition holds, but we can say something in the peak-load context where demand is positively correlated with costs. Suppose the firm sets equal mark-ups, $p_i^{NC} - c_i$, when there is no regulation. In the linear model this occurs when $a_i - c_i = \text{constant}$. With equal mark-ups the sufficient condition for $W(AR) \leq W(NC)$ is simply that $Q(AR) \leq Q(NC)$. This is intuitive: if output remains equal to Q(NC) when AR is imposed then it will be badly distributed, because under NC the fact that mark-ups are equal implies optimal distribution. It is not possible to state general conditions under which $Q(AR) \leq Q(NC)$, but the discussion above suggests that if $\bar{p} = E(c)$ then the firm is not interested in maximizing output, rather its focus is on the distribution of output. In particular the firm will seek to have high prices in markets with above average cost, so that sales there are below average, even if sales would be above average in those markets if prices were unconstrained.

We now explore the possible welfare effects using the linear model. To ensure that the constraint binds, $\bar{p}$ must be low enough. It is assumed that

$$\bar{p} < \frac{\sigma_a^2 - \sigma_c^2}{2[E(a) - E(c)]} + \frac{E(a) + E(c)}{2}. \tag{4.11}$$

where the right-hand side is equal to average revenue earned under no constraint at prices given in (4.8).

In the linear case the first-order conditions for prices are:

$$p_i = \frac{a_i + c_i}{2} - \theta\frac{(\bar{p} - c_i)}{2} \qquad i = 1, \ldots, n \tag{4.12}$$

where $\theta \equiv \lambda_{AR}/(1 - \lambda_{AR})$. Substituting (4.12) into the price constraint gives:

$$\theta = \sqrt{\frac{\sum(a_i - \bar{p})^2}{\sum(\bar{p} - c_i)^2}} - 1 = \sqrt{\frac{\sigma_a^2 + [E(a) - \bar{p}]^2}{\sigma_c^2 + [\bar{p} - E(c)]^2}} - 1, \tag{4.13}$$

which is derived in Appendix 4.1. The assumption that (4.11) holds implies that

$\sigma_a^2 + [E(a) - \bar{p}]^2 > \sigma_c^2 + [\bar{p} - E(c)]^2$, so $\theta > 0$ and $0 < \lambda_{AR} < 1$.

The second-order conditions require calculation of the minors of the bordered Hessian. The minor of order $j = 3, \ldots, n$ is

$$(-1)^{j-1}(1 - \lambda_{AR})^{j-2}(1 + \theta)^2 \sum_{i=1}^{j-1} (\bar{p} - c_i)^2.$$

This alternates in sign since $0 < \lambda_{AR} < 1$, and when $j = 3$ it is positive, so the second-order conditions hold.

If costs are uniform (and equal to c) then we require $\bar{p} > c$ to ensure the firm's participation, and from (4.12) p_i is below the unconstrained profit-maximizing level, $(a_i + c)/2$, in each market. When costs differ, however, it might be the case that in some market, say market j, $c_j > \bar{p}$ implying that p_j is *above* the unregulated level, even though there is regulation.[3] Why does the firm want to push p_j above $(a_j + c_j)/2$ when $c_j > \bar{p}$? The reason is that the firm wants to sell as much as possible in markets where $\bar{p} > c_i$. By reducing demand in market j it minimizes the loss from selling in this market and relaxes the price constraint for other profitable markets. This incentive is independent of the strength of demand, which is measured by a_j.

Total output under average revenue regulation is

$$Q(AR) = \sum_{i=1}^{n} (a_i - p_i) = \tfrac{n}{2}[E(a) - E(c)] + \tfrac{\theta n}{2}[\bar{p} - E(c)] \tag{4.14}$$

which exceeds the unconstrained output level as long as $\bar{p} > E(c)$. Again with uniform costs this condition holds, but if costs vary then output would be at or below the unregulated level if the price cap, $\bar{p}$, is sufficiently tight, i.e. if $\bar{p} \leq E(c)$.

We can now use the welfare bound (4.1′) to derive sufficient conditions for $W(AR) \leq W(NC)$. The unregulated profit margin is $(a_i - c_i)/2$ and the difference between output under AR and output under no constraint in market i is

[3]Bradley and Price (1988, p 104) note that it is possible to have one price above the unconstrained level.

$\Delta q_i = \theta(\bar{p} - c_i)/2$. Substituting these into the upper bound gives

$$\frac{\theta}{4}\sum_{i=1}^{n}(a_i - c_i)(\bar{p} - c_i) \geq W(AR) - W(NC). \tag{4.15}$$

If the left-hand side is zero then $W(AR) \leq W(NC)$. There are two interesting cases where the left-hand side is zero.[4] First, this occurs when the unregulated firm chooses to set uniform mark-ups, i.e. $\sigma_c^2 = \sigma_a^2 = \sigma_{ac} > 0$, and $\bar{p} = E(c)$, since in this case $a_i - c_i = k$ where k is a constant, and the left-hand side of (4.15) is $\theta kn[\bar{p} - E(c)]/4 = 0$. We call this case (i). Total output is equal to the unconstrained level because $\bar{p} = E(c)$, and its poor distribution under AR implies that $W(AR) < W(NC)$. Without regulation the equal mark-up policy leads to equal sales in all markets, so this is a case of a "shifting peak". Under AR sales are cut in markets where $c_j > E(c)$, even though these are exactly the markets where $a_j > E(a)$.

Case (ii) is more important and general. Suppose that the regulator sets $\bar{p}$ equal to average revenue with marginal cost pricing, i.e.

$$\bar{p} = \frac{\sum_i c_i(a_i - c_i)}{\sum_i(a_i - c_i)}.$$

Substituting this directly into (4.15) gives $0 \geq W(AR) - W(NC)$. We must check that the firm can still make positive profits in both cases. In general under AR profits are:

$$\Pi = \sum_{i=1}^{n}(\bar{p} - c_i)(a_i - p_i) = \frac{1}{2}\sum_{i=1}^{n}(\bar{p} - c_i)(a_i - c_i + \theta(\bar{p} - c_i)). \tag{4.16}$$

In cases (i) and (ii) $\Pi = \theta\sum_i(\bar{p} - c_i)^2/2 > 0$ since there is some cost variation. This is intuitive for case (ii). Pricing at marginal cost would give the firm zero profits, and other prices are also available. As long as fixed costs are small the firm is willing to participate.

This result is disturbing. If the firm was initially setting first-best prices then

[4] Of course the left-hand side of (4.15) is also equal to zero if the constraint does bind, so $\theta = 0$.

giving it freedom to set relative prices, subject to the constraint that average revenue is at most equal to that with marginal cost pricing, generates a reduction in welfare below not only the first-best level but also the level with no regulation at all. No unreasonable assumptions were necessary to obtain this result, and it depends only on the value of $\bar{p}$, not on the initial relative prices.

We can be more precise about the magnitude of the distortion created by AR if we use the exact expression for W(AR). Using (4.12) in the general expression for welfare gives:

$$
\begin{aligned}
W(AR) &= \tfrac{3}{8}\sum_{i=1}^{n}(a_i - c_i)^2 + \tfrac{2\theta}{8}\sum_{i=1}^{n}(\bar{p} - c_i)(a_i - c_i) - \tfrac{\theta^2}{8}\sum_{i=1}^{n}(\bar{p} - c_i)^2 \\
&= W(NC) + \tfrac{2\theta}{8}\sum_{i=1}^{n}(\bar{p} - c_i)(a_i - c_i) - \tfrac{\theta^2}{8}\sum_{i=1}^{n}(\bar{p} - c_i)^2. \qquad (4.17)
\end{aligned}
$$

Thus in cases (i) and (ii) the second term in (4.17) is zero and the welfare difference is:

$$W(AR) - W(NC) = -\tfrac{\theta^2}{8}\sum_{i=1}^{n}(\bar{p} - c_i)^2.$$

We note that the summation terms in (4.17) can be expressed as:

$$\sum_{i=1}^{n}(\bar{p} - c_i)(a_i - c_i) = n\{\sigma_c^2 - \sigma_{ac} + (\bar{p} - E(c))[E(a) - E(c)]\} \qquad (4.18a)$$

$$\sum_{i=1}^{n}(\bar{p} - c_i)^2 = n\{\sigma_c^2 + [\bar{p} - E(c)]^2\}. \qquad (4.18b)$$

In cases (i) and (ii) we set (4.18a) to zero, and using the resulting expression for $\bar{p}$ in (4.14) total output is

$$Q = \tfrac{n}{2}[E(a) - E(c)] + \tfrac{\theta}{2}\cdot\frac{n[\sigma_{ac} - \sigma_c^2]}{[E(a) - E(c)]}.$$

In case (i) output equals its unconstrained level, whereas in case (ii) it can be above or below Q(NC).

An example will help to illustrate cases (i) and (ii) together. Let $n = 2$, $a_1 = 1$, $a_2 = 1.2$, $c_1 = 0.2$ and $c_2 = 0.4$. This implies $E(a) = 1.1$, $E(c) = 0.3$ and $\sigma_a^2 = \sigma_c^2 = \sigma_{ac} = 0.01$. The unconstrained prices are $p_1(NC) = 0.6$ and $p_2(NC) = 0.8$. Substituting the values of E(a), E(c) and the variances and covariance into (4.9) yields $W(NC) = 0.48$. Total output with no regulation is 0.8, and because the variances are equal and there is perfect correlation between demand and costs this output is optimally distributed when there is no constraint on pricing. When prices equal marginal costs average revenue is 0.3, so $\bar{p} = 0.3$, which also equals E(c). This yields $p_1 = 0.247$, $p_2 = 1.153$, $\theta = 7.062$, $\lambda_{AR} = 0.876$ and $W(AR) = 0.355$, which is 26 per cent *below* the unconstrained level. The AR constraint induces the firm to set price in the higher cost market well above its unconstrained level, reducing sales here almost to zero. The welfare comparison between AR and NC depends simply on the distribution of output. By inducing a large difference in margins AR produces an outcome that is worse than no regulation. If before AR was imposed the firm was pricing uniformly at $\bar{p} = 0.3$ then welfare would be 0.63, which is very close to the first-best level of 0.64, and the introduction of AR causes welfare to fall significantly. But the distortion depends only on the value of $\bar{p}$ and not on whether the initial prices were uniform.

We shall present two further sets of conditions sufficient for $W(AR) < W(NC)$. In case (iii) we have conditions that give a strictly negative upper bound to the welfare difference. Suppose that $\bar{p} = E(c)$ and $\sigma_{ac} > \sigma_c^2 > 0$. The latter condition implies that $\sigma_{ac}/\sigma_a\sigma_c > \sigma_c/\sigma_a$ where the left-hand side is the correlation between a and c. Since this is bounded above by unity we have $\sigma_a > \sigma_c$. This is similar to case (i) except that there is more variation in a than in c. Substituting (4.18a) into (4.17) gives:

$$\tfrac{\theta}{4}n\{\sigma_c^2 - \sigma_{ac} + (\bar{p} - E(c))[E(a) - E(c)]\} \geq W(AR) - W(NC).$$

Under the conditions of case (iii) the left-hand side is strictly negative, so

$W(AR) < W(NC)$. When $\bar{p} = E(c)$ the output level given in (4.14) is equal to the unregulated level. At least one market, say j, has $c_j > \bar{p}$ and so $p_j > (a_j + c_j)/2$. We can also substitute $\bar{p} = E(c)$ into the general expression for welfare, (4.17), and into (4.18a & b), to obtain

$$W(AR) = W(NC) + \tfrac{\theta}{4}n[\sigma_c^2 - \sigma_{ac}] - \tfrac{\theta^2}{8}n\sigma_c^2$$

which is below W(NC) when $\sigma_{ac} > \sigma_c^2 > 0$. The intuition is again that the output level is the same as in the unregulated case but the fact that price in at least one market is above the unconstrained level means that the distribution of output is poor.

We need to show that profits with regulation are positive. From (4.16)

$$2\Pi = \sum_{i=1}^{n} (E(c) - c_i)\{a_i - E(a) + E(a) - E(c) + (1+\theta)[E(c) - c_i]\}$$

$$= n[(1+\theta)\sigma_c^2 - \sigma_{ac}]. \tag{4.19}$$

When $\bar{p} = E(c)$ we have $1 + \theta = \sqrt{\sigma_a^2 + [E(a) - E(c)]^2}/\sigma_c$. If profits are negative then $\sigma_{ac} > (1+\theta)\sigma_c^2$, which implies $\sigma_{ac}/\sigma_a\sigma_c > \sqrt{\sigma_a^2 + [E(a) - E(c)]^2}/\sigma_a$. This cannot hold because the correlation of a with c, $\sigma_{ac}/\sigma_a\sigma_c$, is bounded above by unity, while the right-hand side is strictly greater than unity since $[E(a) - E(c)]^2 > 0$. Thus the firm is profitable under the conditions stated.

A small change to the example used for cases (i) and (ii) can be used to illustrate case (iii). The only difference is that $a_2 = 1.3$ rather than 1.2. This implies that $\sigma_{ac} = 0.15$ and $\sigma_a^2 = 0.0225$, while σ_c^2 remains at 0.01. The price cap is $\bar{p} = E(c) = 0.3$. AR yields $p_1 = 0.218$ and $p_2 = 1.231$ and $W(AR) = 0.379 < 0.544 = W(NC)$. Welfare with uniform pricing at $\bar{p} = 0.3$ is 0.715.

In case (iv) again we have $\bar{p} = E(c)$ but we assume that $\sigma_a^2 = 0$ and $a - E(c) > 3\sigma_c > 0$. To show that these assumptions are sufficient for $W(AR) <$

W(NC), we cannot use the welfare bound. Instead we look directly at (4.17) and (4.18a & b). When $\sigma_a^2 = \sigma_{ac} = 0$ and $\bar{p} = E(c)$ both (4.18a) and (4.18b) are equal to $n\sigma_c^2$. Substituting into (4.15) gives $W(AR) = W(NC) + \frac{\theta n}{8}(2-\theta)\sigma_c^2$, so $W(AR) < W(NC)$ if $\theta > 2$. Under the conditions we have $\theta = \{a - E(c)\}/\sigma_c - 1$, so as long as $a - E(c) > 3\sigma_c$ we have $\theta > 2$. Using (4.19) profits are $n(1+\theta)\sigma_c^2/2 > 0$ under the conditions of case (iv). Again because $\bar{p} = E(c)$ output is equal to the unregulated level and is badly distributed.

The example of case (iv) is similar to those for the other cases, except that $a_2 = a_1 = 1$. The price cap remains at $\bar{p} = E(c) = 0.3$. The condition $a - E(c) > 3\sigma_c$ is satisfied since $3\sigma_c = 0.3$ and $a - E(c) = 0.7$.[5] Prices under AR are $p_1 = 0.3$ and $p_2 = 1$, so demand in market 2 is exactly zero. The welfare levels are $W(AR) = 0.315$ and $W(NC) = 0.375$. Welfare with uniform pricing at 0.3, $W(\bar{p})$, is 0.49. Profits when both prices equal 0.3 are zero.

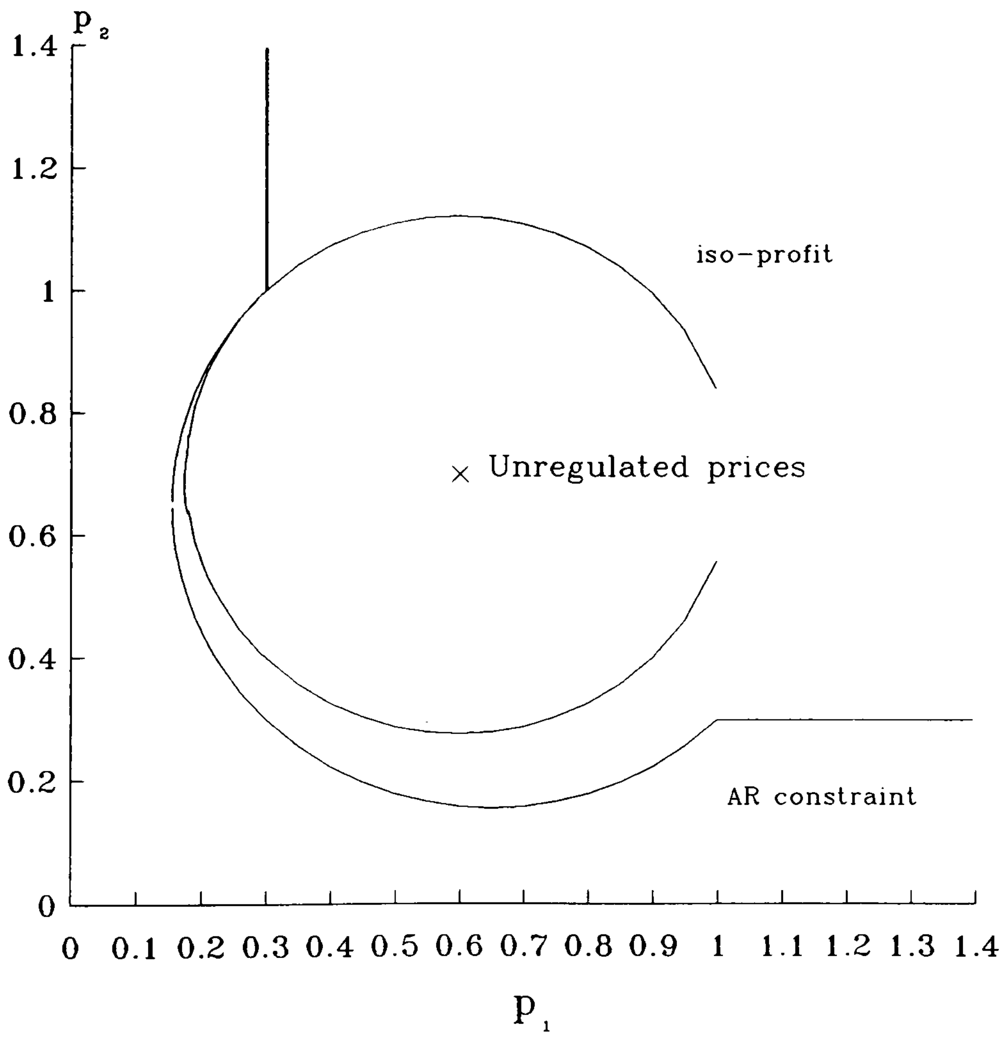


Figure 4.2 Distortions under Average Revenue Regulation

[5]If costs were $c_1 = 0.6$ and $c_2 = 0.8$, with $\bar{p} = 0.7$ then we would have $a - E(c) - 3\sigma_c = 0$, yielding welfare exactly equal to the unconstrained level. The higher the mean cost, E(c), the more likely it is that the condition does not hold. This is intuitive because the greater costs are the less the large reductions in output that can occur with AR matters for welfare.

Figure 4.2 illustrates the example for case (iv). To find an expression for the AR constraint the equation $(\bar{p}-p_1)(a_1-p_1)+(\bar{p}-p_2)(a_2-p_2)=0$ is solved for p_2. There are two roots of this equation for any given value of p_1. Only one of the roots is relevant when $p_1>\bar{p}$, since it is not possible to have both prices exceeding $\bar{p}$. The solution is:

$$p_2=\frac{a_2+\bar{p}\pm\sqrt{(a_2-\bar{p})^2-4(\bar{p}-p_1)(a_1-p_1)}}{2}.$$

The AR constraint is a circle with points cut-off to the north-east of the point where $p_1=p_2=\bar{p}$. Note that when $p_2\geq 1$ the highest value of p_1 that can be set is 0.3, and similarly $p_2=0.3$ when $p_1\geq 1$. The best prices for the firm are $p_1=0.3$ and $p_2=1$, where the highest iso-profit curve touches the AR constraint. If there was no cost variation, e.g. if $c_1=c_2=0.2$, then the centre of the iso-profit circle would be shifted down by 0.1 and the firm would choose to set both prices equal to $\bar{p}=0.3$. It is the variation in costs that pushes up the iso-profit curve in Figure 4.2 and causes the value of p_2 to jump from 0.3 to 1.

In the four cases we have considered regulation is "tight". The price cap is equal to the (weighted or simple) average marginal cost. Since welfare is lower under AR than under no regulation complete abolition of the cap would raise welfare. The firm would earn greater profits and consumers in the high-cost market would gain, while those consuming in the low-cost market lose from a higher price. We can also consider the effect of a marginal relaxation of the price cap. In the example for case (iv) let $\bar{p}=0.4$. Since $\bar{p}=c_2$ this implies that p_2 falls from 1 to 0.7, which is the unconstrained level. The new value of p_1 is 0.276, so both prices *fall* with the *rise* in $\bar{p}$. Since profits also rise the relaxation of the cap has generated a Pareto improvement. The value of W(AR) when $\bar{p}=0.4$ is 0.452 which exceeds $W(NC)=0.375$, but is still below $W(\bar{p})=0.48$.

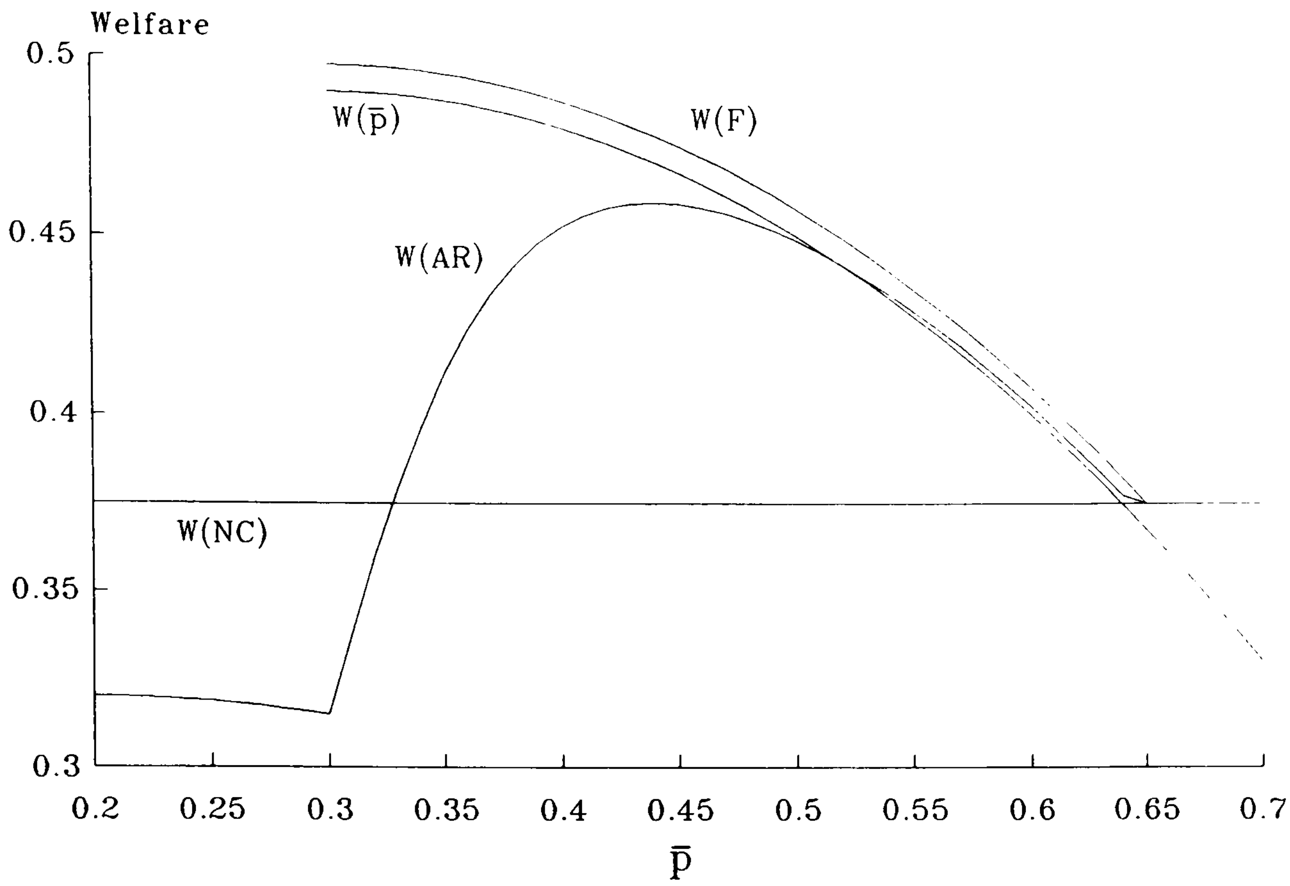


Figure 4.3 Welfare under AR with different values of $\bar{p}$

Figure 4.3 shows the values of W(AR) for different values of $\bar{p}$. The value of W(NC) is 0.375. When $\bar{p} < 0.3$ the firm simply sets $p_2 = 1$ and $p_1 = \bar{p}$, so relaxation of the cap lowers welfare in this region. But as $\bar{p}$ increases in the region [0.3, 0.35] both prices fall and there is a Pareto improvement. Welfare continues to rise as $\bar{p}$ increases in the region [0.35, 0.44], but without causing a Pareto improvement. Welfare reaches its maximum when $\bar{p} = 0.44$, and thereafter it falls back towards W(NC) = 0.375.

The result that relative prices can be highly inefficient bears a resemblance to the Averch-Johnson effect with rate-of-return regulation (Averch and Johnson, 1962). With AR the firm faces a nonlinear constraint with upward-sloping parts. Similarly under static rate-of-return regulation the firm faces an ellipse-shaped constraint in capital-labour space and chooses an inefficiently large ratio of capital to labour. The use of capital and thus the distortion increases as the allowed rate-of-return approaches the cost of capital (Baumol and Klevorick, 1970), while a tightening of the AR cap can result in both prices being higher and a lower level of welfare.

All four sets of sufficient conditions have required costs to vary. This suggests that a *necessary* condition for W(AR) < W(NC) is that $\sigma_c^2 > 0$. We now prove that this is so. The value of θ when $\sigma_c^2 = 0$ is

$$\theta = \frac{\sqrt{\sigma_a^2 + [E(a) - \bar{p}]^2}}{\bar{p} - c} - 1 = \frac{Y - (\bar{p} - c)}{\bar{p} - c}$$

where $Y^2 \equiv \sigma_a^2 + [E(a) - \bar{p}]^2$. As long as (4.11) holds $Y > \bar{p} - c$. If (4.11) does not hold then $\theta = 0$ and the unconstrained prices given in (4.8) are set. We can use the value of θ, together with (4.18a) and (4.18b) to find welfare when $\sigma_c^2 = 0$:

$$W(AR) = W(NC) + \tfrac{2n}{8}[Y - (\bar{p} - c)](E(a) - c) - \tfrac{n}{8}[Y - (\bar{p} - c)]^2.$$

This implies that W(AR) < W(NC) if and only if

$$2[Y - (\bar{p} - c)](E(a) - c) - [Y - (\bar{p} - c)]^2 < 0.$$

Suppose this holds. Then

$$-\{E(a) - c - [Y - (\bar{p} - c)]\}^2 + [E(a) - c]^2 < 0$$

which implies that $Y < \bar{p} - c$. But this is not consistent with the fact that the constraint binds. Thus welfare exceeds the unconstrained level when there is no cost variation. This is intuitive: prices are below their unconstrained levels when costs are uniform, and total output is therefore above the unconstrained level. Some prices can, however, be below c so we required a formal proof that W(AR) > W(NC).

When the price cap is relaxed we expect that all prices would rise. This is confirmed in the case where $\sigma_c^2 = 0$, although as we have noted this need not happen when there is cost variation. Substituting for θ into the pricing equation (4.12) yields

$$p_i = \frac{a_i + \bar{p} - \sqrt{\sigma_a^2 + [E(a) - \bar{p}]^2}}{2},$$

so prices do not depend on the cost level, and the difference between p_i and p_j is $(a_i - a_j)/2$. Relaxing the price cap means raising $\bar{p}$, giving

$$\frac{dp_i}{d\bar{p}} = \frac{1}{2}\left(1 + \frac{E(a) - \bar{p}}{\sqrt{\sigma_a^2 + [E(a) - \bar{p}]^2}}\right) > 0,$$

since $E(a) > \bar{p}$, so all prices rise by the same absolute amount. This certainly lowers welfare when all prices exceed c. Thus with no differences in costs the properties of AR regulation are as expected.

We now briefly compare W(AR) with welfare under uniform pricing at $\bar{p}$, $W(\bar{p})$. Armstrong and Vickers (1991) show by example that this comparison can go either way in the linear case when costs are uniform. The nature of this ambiguity is explored a little further. The difference between W(AR) and $W(\bar{p})$ is the sum of the differences between profits and consumer surpluses under the two regimes:

$$W(AR) - W(\bar{p}) = \Pi(AR) - \Pi(\bar{p}) + V(AR) - V(\bar{p})$$

$$= \sum_{i=1}^{n} (\bar{p} - c_i)(\bar{p} - p_i) - \frac{1}{2}\sum_{i=1}^{n} (\bar{p} - p_i)^2.$$

where we have assumed that the constraint binds. We have seen that when $\bar{p}$ is low the loss to consumer surplus from AR rather than uniform pricing, $\sum_i(\bar{p} - p_i)^2/2$, is large when costs differ, and W(AR) can be below W(NC), which in turn is below $W(\bar{p})$. But as $\bar{p}$ rises any distortion from the imposition of AR falls. Thus we expect that the probability that $W(AR) > W(\bar{p})$ increases as the constraint is relaxed. In Figure 4.3 we illustrate the comparison between $W(\bar{p})$ and W(AR) for different values of $\bar{p}$ using the same example as for case (iv). This confirms that for low values of the cap $W(\bar{p}) > W(AR)$ and it is only for large values of $\bar{p}$ that allowing price differentiation subject to the AR constraint raises welfare.

We conclude this subsection with a summary of the results. AR has the potential to induce gross distortions in pricing when costs differ. Some prices can be above their unconstrained levels, the distortion can get worse as regulation is tightened and welfare can be below its unregulated level. These results depend only on the height of the price cap and not on whether prices were previously uniform or not. Allowing price freedom is only better than requiring uniform pricing at $\bar{p}$ when the constraint is relatively loose.

4.3.2 Fixed Weights Regulation

In Chapter 3 we showed that regulation based on a price cap with fixed weights had reasonable welfare properties. In this subsection we explore these properties further using the linear model of the previous subsection.

We assume that the firm's initial prices are uniform and equal to $\bar{p}$. It is also assumed that profits with uniform pricing are not negative. The generalization to the case where initial prices are not uniform is straightforward. The weights in the price index are the initial quantities $\{a_i - \bar{p}\}$ and the firm's Lagrangian is:

$$\sum_{i=1}^{n} (p_i - c_i)(a_i - p_i) + \lambda_F \sum_{i=1}^{n} [\bar{p} - p_i](a_i - \bar{p})$$

where λ_F is the multiplier on the price constraint. Prices are given by:

$$p_i = \frac{a_i + c_i}{2} - \lambda_F \frac{(a_i - \bar{p})}{2}. \tag{4.20}$$

Since we assume that $a_i \geq \bar{p}$ for all i, i.e. all markets are served with uniform pricing, (4.20) indicates that $p_i \leq (a_i + c_i)/2$. This suggests that we could not have a case where $W(F) < W(NC)$. Substituting (4.20) into the price constraint gives

$$\lambda_F = 1 - \frac{\sum_i (\bar{p} - c_i)(a_i - \bar{p})}{\sum_i (a_i - \bar{p})^2} = 1 - \frac{\Pi(\bar{p})}{\sum_i (a_i - \bar{p})^2}.$$

Since we have assumed $\Pi(\bar{p}) \geq 0$, $\lambda_F \leq 1$. Output is

$$Q(F) = \frac{n}{2}[E(a) - E(c)] + \frac{\lambda_F}{2}n[E(a) - \bar{p}]$$

which is greater than the unrestricted level since $E(a) > \bar{p}$.

Both consumers and the firm are better-off with regulation by F than with uniform pricing at $\bar{p}$, as was shown in Chapter 3 for the general case where initial prices are not necessarily uniform. Welfare with fixed weights regulation is obtained by substituting the pricing equation (4.20) into the expression for welfare, giving

$$W(F) = \frac{3}{8}\sum_{i=1}^{n}(a_i - c_i)^2 + \frac{2\lambda_F}{8}\sum_{i=1}^{n}(a_i - c_i)(a_i - \bar{p}) - \frac{\lambda_F^2}{8}\sum_{i=1}^{n}(a_i - \bar{p})^2$$

$$= W(NC) + \frac{2\lambda_F}{8}\Pi(\bar{p}) + \frac{\lambda_F(2-\lambda_F)}{8}\sum_{i=1}^{n}(a_i - \bar{p})^2. \tag{4.21}$$

It is simple to show that under our assumptions $W(F) \geq W(NC)$. With $\Pi(\bar{p}) \geq 0$, $\lambda_F \leq 1$ and the second and third terms in the above expression are non-negative.

What happens when a tight price cap is imposed? We know that AR can be associated with inefficiency under such conditions. Let $\bar{p}$ be the smaller root of $\Pi(\bar{p}) = 0$. Since we can write $\Pi(\bar{p}) = n[(\bar{p} - E(c))(E(a) - \bar{p}) - \sigma_{ac}] = 0$, we have

$$\bar{p} = \frac{E(a) + E(c) - \sqrt{[E(a) - E(c)]^2 - 4\sigma_{ac}}}{2} \tag{4.22}$$

which equals $E(c)$ when $\sigma_{ac} = 0$, and exceeds $E(c)$ for $\sigma_{ac} > 0$. Then $\lambda_F = 1$ and

$$W(F) = \frac{3}{8}\sum_{i=1}^{n}(a_i - c_i)^2 + \frac{1}{8}\sum_{i=1}^{n}(a_i - \bar{p})^2 = W(NC) + \frac{1}{8}\sum_{i=1}^{n}(a_i - \bar{p})^2.$$

Thus $W(F) > W(NC)$. Given $\Pi(\bar{p}) = \sum_i(\bar{p} - c_i)(a_i - \bar{p}) = 0$ W(F) can be written as

$$=\frac{1}{2}\sum_{i=1}^{n}(a_i-c_i)^2-\frac{1}{8}\sum_{i=1}^{n}(\bar{p}-c_i)^2=W^*-\frac{1}{8}n[\sigma_c^2+(E(c)-\bar{p})^2]$$

where W^* is the first-best welfare level, so when $\sigma_c^2>0$, $W(F)<W^*$. We would, however, expect welfare to be close to its maximum level. In the example of case (iv) of the previous subsection, where $a_1=a_2=1$, $c_1=0.2$ and $c_2=0.4$, $\bar{p}=0.3$ and $\sigma_c^2=0.01$, we have $\Pi(\bar{p})=0$ and $W(F)=0.4975$, just below $W^*=0.5$ and considerably above $W(AR)=0.315$. When $\sigma_c^2=0$ we have $W(F)=W^*$ whenever $\bar{p}=c$.

We now compare W(F) with W(AR). It has not been possible to characterize the precise conditions under which one is greater than another.[6] In the example of case (iv) $W(F)>W(AR)$ for values of $\bar{p}\geq 0.3$, as illustrated in Figure 4.3. For lower values of $\bar{p}$, however, F is not feasible because it generates negative profits, whereas AR is feasible since the firm drives demand in the high-cost market to zero. The reason for this is clear. AR allows the firm more flexibility and thus enables it to make profits in situations where the firm cannot make profits under F.

Another example where $W(AR)>W(F)$ is the following. Let $a_1=1$, $a_2=0.4$, $c_1=0.2$, $c_2=0.4$ and $\bar{p}=0.3$. The correlation between a and c is -1. These conditions imply $\bar{p}=E(c)$ but since $\sigma_{ac}<0$ the conditions of case (iii) do not hold. The first-best level of demand in market 2 is 0 because $a_2=c_2$. Under AR the firm sets $p_1=0.3$ and $p_2=0.4$, driving demand in market 2 to zero. In the examples in section 4.3.1 such extreme pricing is sub-optimal, but here it is beneficial. Welfare under AR is simply welfare in market 1 which is 0.315. Under F, however, the firm sets $p_1=0.292$ and $p_2=0.356$, so demand in market 1 is still positive, yielding $W(F)=0.3148<W(AR)$. Small variations in the parameters of the example do not affect the result that $W(AR)>W(F)$.

Although we have been able to find cases where AR is preferable to F these are rather extreme. The examples rely either on a price cap being so tight that the firm

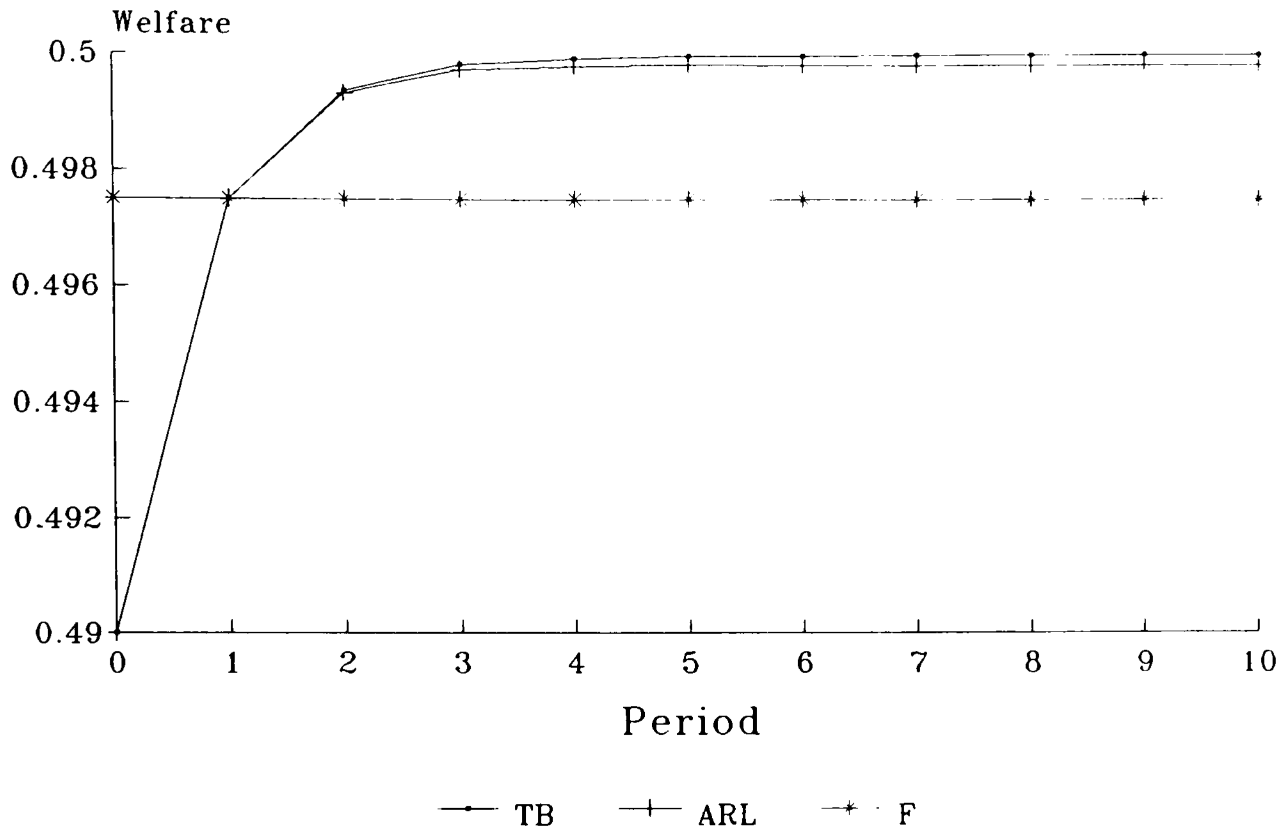
[6]When the firm sets a two-part tariff $W(F)>W(AR)$ without ambiguity. This is proved in the next chapter.

cannot make a profit at all under F, or on a case where the effect of the extreme pricing induced by a tight price cap under AR is beneficial because it reduces to zero the demand of a product that generates negative social surplus with positive production levels. The general conclusion of the analysis is that while F is not always better than AR, at least F is more robust in the sense that under reasonable assumptions it generates higher welfare than both uniform pricing and no regulation, and tightening the cap raises welfare towards the first-best level.

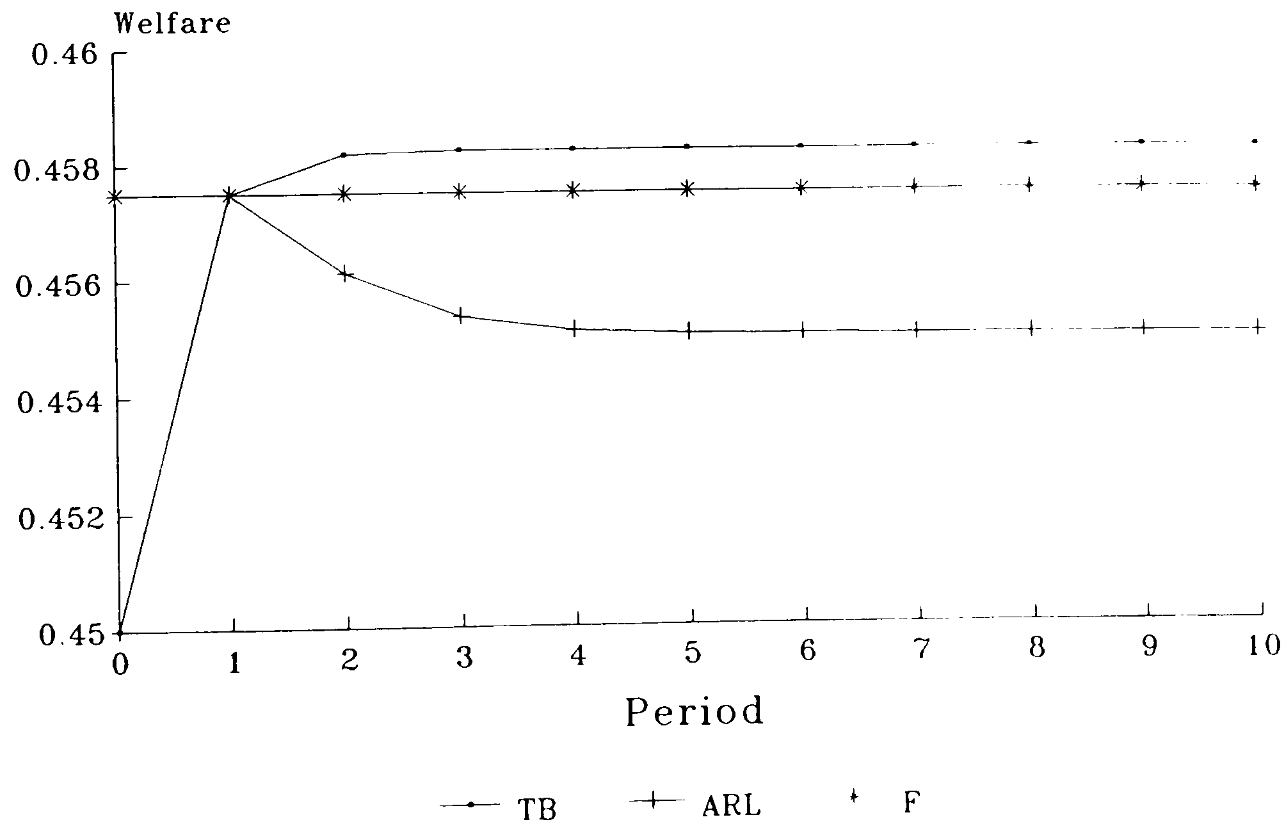
4.4 Other Price Caps

In this section we briefly consider the welfare properties of the tariff basket (TB) and average revenue (lagged) (ARL) schemes that were discussed in Chapter 3. The scheme that is a combination of TB and the Paasche Price Index scheme, characterized in equation (3.18), is also assessed. We use simple linear examples related to the ones consider in section 4.3 to explore some possible outcomes and compare them with AR and F. Appendix 4.2 contains more details of the formulae that underlie the calculations.

For tractability it is assumed that the discount factor of the firm is zero. The example that we start with is the same as that for case (iv) in 4.3.1 with $\bar{p} = 0.3$. We noted earlier that the maximum level of welfare is $W^* = 0.5$, $W(\bar{p}) = 0.49$, $W(NC) = 0.375$, $W(AR) = 0.315$ and $W(F) = 0.4975$. The mixed constraint (M) generates $W(M) = 0.499949$, which is very close to W^*. When a myopic firm faces TB regulation the outcome in the first period is that same as that under F, but thereafter welfare rises as the firm chooses prices that raise profits and increase consumer surplus. When ARL is applied the first period outcome is also the same as in TB and F, and thereafter consumer surplus falls while the firm takes advantage of the extra flexibility in the cap relative to TB to earn higher profits. The time paths of welfare for TB, ARL and F when $\bar{p} = 0.3$ are illustrated in Figure 4.4(a). The net effect under ARL of the decline in consumer surplus and rise in profits from period 2 onwards is a rise in welfare above W(F), but TB generates a superior welfare path. Welfare levels in period 1 under TB, ARL and F are below the level with M, but in steady state both TB and ARL have higher welfare levels.



(a) $\bar{p} = 0.3$



(b) $\bar{p} = 0.5$

Figure 4.4 Time Paths of Welfare under TB, ARL and F

A relaxation of the constraint to $\bar{p} = 0.5$ changes the welfare ranking. The value of W(AR) rises to 0.4486, while W(F) = 0.4575 and W($\bar{p}$) = 0.45. W(M) falls below W(F) to 0.4571. The time paths of welfare under TB, ARL and F are illustrated in Figure 4.4(b). Again TB is the best scheme. With $\bar{p} = 0.5$ F is preferable to ARL because W(ARL) is lower than W(F), except in the first period when they are equal. ARL is still preferable to AR since W(ARL) always exceeds W(AR).

In the example in the previous section where $a_2 = c_2 = 0.4$ and $\bar{p} = 0.3$ welfare under AR exceeds W(F) because it is desirable to drive demand in market 2 to zero. In this case we had W(AR) = 0.315 and W(F) = 0.3148. We can also calculate that W(M) = 0.315685, so of the static schemes M is the best, followed by AR and F. Under TB, however, the welfare level is only lower than W(AR) and W(M) in the first period. From period 2 onwards TB generates higher welfare than both. ARL has welfare that is above AR after the first period, tending to W(AR) in steady-state. These results are illustrated in Figure 4.5.

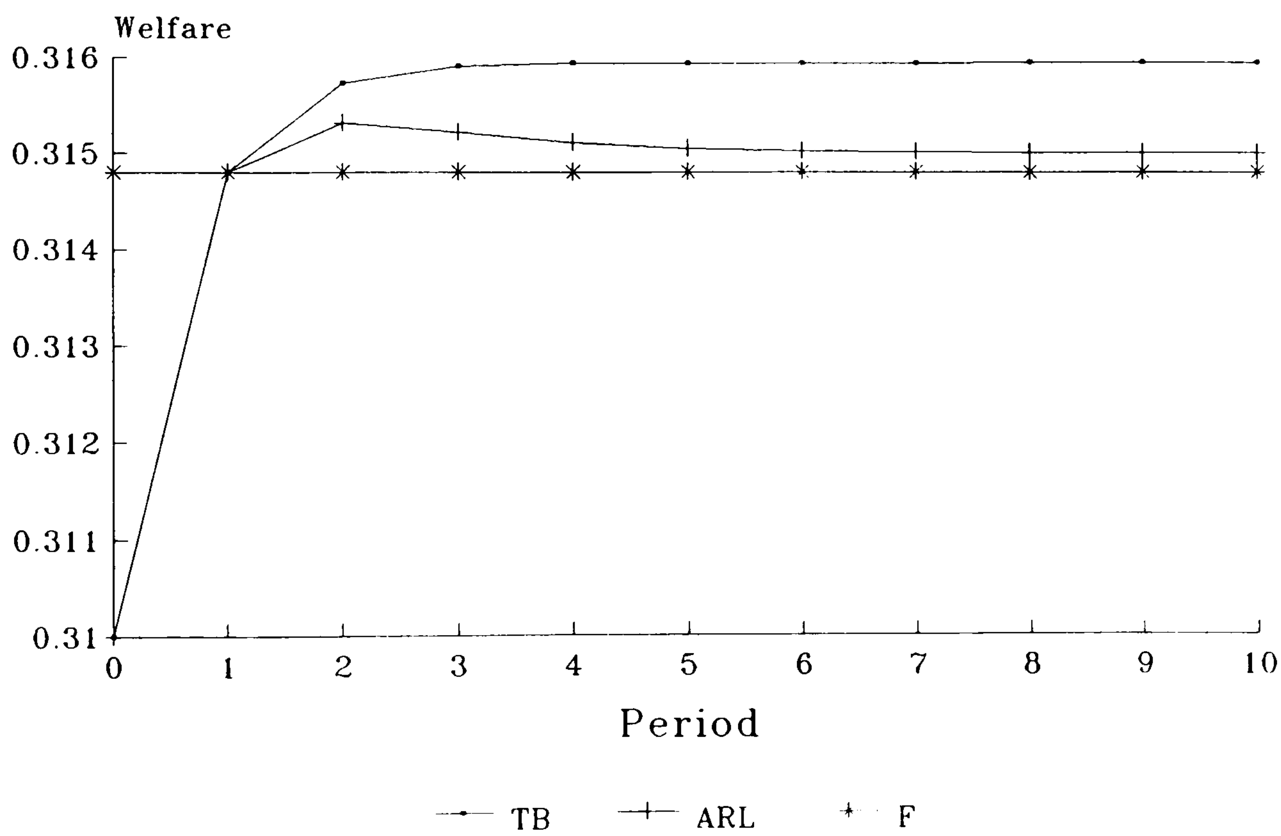


Figure 4.5 Time Paths of Welfare when W(AR) > W(F).

4.5 Conclusions

In this chapter we have used a simple model to explore some welfare effects of allowing a monopolist to set different prices in different markets, with the focus on the case where the costs of serving markets vary. The aim was two-fold. First, we sought to provide a framework that encompasses the standard results on price discrimination in the linear model with no new markets opened up and no price level regulation, and to extend this to the case where costs differ. Second, when price level regulation was considered we showed that under reasonable conditions regulation based on average revenue can have highly distortionary effects. Other price cap schemes that have more reasonable properties were compared with average revenue regulation, and although examples can be found where they generate lower welfare than average revenue regulation, they generally are preferable on welfare grounds.

Appendix 4.1

We prove that θ is given by (4.13). The AR constraint implies

$$\sum_{i=1}^{n} (\bar{p} - p_i)(a_i - p_i) = 0$$

$$\rightarrow \quad \sum_{i=1}^{n} [2\bar{p} - a_i - c_i + \theta(\bar{p} - c_i)](a_i - c_i + \theta(\bar{p} - c_i)) = 0$$

$$\rightarrow \quad \sum_{i=1}^{n} [(1+\theta)(\bar{p} - c_i) - (a_i - \bar{p})](a_i - \bar{p} + (1+\theta)(\bar{p} - c_i)) = 0$$

$$\rightarrow \quad (1+\theta)^2 \sum_{i=1}^{n} (\bar{p} - c_i)^2 = \sum_{i=1}^{n} (a_i - \bar{p})^2$$

$$\rightarrow \quad \theta = \sqrt{\frac{\sum_i (a_i - \bar{p})^2}{\sum_i (\bar{p} - c_i)^2}} - 1.$$

It is the positive root that is relevant. Subtracting and adding E(a) in the numerator of the fraction and doing the same with E(c) in the denominator yields (4.13).

Appendix 4.2

We change notation in this appendix back to that of the earlier chapters. Subscripts now indicate time periods and superscripts denote the market. The TB constraints are linear, enabling us to write p_t^2 as a function of p_t^1 and to substitute this into the profit function, yielding a choice problem for the myopic firm in one variable, p_t^1. This gives the following expressions for prices:

$$p_t^1 = \frac{1}{2\left(1 + \left(q_{t-1}^1/q_{t-1}^2\right)^2\right)}\left\{a^1 + c^1 + (2p_{t-1}^2 - a^2 - c^2)\frac{q_{t-1}^1}{q_{t-1}^2} + 2p_{t-1}^1\left(\frac{q_{t-1}^1}{q_{t-1}^2}\right)^2\right\}$$

and

$$p_t^2 = p_{t-1}^2 + (p_{t-1}^1 - p_t^1)\frac{q_{t-1}^1}{q_{t-1}^2}.$$

Similarly with ARL we can substitute out for p_t^2 giving:

$$p_t^1 = \frac{1}{2\left(1 + \left(q_{t-1}^1/q_{t-1}^2\right)^2\right)}\left\{a^1 + c^1 + (2\bar{p} - a^2 - c^2)\frac{q_{t-1}^1}{q_{t-1}^2} + 2\bar{p}\left(\frac{q_{t-1}^1}{q_{t-1}^2}\right)^2\right\}$$

and

$$p_t^2 = \bar{p} + (\bar{p} - p_t^1)\frac{q_{t-1}^1}{q_{t-1}^2}.$$

The mixed constraint is given by equation (3.18):

$$\sum_{i=1}^{2} [p_{t-1}^i - p_t^i](q_{t-1}^i + q_t^i) \geq 0.$$

With linear demands we showed in Appendix 3.1 that this is the same as requiring consumer surplus not to fall below its initial level. This is represented here by

$$\frac{(a^1-p^1)^2}{2}+\frac{(a^2-p^2)^2}{2}\geq V(\bar{\mathbf{p}}),$$

where we do not need to date the prices. The Lagrangian is:

$$\mathcal{L}=\sum_{i=1}^{n}(p^i-c^i)(a^i-p^i)+\lambda_M\left\{\frac{(a^1-p^1)^2}{2}+\frac{(a^2-p^2)^2}{2}-V(\bar{\mathbf{p}})\right\}.$$

This yields

$$p^i=\frac{(1-\lambda_M)a^i+c^i}{2-\lambda_M}\quad \text{for } i=1,2 \text{ and}$$

$$\lambda_M=2-\frac{\sqrt{V^*}}{\sqrt{V(\bar{\mathbf{p}})}}$$

where $V^*\equiv\sum_i(a^i-c^i)^2/2$ is the level of consumer surplus with marginal cost pricing. When $V(\bar{\mathbf{p}})=V^*$ we have $\lambda_M=1$ and $p^i=c^i$ for all i. Some manipulations yield

$$W(M)=2\sqrt{V^*}\sqrt{V(\bar{\mathbf{p}})}-V(\bar{\mathbf{p}}).$$

This is used in section 4.4.

CHAPTER 5

TWO-PART TARIFFS AND PRICE CAPS

5.1 Introduction

In this chapter we analyze price caps when the regulated firm sets a two-part tariff. An access fee is charged to each customer, while the marginal price of a unit of the product is constant. The vast majority of domestic customers of the utilities in the UK face such a tariff. Gas, electricity and water tariffs are all two-part, and for most domestic telecommunications users the tariff is effectively two-part, although more sophisticated nonlinear tariffs are being introduced. Revenue received from access charges allows some of the fixed costs of network operation to be financed and thus allows product prices to be closer to marginal costs than with linear pricing. In addition there are usually some costs that are driven by the number of customers rather than by the volumes they consume. Metering, billing and connection costs are examples. Setting an access charge allows some of these non-volume customer-related costs to be covered. Two-part tariffs, like other types of price discrimination, are feasible only if the customer cannot resell the product. A more sophisticated type of pricing is a nonlinear tariff, where the firm relates the bill that the customer pays to the quantity consumed in a nonlinear way and marginal prices are not constant over the whole range of the tariff. Such pricing is common for non-domestic customers and some larger domestic customers now face or can choose nonlinear tariffs. Nonlinear pricing will be discussed in the next chapter.

In general with a two-part tariff the number of customers depends on both the access charge and the marginal price, and we shall examine this case in Chapter 7, but here it is assumed that the number of customers is constant. This is a common assumption in the regulation literature (see the literature discussed in Chapter 2). One possible interpretation of this assumption is that consumers are identical, but this is not necessary. Although the assumption is not fully realistic its adoption significantly simplifies the analysis of the dynamics. We also focus on a model with linear demand

and quadratic costs that allows for straightforward comparison between the different price cap schemes introduced in Chapter 3. The ambiguities that were noted in that chapter are resolved, and many conjectures that could not be proved there can be demonstrated here. The use of a linear model enables us to characterize completely the dynamic behaviour of prices, profits and consumer surplus for each price cap. It also allows clear welfare comparisons to be made.

There is a small existing literature on dynamic price caps with two-part tariffs. Vogelsang (1989a, 1991) notes that a price cap scheme identical to the tariff basket scheme discussed below generates marginal cost pricing in the limit. Sappington and Sibley (1992) consider both the average revenue (lagged) scheme (ARL) and the fixed weights scheme. They show that ARL can cause distortions, and in particular that it can lower the present value of consumer surplus relative to that earned at the initial prices. Their analysis is based on a finite horizon and is mainly numerical. In our model there is an infinite horizon, so the price caps are "pure". The assumption of an infinite horizon is not a realistic way to model actual price caps, but it does allow us to compare the dynamic properties of the different price caps without having to resort to numerical methods and it facilitates comparison with the results from the literature on dynamic regulatory schemes discussed in Chapter 2.

We shall consider in turn the five main price caps analyzed in Chapter 3, i.e. the tariff basket (TB) scheme, the fixed weights (F) scheme, the average revenue (AR) scheme, the ARL scheme and the Paasche Price Index (PPI) scheme. It turns out that AR and PPI are dominated by others in welfare terms, when welfare is the unweighted sum of consumer surplus and profits. The comparison of the remaining schemes, TB, F and ARL depends on two parameters, the discount factor and the gradient of the marginal cost curve (relative to the gradient of the inverse demand function). By considering all possible values of these two parameters we can characterize the precise conditions under which a particular scheme is best.

The plan of the chapter is as follows. In 5.2 we analyze and compare the five main price caps. In 5.3 two hybrid caps based on those discussed in 3.3 are discussed. Conclusions are in 5.4.

5.2 Analysis of the Price Caps

5.2.1 The Tariff Basket Price Cap

When the firm sets a two-part tariff the tariff basket constraint is

$$(A_{t-1} - A_t)N_{t-1} + \sum_{i=1}^{n} (p^i_{t-1} - p^i_t)q^i_{t-1} \geq 0 \tag{5.1}$$

where A is the access charge and N is the number of customers.[1] The revenue that would be obtained at the period t prices if the number of customers and demand for the products remained at their $t-1$ levels must be no greater than actual revenue in period $t-1$. In this chapter we assume without loss of generality that N is fixed at 1 and, for notational simplicity, that there is only one product. The firm's problem under TB is then:

$$\underset{\{A_t,\, p_t\}}{\text{Max}} \sum_{t=1}^{\infty} \beta^{t-1}\{A_t + \Pi(p_t)\} \quad \text{s.t. } A_{t-1} - A_t + (p_{t-1} - p_t)q(p_{t-1}) \geq 0 \ \text{ for } t \geq 1$$

where $\Pi(p_t) \equiv p_t q(p_t) - C(q_t)$ is operating profit without the access charge and $0 \leq \beta < 1$. A sufficient condition for the constraints to bind is that $V(p_0) - A_0 > 0$. We know from Proposition 3.1 that the sequence of consumer surpluses is non-decreasing. A constraint would be slack only if the firm priced at marginal cost and extracted all the consumer surplus through A. But this contradicts the fact that $V(p_t) - A_t > 0$, by the properties of the constraint. Thus in all interesting cases the constraints bind.

With (5.1) binding we can write it as:

$$A_t = A_{t-1} + (p_{t-1} - p_t)q_{t-1} = A_0 + \sum_{i=1}^{t} (p_{i-1} - p_i)q_{i-1} \tag{5.1$'$}$$

[1]The notation reverts to that used in Chapters 2 and 3, so superscripts denote markets or products and subscripts denote time periods.

where the second expression follows from repeated substitution. As in Chapters 3 and 4 regulation is imposed from period 1 onwards, and p_0 and A_0 are given. The similarity with the subsidy scheme in (2.26) should be noted. If p_t is below p_{t-1} the firm is compensated for the loss of revenue on existing sales, $(p_{t-1} - p_t)q_{t-1}$, by being allowed to increase its access charge by exactly that amount, and because of the autoregressive nature of the constraint all future access charges will be higher by that amount.

In Figure 5.1 we show the income sequence for the price sequence $\{p_{t-1}, p_t, p_{t+1}\}$. Period $t-1$ income is the access charge A_{t-1} (not shown) plus operating profits $B+D+F$ (less fixed costs). In period t the access charge is $A_{t-1}+B$, and operating profits are $D+E+F+G$ (less fixed costs), so income is $E+G$ higher than in period $t-1$. Consumer surplus is higher in period t than in $t-1$ by the triangle I. In period $t+1$ the firm's income increases by area H, which is the sum of the increase in the access charge, $D+E$, and the change in operating profits, $H-(D+E)$.

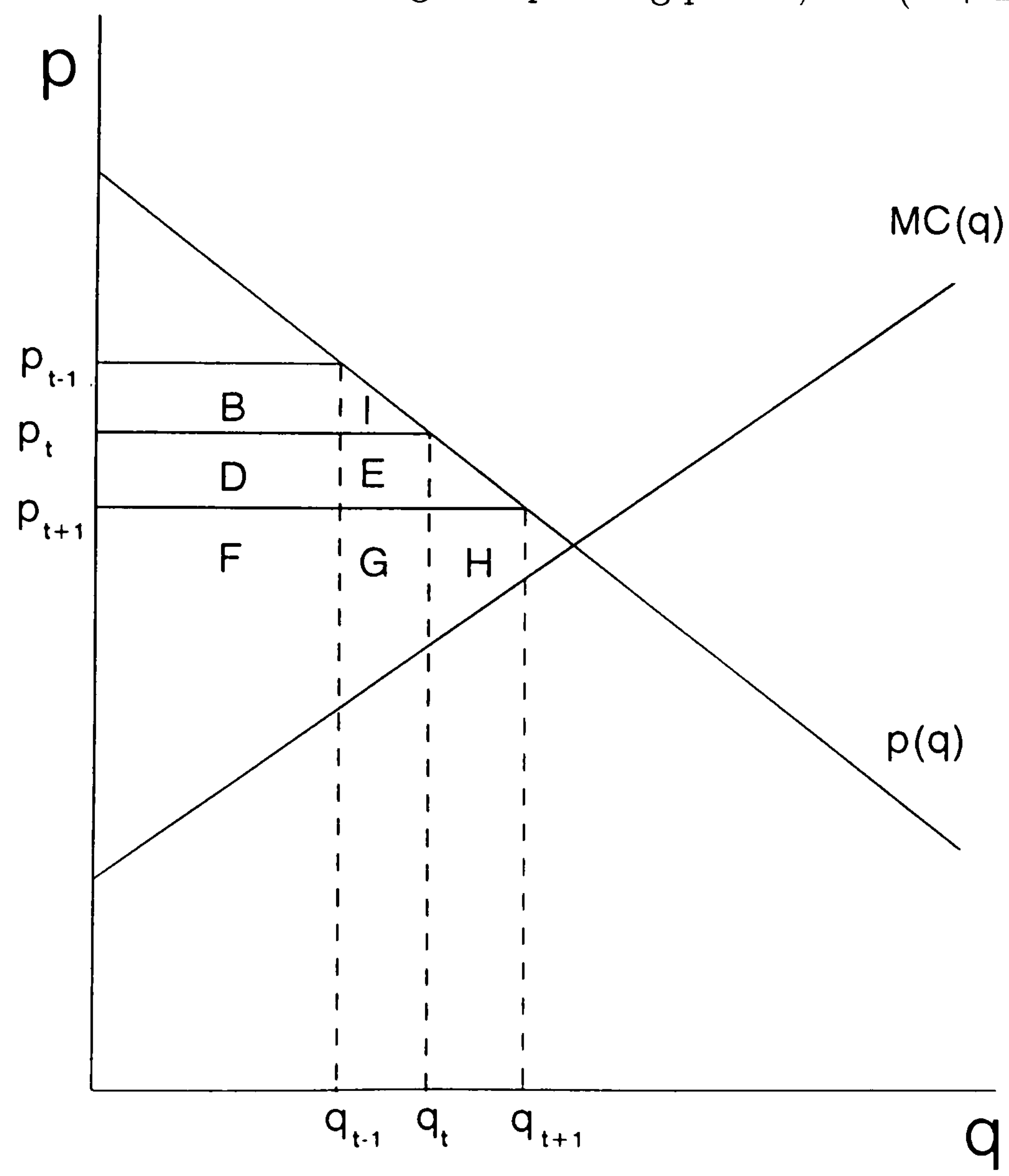


Figure 5.1 The Tariff Basket Scheme

TB regulation with a two-part tariff is closely related to the Loeb and Magat (L-M) scheme discussed in 2.3.1. Under L-M the transfer is equal to $V(p_t)$. This strictly exceeds the transfer under TB, A_t, because otherwise the TB constraint would not bind. In Figure 5.1 the L-M transfer for period t is higher than A_t by area I.

Under TB the present value of the firm's income stream is:

$$Z = \frac{A_0}{1-\beta} + \frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}(p_{t-1} - p_t)q_{t-1} + \sum_{t=1}^{\infty}\beta^{t-1}\{p_t q_t - C_t\} \tag{5.2}$$

given that the constraints bind and using the second expression for A_t in (5.1′). The firm chooses the price sequence $\{p_t\}_{t=1}^{\infty}$ to maximize this expression. The first-order condition for the choice of p_t yields:

$$q_t + [p_t - MC_t]q'(p_t) - \frac{1}{1-\beta}q_{t-1} + \frac{\beta}{1-\beta}[(p_t - p_{t+1})q'(p_t) + q_t] = 0. \tag{5.3}$$

The effect of a small increase in p_t is three-fold. Operating profits in period t rise by $q_t + [p_t - MC_t]q'(p_t)$. The access charge, A_t, falls by q_{t-1}, and this reduces all future access charges by q_{t-1}, lowering the present value of income by $q_{t-1}/(1-\beta)$. Finally there is the direct effect on A_{t+1} given by the term in square brackets. This affects all future values of A, and each unit of this change has present value in period t of $\beta/(1-\beta)$ units.

Using (5.3) it is straightforward to show that in steady state price equals marginal cost. This is not surprising. We showed in Chapter 3 that TB in general leads to Ramsey pricing. When there is an access charge that acts like a transfer Ramsey pricing is the same as marginal cost pricing. In Chapter 2.5.3 we showed that the Finsinger-Vogelsang scheme (F-V) also generates efficient pricing. In fact TB and F-V are closely related. We shall show this by rewriting the objective function (5.2). In general the present value of a sequence $\{X_t\}_{t=1}^{\infty}$ can be expressed as:

$$\sum_{t=1}^{\infty}\beta^{t-1}X_t = \frac{X_0}{1-\beta} + \frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}\{X_t - X_{t-1}\}. \tag{5.4}$$

This follows because $X_t = X_0 + (X_1 - X_0) + (X_2 - X_1) + \ .\ .\ + (X_t - X_{t-1})$. Using (5.4) we can write the firm's objective function as:

$$Z = \frac{A_0 + \Pi_0}{1-\beta} + \frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}\{A_t - A_{t-1} + \Pi_t - \Pi_{t-1}\}.$$

This holds whatever the form of the price cap. The present value of the firm is the present value of its initial income $A_0 + \Pi_0$, plus the value of future changes in its income. The present value of the change in income in period j is $\beta^j/(1-\beta)$ times the change in income between j and $j-1$. This is equivalent to a perpetuity that starts paying out at date j. In the TB case we can substitute for $A_t - A_{t-1}$ from (5.1′), giving:

$$Z = \frac{A_0 + \Pi_0}{1-\beta} + \frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}\{(p_{t-1} - p_t)q_{t-1} + \Pi_t - \Pi_{t-1}\}. \tag{5.5}$$

Under F-V the present value of the firm's income is $\sum_{t=1}^{\infty}\beta^{t-1}\{(p_{t-1} - p_t)q_{t-1} + \Pi_t - \Pi_{t-1}\}$. Clearly the sequence $\{p_t\}$ that maximizes Z also maximizes the firm's present value under F-V. It follows that the first-order conditions under F-V (characterized by (2.24)) are the same as (5.3), and a little algebra shows this, even though our interpretations of them are quite different because the access charge and transfer are different in each case.

We now illustrate the TB scheme with an example with linear demand and quadratic costs. These functional forms yield linear first-order conditions and allow full characterization of the dynamics. To get some information on the dynamic behaviour of prices close to steady state we could have taken a linear approximation to (5.3) around the steady state (Stokey and Lucas, 1989, Chapter 6.4). But this involves considerable notation and the main insights can be achieved by assuming linearity. Demand is $q(p_t) = 1 - p_t$ and costs are $C_t = cq(p_t) + (\delta/2)q(p_t)^2$ plus a constant, where $c < 1$ and $\delta > -1$. These assumptions ensure that marginal cost pricing induces a welfare maximum and not a minimum, and that the welfare maximizing price, $p_* = (c+\delta)/(1+\delta)$, is unique and lies strictly between 0 and 1. Since we have normalized the slope of demand to be -1 the cost parameter δ is essentially the gradient of marginal cost relative to the slope of the demand curve.

Using these functional forms in the first-order condition (5.3) gives a second-order difference equation – an Euler equation – for prices:

$$\beta p_{t+1} - [2 + \delta(1-\beta)]p_t + p_{t-1} = -(c+\delta)(1-\beta). \tag{5.3$'$}$$

When p is constant (5.3$'$) implies $p = (c+\delta)/(1+\delta) = p_*$, confirming that marginal cost pricing occurs in the limit. The technique for solving Euler equations is discussed by Sargent (1987, pp 199-204). Equation (5.3$'$) has a general solution of the form $p_t = p_* + G(\chi_1)^t + H(\chi_2)^t$, where χ_1 and χ_2 are the two roots and G and H are constants to be determined by the boundary conditions of the problem. The roots are:

$$\chi_i = \frac{2 + \delta(1-\beta) \pm \sqrt{\left\{4(1+\delta)(1-\beta) + \delta^2(1-\beta)^2\right\}}}{2\beta} \qquad i = 1, 2.$$

Since $\beta < 1$ and $\delta > -1$ the term under the square root sign is positive, so both roots are real. We have the following proposition.

Proposition 5.1 *With linear demand and quadratic costs,* p_t *moves monotonically according to the price path given by* $p_t = p_* + (p_0 - p_*)(\chi_1)^t$*, where* χ_1 *satisfies the following properties:*

(i) $\frac{1}{2+\delta} < \chi_1 < 1$ *for* $\beta > 0$.

(ii) $\partial\chi_1/\partial\beta > 0$ *for* $\beta > 0$.

(iii) $\lim_{\beta \to 0} \chi_1 = \frac{1}{2+\delta}, \quad \lim_{\beta \to 1} \chi_1 = 1.$

(iv) $\partial\chi_1/\partial\delta < 0$.

Proof. See Appendix 5.1.

The two boundary conditions are that p_0 is given and the transversality condition. We postpone detailed discussion of transversality conditions until Chapter 7 when we consider the effects of demand growth. The main point to note is that if p_t follows an explosive time path the transversality condition is not satisfied. Thus we can rule out the root χ_2, which exceeds 1, by setting $H = 0$ in the expression for the general solution for p_t. This leaves us with the solution $p_t = p_* + (p_0 - p_*)(\chi_1)^t$. From now on we shall drop the subscript on χ_1.

Why does price not fall immediately to marginal cost? Figure 5.2 illustrates the reason. Suppose p_0 is above marginal cost and p_1 is set equal to p_*. The firm increases its income by $B + D$. If instead it chose price $\hat{p}$ slightly above p_* it would gain a first-order increase in income of E and make a second-order loss of D. This argument is even stronger when marginal cost is constant. In this case if $\dot{p}_1 = MC$ the firm earns no extra income above $A_0 + \Pi_0$, but any higher value of p_1 would give it higher income.

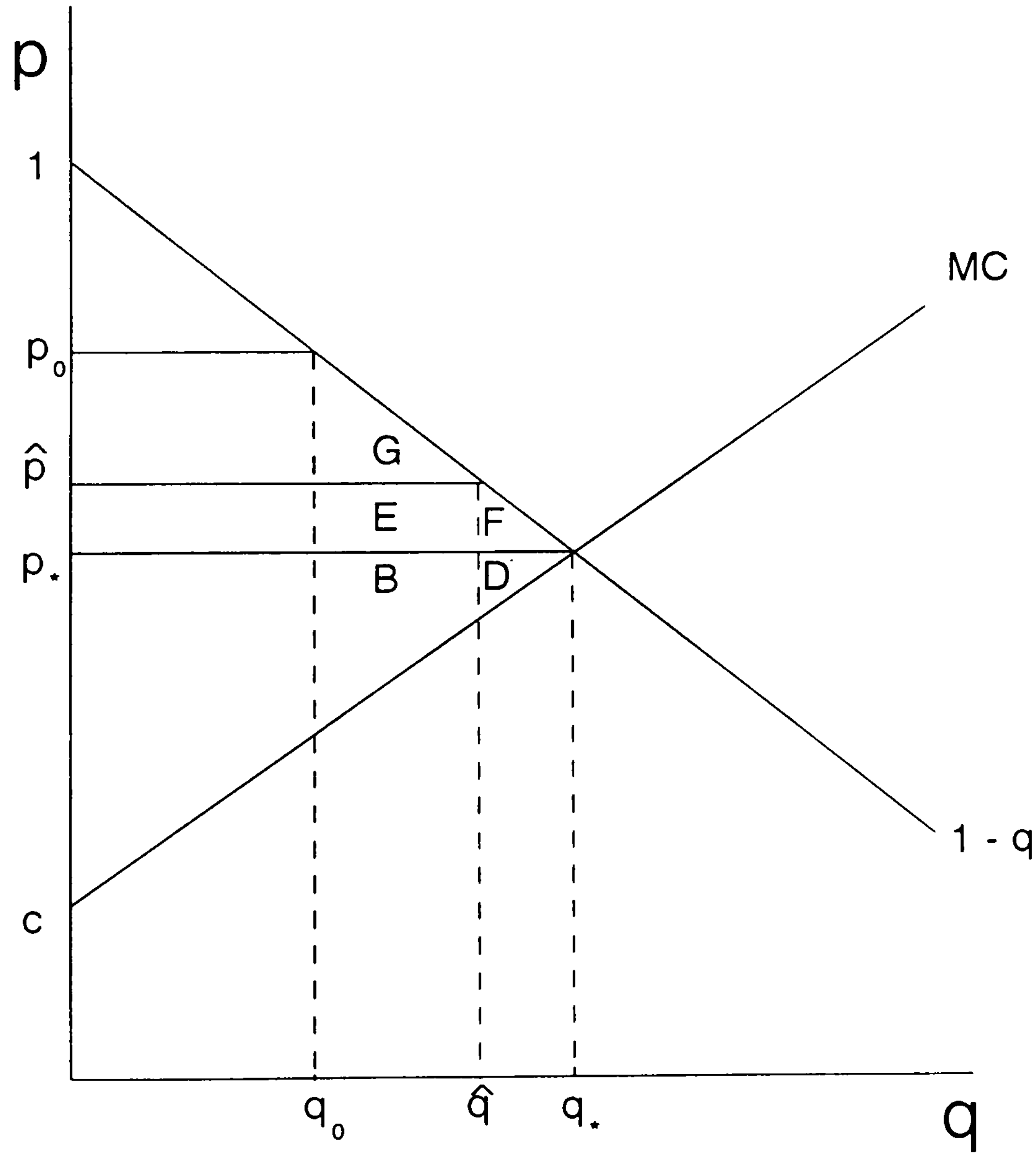


Figure 5.2 Convergence is not immediate under TB

Why does an increase in the discount factor slow down the rate of decline of prices? Each period the firm's income is less than the income it would earn from the L-M scheme by the change in net consumer surplus. A firm that values the future will reduce price slowly and will "give away" only a sequence of very small triangles to consumers. For β very close to 1 the triangles lost to consumers almost disappear. This confirms the conjecture we made in 3.2.1 that a far-sighted firm faced with TB will not be as generous to consumers as a myopic one. By abstaining from myopic profit-maximization the firm strategically shifts future constraints in its favour.

Using (5.5) we can find an expression for the present value of the firm's income at the start of period 0. This is

$$Z(\mathrm{TB}) = \frac{A_0 + \Pi(p_0)}{1-\beta} + \frac{(p_0 - p_*)^2}{(1-\beta)} \cdot \frac{(\chi + \delta)}{2}. \tag{5.6}$$

It is straightforward to show that $\chi + \delta > 0$, even when $\delta < 0$. This confirms that the firm is better off than when it keeps the access charge and price fixed at the period 0 values. If the firm was subject to the L-M scheme (modified so that the transfer is $V(p_t) - V(p_0)$) then it would earn its initial income for ever, plus the present value of the increment to social surplus caused by pricing at marginal cost from period 1 onwards. Its present value would be:

$$Z(\text{L-M}) = \frac{A_0 + \Pi(p_0)}{1-\beta} + \frac{(p_0 - p_*)^2}{(1-\beta)} \cdot \frac{(1 + \delta)}{2}$$

which is higher than Z(TB) because $\chi < 1$. However, the closer β is to 1 the closer χ is to 1, by Proposition 5.1, and the closer Z(TB) is to its supremum of Z(L-M).

We can also derive an explicit expression for the present value of consumer surplus under the assumption that consumers and the firm share the same discount factor. When demand is $q = 1 - p$, net consumer surplus is $CS \equiv (1-p)^2/2 - A$. The present value of the consumer surplus sequence is

$$\begin{aligned}
M(TB) &= \sum_{t=1}^{\infty} \beta^{t-1}\{V(p_t) - A_t\} \\
&= \frac{CS_0}{1-\beta} + \frac{1}{1-\beta}\sum_{t=1}^{\infty} \beta^{t-1}\{V(p_t) - A_t - V(p_{t-1}) + A_{t-1})\}.
\end{aligned}$$

From the TB constraint (5.1′) we have $V_t - A_t - V_{t-1} + A_{t-1} = (p_{t-1} - p_t)^2/2$, which is non-negative, confirming that the consumer surplus sequence is non-decreasing. Since $p_t = p_* + (p_0 - p_*)\chi^t$ it follows that $(p_{t-1} - p_t)^2 = (p_0 - p_*)^2\chi^{2t-2}(1-\chi)^2$ and

$$\begin{aligned}
M(TB) &= \frac{CS_0}{1-\beta} + \frac{(p_0 - p_*)^2}{2(1-\beta)}\sum_{t=1}^{\infty} \beta^{t-1}\chi^{2t-2}(1-\chi)^2 \\
&= \frac{CS_0}{1-\beta} + \frac{(p_0 - p_*)^2(1-\chi)^2}{2(1-\beta)(1-\beta\chi^2)} \qquad (5.7)
\end{aligned}$$

We note that $\beta\chi^2 < 1$ since both β and χ are below 1.

Having derived the present values of profits and consumer surplus we can find the present value of total welfare, assuming that the regulator shares the same discount factor as the firm and consumers. One way to do this is to add Z(TB) and M(TB). It is more helpful for subsequent analysis, however, to derive a more general expression. Welfare in period t is

$$W_t = \frac{(1-p_t)^2}{2} + (p_t - c)(1-p_t) - \frac{\delta}{2}(1-p_t)^2$$

since the access charge cancels out when there are equal weights in the welfare function. Some rearrangement gives:

$$W_t = \frac{(1-c)(1-p_*)}{2} - \frac{(1+\delta)(p_* - p_t)^2}{2}. \qquad (5.8)$$

By inspection W_t is maximized when $p_t = p_*$. The expression (5.8) is illustrated in Figure 5.2 for $p_t = p_0$. The first part is the maximum feasible value of W_t, which is the triangle with height $1-c$ and length $q_* = 1-p_*$. The second part is the loss of welfare from pricing at p_0 rather than p_*, and is the triangle with area $B+E+G+D+F$. The further p_t is from p_* (in either direction) the lower welfare is. Under TB we know from Proposition 5.1 that $p_t - p_* = (p_0 - p_*)\chi^t$, so the present value of welfare is:

$$W(TB) = \frac{(1-c)(1-p_*)}{2(1-\beta)} - \frac{(1+\delta)(p_0-p_*)^2}{2} \cdot \frac{\chi^2}{1-\beta\chi^2} \tag{5.9}$$

Equation (5.9) will be used later to compare welfare under different price caps.

We have seen how the attractive properties of TB shown in Chapter 3 hold in the two-part tariff case. Consumers benefit each period, the firm is better off than with fixed prices, and pricing is efficient in the long run. However a far-sighted firm can manipulate future constraints, which slows down the approach of the product price to marginal cost and transfers more of the available surplus to the firm. Under the fixed weights scheme the firm cannot manipulate future constraints, and it is to this that we now turn.

5.2.2 Fixed Weights Regulation

The fixed weights case has been analyzed briefly by Sappington and Sibley (1992). We focus here on the linear example and on the comparison of F with TB, which Sappington and Sibley do not consider. The general constraint is

$$(A_0 - A_t)N_0 + \sum_{i=1}^{n} (p_0^i - p_t^i)q_0^i \geq 0$$

which reduces to

$$A_0 - A_t + (p_0 - p_t)q(p_0) \geq 0$$

when the number of customers is fixed and there is only one product. We assume that the constraint binds. Substituting in for A_t from the binding constraint, income is $A_0 + (p_0 - p_t)q(p_0) + p_t q(p_t) - C(q(p_t))$, and the first-order condition is:

$$q(p_t) + [p_t - MC_t]q'(p_t) - q(p_0) = 0. \tag{5.10}$$

As long as operating profits are strictly concave in p_t the solution to (5.10) is the unique profit-maximizing value of p_t. As in Chapter 3, the firm's problem is static, so p_t and A_t are constant from period 1 onwards. The value of A_0 has no effect on the pricing decision. The price under F is the same as the first period price that would be chosen under TB if the firm was myopic.

We shall focus on the linear case. The first-order condition (5.10) yields:

$$p_t = \frac{c + \delta + p_0}{2 + \delta} = p_* + \frac{(p_0 - p_*)}{2 + \delta}.$$

The coefficient of $(p_0 - p_*)$ is the same as the value of χ when the firm facing TB is myopic. The present value of the firm's income stream is

$$Z(F) = \frac{A_1 + \Pi(p_1)}{1 - \beta} = \frac{A_0 + \Pi(p_0)}{1 - \beta} + \frac{(p_0 - p_*)^2}{2(1 - \beta)}\left(\delta + \frac{1}{2 + \delta}\right). \tag{5.11}$$

In Proposition 3.2 we showed that the firm prefers TB to F, and this is confirmed by inspection of (5.11) and (5.6). By Proposition 5.1, $\chi \geq 1/(2 + \delta)$ and we have $Z(TB) \geq Z(F)$, with strict inequality when $\beta > 0$.

In 3.2.2 we stated that consumers' preferences between TB and F are likely to depend on the discount factor. Under F the product price drops further in period 1 than it does under TB, giving the consumer a larger net gain in surplus in that period. Thereafter the firm keeps the price constant under F, but gradually lowers it under TB towards p_*, so there are subsequent gains to consumers under TB but not under F. The present value of consumer surplus under F is

$$M(F) = \frac{CS_0}{1-\beta} + \frac{(p_0 - p_1)^2}{2(1-\beta)} = \frac{CS_0}{1-\beta} + \frac{(p_0 - p_*)^2(1+\delta)^2}{2(1-\beta)(2+\delta)^2}$$

and so, subtracting (5.7),

$$M(F) - M(TB) \overset{\text{sgn}}{=} \frac{(1+\delta)^2}{(2+\delta)^2} - \frac{(1-\chi)^2}{1-\beta\chi^2}.$$

If marginal cost is constant, i.e. $\delta = 0$, consumers prefer F whatever the value of β. The proof is by contradiction. Let $M(F) - M(TB) < 0$. Then $1/4 - (1-\chi)^2/(1-\beta\chi^2) < 0$. When $\delta = 0$ the characteristic equation obtained from (5.3′) is $\beta\chi^2 - 2\chi + 1 = 0$, so $1 - \beta\chi^2 = 2(1-\chi)$. The condition then becomes $1/4 - (1-\chi)/2 < 0$, which is equivalent to $\chi < 0.5$. This cannot hold since χ is bounded below by 0.5 when $\delta = 0$. More generally numerical analysis shows that $M(F) - M(TB)$ is positive for all positive values of δ. Consumers can prefer TB to F, but only when δ is implausibly low – for example when $\beta = 0.95$, TB gives higher consumer surplus when $\delta < -0.78$.

When F is imposed, welfare is above the period-0 level because both consumers and the firm gain. By substituting $p_t = p_* + (p_0 - p_*)/(2+\delta)$ into the expression for period t welfare (5.8) we can calculate the present value of welfare:

$$W(F) = \frac{(1-c)(1-p_*)}{2(1-\beta)} - \frac{(1+\delta)(p_0 - p_*)^2}{2(2+\delta)^2(1-\beta)}. \tag{5.12}$$

The difference between W(TB) and W(F) is

$$W(TB) - W(F) = \frac{(1+\delta)(p_0 - p_*)^2}{2}\left\{\frac{1}{(2+\delta)^2(1-\beta)} - \frac{\chi^2}{1-\beta\chi^2}\right\}. \tag{5.13}$$

We need to find the sign of the expression in curly brackets in (5.13). TB generates higher welfare when $\delta = 0$, despite the fact that consumers prefer F in this case. In this case the expression to be signed is: $1/4(1-\beta) - \chi^2/(1-\beta\chi^2)$. Using the characteristic equation $\beta\chi^2 - 2\chi + 1 = 0$, some rearrangement shows that this has the same sign as

$-2\chi^2 + 3\chi - 1 = (2\chi - 1)(1 - \chi)$, which is positive since $0.5 \leq \chi < 1$.

More generally we can discover numerically under what circumstances one scheme yields higher welfare than the other. The sign of the welfare difference in (5.13) is a function of β and δ only, as χ is itself a function of these two parameters. In Figure 5.3 we illustrate in (β, δ) space the curve that gives equal welfare under the two schemes. Above this curve we have W(F) > W(TB). Intuitively, as δ increases the deviation of the price under F from p_* falls by more than the deviation of the initial TB price from p_*, and welfare is higher under F than TB. But as β increases the fact that under F price does not alter from its period 1 level, while TB generates continuous welfare increases, counts against F. This is despite the fact that a higher value of β slows down the rate of convergence under TB. Higher values of δ are needed when β increases to keep W(F) = W(TB). Thus F is preferred to TB when δ is high and β is low, but TB is preferred when marginal cost is relatively flat, or when the firm values future income highly.

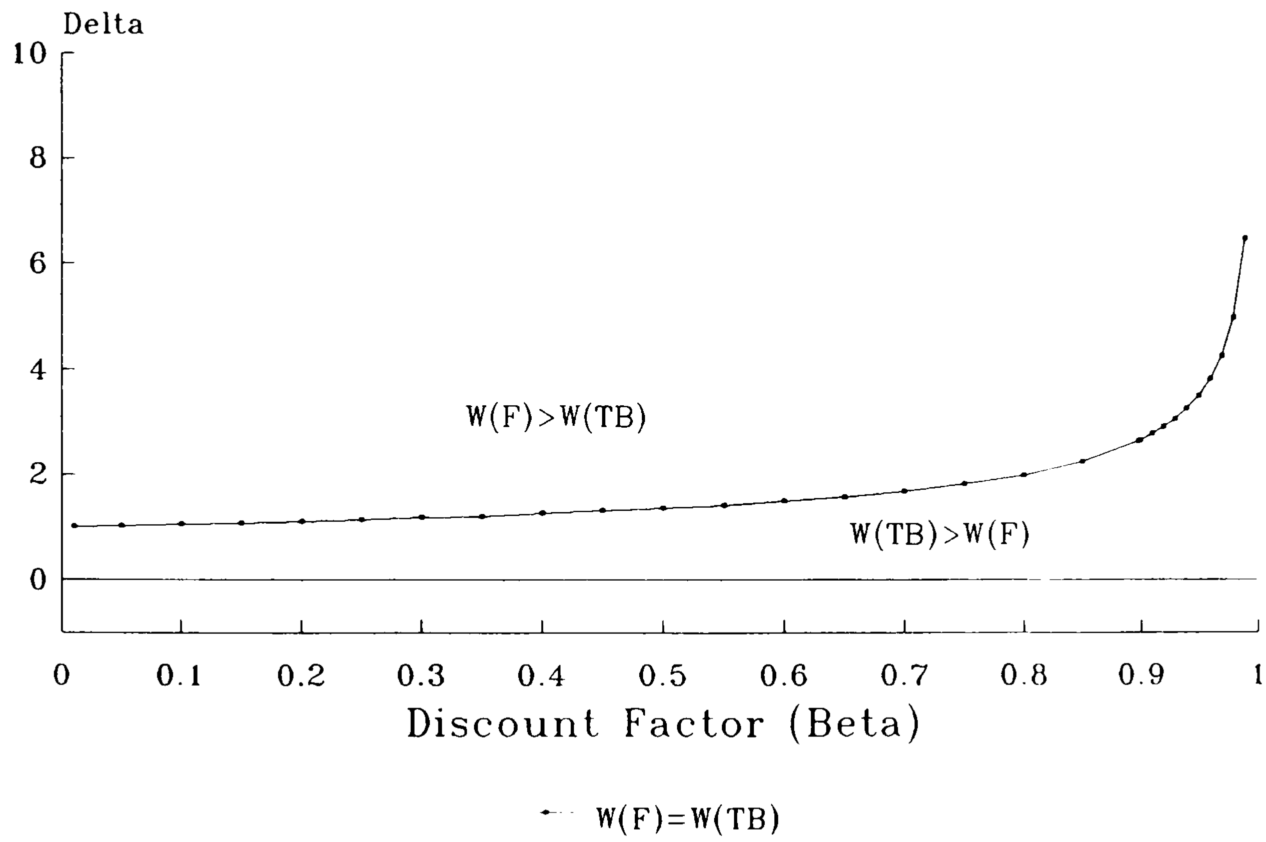


Figure 5.3 Welfare under TB and F

5.2.3 Average Revenue Regulation

We now turn to average revenue regulation. Like fixed weights regulation, AR is a static scheme. The general constraint is

$$\frac{A_t N_t + \sum_{i=1}^{n} p_t^i q_t^i}{\sum_{i=1}^{n} q_t^i} \leq \bar{p} \tag{5.14}$$

where $\bar{p} = (A_0 N_0 + \sum_{i=1}^{n} p_0^i q_0^i)/\sum_{i=1}^{n} q_0^i$. We note that the access charge and the product prices do not enter symmetrically into the constraint. Output, $\sum_{i=1}^{n} q_t^i$, and the number of customers, N_t, cannot be meaningfully added together, and it is the volume of sales that is used as the denominator of (5.14).

For our simplified case the constraint (5.14) is $A_1 \leq (\bar{p} - p_1)q(p_1)$, where prices under AR are p_1 and A_1. Income is $A_1 + p_1 q(p_1) - C(q(p_1))$. With a binding constraint we can substitute for A_1, so income is $\bar{p}q(p_1) - C(q(p_1))$. We need to think carefully about the form of the cost function. If it is $C(q) = cq + K$, then profits are $(\bar{p} - c)q(p_1) - K$. Since it earns a constant margin of $\bar{p} - c$ on extra sales, the firm wants to expand its output indefinitely. This is not feasible, and there must be some features of the firm's problem that prevent it from expanding output too far. We consider four constraints on the problem. First, marginal cost might be increasing rather than constant. Second, there might be a capacity constraint. Third, the maximum feasible output level might be determined by the consumer's participation constraint. Finally there might be a non-negativity restriction on the product price.

We focus on the case where marginal cost is strictly increasing, before discussing the other constraints. Given a downward-sloping demand function that crosses the marginal cost curve, there is a unique price, p_*, where $p_* = MC(q(p_*))$. The firm's problem is to choose p_1 to maximize $\bar{p}q(p_1) - C(q(p_1))$, and the first-order condition is

$$\bar{p} = MC(q(p_1)). \tag{5.15}$$

The second-order condition requires that marginal cost is increasing, which we have assumed. The firm sets output where marginal revenue, $\bar{p}$, equals marginal cost. When $\bar{p} > p_*$ we have $MC(q(p_1)) = \bar{p} > p_* = MC(q(p_*))$, so $q(p_1) > q(p_*)$, i.e. output is above the first-best level. The firm cuts the marginal price to increase the access charge.

The effect on consumers of AR is found by using the standard convexity result and applying the constraint. Convexity implies $V(p_0) \geq V(p_1) + (p_1 - p_0)q(p_1)$. Subtracting A_0 from both sides, and adding and subtracting A_1 from the right-hand side gives $V(p_0) - A_0 \geq V(p_1) - A_1 + A_1 + (p_1 - p_0)q(p_1) - A_0$. If the constraint binds then $A_1 + p_1 q(p_1) = \bar{p}q(p_1)$, so

$$CS_0 - CS_1 \;\geq\; (\bar{p} - p_0)q(p_1) - A_0 = A_0\left(\frac{q(p_1) - q(p_0)}{q(p_0)}\right) \tag{5.16}$$

where $CS_i \equiv V(p_i) - A_i$ is net consumer surplus and we have used $\bar{p} = A_0/q_0 + p_0$. Thus one sufficient condition for consumer surplus to be lower under AR than at the initial prices is that $A_0 = 0$. Another sufficient condition is that $p_1 < p_0$ when $A_0 > 0$, and a sufficient condition for this is that $\bar{p} > p_0 > p_*$.

Thus consumers generally lose from the introduction of AR regulation, while the firm makes higher profits than in period 0. As in the linear pricing case of Chapters 3 and 4 the effect on total welfare of the imposition of AR regulation is ambiguous. If $p_0 > p_*$, as we generally assume, then $q(p_0) < q(p_*)$, but AR induces $q(p_1) > q(p_*)$, so output moves from being too low to an excessive level. If the marginal cost function is steep relative to demand then a welfare increase is likely, while if marginal cost is relatively flat a loss in welfare is probable, and we shall show this in the linear case later in this section.

The effect of a relaxation in the price cap is to increase the incentive for the firm to raise output, so welfare falls as $\bar{p}$ rises. Formally, the effect on welfare of a increase in $\bar{p}$ is $\frac{dW}{d\bar{p}} = [p_1 - MC(q(p_1))]q'(p_1)\frac{dp_1}{d\bar{p}} < 0$ because $\frac{dp_1}{d\bar{p}} = 1/[MC'(q(p_1))q'(p_1)] < 0$ and price is below marginal cost. Thus, as long as $\bar{p} \geq p_*$, welfare rises monotonically as the cap becomes tighter. The following proposition is used later.

Proposition 5.2 Inefficiency of pricing under AR. *The firm sets* p_1 *equal to* p_* *if and only* $\bar{p} = p_*$.

Proof. First, we prove sufficiency. Substituting $\bar{p} = p_*$ into the first-order condition (5.15) gives $p_* = MC(q(p))$. But since $p_* = MC(q(p_*))$, and this is the unique marginal cost price, we must have $p_1 = p_*$. Second, we show necessity. Suppose $p_1 = p_*$. The first-order condition (5.15) implies that $\bar{p} = MC(q(p_1)) = MC(q(p_*)) = p_*$. □

We return to the increasing marginal cost case later, but first we discuss the other constraints on the firm's ability or desire to increase output indefinitely. The limiting form of increasing marginal costs is when there is a capacity constraint. Effectively marginal cost is infinite when demand equals capacity. We revert to the assumption that below the capacity level, denoted by K, marginal cost is constant and equal to c, and also assume $\bar{p} > c$. The Lagrangian for the firm's problem is

$$A_1 + (p_1 - c)q(p_1) + \lambda[(\bar{p} - p_1)q(p_1) - A_1] + \xi[K - q(p_1)]$$

where λ is the multiplier on the AR constraint and ξ is the multiplier on the constraint that capacity cannot be exceeded. We assume that the price at which demand equals K is greater than zero, and that at this price all consumers still want to purchase (i.e. neither the participation constraint nor the non-negativity constraint binds). The Kuhn-Tucker conditions for A_1, p_1, λ and ξ are:

$$1 - \lambda \leq 0, \text{ and } A_1 \geq 0 \tag{5.17a}$$

$$(1-\lambda)q(p_1) + [(1-\lambda)p_1 - c + \lambda\bar{p} - \xi]q'(p_1) \leq 0, \text{ and } p_1 \geq 0 \tag{5.17b}$$

$$(\bar{p} - p_1)q(p_1) - A_1 \geq 0, \text{ and } \lambda \geq 0 \tag{5.17c}$$

$$K - q(p_1) \geq 0, \text{ and } \xi \geq 0 \tag{5.17d}$$

with complementary slackness in each case. The interesting case is where A_1 and p_1 are positive. When $A_1 > 0$ we have $\lambda = 1$ by complementary slackness in (5.17a), so the AR constraint, (5.17c), binds. Substituting $\lambda = 1$ into (5.17b) implies $\bar{p} = c + \xi$. Since $\bar{p} > c$ by assumption, this implies that $\xi > 0$ and the capacity constraint binds. The multiplier ξ represents the shadow cost to the firm of the capacity constraint. If it had an extra unit of capital it could sell another unit of output and would earn a rent of $\xi = \bar{p} - c$ on this.

When demand intersects capacity at a price above c AR generates the socially optimal outcome, *given the capacity constraint.* Output equals capacity, and p_1 is equal to short-run marginal social cost. But if demand intersects capacity at a price that is below c, price is *not* equal to short-run marginal social cost. In social terms the firm has excess capacity, but AR regulation encourages the firm to use all of this capacity. The socially optimal price in this case is c.

In both cases the firm might want to adjust its capacity in the long run. The private incentive to increase K is equal to the marginal benefit, $\bar{p} - c$, less the marginal cost of capacity. If the marginal cost of capacity was constant and below $\bar{p} - c$ then the firm would want to expand capacity indefinitely (subject to the other constraints on output maximization), so we must assume that marginal capacity costs are increasing to obtain an interior solution. In equilibrium the firm will choose K where the marginal capacity cost just equals $\bar{p} - c$. Usually the chosen capacity will be excessive. At the socially optimal capacity, price is equal to c plus marginal capacity cost, and demand is equal to capacity. Call this price p_*. But under AR the firm equates the sum of c and marginal capacity cost to $\bar{p}$, not p_*. If $\bar{p} > p_*$ then the firm has an incentive to build excessive capacity. This is essentially the same argument we had for excess output in the increasing marginal costs case. Only if $\bar{p} = p_*$, as in Proposition 5.2, will the incentive to build capacity be correct.

The other constraint on the ability to expand output that we shall consider is the participation constraint.[2] If customers are of different types then it is usually profitable to exclude some low-demand customers (see Chapter 7). When consumers are identical the participation constraint is $V(p_1) - A_1 \geq 0$. Adding this to the standard AR problem, with constant marginal cost for simplicity, gives the Lagrangian:

$$A_1 + (p_1 - c)q(p_1) + \lambda[(\bar{p} - p_1)q(p_1) - A_1] + \theta[V(p_1) - A_1]$$

where θ is the multiplier on the participation constraint. When A_1 and p_1 are positive and both constraints bind the first-order conditions for A_1, p_1, λ and θ are

$$1 - \lambda - \theta = 0 \tag{5.18a}$$

$$(1 - \lambda - \theta)q(p_1) + [(1 - \lambda)p_1 - c + \lambda\bar{p}]q'(p_1) = 0 \tag{5.18b}$$

$$(\bar{p} - p_1)q(p_1) - A_1 = 0 \tag{5.18c}$$

$$V(p_1) = A_1. \tag{5.18d}$$

The AR constraint binds as long as $\bar{p}$ is below the maximum average revenue that would be earned without regulation, which is $c + V(c)/q(c)$. It is clear that the consumer's participation constraint binds. As the firm expands output to earn rents of $\bar{p} - c > 0$ on each unit it pushes up the access charge as much as the constraint allows, and lowers consumer surplus until it equals zero. Putting (5.18a) into (5.18b) gives:

$$p_1 = c - \frac{\lambda}{1 - \lambda}(\bar{p} - c). \tag{5.19}$$

Equation (5.18a) and the fact that $\theta > 0$ implies that $\lambda < 1$, so equation (5.19) implies that $p_1 < c$. Output is inefficiently high. This can be compared with the excess capitalization result in the Averch-Johnson (1962) model.[3] Under rate-of-return

[2]The fourth constraint, that $p_1 \geq 0$, is simple to analyze. If it binds then the firm sets $p_1 = 0$, demand is $q(0)$, which we assume to be finite, and all of the firm's revenue comes from the access charge, which equals $\bar{p}q(0)$.

[3]Equation (5.19) has the same superficial form as the standard Averch-Johnson equation. This is $\partial R/\partial K = r - \lambda(s - r)/(1 - \lambda)$, where $\partial R/\partial K$ is marginal revenue from extra capital, r is the true cost of capital, s is the allowed rate of return where $s > r$, and λ is the multiplier on the rate-of-return constraint. Sherman and Visscher (1982) analyze rate-of-return regulation with a two-part tariff.

regulation of the Averch-Johnson type the firm is allowed to earn a constant return per unit of capital, which gives it an incentive to maximize capital. Under a price cap based on average revenue with a constant marginal cost the firm earns a fixed margin on each unit of output, and so has an incentive to maximize sales.

Comparison of AR with TB and F.

We now consider how AR performs relative to TB and F when demand is linear and costs are quadratic. The following simplifying assumptions will be made. First, marginal cost is strictly increasing but not vertical ($\delta > 0$). This implies that the second-order sufficient condition for AR holds. Second, we assume that $(2\delta + c)/(2\delta + 1) > \bar{p}$, which ensures that the consumer's participation constraint does not bind at the equilibrium. These two assumptions imply that (5.15) determines p_1. Substituting into (5.15) from $q = 1 - p$ and $C = c(1-p)^2 + \delta(1-p)^2/2$ gives $\bar{p} = c + \delta(1 - p_1)$, so

$$p_1 = p_* - \frac{(\bar{p} - p_*)}{\delta} \tag{5.20}$$

where we have used the fact that $p_* = (c + \delta)/(1 + \delta)$. As expected p_1 is below p_* whenever $\bar{p} > p_*$. The participation constraint is $V(p_1) - A_1 \geq 0$. We have

$$V(p_1) - A_1 = \frac{(1-p_1)^2}{2} - (\bar{p} - p_1)(1 - p_1) \overset{\text{sgn}}{=} 1 + p_1 - 2\bar{p} \overset{\text{sgn}}{=} 2\delta + c - (2\delta + 1)\bar{p},$$

so this is positive given our second assumption. The third assumption we make is that $A_0 = 0$, so that $\bar{p} = p_0$. We could allow for an initial access charge that is positive, but this would make the comparison of AR with TB and F depend on A_0 as well as β and δ. With this assumption we make it more likely that AR is preferable to the other schemes as an increase in A_0 raises $\bar{p}$, which lowers welfare by increasing output further beyond $q(p_*)$. The welfare properties of TB and F, however, do not depend on A_0.

The firm does better with AR than either TB or F. The present value of profits can be shown to be:

$$Z(\mathrm{AR}) \quad = \frac{\Pi(p_0)}{1-\beta} + \frac{(p_0 - p_*)^2}{(1-\beta)} \cdot \frac{(1+\delta)}{2} + \frac{(p_0 - p_*)^2}{(1-\beta)} \cdot \frac{(1+\delta)}{2\delta} \tag{5.21}$$

Note that the first two terms are the same as Z(L-M). Since $\delta > -1$ the last term is positive, and the firm does better under AR than under the modified L-M scheme. By transitivity the firm prefers AR to TB and F. We illustrate (5.21) in Figure 5.4. The firm earns profits per period of the triangle p_0Ec. The access charge is p_0EHp_1. Apart from the factor $1/(1-\beta)$ the three terms in (5.21) are represented by p_0DGc ($\Pi(p_0)$), DFG, and DEF, so Z(AR) exceeds Z(L-M) by DEF.

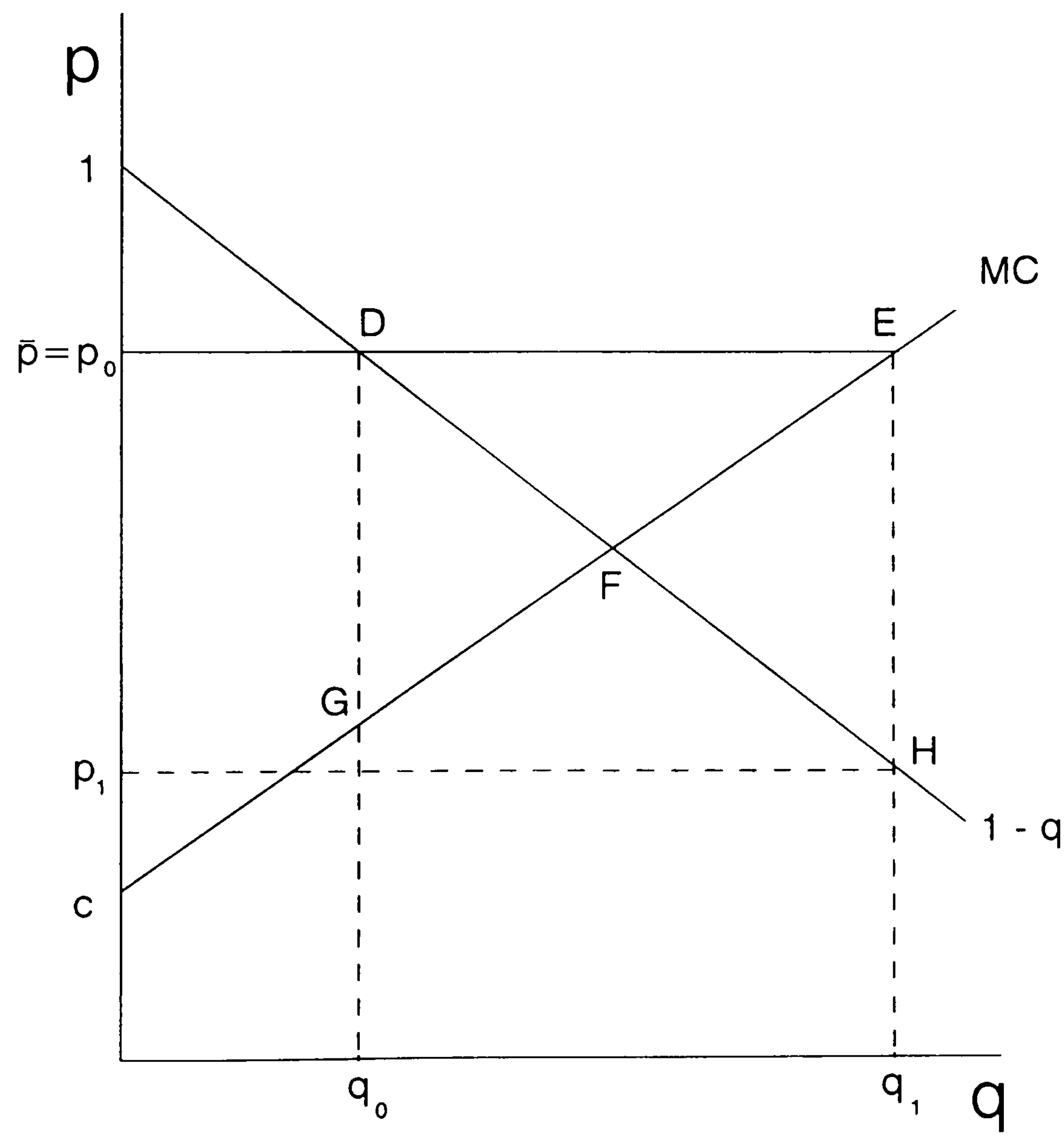


Figure 5.4 Average Revenue Regulation

We already know from the general analysis that consumers are worse off under AR than in period 0, and thus are worse off than under TB or F. The expression for the present value of consumer surplus is:

$$M(AR) = \frac{CS_0}{1-\beta} - \frac{(p_0 - p_1)^2}{2(1-\beta)} = \frac{CS_0}{1-\beta} - \frac{(p_0 - p_*)^2(1+\delta)^2}{2(1-\beta)\delta^2}$$

In Figure 5.4 consumer surplus falls by the area DEH relative to surplus at p_0.

The present value of welfare under AR is found by substituting into (5.8):

$$W(AR) = \frac{(1-c)(1-p_*)}{2(1-\beta)} - \frac{(1+\delta)(p_0 - p_*)^2}{2(1-\beta)\delta^2}. \tag{5.22}$$

We shall make three welfare comparisons. First, we see under what conditions AR generates an increase in welfare above the period 0 level. The difference in welfare is:

$$W(AR) - W(p_0) = \frac{(1+\delta)(p_0 - p_*)^2}{2(1-\beta)}\left\{1 - \frac{1}{\delta^2}\right\}.$$

This has the same sign as $(\delta - 1)$. With a capacity constraint AR can be fully efficient, so it is intuitive that the sign of the welfare change depends positively on δ. If $\delta = 1$ then the two triangles DFG and EHF in Figure 5.4 are of equal size and the welfare change is zero. For higher values of δ the triangle EHF becomes smaller than DFG.

The second comparison is with W(F) given in (5.12). The difference is

$$W(F) - W(AR) = \frac{(1+\delta)(p_0 - p_*)^2}{2(1-\beta)}\left\{\frac{1}{\delta^2} - \frac{1}{(2+\delta)^2}\right\}.$$

This expression has the same sign as the term in curly brackets. As this is strictly positive we conclude that F dominates AR. Since F always gives a price that is closer to p_* than under AR, it generates higher welfare. Of course if $A_0 > 0$ then the welfare performance of AR would be even worse, without affecting W(F).

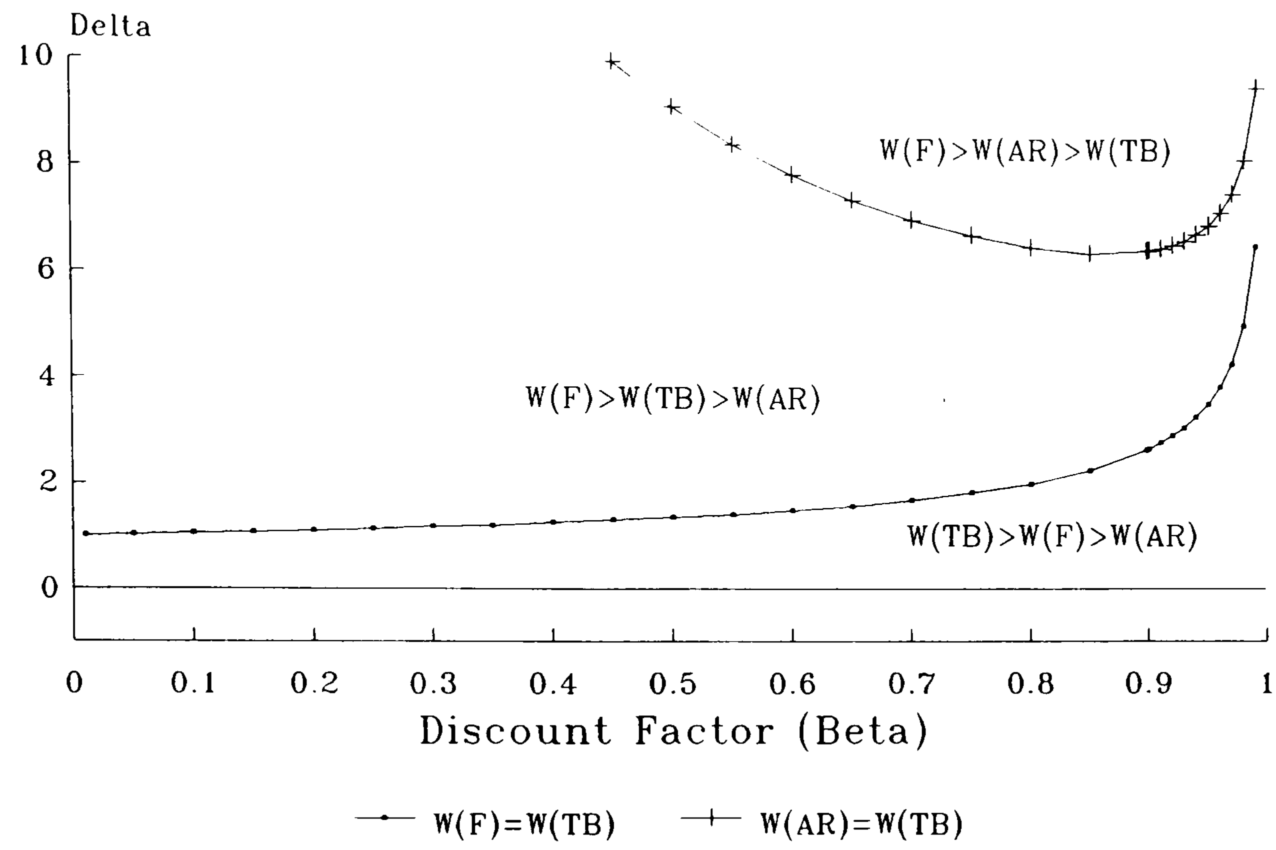


Figure 5.5 Welfare under TB, F and AR

Although AR is dominated by F, it is not dominated by TB. The difference in the present values of welfare is:

$$\mathrm{W(TB)} - \mathrm{W(AR)} = \frac{(1+\delta)(\mathrm{p}_0 - \mathrm{p}_*)^2}{2}\left\{\frac{1}{\delta^2(1-\beta)} - \frac{\chi^2}{1-\beta\chi^2}\right\}$$

The sign of the term in curly brackets depends on β and δ. In Figure 5.5 we illustrate in (β, δ) space the curve that gives W(TB) = W(AR). We also repeat the curve from Figure 5.3 showing equal welfare under TB and F. The AR scheme is more likely to give higher welfare the higher δ is, so above the W(TB) = W(AR) curve AR is preferred, while below it TB is preferred. The relationship is parabolic, with TB tending to dominate for both high and low values of β. The value of δ needs to be quite high for AR to be preferable to TB. This requires $\delta > 6$, so marginal cost has to have at least six times the slope of the demand curve, which seems implausible except in the case of a capacity constraint. But in all cases where AR is preferred to TB, F is even better.

5.2.4 Average Revenue (Lagged) Regulation

Chapter 3 suggested that ARL was a hybrid of TB and AR. The firm faces the constraint that average revenue calculated using lagged output should not exceed $\bar{p}$, where $\bar{p}$ is actual average revenue in period 0:

$$\frac{A_t N_{t-1} + \sum_{i=1}^{n} p_t^i q_{t-1}^i}{\sum_{i=1}^{n} q_{t-1}^i} \leq \bar{p} \quad \text{or}$$

$$A_t + (\bar{p} - p_t)q(p_{t-1}) \geq 0 \tag{5.23}$$

in our simplified case. As in the AR case we assume that the consumer's participation constraint never binds, so consumer surplus is always strictly positive. This implies that the ARL constraint (5.23) always binds. Substituting for A_t from the binding constraint (5.23) into the profit function gives the following problem for the firm:

$$\underset{\{p_t\}}{\text{Max}} \sum_{t=1}^{\infty} \beta^{t-1}\Big[(\bar{p} - p_t)q(p_{t-1}) + p_t q(p_t) - C(q(p_t))\Big].$$

The Euler equation is:

$$- q_{t-1} + q_t + (p_t - MC_t)q'(p_t) + \beta(\bar{p} - p_{t+1})q'(p_t) = 0. \tag{5.24}$$

A small rise in p_t cuts A_t by q_{t-1}, changes period t operating profits and, by cutting q_t, reduces next period's access charge by $(\bar{p} - p_t)q'(p_t)$. Note that (5.24) is the same as (5.3) for TB when $\beta = 0$. If a steady state is reached then the price is characterized by:

$$p = MC(q(p)) - \frac{\beta[\bar{p} - MC(q(p))]}{1 - \beta},$$

Thus if $\bar{p} > MC$ the steady state price is below marginal cost, as in the AR case. In steady state actual average revenue equals $\bar{p}$, giving an access charge of $A = (\bar{p} - p)q(p)$.

To say more about ARL we use the linear-quadratic example. In the cases of TB and F we did not need to make any assumptions about β and δ beyond the standard ones that $\beta \in [0,\ 1)$ and $\delta > -1$. For AR to work, however, we had to assume $\delta > 0$. We make a less restrictive assumption for ARL:

Assumption 1. $\delta > \beta - 1$.

The purpose of this will become clear later. To ensure that the participation constraint is always satisfied we assume also that

$$\frac{1-\beta+2\delta+c}{2-\beta+2\delta} > p_0.$$

Again we note that this is less restrictive than the parallel assumption for AR that $(2\delta + c)/(2\delta + 1) > p_0$. We assume that $\bar{p} = p_0$.

The Euler condition (5.24) for the firm's problem becomes

$$\beta p_{t+1} - (2+\delta)p_t + p_{t\text{-}1} = -[c + \delta - \beta p_0]. \tag{5.24$'$}$$

Using the fact that $c + \delta = (1+\delta)p_*$, the steady-state solution, p_{ARL}, to (5.24′) is

$$p_{ARL} = p_* - \frac{\beta(p_0 - p_*)}{1+\delta-\beta}. \tag{5.25}$$

It can be checked that consumer surplus is positive at p_{ARL} given our assumption about the upper bound of p_0. Under Assumption 1 the denominator of the second term in (5.25) is positive, and the steady-state price is below the optimal price p_* when $p_0 > p_*$. The following proposition shows its relationship to the price chosen under AR, which we denote by p_{AR} here.

Proposition 5.3 *When* $p_0 > p_*$ *price in steady state under ARL lies between the price* *chosen under AR and* p_*.

Proof. The difference between p_{ARL} and p_{AR} is:

$$p_{ARL} - p_{AR} = (p_0 - p_*)\left\{\frac{1}{\delta} - \frac{\beta}{1+\delta-\beta}\right\} = (p_0 - p_*)\frac{(1-\beta)(1+\delta)}{\delta(1+\delta-\beta)}.$$

This has the same sign as $p_0 - p_*$, which is positive by assumption. Since $p_{ARL} < p_*$ it follows that the steady-state price under ARL is closer to p_* than under AR. □

The full solution to (5.24′) is of the form $p_t = p_{ARL} + (p_0 - p_{ARL})\mu^t$ where μ is:

$$\mu = \frac{2+\delta-\sqrt{4(1+\delta-\beta)+\delta^2}}{2\beta}.$$

We have the following proposition.

Proposition 5.4 *With linear demand and quadratic costs, and under Assumption 1, price moves monotonically to* p_{ARL}. *The price path is given by* $p_t = p_* + (p_0 - p_*)\mu^t$, *where* μ *satisfies the following properties:*

(i) $\frac{1}{2+\delta} < \mu < 1$ *for* $\beta > 0$.

(ii) $\partial\mu/\partial\beta > 0$ *for* $\beta > 0$.

(iii) $\lim_{\beta \to 0} \mu = \frac{1}{2+\delta}$, $\lim_{\beta \to 1} \mu = 1 + \frac{1}{2}\{\delta - \sqrt{4\delta+\delta^2}\}$.

(iv) $\partial\mu/\partial\delta < 0$.

Proof. See Appendix 5.2.

Assumption 1 is sufficient for μ to be real. It is also necessary and sufficient for μ to be less than 1. We should note the similarity between Propositions 5.1 and 5.4. The coefficients of adjustment, χ and μ, under TB and ARL are equal if $\delta = 0$ or if $\beta = 0$. In Appendix 5.3 we show that $(\chi - \mu)$ has the same sign as δ.

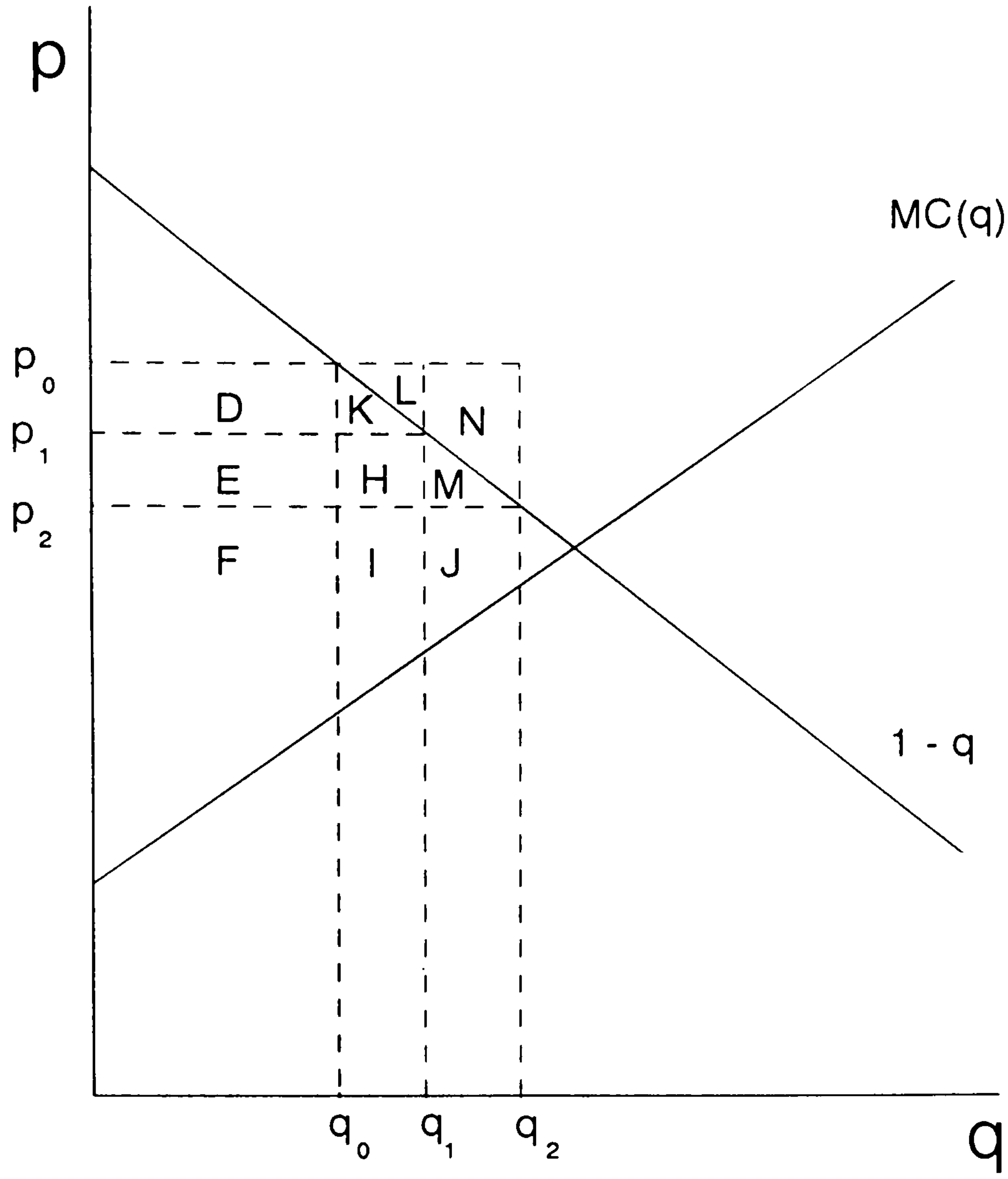


Figure 5.6 Average Revenue (Lagged) Regulation

The time paths of prices and incomes are illustrated in Figure 5.6. Period 0 profits are represented by $D + E + F$. In period 1 the firm's income increase by $H + I$. The access charge rise by D while the change in operating profits is $H + I - D$. In period 2 income rises by $K + L + J$. The access charge in period 2 is higher than A_1 by $E + H + K + L$, and operating profits change by $J - (E + H)$. Note that if the same price path had been chosen under TB, period 1 income would be the same but period 2 income would be lower by $K + L = 2K$. If the firm was regulated by AR but chose the same sequence for the product price as under ARL then its period 1 access charge would be higher by $K + L$, and its period 2 access charge would be higher than the ARL level by $M + N$.

Figure 5.6 suggests that the firm is better off with ARL than with TB or F, but that AR is even more profitable. This is formalized in the following proposition.

Proposition 5.5 The firm's ranking of AR, ARL, TB and F.

The firm prefers AR to ARL to TB to F when $\bar{p} = p_0 > p_*$.

Proof. We show that the present value of the firm under ARL is bounded below by Z(TB) and above by Z(AR). We look first at the lower bound. Suppose the firm is regulated by ARL, but it chooses the TB sequence $\{p_t\}$, allowing the access charge to be set by the ARL formula. The access charge under TB is

$$A_t^{TB} = \sum_{i=1}^{t} (p_{i-1} - p_i) q(p_{i-1}),$$

while that under ARL is

$$\begin{aligned} A_t^{ARL} &= (p_0 - p_t) q(p_{t-1}) = (p_0 - p_1 + p_1 - p_2 + \dots + p_{t-1} - p_t) q(p_{t-1}) \\ &= q(p_{t-1}) \sum_{i=1}^{t} (p_{i-1} - p_i). \end{aligned}$$

The difference is:

$$A_t^{ARL} - A_t^{TB} = \sum_{i=1}^{t} (p_{i-1} - p_i)[q(p_{t-1}) - q(p_{i-1})].$$

Under TB price falls monotonically towards p_* by Proposition 5.1, so $p_{i-1} - p_i \geq 0$ and $q(p_{t-1}) \geq q(p_{i-1})$ for $i \leq t$. Thus the ARL access charge is above that allowed under TB, and the firm prefers ARL to TB. Since TB is preferred to F the firm also prefers ARL to F. A similar argument shows that the firm prefers AR to ARL. If the firm is regulated by AR, but chooses the ARL prices given by Proposition 5.4 then its access charges under AR would be $A_t^{AR} = (p_0 - p_t) q(p_t)$, while its ARL access charge would be $A_t^{ARL} = (p_0 - p_t) q(p_{t-1})$. The difference is $(p_0 - p_t)[q(p_t) - q(p_{t-1})]$ which is positive since price declines monotonically under ARL by Proposition 5.4. □

The effect on consumers of ARL is complex. Since the period 1 constraint is the same as that under F or TB we know that $CS_1 > CS_0$. But in steady state actual average revenue equals $\bar{p}$ and (5.16) implies that steady-state surplus is below CS_0. A little algebra shows that consumer surplus is:

$$CS_t = CS_0 + \frac{(p_0 - p_{ARL})^2}{2}\{\mu^{2t} - 2\mu^{2t-1} + 2\mu^{t-1} - 1\}$$

where $p_0 - p_{ARL} = (1+\delta)(p_0 - p_*)/(1+\delta-\beta)$. When $t = 1$ the expression in curly brackets is $(\mu - 1)^2 > 0$, confirming that $CS_1 > CS_0$. As t increases the term in curly brackets tends to -1, which implies steady-state surplus is below CS_0. The present value of consumer surplus is:

$$M(ARL) = \frac{CS_0}{1-\beta} + \frac{(1+\delta)^2(p_0 - p_*)^2}{2(1+\delta-\beta)^2}\left\{\frac{\mu^2 - 2\mu}{1-\beta\mu^2} + \frac{2}{1-\beta\mu} - \frac{1}{1-\beta}\right\}$$

We can analyze the term in curly brackets numerically. When $\delta = 0$ it is positive for $\beta < 0.82$ and negative otherwise, so $M(ARL) > CS_0/(1-\beta)$ for values of β below 0.82. The higher δ is the lower the cut-off value of β at which $M(ARL) = CS_0/(1-\beta)$, so consumers are less likely to gain overall the higher δ is. Rises in β also reduce the probability that the consumer gains. This is intuitive – as β approaches 1 the ARL outcome approaches that under AR, where consumers definitely lose.

The relative consequences of ARL, TB, F and AR for consumers and the firm can also be discovered. We first compare ARL and TB. Taking the standard case with $\delta \geq 0$, ARL pushes p_t down further and faster than TB, as $\mu \leq \chi$ and $p_{ARL} < p_*$. In period 1 consumers are better off under ARL than they would be under TB. In subsequent periods $V(p_t)$ is higher under ARL because of the lower price, but the ability to set a higher access charge lowers net consumer surplus, and in the limit consumer surplus is below CS_0. For $\delta \geq 0$ we find that $M(TB) > M(ARL)$. From 5.2.2 we have $M(F) > M(TB)$ for $\delta \geq 0$. It is also not surprising to find that $M(ARL) > M(AR)$, since ARL is intermediate between TB and AR. Thus consumers prefer F to TB to ARL to AR when $\delta \geq 0$. Proposition 5.5 showed that firm has exactly the opposite ranking of the schemes.

The effect of ARL on welfare is also complex and we have to rely on numerical analysis. By substituting for p_t into the general expression for welfare, (5.8), we get

$$W_t = \frac{(1-c)(1-p_*)}{2} - \frac{(1+\delta)(p_0-p_*)^2(\beta-(1+\delta)\mu^t)^2}{2(1+\delta-\beta)^2}.$$

The present value of welfare is:

$$W(ARL) = \frac{(1-c)(1-p_*)}{2(1-\beta)} - \frac{(1+\delta)(p_0-p_*)^2}{2(1+\delta-\beta)^2}\left\{\frac{\beta^2}{1-\beta} - \frac{2\beta(1+\delta)\mu}{1-\beta\mu} + \frac{(1+\delta)^2\mu^2}{1-\beta\mu^2}\right\}.$$

In Figure 5.7 we show the values of β and δ that yield W(ARL) = W(TB), W(ARL) = W(F), W(F) = W(TB) and W(AR) = W(TB), and the regions in which ARL, TB and F are best, denoted by the labels "ARL", "TB" and "F".

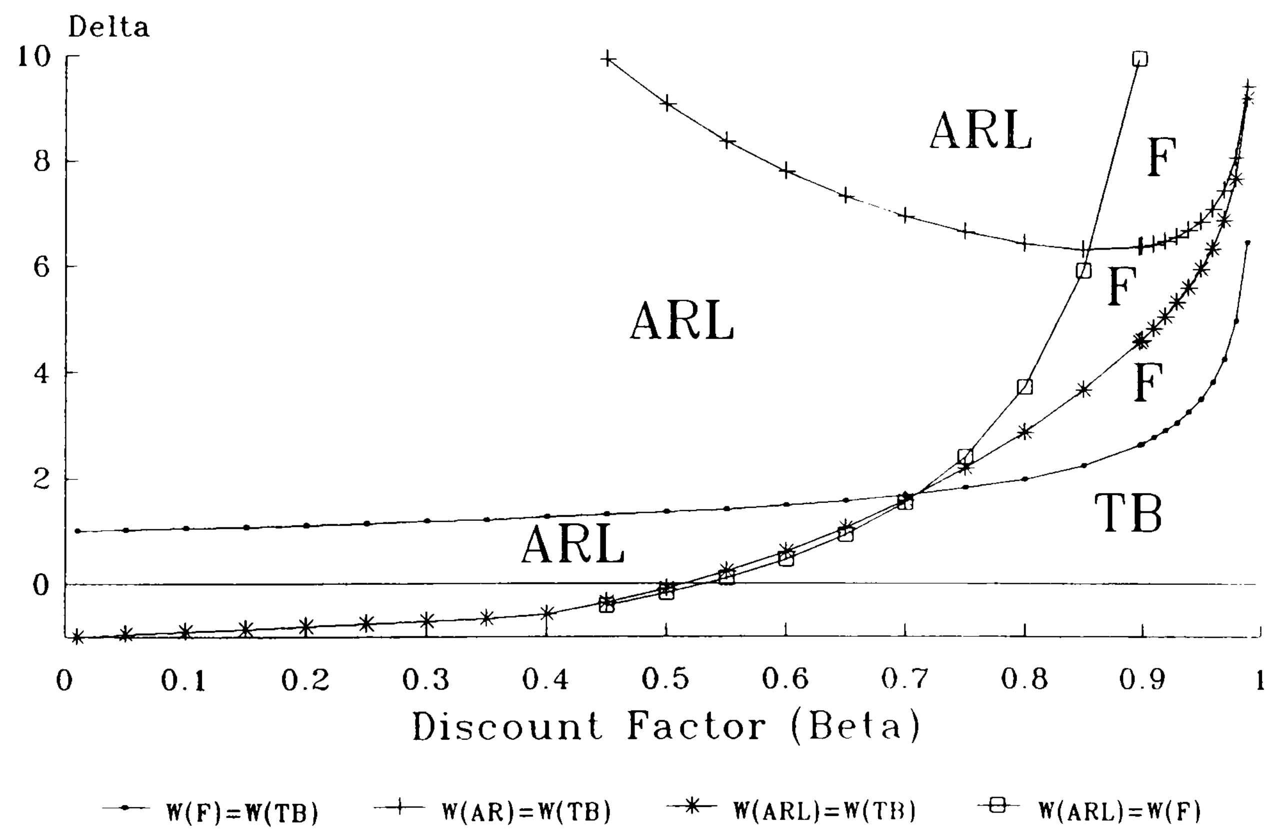


Figure 5.7 Welfare under TB, F and ARL

Values of δ above W(ARL) = W(TB) and W(ARL) = W(F) give welfare that is higher under ARL. Since ARL is related to AR it is not surprising that higher values of δ increase the relative advantage of ARL. When $\beta \simeq 0.72$ and $\delta \simeq 1.76$, welfare is identical under ARL, TB and F. There is no curve on Figure 5.7 representing combinations of β and δ where the regulator is indifferent between ARL and AR. This is because ARL dominates AR. In 5.2.3 we also showed that F dominates AR in a welfare sense. Roughly speaking TB is best when there is little discounting and marginal cost is flat, F is best for high values of β and steep marginal cost, while ARL is best for low values of β.

Of course the regulator is unlikely to know the coefficients of the demand and cost functions and their functional forms, so we must be wary of precise welfare comparisons. But the analysis is suggestive. In practice the regulatory period is one year, so the firm's discount factor β is equal to 1 divided by (1 plus the cost of capital). Regulators in Britain typically allow the real cost of capital for new investment to be no more than 7.5 per cent (and in many cases a significantly lower figure is used).[4] This gives a lower bound to β of about 0.93, and this means that TB or F appear to be preferable to ARL. Our analysis of AR showed that it was optimal for the short run when there is a capacity constraint, as long as the capital stock is not too large, but in the longer run this will generate excessive incentives to invest in capital. Thus our conclusion is that, for reasonable values of β, TB and F are the best schemes. TB is preferred when marginal cost is flat and F is preferred when it is steep, relative to demand.

5.2.5 The Paasche Price Index Scheme

The PPI scheme provides an interesting contrast to the other four schemes. It turns out that TB dominates it and that it is rarely feasible, so the analysis here will be brief. The problem is

[4]This is the upper limit of the range that the Monopolies and Mergers Commission (1993) determined for British Gas. See Armstrong, Cowan and Vickers (1994, Chapter 6) for a discussion of the cost of capital for regulated utilities.

$$\underset{\{A_t,\, p_t\}}{\text{Max}} \sum_{t=1}^{\infty} \beta^{t-1}\{A_t + \Pi(p_t)\} \qquad \text{s.t. } A_{t-1} - A_t + (p_{t-1} - p_t)q(p_t) \geq 0.$$

The constraint differs from the TB constraint (5.1) by using current quantities, $q(p_t)$, rather than lagged quantities as the weights. We go immediately to the linear example and assume for expository purposes that $\beta = 0$. The following additional assumptions are made. First, $A_0 = 0$ to allow comparison with the other schemes. Second, $\delta > 1$, i.e. marginal cost is steeper than the demand curve. Income in period t with a binding constraint is $I_t = A_{t-1} + (p_{t-1} - c)(1 - p_t) - \frac{\delta}{2}(1 - p_t)^2$. The first-order condition gives the following first-order difference equation for p:

$$\delta p_t + p_{t-1} = c + \delta.$$

The second-order sufficient condition is that $-\delta < 0$, and this holds since $\delta > 1$. The steady-state price is $p = (c + \delta)/(1 + \delta) = p_*$. In Chapter 3 we showed that a steady state, if it exists, gives Ramsey prices, which is the marginal cost price in this case. The full solution to the difference equation is $p_t = p_* + (p_0 - p_*)(-1/\delta)^t$. The adjustment coefficient is negative but less than 1 in absolute value, so the price path has damped oscillations and tends to p_* as t increases. The time path of prices is identical to that in the cobweb model of agricultural markets, where stability requires that the supply curve has a greater absolute slope than the demand curve, i.e. $\delta > 1$.

When the firm acts strategically the analysis is more subtle. In particular convergence to the steady state is not guaranteed. The following proposition is very general. The only assumption we make is that the demand curve slopes down.

Proposition 5.6 Non-convergence of the PPI scheme

For β close to 1 the product price under PPI does not converge to a steady state.

Proof. The present value of the firm's income under PPI, using (5.4), is

$$Z(PPI) = \frac{A_0 + \Pi_0}{1-\beta} + \frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}\{(p_{t-1} - p_t)q(p_t) + \Pi_t - \Pi_{t-1}\}.$$

We focus on the term after the summation sign. Consider the following cycle. In period 1 the firm sets p_1 below p_0. In period 2 price rises to somewhere between p_0 and p_1. Period 3's price is equal to p_0, and thereafter price is constant. The difference in the summation term between following this cycle, and keeping price constant at p_0, is

$$\begin{aligned}&(p_0 - p_1)q(p_1) + \Pi(p_1) - \Pi(p_0) + \beta[(p_1 - p_2)q(p_2) + \Pi(p_2) - \Pi(p_1)]\\ &+ \beta^2[(p_2 - p_0)q(p_0) + \Pi(p_0) - \Pi(p_2)].\end{aligned} \tag{5.27}$$

We make the counterfactual assumption that there is no discounting. When $\beta = 1$ the operating profit terms $\Pi(.)$ in (5.27) cancel out, and (5.27) becomes

$$(p_0 - p_2)[q(p_1) - q(p_0)] + (p_2 - p_1)[q(p_1) - q(p_2)].$$

Since $p_0 > p_2 > p_1$ and $q(p_1) > q(p_2) > q(p_0)$ this is strictly positive. By continuity (5.27) is positive for values of β close to 1. □

Proposition 5.6 implies that we need to assume that β is low enough to guarantee convergence in our linear example. We shall assume that $\beta < (\delta - 1)/(3 + \delta)$. This is a severe restriction on β. For example if $\delta = 2$ then β must be below 0.2 to ensure convergence. If the regulatory period is one year then this requires a cost of capital of 400 per cent. We can also write the inequality as a restriction on δ for a given value of β. This is $\delta > (1 + 3\beta)/(1 - \beta)$. Thus δ must be significantly above 1. For example when $\beta = 0.93$ we require $\delta > 54$. For AR and ARL to work we needed much weaker assumptions ($\delta > 0$ and $\delta > \beta - 1$ respectively).

The Euler and transversality conditions yield a solution for p_t of the form $p_t = p_* + (p_0 - p_*)\phi^t$ where the root ϕ is:

$$\phi = \frac{2\beta - \delta(1-\beta) + \sqrt{\delta^2(1-\beta)^2 - (1+\delta)4\beta(1-\beta)}}{2\beta}.$$

The assumption that $\delta > (1+3\beta)/(1-\beta)$ is sufficient for the root to be real and is both necessary and sufficient for $-1 < \phi < 0$, which is needed to ensure convergence. We can also show that $|\phi| > 1/\delta$, so the fluctuations of price are greater when $\beta > 0$ than when the firm is myopic.

The firm prefers PPI to all of the other schemes. The intuition for this is that the firm can always choose the AR price, p_1, and stay there. But PPI gives it further flexibility. Since AR is preferred by the firm to ARL, TB and to F, by transitivity PPI is also preferred to ARL, TB and F. The firm's value under PPI can be shown to be:

$$Z(PPI) = \frac{\Pi(p_0)}{1-\beta} + \frac{(p_0 - p_*)^2}{(1-\beta)} \cdot \frac{(2+\delta-\phi)}{2}.$$

We can compare this with the present value of income under AR, given in (5.21). The difference, $Z(PPI) - Z(AR)$, has the same sign as $-\phi - 1/\delta$. Since $|\phi| > 1/\delta$ this is positive for $\beta > 0$ and is 0 for $\beta = 0$.

We noted in Chapter 3 that when the constraint binds consumer surplus declines each period. This is confirmed in the two-part tariff case. Using the constraint we find that $CS_t - CS_{t-1} = -(p_{t-1} - p_t)^2/2$. To ensure that the consumer's participation constraint is never breached we need to assume that p_0 is close enough to p_*. Since consumers are worse off than with constant pricing at p_0, they prefer TB and F to PPI. We can also show that consumers prefer AR and ARL to PPI. When $\beta > 0$ PPI gives a period-1 outcome for consumers that is worse than under AR, and thereafter consumer surplus falls even further. Thus PPI is worse for consumers than AR, which in turn is worse than ARL.

PPI is a poor scheme since its feasibility requires a restrictive assumption about the relationship between δ and β, and consumer surplus declines each period. It is perhaps surprising to find that it ensures that welfare increases each period (like the TB scheme). Welfare, given in (5.8), is higher the smaller the absolute value of the gap between p_t and p_*. Under PPI this gap is $(p_0 - p_*)\phi^t$ which declines in absolute value as t increases since $|\phi| < 1$. The present value of welfare is

$$W(PPI) = \frac{(1-c)(1-p_*)}{2(1-\beta)} - \frac{(1+\delta)(p_0 - p_*)^2}{2} \cdot \frac{\phi^2}{1-\beta\phi^2}.$$

We shall compare W(PPI) with W(TB) given in (5.9). We have:

$$W(TB) - W(PPI) \overset{\text{sgn}}{=} \frac{-\chi^2}{1-\beta\chi^2} + \frac{\phi^2}{1-\beta\phi^2} \overset{\text{sgn}}{=} \phi^2 - \chi^2 = (\phi - \chi)(\phi + \chi).$$

Now $\phi - \chi < 0$ since $\phi < 0$ and $\chi > 0$, so $W(TB) - W(PPI)$ has the opposite sign to $(\phi + \chi)$. In Appendix 5.4 we show that this is negative, so it follows that $W(TB) > W(PPI)$. Thus TB dominates PPI in welfare terms. The regulator will never want to impose PPI when she can use TB, and in many cases (where δ is low or β is high) implementation of PPI is not even feasible.

5.3 Some Related Price Caps

In Chapter 3.3 we considered combinations of TB and PPI, and of ARL and AR and briefly described some of their properties. We consider (i) a hybrid of TB and PPI and (ii) a hybrid of ARL and AR. The combination of TB and PPI given in (3.18) can be adapted to the two-part tariff case:

$$A_{t-1} - A_t + (p_{t-1} - p_t)\frac{[q(p_t) + q(p_{t-1})]}{2} \geq 0. \tag{5.28}$$

Instead of weighting $p_{t-1} - p_t$ by current or lagged output we weight by the average of the two. Implicitly the change in the access charge is weighted by the average number of customers, which is 1. With a binding constraint the Euler condition is:

$$\beta(q_{t+1} - q_t) + q_t - q_{t-1} +$$
$$(p_{t-1} - MC_t + (1-\beta)(p_t - MC_t) - \beta(p_{t+1} - MC_t)q'(p_t) = 0. \tag{5.29}$$

If a steady state is reached then there is marginal cost pricing.

We saw in the previous section that restrictive assumptions were necessary to ensure convergence of the PPI scheme. Since the hybrid scheme has PPI as a component it is not surprising to find that this also does not guarantee convergence. One special case is when demand is linear. In Chapter 3 we saw that the hybrid scheme ensures that consumer surplus remains constant in the linear case. The left-hand side of (5.28) is equal to $CS_t - CS_{t-1}$, so $A_t - A_{t-1} = V_t - V_{t-1}$. Using the general formula (5.4) the present value of income for the hybrid scheme when demand is linear is then:

$$Z(HY_1) = \frac{A_0 + \Pi_0}{1-\beta} + \frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}\{V_t - V_{t-1} + \Pi_t - \Pi_{t-1}\}$$

$$= \frac{A_0 - V(p_0)}{1-\beta} + \sum_{t=1}^{\infty}\beta^{t-1}\{V_t + \Pi_t\}.$$

Each period the firm's income is equal to total welfare, $V_t + \Pi_t$, less the initial level of consumer surplus, $V(p_0) - A_0$. Since the latter is fixed the firm maximizes profits by maximizing welfare. Price is set at p_* immediately. Thus, at least when the weights in the welfare function are equal, the hybrid constraint is optimal when demand is linear. Effectively the scheme is the same as the modified L-M scheme, where the firm receives a transfer from the regulator of $V(p_t) - [V(p_0) - A_0]$. To implement this the regulator has to know that demand is linear, and what demand will be at particular prices.

Although the hybrid scheme is optimal with linear demand (at least as long as the welfare weights are equal), when demand is nonlinear there are serious problems with convergence. PPI did not converge when β is high, whatever the form of the demand function. The hybrid scheme is not as bad, but convergence is still not guaranteed when demand is nonlinear. We first show that the scheme is effectively equivalent to the scheme of Vogelsang (1988) which was discussed in Chapter 2.5.3. The transfer in Vogelsang's scheme is equal to $(p_{t-1} - p_t)[q_{t-1} + q_t]/2 - \Pi(p_{t-1})$. In the hybrid scheme the access charge is autoregressive and lagged profits are not subtracted. From (5.28) we have $A_t = A_{t-1} + (p_{t-1} - p_t)[q_{t-1} + q_t]/2$. The present value of income in Vogelsang's scheme is:

$$Z(Vo) = \sum_{t=1}^{\infty} \beta^{t-1}\left\{\frac{(p_{t-1} - p_t)(q_{t-1} + q_t)}{2} + \Pi_t - \Pi_{t-1}\right\}$$

while for the first hybrid scheme it is, using (5.4),

$$Z(HY_1) = \frac{A_0 + \Pi_0}{1-\beta} + \frac{1}{1-\beta}\sum_{t=1}^{\infty} \beta^{t-1}\left\{\frac{(p_{t-1} - p_t)(q_{t-1} + q_t)}{2} + \Pi_t - \Pi_{t-1}\right\}.$$

Apart from the constant term this is simply $Z(Vo)/(1-\beta)$, and the firm will choose exactly the same price path under the two schemes.[5]

This effective equivalence means that the convergence problems noted by Vogelsang for his scheme when demand is nonlinear also apply to the hybrid price-cap scheme. Vogelsang (1988) presents a formal proof of nonconvergence for values of β close to 1 when demand is either convex or concave in the region of p_* and we shall not repeat that here. If a far-sighted firm has its price at p_* then it has a strong incentive to move away from optimal pricing when demand is nonlinear. It sacrifices profits in the first period it moves away from p_*, but can make higher profits thereafter. The argument is similar to the one used to prove Proposition 5.6, but relies on local convexity (or concavity). When the firm is myopic it does not engage in strategic behaviour and convergence is more likely. We can use a simple example to show that convergence can indeed occur when $\beta = 0$. Suppose the demand function has a constant elasticity of 1, i.e. $q = 1/p$, and that marginal cost is constant at c. Substitution into the Euler condition (5.29) gives a nonlinear first-order difference equation that can be

[5]The price paths for both schemes are implicitly determined by the Euler condition (5.29). Vogelsang (1988) writes this incorrectly in his equation (14).

written as $p_t = \sqrt{2cp_{t-1} - p_{t-1}^2}$. Simulations show that p_t converges to c as long as $c > p_0/2$. But the lack of any guarantee of convergence when $\beta > 0$ and demand is nonlinear makes the hybrid scheme unattractive, even though it performs optimally when demand is linear.

In Chapter 3.4 we mentioned that a mixture of ARL and AR, with weights $1/1-\beta$ and $-\beta/(1-\beta)$ respectively, generates Ramsey pricing. In the two-part tariff case we would expect that marginal cost pricing is induced. This is indeed true, but a slightly more general result also holds. This is that the weighted average scheme generates exactly the same price path as the TB scheme for $\beta > 0$, and the present value of the firm is the same. The hybrid constraint gives an access charge of

$$A_t = (\bar{p} - p_t)\left(\frac{1}{1-\beta}q(p_{t-1}) - \frac{\beta}{1-\beta}q(p_t)\right).$$

The present value of the firm is then:

$$Z(HY_2) = \sum_{t=1}^{\infty}\beta^{t-1}\left\{(\bar{p} - p_t)\left(\frac{1}{1-\beta}q(p_{t-1}) - \frac{\beta}{1-\beta}q(p_t)\right) + \Pi(p_t)\right\}.$$

Since $\bar{p} = A_0/q_0 + p_0$ the present value of the access charges is:

$$\sum_{t=1}^{\infty}\beta^{t-1}A_t = \frac{1}{1-\beta}\left[A_0 + (p_0 - p_1)q_0 - \beta(\bar{p} - p_1)q_1 + \beta(\bar{p} - p_2)q_1 - \beta^2(\bar{p} - p_2)q_2 + \dots\right].$$

Cancelling terms gives:

$$\sum_{t=1}^{\infty}\beta^{t-1}A_t = \frac{1}{1-\beta}\left[A_0 + (p_0 - p_1)q_0 + \beta(p_1 - p_2)q_1 + \beta^2(p_2 - p_3)q_2 + \dots\right]$$

which is exactly the same as the present value of access charges in the TB case (see equation (5.2)). Thus the Euler condition for the firm is the same as (5.3), and the present value of the firm is the same as in the TB case. We should note, however, that the access charges differ from those under TB, although their present values are equal.

To implement this scheme the regulator needs to know the firm's discount factor.

5.4 Conclusions

By making the simplifying assumptions of a fixed number of customers and linear demand and marginal cost we have been able to provide a complete characterization of the implications of each price cap for the firm, consumers and welfare. The firm has the following ranking: PPI $\succeq$ AR $\succeq$ ARL $\succeq$ TB $\succeq$ F, while for non-decreasing marginal cost consumers have exactly the opposite ranking of the schemes.

The PPI scheme is often not feasible and is dominated in welfare terms by TB. The AR scheme is feasible under some assumptions but its tendency to induce excess output counts against it. AR is optimal in the short-run when there is a capacity constraint, but the same problems of excess output occur in the longer run when capacity is variable, and when marginal cost is finite it is dominated by both F and ARL. The scheme that is a hybrid of TB and PPI is optimal when demand is linear, but the lack of convergence when demand is nonlinear means that it is far from optimal in general.

The remaining three schemes are TB, F and ARL. ARL is only best for implausibly low values of the discount factor, so the real choice is between TB and F. In any case both schemes have the advantage over ARL that they guarantee that consumers gain. TB induces efficient pricing in the steady state, but the more strategic the firm is the slower the rate of convergence. F moves close to efficient pricing in the first period, but thereafter there are no further welfare increases. TB is better when marginal cost is fairly flat, while F is better for steep marginal cost. When the discount factor increases the relative advantage of TB grows. TB and F appear to be robust price cap schemes that are generally feasible and have desirable welfare properties.

Appendix 5.1 Proof of Proposition 5.1.

(i) We first show that $\chi_1 < 1$ for $\beta > 0$. The proof is by contradiction. If $\chi_1 \geq 1$, $2+\delta(1-\beta) - \sqrt{\left\{4(1+\delta)(1-\beta)+\delta^2(1-\beta)^2\right\}} \geq 2\beta$. Rearranging gives $(2+\delta)(1-\beta) \geq \sqrt{\left\{4(1+\delta)(1-\beta)+\delta^2(1-\beta)^2\right\}}$. Squaring both sides yields: $(4(1+\delta)+\delta^2)(1-\beta)^2 \geq 4(1+\delta)(1-\beta)+\delta^2(1-\beta)^2$. Subtracting $\delta^2(1-\beta)^2$ from both sides, and dividing by $4(1+\delta)(1-\beta)$, which we can do because $\delta > -1$ and $\beta < 1$, gives: $1-\beta \geq 1$. This is a contradiction since $\beta > 0$. An analogous proof (which we omit) shows that the other root, χ_2, > 1. To show that $\chi_1 > 1/(2+\delta)$ for $\beta > 0$ we again assume the opposite and find a contradiction. So we suppose

$$\frac{2+\delta(1-\beta) - \sqrt{4(1+\delta)(1-\beta)+\delta^2(1-\beta)^2}}{2\beta} \leq \frac{1}{2+\delta}.$$

Remembering that $4(1+\delta)(1-\beta)+\delta^2(1-\beta)^2 = [2+\delta(1-\beta)]^2 - 4\beta$ we have $(2+\delta)(2+\delta(1-\beta)) - 2\beta \leq (2+\delta)\sqrt{[2+\delta(1-\beta)]^2 - 4\beta}$. Squaring both sides gives $(2+\delta)^2(2+\delta(1-\beta))^2 - 4\beta(2+\delta)(2+\delta(1-\beta)) + 4\beta^2 \leq (2+\delta)^2\left\{[2+\delta(1-\beta)]^2 - 4\beta\right\}$. Subtracting $(2+\delta)^2(2+\delta(1-\beta))^2$ from both sides, and dividing the remaining terms by 4β gives $(2+\delta)^2 - (2+\delta)(2+\delta-\delta\beta) + \beta = \beta(\delta^2+2\delta+1) \leq 0$ which implies (taking $\beta > 0$) that $\delta^2+2\delta+1 \leq 0$. But $\delta^2+2\delta+1 = (1+\delta)^2 > 0$ since $\delta < -1$. Thus $\chi > 1/(2+\delta)$ for $\beta > 0$.

(ii) Differentiating χ_1 with respect to β gives:

$$4\beta^2\frac{\partial\chi_1}{\partial\beta} =$$

$$-2\beta\delta + \frac{4\beta(1+\delta)+2\beta\delta^2(1-\beta)}{\sqrt{4(1+\delta)(1-\beta)+\delta^2(1-\beta)^2}} - 2\left(2+\delta(1-\beta) - \sqrt{4(1+\delta)(1-\beta)+\delta^2(1-\beta)^2}\right)$$

We want to show that the right-hand side is positive. Suppose not. Dividing the resulting inequality by 2, multiplying by the square root term and rearranging gives:

$$\frac{2+\delta(1-\beta) - \sqrt{4(1+\delta)(1-\beta)+\delta^2(1-\beta)^2}}{2\beta} \leq \frac{1}{2+\delta}.$$

We already know from the proof of (i) that this cannot hold for $\beta > 0$, so we know that $\partial\chi_1/\partial\beta > 0$ for $\beta > 0$.

(iii) Define the numerator of χ_1 as $g(\beta)$ and the denominator $f(\beta)$. Using L'Hôpital's Rule

$$\lim_{\beta \to 0} \chi_1 = \lim_{\beta \to 0} \frac{f'(\beta)}{g'(\beta)} = \lim_{\beta \to 0} \frac{-\delta + \dfrac{2(1+\delta) + \delta^2(1-\beta)}{\sqrt{4(1+\delta)(1-\beta) + \delta^2(1-\beta)^2}}}{2} = \frac{1}{2+\delta}.$$

The proof that $\lim_{\beta \to 1} \chi_1 = 1$ is by inspection.

(iv) Differentiating χ with respect to δ gives:

$$\frac{\partial \chi}{\partial \delta} = \frac{(1-\beta)}{2\beta}\left\{1 - \frac{2+\delta(1-\beta)}{\sqrt{(2+\delta(1-\beta))^2 - 4\beta}}\right\} < 0.$$

The general solution to (5.3′) is of the form $p_t = p_* + G(\chi)^t + H(\chi_2)^t$. Since $\chi_2 > 1$ we must have $H = 0$ to satisfy the transversality condition. The coefficient G is then given by $(p_0 - p_*)$. Monotonicity of the price path follows from the fact that $0 < \chi_1 < 1$. □

Appendix 5.2 Proof of Proposition 5.4.

The two roots of (5.24′) are:

$$\mu_i = \frac{2+\delta \pm \sqrt{(2+\delta)^2 - 4\beta}}{2\beta} \qquad i = 1, 2$$

If the larger root, μ_2, exceeds 1 then we can rule it out by the transversality condition. To show that $\mu_2 > 1$ we suppose $\mu_2 \le 1$. This implies $\sqrt{(2+\delta)^2 - 4\beta} \le 2\beta - (2+\delta)$. Squaring both sides gives: $(2+\delta)^2 - 4\beta \le 4\beta^2 - 4\beta(2+\delta) + (2+\delta)^2$. This implies that $1 + \delta - \beta \le 0$, which contradicts Assumption 1. Thus we can rule out μ_2.

Property (i). To prove that $\mu_1 = \mu < 1$ suppose not. Then $2\beta - (2+\delta) \geq \sqrt{(2+\delta)^2 - 4\beta}$. Squaring both sides gives: $4\beta^2 - 4\beta(2+\delta) + (2+\delta)^2 \geq (2+\delta)^2 - 4\beta$, which again implies that $0 \geq 1 + \delta - \beta$. Thus $\mu < 1$. To show that $\mu > 1/(2+\delta)$ for $\beta > 0$ suppose not. If $\mu \leq 1/(2+\delta)$ then $(2+\delta)^2 - 2\beta \leq (2+\delta)\sqrt{(2+\delta)^2 - 4\beta}$. Squaring we get: $(2+\delta)^4 - 4\beta(2+\delta)^2 + 4\beta^2 \leq (2+\delta)^4 - 4\beta(2+\delta)^2$ which implies that $\beta \leq 0$. This contradicts the assumption that $\beta > 0$. Thus the price moves monotonically towards p_{ARL}.

Property (ii). The derivative of μ with respect to β is

$$\frac{\partial \mu}{\partial \beta} = \frac{1}{4\beta^2}\left\{\frac{(2+\delta)\left[2+\delta-\sqrt{(2+\delta)^2-4\beta}\right]-2\beta}{\sqrt{(2+\delta)^2-4\beta}}\right\}.$$

This has the same sign as $\mu - 1/(2+\delta)$. Since this is positive for $\beta > 0$, $\partial\mu/\partial\beta > 0$.

Property (iii). Defining the numerator of μ as $f(\beta)$ and the denominator as $g(\beta)$, and using L'Hôpital's Rule

$$\lim_{\beta \to 0} \mu = \lim_{\beta \to 0} \frac{f'(\beta)}{g'(\beta)} = \lim_{\beta \to 0} \frac{\frac{2}{\sqrt{4(1+\delta-\beta)+\delta^2}}}{2} = \frac{1}{2+\delta}.$$

By inspection $\lim_{\beta \to 1} \mu = 1 + \frac{1}{2}\{\delta - \sqrt{4\delta + \delta^2}\} \leq 1$ for $\delta \geq 0$. Note that we must have $\delta \geq 0$ to take this limit in order to satisfy Assumption 1.

Property (iv). The derivative of μ with respect to δ is:

$$\frac{\partial \mu}{\partial \delta} = \frac{1}{2\beta}\left\{1 - \frac{2+\delta}{\sqrt{(2+\delta)^2 - 4\beta}}\right\} < 0.$$ □

Appendix 5.3 Proof that the sign of $(\chi - \mu)$ is the same as the sign of δ.

The characteristic equation for TB is $\beta\chi^2 - (2 + \delta - \delta\beta)\chi + 1 = 0$ and for ARL we have $\beta\mu^2 - (2+\delta)\mu + 1 = 0$. Subtracting the second equation from the first yields $(\chi - \mu)[2 + \delta - \beta(\chi + \mu)] = \delta\beta\chi$. If the expression in square brackets is positive then the sign of $(\chi - \mu)$ is the same as the sign of δ. We have: $2 + \delta > 1 + \beta = \beta(1/\beta + 1) > \beta(\chi + \mu)$. The first inequality follows from Assumption 1, and the second from the facts that $\chi < 1$ and $\mu < 1$. Thus $(\chi - \mu) \overset{\text{sgn}}{=} \delta$. $\square$

Appendix 5.4 Proof that TB gives higher welfare than PPI.

We need to show that $\phi + \chi < 0$. Suppose the opposite, so

$$\phi + \chi = \frac{1}{2\beta}\left\{2(1+\beta) + \sqrt{(2\beta - \delta(1-\beta))^2 - 4\beta} - \sqrt{(2 + \delta(1-\beta))^2 - 4\beta}\right\} \geq 0$$

and

$$2(1+\beta) + \sqrt{(2\beta - \delta(1-\beta))^2 - 4\beta} \geq \sqrt{(2 + \delta(1-\beta))^2 - 4\beta}.$$

Squaring both sides implies that

$$(1+\beta)\left\{2\beta - \delta(1-\beta) + \sqrt{(2\beta - \delta(1-\beta))^2 - 4\beta}\right\} \geq 0. \tag{A.1}$$

The term in curly brackets is the numerator of ϕ. To show that this is negative suppose the opposite, i.e. $2\beta - \delta(1-\beta) + \sqrt{(2\beta - \delta(1-\beta))^2 - 4\beta} \geq 0$. Subtracting $2\beta - \delta(1-\beta)$ from both sides and squaring gives the implication that $-4\beta \geq 0$, which cannot hold for $\beta > 0$. Thus we know that $\phi < 0$ and $2\beta - \delta(1-\beta) + \sqrt{(2\beta - \delta(1-\beta))^2 - 4\beta} < 0$ which contradicts the implication (A.1). Thus $\phi + \chi < 0$. $\square$

CHAPTER 6

NONLINEAR PRICING AND PRICE CAPS

6.1 Introduction

Large customers of utilities are often offered tariffs that are more sophisticated than the simple two-part tariffs that are set for most small customers. British Gas, for example, has an effective two-part tariff for domestic customers who consume less than 2,500 therms per year (average domestic consumption is about 640-670 therms a year, depending on the weather), but for marginal purchases above 2,500 therms customers pay lower prices.[1] British Telecom introduced quantity discounts into its price list in 1993 for customers with quarterly call bills over £75. Nonlinear pricing enables firms to extract more surplus from customers than when using linear or two-part tariffs, and can often enhance efficiency.

A nonlinear tariff can be denoted by $T(q)$ where T is the payment made by the customer for quantity q. Two-part and uniform tariffs are special cases of this where $T = A + pq$ and $T = pq$ respectively. Like other forms of price discrimination nonlinear pricing is feasible when the firm can prevent resale between customers, as is typically the case in the utilities. The firm is assumed to know the distribution of customer types, but is unable to identify which type a customer is, unlike in the cases of first- and third-degree price discrimination. Different customers must thus be given *incentives* to choose the packages designed for them.

Wilson (1993) and Tirole (1988, Chapter 3) provide useful discussions of nonlinear pricing. A standard result is that the marginal price for the largest customer should be equal to marginal cost. This holds for both profit- and welfare-maximization. It is an "absence of distortion at the top" result that is familiar from the literature on mechanism design – see Chapter 2.3.2. If the largest customer faced a marginal price above marginal cost then by shading the price of an additional unit the firm could sell an extra unit at a profit and the consumer would be better off. But smaller customers

[1] One therm is equivalent to 29.31 kilowatt hours.

face high marginal prices and do not consume the efficient quantities, as the firm seeks to dissuade the larger customers from opting for the bundles designed for smaller customers. Sharkey and Sibley (1993) have shown that when the regulator gives higher weight to the utility of the largest customer, the welfare-maximizing tariff has the largest customer consuming more than the efficient quantity when the firm sets a menu of optional two-part tariffs. We demonstrate below that when the firm is subject to average revenue regulation the firm also induces the largest customer to over-consume.

A common form of nonlinear pricing is when the customer is given a menu of optional two-part tariffs. Mobile telephony in Britain provides a good example of this. A simple version of this is where the customer can choose between tariff 1 denoted by $\{A^1, p^1\}$ and tariff 2, $\{A^2, p^2\}$, where $A^1 < A^2$ and $p^1 > p^2$. Customers who want to use the phone infrequently, perhaps only for emergencies, choose tariff 1 that has the lower access charge and the higher marginal charge, while larger customers, who tend to be business users, choose tariff 2. In fixed-link telephony British Telecom offers optional two-part tariffs. Customers can obtain a 10 per cent discount on all calls for a payment of £12 per quarter. Willig (1978) shows that optional two-part tariffs can generate a Pareto improvement, as long as all customers are final consumers. Suppose that the firm initially offers a standard two-part tariff. A new two-part tariff that just enables large customers to purchase the same quantity for the same total outlay as under the old tariff is offered, in addition to the original tariff. The new tariff has a lower marginal price (that is still above marginal cost) and a higher access charge. Large customers will choose to go on the new tariff and will consume more, raising profits because the price of the marginal sales exceeds marginal cost. They will be better off because they could have chosen to remain on the old tariff, while customers who do not choose the new tariff have no change in their welfare.[2] Usually when the nonlinear tariff exhibits quantity discounts the firm can achieve the same outcome with a menu of two-part tariffs as with a full nonlinear tariff.

[2]If the product is an intermediate good then it is possible that such an optional tariff will distort downstream competition and not raise welfare - see Ordover and Panzar (1980, 1982).

Much of the analysis of nonlinear pricing has been under the assumption that the firm is either an unregulated profit-maximizer, or is required by a regulator who has same information set as the firm to set a welfare-maximizing tariff. As was discussed in Chapter 2 much insight can be obtained about regulatory practice by assuming that the regulator does not have as much information as the firm. Taking account of this information asymmetry, as well as the information asymmetry between the firm and its customers that is the reason for nonlinear pricing, is complex. Laffont and Tirole (1990, 1993, Chapter 4) integrate a model of nonlinear pricing with two types of customer with their hidden action and information model of regulation. Srinagesh (1986) presents a model of regulation of nonlinear pricing subject to a rate-of-return constraint. The approach here is to assume that the regulator sets a price cap, in which costs play no role. Armstrong, Cowan and Vickers (1995) consider a similar problem to the one analyzed in the first part of this section, but assume that the distribution of customer types is continuous and have a more restrictive utility function than that used below.

In this chapter we shall use simple two-type models to obtain the main results. In section 6.2 we consider how nonlinear pricing is used to maximize: (i) welfare subject to a profit constraint; (ii) profits with no constraint, and (iii) profits under average revenue regulation. This two-type model of average revenue regulation was written before the model of Armstrong, Cowan and Vickers (1995). That paper is summarized at the end of section 6.2. In section 6.3 we explore how a tariff basket constraint might be applied when there is nonlinear pricing, with the focus on the case where the firm offers optional two-part tariffs. Section 6.4 contains concluding remarks.

6.2 Nonlinear Pricing and Average Revenue Regulation

In this section we present a simple model of nonlinear pricing with average revenue regulation. The underlying model is that used by Laffont and Tirole (1990).[3] There are two types of customer, and for simplicity we assume that there are equal numbers of each type. Allowing an unequal distribution of customer types is

[3] See also Tirole (1988, pp 153-154) and Fudenberg and Tirole (1991, pp 246-250).

straightforward but does not affect the main results. The firm knows the distribution of customers but cannot tell which type a particular customer is. Type i ($i = 1, 2$) customers have gross utility function $U^i(q^i)$ where $U^i(.)$ is an increasing, concave function with $U^i(0) = 0$. The customer pays T^i to the firm in order to consume q^i, so consumer surplus is $U^i(q^i) - T^i$. Customer 2 is the high-demand consumer, and when both consume the same quantity he has higher total and marginal utility, i.e. $U^2(q) > U^1(q)$ and $dU^2(q)/dq > dU^1(q)/dq$. Consumers have the same weights in social welfare. It is assumed that the firm wants to serve both types.

The firm chooses quantities and payments to satisfy two individual rationality constraints (IR^i, $i = 1, 2$) and two incentive compatibility constraints (IC^i, $i = 1, 2$):

$$(IR^i) \quad U^1(q^1) - T^1 \geq 0 \qquad \text{and} \quad U^2(q^2) - T^2 \geq 0 \tag{6.1}$$

$$(IC^i) \quad U^1(q^1) - T^1 \geq U^1(q^2) - T^2 \qquad \text{and} \quad U^2(q^2) - T^2 \geq U^2(q^1) - T^1. \tag{6.2}$$

The constraints in (6.1) require that both types are willing to consume while the constraints in (6.2) require that neither type wants to choose the package that is designed for the other type. Typically only IR^1 and IC^2 bind. If IR^1 and IC^2 are satisfied, then IR^2 is also satisfied. The proof of this is straightforward. When IC^2 holds $U^2(q^2) - T^2 \geq U^2(q^1) - T^1$. Since $U^2(q^1) > U^1(q^1)$ as long as $q^1 > 0$ it follows that $U^2(q^2) - T^2 > U^1(q^1) - T^1$. If IR^1 holds then the right-hand side of this inequality is nonnegative and IR^2 is slack. This makes economic sense: if customer 2 chose the bundle $\{q^1, T^1\}$ designed for customer 1 then he would receive strictly positive surplus. He can do even better by choosing the bundle designed for him.

The standard way to solve the two-type model is to ignore IC^1, and then show that IC^1 is satisfied by the solution to the under-constrained problem. If IR^1 and IC^2 bind then a sufficient condition for IC^1 to hold is that $q^2 > q^1$. With IR^1 binding the type 1 consumer has zero surplus and IC^1 is $0 \geq U^1(q^2) - T^2$. Substituting for T^2 from the binding IC^2 in (6.2) gives

$$0 \geq \left[U^1(q^2) - U^1(q^1)\right] - \left[U^2(q^2) - U^2(q^1)\right] = \int_{q^1}^{q^2} \left(\frac{dU^1(\tilde{q})}{dq} - \frac{dU^2(\tilde{q})}{dq} \right) d\tilde{q}$$

which is negative when $q^2 > q^1$ since $dU^1(\tilde{q})/dq < dU^2(\tilde{q})/dq$. Thus to check that IC^1 is satisfied all that has to done is to confirm that $q^2 > q^1$.

We first consider the Ramsey problem of choosing a nonlinear tariff to maximize welfare subject to a zero profit constraint. Assuming that IR^1 binds we have $T^1 = U^1(q^1)$. Customer 1's surplus is fully extracted by the firm. If IC^2 also binds then $T^2 = U^2(q^2) - U^2(q^1) + U^1(q^1)$. Customer 2 earns a positive "information rent" of $U^2(q^1) - U^1(q^1)$, which is increasing in q^1. Aggregate consumer surplus is thus given by $U^2(q^1) - U^1(q^1)$. Revenue is $T^1 + T^2 = 2U^1(q^1) + U^2(q^2) - U^2(q^1)$. The Ramsey problem gives the Lagrangian

$$\mathcal{L} = U^2(q^1) - U^1(q^1) + \mu\left[2U^1(q^1) + U^2(q^2) - U^2(q^1) - cq^1 - cq^2 - K\right]$$

where we have assumed a constant marginal cost of c and fixed costs of K. The first-order conditions yield:

$$\frac{dU^1(q^1)}{dq} = c + \frac{(\mu - 1)}{\mu}\left[\frac{dU^2(q^1)}{dq} - \frac{dU^1(q^1)}{dq}\right] \tag{6.3a}$$

$$\frac{dU^2(q^2)}{dq} = c. \tag{6.3b}$$

If the zero profit constraint binds then $\mu > 1$ and customer 1's marginal utility exceeds c, so he under-consumes relative to the first-best level.[4] The term in square brackets in (6.3a) is the marginal increase in customer 2's information rent as q^1 is increased. The

[4] If fixed costs are small then the zero profit constraint might not bind. A Coase two-part tariff (Coase, 1946) with marginal price equal to c would be optimal when the net surplus that customer 1 earns is sufficient to enable him to pay enough of the fixed costs, with the remainder being financed by customer 2.

high-demand customer, however, consumes the efficient quantity. Equations (6.3a) and (6.3b) imply that $q^2 > q^1$. The high-demand customer consumes his efficient quantity while the low-demand customer under-consumes, and the former's efficient quantity is larger than the latter's. The regulator is prepared to reduce the quantity that customer 1 consumes below the first-best level to make it less attractive for customer 2 to choose 1's bundle. If customer 2 can be persuaded to choose the bundle with high quantity then the regulator can set a high value of T^2 to provide revenue with which to finance the fixed costs.

As μ increases towards infinity the first-order conditions (6.3a) and (6.3b) tend to those that apply when the firm is an unconstrained profit-maximizer:

$$\frac{dU^1(q^1)}{dq} = c + \left[\frac{dU^2(q^1)}{dq} - \frac{dU^1(q^1)}{dq}\right] \tag{6.4a}$$

$$\frac{dU^2(q^2)}{dq} = c. \tag{6.4b}$$

Equation (6.4a) has an intuitive interpretation. A unit increase in q^1 raises T^1 by $dU^1(q^1)/dq$, which is marginal revenue earned from customer 1. The marginal cost of raising q^1 is the incremental production cost, c, plus the increase in the information rent that customer 2 earns, which is $dU^2(q^1)/dq - dU^1(q^1)/dq$. The increase in the information rent represents a cost to the firm because it must reduce T^2 to ensure that customer 2's incentive compatibility constraint still holds, since T^2 is equal to $U^2(q^2)$ less the information rent. Customer 2 still consumes the efficient quantity but earns less surplus than in the Ramsey problem as q^1 is driven down to allow T^2 to increase. For later use we note that if marginal cost is reduced the firm will increase sales to both types of customer. This result is straightforward and depends only on the second-order conditions holding.

We now consider the regulated firm's problem. The average revenue constraint is $\bar{p}(q^1 + q^2) \geq T^1 + T^2$. It is assumed that $\bar{p} > c$. The Lagrangian for the firm is:

$$\begin{aligned}\mathcal{L} &= T^1 + T^2 - c(q^1 + q^2) + \lambda[\bar{p}(q^1 + q^2) - T^1 - T^2] \\ &\quad + \omega[U^1(q^1) - T^1] + \phi[U^2(q^2) - T^2 - U^2(q^1) + T^1]\end{aligned}$$

where λ is the multiplier on the average revenue constraint, ω is the multiplier on IR^1 and ϕ is the multiplier on IC^2. We note that if we define revenue $R \equiv T^1 + T^2$ and aggregate output $Q \equiv q^1 + q^2$ then the first line of the Lagrangian can be written as

$$R - cQ + \lambda[\bar{p}Q - R] = (1 - \lambda)\left\{R - \left[c - \frac{\lambda}{1-\lambda}(\bar{p} - c)\right]Q\right\}.$$

As long as $0 < \lambda < 1$ it is as if the firm is an unregulated profit-maximizer facing a lower marginal cost of $c - \lambda(\bar{p} - c)/(1 - \lambda)$.

Assuming interior solutions for T^1, T^2, q^1 and q^2, the associated first-order conditions are:

$$\frac{\partial \mathcal{L}}{\partial T^1} = 1 - \lambda - \omega + \phi = 0 \tag{6.5a}$$

$$\frac{\partial \mathcal{L}}{\partial T^2} = 1 - \lambda - \phi = 0 \tag{6.5b}$$

$$\frac{\partial \mathcal{L}}{\partial q^1} = -c + \lambda\bar{p} + \omega\frac{dU^1(q^1)}{dq} - \phi\frac{dU^2(q^1)}{dq} = 0 \tag{6.5c}$$

$$\frac{\partial \mathcal{L}}{\partial q^2} = -c + \lambda\bar{p} + \phi\frac{dU^2(q^2)}{dq} = 0. \tag{6.5d}$$

We look for a solution where IR^1, IC^2 and the AR constraint all bind with positive values for ω, ϕ and λ, and we then check that IC^1 is satisfied. We first show that $\lambda < 1$. Adding (6.5a) and (6.5b) gives $2(1 - \lambda) = \omega$. If $\omega = 0$ then (6.5a) and (6.5b) would imply that $\phi = 0$ and $\lambda = 1$. Using these in (6.5c) and (6.5d) gives $\bar{p} = c$, which contradicts the assumption that $\bar{p} > c$. Thus $\omega > 0$ and $\lambda < 1$. In Chapter 5 we showed that with one type of customer and constant marginal costs the firm expands output until the participation constraint binds, so the fact that one of the participation constraints binds here is not surprising. Substituting $\omega = 2(1-\lambda)$ and $\phi = 1-\lambda$ into (6.5c) and (6.5d) gives

$$\frac{dU^1(q^1)}{dq} = c - \frac{\lambda}{1-\lambda}(\bar{p}-c) + \left[\frac{dU^2(q^1)}{dq} - \frac{dU^1(q^1)}{dq}\right] \tag{6.6a}$$

$$\frac{dU^2(q^2)}{dq} = c - \frac{\lambda}{1-\lambda}(\bar{p}-c). \tag{6.6b}$$

The term in square brackets in (6.6a) is the marginal information rent, which is positive, so $dU^1(q^1)/dq > dU^2(q^2)/dq$, and we can confirm that $q^2 > q^1$ and IC^1 is satisfied.

Equation (6.6b) implies:

Proposition 6.1 *The high-demand customer consumes more than the efficient quantity.*

Proof. This follows simply from (6.6b) and the facts that $\bar{p} > c$ and $0 < \lambda < 1$. □

To see why the high-demand customer over-consumes we note that profits are $(\bar{p}-c)[q^1+q^2]$, so any feasible move that raises total sales is profitable. With IR^1 and IC^2 binding the AR constraint requires that

$$T^1 + T^2 = 2U^1(q^1) + U^2(q^2) - U^2(q^1) \leq \bar{p}[q^1+q^2].$$

Suppose that (6.6a) holds, but the firm sets customer 2's marginal utility equal to c. The first-order condition for q^1, equation (6.6a), and the fact that $\bar{p} > c$ imply that

$$\bar{p} > c > 2\frac{dU^2(q^1)}{dq} - \frac{dU^1(q^1)}{dq}. \tag{6.7}$$

Suppose that q^2 is increased by dq^2. Totally differentiating the binding AR constraint gives

$$\left[2\frac{dU^2(q^1)}{dq} - \frac{dU^1(q^1)}{dq} - \bar{p}\right] dq^1 = \left\{\bar{p} - \frac{dU^2(q^2)}{dq}\right\} dq^2 = (\bar{p} - c)dq^2 > 0 \tag{6.8}$$

where the second equality occurs because customer 2's marginal utility is assumed to be equal to c, and $(\bar{p} - c)dq^2 > 0$ by assumption. Using (6.7) we infer that the term in square brackets in (6.8) that multiplies dq^1 is negative, so $dq^1 < 0$. Total output, however, increases since

$$dq^1 + dq^2 = \frac{(\bar{p} - c)dq^2}{2\frac{dU^2(q^1)}{dq} - \frac{dU^1(q^1)}{dq} - \bar{p}} + dq^2 = \frac{\left(2\frac{dU^2(q^1)}{dq} - \frac{dU^1(q^1)}{dq} - c\right)}{\left(2\frac{dU^2(q^1)}{dq} - \frac{dU^1(q^1)}{dq} - \bar{p}\right)} dq^2 > 0.$$

Thus profits increase with the move to lower the marginal utility of customer 2 below c by raising q^2, because total output increases and each extra unit of output raises profits by $\bar{p} - c > 0$.

We noted above that we can see the firm's problem as being like that of an unregulated monopolist with lower costs, and inspection of (6.6a) and (6.6b) confirms this. A reduction in marginal cost raises the quantities that the unregulated firm offers, so both customers purchase more than when the firm faces no regulation. The fact that q^1 is higher than with no regulation implies that customer 2's surplus, and hence overall consumer surplus, is higher under AR than with no regulation, since the information rent increases with q^1. Figure 6.1 illustrates the comparison between the case of no regulation and average revenue regulation. The packages that the firm offers when there is no regulation are A and B. Customer 2 is indifferent between packages A and B, customer 1 strictly prefers A to B and is indifferent between A and not consuming. When average revenue regulation is imposed the firm offers packages C and D. Customer 1 chooses C but still obtains zero surplus. Customer 2 chooses D, and is better off than with package B.

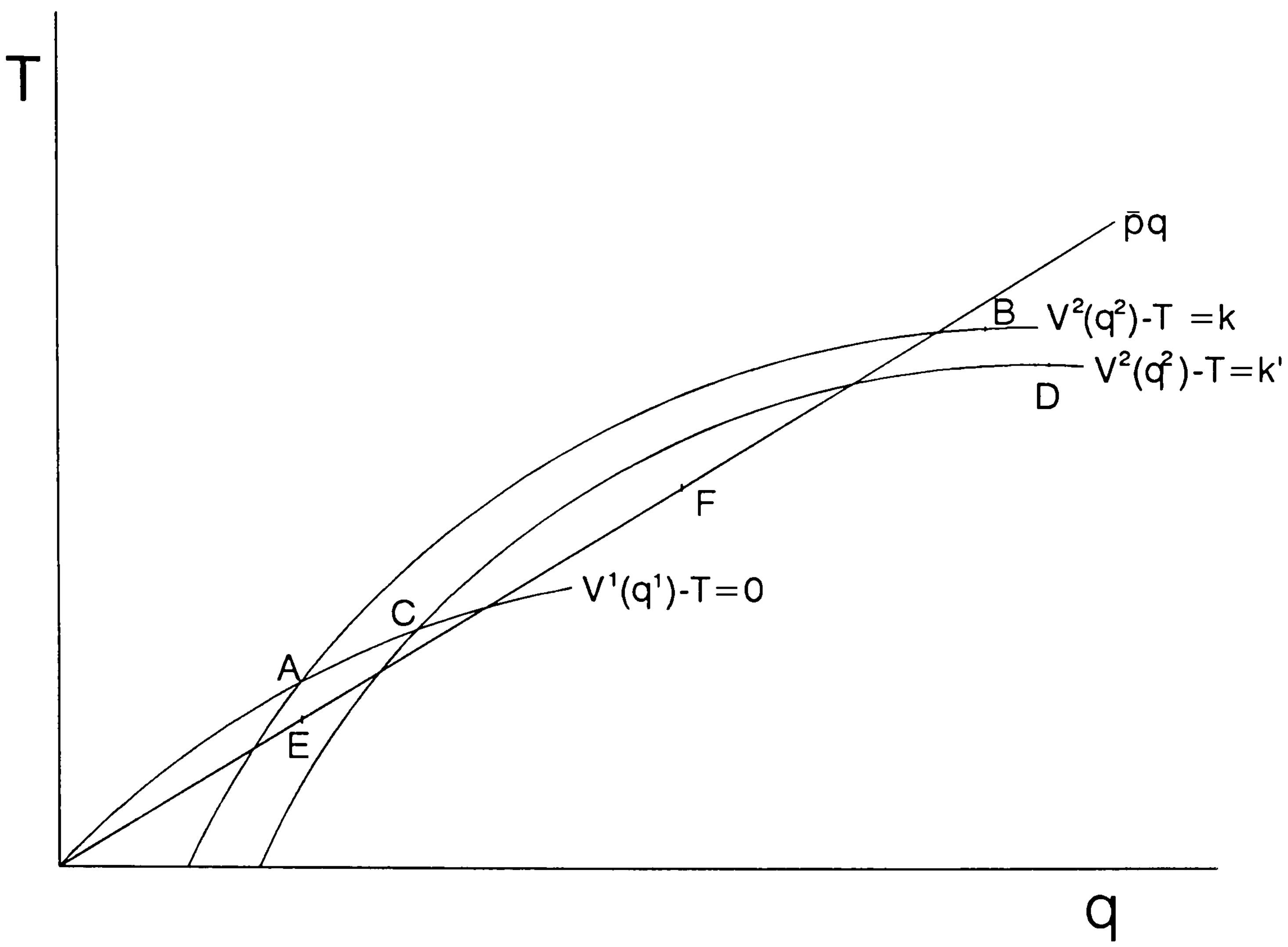


Figure 6.1 Nonlinear Pricing and Average Revenue Regulation

In previous chapters we have shown that average revenue regulation makes customers worse off than with uniform pricing at $\bar{p}$. This also holds in the nonlinear pricing case. We can see this intuitively in the case illustrated in Figure 6.1. With uniform pricing at $\bar{p}$ both customers are better off, since packages such as E and F are preferable for customers 1 and 2 respectively than C and D. More formally we have:

Proposition 6.2 *Aggregate consumers surplus is lower with regulated nonlinear pricing than with uniform pricing at* $\bar{p}$.

Proof. We denote the quantities chosen under the uniform tariff $T = \bar{p}q$ by $q^1(\bar{p})$ and $q^2(\bar{p})$, while quantities chosen under the nonlinear tariff with AR regulation are denoted by a subscript "AR". By concavity of the $U^i(.)$:

$$U^1(q^1(\bar{p})) + U^2(q^2(\bar{p})) \geq$$

$$U^1(q^1_{AR}) + \frac{dU^1(q^1(\bar{p}))}{dq}(q^1(\bar{p}) - q^1_{AR}) + U^2(q^2_{AR}) + \frac{dU^2(q^2(\bar{p}))}{dq}(q^2(\bar{p}) - q^2_{AR}).$$

When the tariff is uniform customer i chooses q^i where marginal utility equals $\bar{p}$, so $dU^i(q^i(\bar{p}))/dq = \bar{p}$ for i = 1, 2. Subtracting $\bar{p}(q^1 + q^2)$ from both sides yields

$$U^1(q^1(\bar{p})) - \bar{p}q^1(\bar{p}) + U^2(q^2(\bar{p})) - \bar{p}q^2(\bar{p}) \geq U^1(q^1_{AR}) - \bar{p}q^1_{AR} + U^2(q^2_{AR}) - \bar{p}q^2_{AR}.$$

The left-hand side is aggregate consumer surplus with the uniform tariff. With the AR constraint binding the right-hand side is $U^1(q^1_{AR}) - T^1 + U^2(q^2_{AR}) - T^2$, which is consumer surplus with the regulated nonlinear tariff. □

The fact that customers lose from allowing nonlinear pricing subject to an average revenue constraint, while the firm gains, suggests that the welfare consequences of AR are ambiguous. This is confirmed by example in the joint work of Armstrong, Cowan and Vickers (1995), and we now turn to the results of that paper. Customer types are indexed by θ which has a continuous distribution on support $[\underline{\theta}, \bar{\theta}]$ with a monotone hazard rate. Consumer surplus for a type-θ customer is $\theta U(q) - T(q)$, so the direct utility function $\theta U(q)$ is not quite as general as thee one used in the two-type model above. The result in Proposition 6.1 is shown to hold in the continuous-types case for weakly convex costs, and Proposition 6.2 also holds. Thus the welfare effect of nonlinear pricing consists in a positive effect on profits, but a negative effect on consumer surplus. As in the third-degree price discrimination case of Chapter 4, allowing nonlinear pricing has a positive effect on total output but leads to unequal marginal utilities. The given output level is best distributed with uniform pricing. When the price cap $\bar{p}$ is close to marginal cost the tariff is almost first-best, and allowing nonlinear pricing subject to an average revenue cap lowers welfare. But, as in the example in Chapter 4, progressive relaxation of the price cap increases the likelihood that regulated nonlinear pricing generates higher welfare than uniform pricing at $\bar{p}$. When the firm can offer an optional nonlinear tariff in addition to the uniform tariff with price equal to $\bar{p}$ the tariff can be characterized. It turns out that the marginal price is simply the minimum of $\bar{p}$ and the pure monopoly price. This optional tariff is Pareto-improving.

6.3 Nonlinear Pricing and Tariff Basket Regulation

It is straightforward to apply an average revenue constraint to nonlinear pricing since the calculation of average revenue does not depend on the type of tariff set. We have argued in previous chapters that the tariff basket (and fixed weight) price cap schemes have desirable properties when the firm practises either linear or two-part pricing. Since nonlinear pricing is widely practised by utilities, and has some desirable welfare properties relative to uniform pricing, it is of interest to discover whether the tariff basket scheme can be applied to nonlinear tariffs.

We characterize three possible cases, which differ in the degree of flexibility that the firm is allowed. In case (i) the regulator simply ignores any quantity discounts that the firm offers in calculating whether the price constraint is satisfied. This is the regime under which British Telecom has operated since 1993 – see Armstrong, Cowan and Vickers (1994, Chapter 7). Suppose the firm currently sets a two-part tariff, {A, p}. The tariff basket constraint in this case is simply

$$(A_{t-1} - A_t)N_{t-1} + (p_{t-1} - p_t)Q_{t-1} \geq 0. \tag{6.9}$$

where N_{t-1} is the number of customers and Q_{t-1} is the volume of sales in the previous period. Any units purchased at a lower marginal price than p_t are counted as having been sold at price p_t for the calculation of the period t+1 constraint. The rationale is presumably that if quantity discounts were allowed in the calculation of the price cap the firm could satisfy the constraint without lowering the prices that smaller customers face.

In case (ii) the tariff is multi-part, i.e. piecewise-linear. Such a pricing schedule is often called a "block-tariff", since it has different prices for different blocks of consumption. Such tariffs are common. An example is a three-part tariff

$$\begin{aligned} T &= A + p^1 q && \text{for } 0 < q \leq y \\ &= A + p^1 y + p^2[q - y] && \text{for } q > y, \end{aligned}$$

where y is the cut-off level of consumption. An access fee of A is paid, and for the first y units the marginal price is p^1, while for units in excess of y a marginal price of p^2 is paid. The British Gas tariff mentioned in 6.1 is of this type with y = 2,500 therms. It is as if smaller customers who are in the first block consume according to the two-part tariff $\{A,\ p^1\}$, while customers in the second block effectively have the two-part tariff $\{A + p^1 y,\ p^2\}$. The tariff basket constraint for a multi-part tariff would be the disaggregated version of (6.9):

$$\sum_{i=1}^{m}\left[(A^i_{t-1} - A^i_t)N^i_{t-1} + (p^i_{t-1} - p^i_t)Q^i_{t-1}\right] \geq 0. \tag{6.10}$$

Here m is the number of blocks in the tariff, N^i is the number of customers who consume in block i, Q^i is their consumption, and A^i is the *effective* access charge that a customer who chooses to be in block i pays. A^i is defined recursively by:

$$A^i = A^{i-1} + (p^{i-1} - p^i)y^i$$

where y^i is the cut-off quantity beyond which price p^i must be paid – see Wilson (1993, p 94). In constructing (6.10) it is as if quantities in different blocks are treated as being different products. The constraint in (6.10) allows the firm to trade off price reductions in one block against price increases in another. One possible disadvantage of such a constraint is that to operate it the definitions of the blocks, e.g. the cut-off number y in the three-part tariff, must be fixed. The firm is given less flexibility over tariff design than under average revenue regulation.

In case (iii) the firm offers a *menu* of two-part tariffs, rather than a given nonlinear tariff, and it is on this case that we shall focus. We use a two-type model, and shall first discuss the IR and IC constraints. The notation reverts to that of earlier chapters, so $V^i(p^i) - A^i$ is type i's surplus when choosing the two-part tariff $\{A^i, p^i\}$ designed for him. Note that the payment T^i in section 6.2 corresponds to $A^i + p^i q^i$ here. Again type 2 is the high-demand customer, with $q^2(p) > q^1(p)$ and $V^2(p) > V^1(p)$ for all values of p. The IR and IC constraints are analogous to (6.1) and (6.2):

$$(IR^i) \quad V^1(p^1) - A^1 \geq 0 \qquad \text{and} \quad V^2(p^2) - A^2 \geq 0$$

$$(IC^i) \quad V^1(p^1) - A^1 \geq V^1(p^2) - A^2 \qquad \text{and} \quad V^2(p^2) - A^2 \geq V^2(p^1) - A^1.$$

As in 6.2 we can show that if IR^1 and IC^2 are satisfied, IR^2 is also satisfied. When IC^2 holds $V^2(p^2) - A^2 \geq V^2(p^1) - A^1$. Since $V^2(p^1) > V^1(p^1)$ it follows that $V^2(p^2) - A^2 > V^1(p^1) - A^1$. If IR^1 holds then the right-hand side of this inequality is non-negative, so IR^2 is slack. There is a simple condition that ensures that IC^1 is satisfied. This is that $p^1 > p^2$, i.e. the larger customer faces a lower marginal price. This follows because when both IR^1 and IC^2 bind, IC^1 is

$$0 = V^1(p^1) - A^1 \geq V^1(p^2) - A^2 = V^1(p^2) - V^2(p^2) + V^2(p^1) - V^1(p^1).$$

The first equality is because IR^1 binds. The inequality is the IC^1 constraint, and the next equality follows from substitution from the binding IC^2 for A^2. The right-hand side can be written as

$$= \int_{p^1}^{p^2} \left(\frac{dV^1(\tilde{p})}{dp} - \frac{dV^2(\tilde{p})}{dp} \right) d\tilde{p}.$$

Since $dV^i(\tilde{p})/dp = -q^i(\tilde{p})$ by Roy's Identity the integrand is $q^2(\tilde{p}) - q^1(\tilde{p}) > 0$. As long as $p^1 > p^2$ the whole expression is negative and IC^1 is satisfied. This also implies that

$A^2 > A^1$. This follows because IC^1 can be rewritten as:

$$A^2 - A^1 \geq V^1(p^2) - V^1(p^1).$$

If $p^2 < p^1$ the right-hand side is positive and $A^2 > A^1$. This is reasonable: if $A^2 \leq A^1$ while $p^1 > p^2$ then no customer would choose tariff 1.

To derive the appropriate price constraint we make use of the convexity of consumer surplus, which implies that

$$V^1(p_t^1) - A_t^1 + V^2(p_t^2) - A_t^2 \geq \sum_{i=1}^{2} [A_{t-1}^i - A_t^i + (p_{t-1}^i - p_t^i)q^i(p_{t-1}^i)]. \tag{6.11}$$

The tariff basket constraint is that the right-hand side of (6.11) is non-negative. This is effectively the same constraint as (6.10). The cost of purchasing the quantities bought in period t-1 at the new prices must be no greater than at the old prices. The Lagrangian for the firm is

$$\mathcal{L} = \sum_{t=1}^{\infty} \beta^{t-1}\left\{ \sum_{i=1}^{2}(A_t^i + [p_t^i - c]q^i(p_t^i) + \lambda_t \sum_{i=1}^{2}[A_{t-1}^i - A_t^i + (p_{t-1}^i - p_t^i)q^i(p_{t-1}^i)]\right\}$$

$$+ \sum_{t=1}^{\infty} \beta^{t-1}\left\{\omega_t[V^1(p_t^1) - A_t^1] + \phi_t[V^2(p_t^2) - A_t^2 - V^2(p_t^1) + A_t^1]\right\}$$

where ω_t is the multiplier on IR_t^1 and ϕ_t is the multiplier on IC_t^2. The first-order conditions for the prices $\{A_t^i\}$ and $\{p_t^i\}$, assuming interior solutions, are:

$$\frac{\partial \mathcal{L}}{\partial A_t^1} = 1 - \lambda_t + \beta\lambda_{t+1} - \omega_t + \phi_t = 0 \tag{6.12a}$$

$$\frac{\partial \mathcal{L}}{\partial A_t^2} = 1 - \lambda_t + \beta\lambda_{t+1} - \phi_t = 0 \tag{6.12b}$$

$$\frac{\partial \mathcal{L}}{\partial p_t^1} = q^1(p_t^1) + (p_t^1 - c)\frac{dq^1(p_t^1)}{dp} - \lambda_t q^1(p_{t-1}^1) + \beta\lambda_{t+1}\left((p_{t+1}^1 - p_t^1)\frac{dq^1(p_t^1)}{dp} + q^1(p_t^1) \right)$$

$$- \omega_t q^1(p_t^1) + \phi_t q^2(p_t^1) = 0 \tag{6.12c}$$

$$\frac{\partial \mathcal{L}}{\partial p_t^2} = q^2(p_t^2) + (p_t^2 - c)\frac{dq^2(p_t^2)}{dp_t^2} - \lambda_t q^2(p_{t-1}^2) + \beta\lambda_{t+1}\left((p_{t+1}^2 - p_t^2)\frac{dq^2(p_t^2)}{dp_t^2} + q^2(p_t^2)\right)$$

$$- \phi_t q^2(p_t^2) = 0 \tag{6.12d}$$

Adding (6.12a) and (6.12b) gives

$$\omega_t = 2(1 - \lambda_t + \beta\lambda_{t+1}) = 2\phi_t.$$

Substituting these back into (6.12c) and (6.12d) yields

$$\frac{\partial \mathcal{L}}{\partial p_t^1} = (1 - \lambda_t + \beta\lambda_{t+1})\{q^2(p_t^1) - q^1(p_t^1)\} + [p_t^1 - c]\frac{dq^1(p_t^1)}{dp} + \lambda_t[q^1(p_t^1) - q^1(p_{t-1}^1)]$$

$$+ \beta\lambda_{t+1}(p_t^1 - p_{t+1}^1)\frac{dq^1(p_t^1)}{dp} = 0 \tag{6.13}$$

and

$$\frac{\partial \mathcal{L}}{\partial p_t^2} = [p_t^2 - c]\frac{dq^2(p_t^2)}{dp} + \lambda_t[q^2(p_t^2) - q^2(p_{t-1}^2)] + \beta\lambda_{t+1}[p_t^2 - p_{t+1}^2]\frac{dq^2(p_t^2)}{dp} = 0. \tag{6.14}$$

As was the case with the dynamic models in Chapter 3 little can be said in general about the dynamics, but two special cases are of interest.

We first characterize the case where $\beta = 0$ and the firm initially practices uniform pricing at $\bar{p}$. This is the fixed weights case. Suppose also that demands are linear, so $V^1(p^1) = (a^1 - b^1p^1)^2/2$ and $V^2(p^2) = (a^2 - b^2p^2)^2/2$. In this case, with $p_{t-1}^1 = p_{t-1}^2 = \bar{p}$ (6.13) and (6.14) can be written as

$$p^1 = c + (1 - \lambda)\frac{1}{b^1}\left[q^2(p^1) - q^1(p^1)\right] + \lambda(\bar{p} - p^1) \tag{6.13'}$$

$$p^2 = c + \lambda(\bar{p} - p^2). \qquad (6.14')$$

Taking the difference between (6.13′) and (6.14′) and rearranging gives:

$$p^1 - p^2 = \frac{1}{b^1}\left[q^2(p^1) - q^1(p^1)\right] > 0.$$

Thus the sufficient condition holds for IC^1 to be satisfied. We note that, as is usual with fixed weights regulation, pricing is not fully efficient since $p^2 > c$ when $\bar{p} > c$.

For the case of tariff basket regulation (where we allow $\beta > 0$) we can characterize the steady state, assuming that one exists. In steady state (6.13) yields

$$p^1 = c - \frac{(1 - \lambda(1-\beta))[q^2(p^1) - q^1(p^1)]}{dq^1(p^1)/dp}.$$

Since $1 - \lambda(1-\beta) = \phi \geq 0$, and $q^2(p^1) - q^1(p^1) > 0$, we have $p^1 \geq c$ in steady state, with $p^1 > c$ as long as the firm wants to practise nonlinear pricing, i.e. if IR^1 and IC^2 bind. In steady state (6.14) implies that $p^2 = c$. Since $p^1 > p^2$ in steady state the sufficient condition for IC^1 to hold is satisfied. Again in steady state the tariff basket mechanism induces efficient pricing. The top customer has a marginal price equal to marginal cost and the low consumer faces a marginal price which exceeds marginal cost.

In this section we have shown that the tariff basket scheme can be imposed in various ways when there is nonlinear pricing. The scheme retains its desirable properties, such as a tendency towards efficient pricing in the limit.

6.4 Conclusions

In this chapter we have explored the properties of average revenue and tariff basket price caps when the firm sets nonlinear tariffs. It was shown that average revenue regulation induces the firm to give the high-demand customer a package with an inefficiently high level of consumption, and the welfare consequences of such regulation are ambiguous. It was also shown that when the tariff basket is imposed on a firm that can practise nonlinear pricing the standard properties of that scheme still hold, i.e. pricing is efficient in the limit, and consumer surplus increases each period.

CHAPTER 7

EXTENSIONS

7.1 Introduction

In this chapter we consider several extensions of the analyses of the previous chapters. The objective is to test the robustness of the tariff basket (TB) scheme. We consider the effect of endogenizing *quality* in section 7.2. In section 7.3 we use a model of two-part tariffs where the number of customers is variable to analyze (i) how a regulator's concerns about *income distribution* can be included in a price cap and (ii) the effect of the presence of *network externalities* on the operation of the price cap. In section 7.4 we consider the effect of *nonstationarities*, in particular we allow for exogenous growth in demand in a simple two-part tariff model. Section 7.5 contains concluding comments.

7.2 Quality

Firms in competitive markets are constrained in their choice of quality by the need to ensure that their price-quality package is attractive. Regulated utilities, however, often have significant flexibility over the quality of service that they can offer. It can be a "gold-plated" or a basic service. One of the objectives of price cap regulation is the provision of incentives to cut costs through the reduction of slack, but capping prices can also induce firms to cut quality. The aim in this section is to show (i) that when quality is not directly regulated the level of quality provision will tend to be too low, and (ii) that under some conditions the tariff basket scheme can be adapted to include quality into the price cap, thus preserving its desirable properties. The focus will be on (ii), since it is already well-known that price cap regulation provides insufficient incentives for quality provision.

There is a considerable literature on monopoly and quality choice, much of it focussing on the case of an unregulated firms. Spence (1975) has the model that is most closely related to the analysis here. He assumes that the good is a search good, i.e. its quality can be assessed before purchase, and he analyzes the quality choices of monopolists that are either unregulated, face rate-of-return regulation or have a simple price cap. Other analyses of quality and monopoly include Besanko, Donnenfeld and White (1987, 1988), Laffont and Tirole (1990, 1993, Chapter 4) and Acton and Vogelsang (1989). Rovizzi and Thompson (1992, 1995) and Bishop, Rovizzi and Thompson (1991) assess the rôle of quality in the regulation of British utilities.

We allow for many products and multiple quality dimensions. The n products are search goods. The quality characteristics are listed in an m x 1 vector denoted by $\mathbf{x}$. The number of characteristics, m, is not necessarily the same as the number of products, n. The quality measures do not have to be aggregated. The representative consumer has a direct utility function $U(\mathbf{q}, \mathbf{x})$ and indirect utility function $V(\mathbf{p}, \mathbf{x})$. Demands are $q^i(\mathbf{p}, \mathbf{x}) = -\partial V/\partial p^i$ for all i. Profits are $\Pi(\mathbf{p}, \mathbf{x}) = \sum_{i=1}^{n} p^i q^i - C[\mathbf{q}, \mathbf{x}]$.

We first characterize the Ramsey problem of choosing prices and quality levels to maximize a weighted sum of consumer surplus and profits subject to a profit constraint, assuming that the regulator has full information. Formally the problem is to choose $\mathbf{p}$ and $\mathbf{x}$ to maximize $V + \alpha\Pi$, where $0 \leq \alpha \leq 1$, subject to $\Pi \geq \bar{\Pi}$. With μ denoting the multiplier on the profit constraint the first-order conditions are:

$$\frac{\partial \Pi}{\partial p^i} = \frac{1}{\alpha + \mu} q^i \qquad i = 1, \ldots, n \tag{7.1}$$

$$\frac{\partial \Pi}{\partial x^j} = \frac{-1}{\alpha + \mu} \frac{\partial V}{\partial x^j} \qquad j = 1, \ldots, m. \tag{7.2}$$

The equations for prices, (7.1), are standard. The new conditions are (7.2). These require that the marginal profitability of each type of quality is negative. For a given price vector a profit-maximizing firm chooses quality where the profit from extra *sales* gained with higher quality just equals the marginal cost of providing increased quality. The regulator, however, also takes account of the extra *utility* derived from non-marginal purchases. She wants the firm to expand quality beyond the profit-maximizing level. There is a positive externality involved in raising quality.

We now demonstrate that when the standard tariff basket formula applies, without allowing for the fact that quality levels are choice variables, the steady-state levels of quality provision will be inefficiently low. The Lagrangian for the firm is

$$\mathcal{L} = \sum_{t=1}^{\infty} \beta^{t-1} \Big\{ \Pi(\mathbf{p}_t, \mathbf{x}_t) + \lambda_t \sum_{i=1}^{n} (p_{t-1}^i - p_t^i) q_{t-1}^i \Big\}.$$

The first-order conditions for prices are the same as in Chapter 3, i.e. (3.3). We focus on the first-order conditions for quality:

$$\frac{\partial \Pi_t}{\partial x_t^j} + \beta \lambda_{t+1} \sum_{i=1}^{n} (p_t^i - p_{t+1}^i) \frac{\partial q_t^i}{\partial x_t^j} = 0. \quad j = 1, \ldots, m \tag{7.3}$$

Along the path to a steady state, if one is reached, the firm's quality choices are determined not only by current marginal profitability – the first term in (7.3) – but also by the effect quality has on future price constraints. An increase in x_t^j will tend to raise sales in period t. This relaxes the period t+1 constraint and means that prices in t+1 can be raised. If, however, a steady state is reached the second term in (7.3) is zero, and the firm simply chooses $\mathbf{x}$ to maximize current profits, given its prices. It sets $\partial \Pi / \partial x^j = 0$, while Ramsey quality-setting requires that the marginal profitability of quality is negative. Prices have the Ramsey property, taking the quality levels as given, but the combination of prices and qualities is inefficient. This is a version of the argument of Spence (1975) that a price-regulated monopolist tends to underinvest in quality, adapted to the case where the firm has multiple products and qualities and faces tariff basket regulation.

A simple linear example can be used to illustrate the possible distortion caused by not including quality in the price cap. Suppose that the firm sells one product and sets a two-part tariff. Consumer surplus is $V(\mathbf{p}, \mathbf{x}) = (x + 1 - p)^2/2 - A$, so quality increases and price reductions are perfect substitutes for the consumer. The fixed fee is denoted by A. Profits are $A + (p - c)(x + 1 - p) - \gamma x^2/2$ where marginal production cost is c and γ is the gradient of the marginal cost of quality improvement. A rise in x shifts the demand function up parallel to its old position at a constant rate, so consumer surplus is convex in x. The costs of quality provision must be convex enough to ensure that second-order conditions hold, and we assume $\gamma > 1$. The first-best outcome has $p = c$, $x = (1 - c)/(\gamma - 1)$ and $A = \gamma(1 - c)^2/2(\gamma - 1)^2$. The quality level is set to balance the marginal increase in consumer surplus with marginal cost pricing, $x + 1 - c$, against the marginal cost of quality, γx, and the access charge is set to cover the cost of quality provision. This generates a first-best welfare level of $W^* = \gamma(1 - c)^2/2(\gamma - 1)$. In Appendix 7.1 we analyze formally how the standard TB formula (without quality adjustment) works under these assumptions. To ensure convergence we have to make the very strong assumption that $\gamma > 1/(1 - \beta)$. The steady state is characterized by $p = c$ and $x = 0$. Pricing is efficient in the limit but there is no incentive for the firm to provide quality. The reason the quality level is driven to zero is that price is equal to marginal cost. With marginal cost pricing the firm wants to set quality as low as possible, since the marginal benefit from increased sales with improved quality is zero. Welfare in steady state is $W = (1 - c)^2/2 < W^*$. If the interest rate is 10 per cent, i.e. $\beta = 0.909$, then $\gamma \geq 11$ ensures convergence. When $\gamma = 11$ the limit of welfare under TB is exactly 10 per cent below W^*.

Under the tariff basket schemes analyzed in Chapters 3, 5 and 6 aggregate consumer surplus rises each period, because the cost of buying the previous period's bundle is constrained to decline (weakly) each period. In the current case, however, where no account is taken of the firm's ability to set its quality level, successive rises in consumer surplus are not guaranteed. How can we adjust the standard TB constraint to take account of quality? Suppose that $V(\mathbf{p}, \mathbf{x})$ is convex in both prices and quality levels. This implies

$$V_t - V_{t-1} \geq \sum_{i=1}^{n}(p^i_{t-1} - p^i_t)q^i_{t-1} + \sum_{j=1}^{m}\frac{\partial V_{t-1}}{\partial x^j_{t-1}}\left[x^j_t - x^j_{t-1}\right]. \tag{7.4}$$

A quality-adjusted price cap can be obtained from (7.4). The constraint is that the right-hand side of (7.4) should be non-negative. The firm is allowed to raise prices, but only if quality is increased sufficiently that in aggregate consumers are better off. We should note that the implementation of such a constraint is informationally demanding. When quality is not a choice variable the TB scheme requires only knowledge of prices and lagged quantities. It uses the fact that the derivatives of $V_{t-1}(.)$ with respect to prices are $(-1$ times) demands in period t-1, which are easily observable. The scheme based on (7.4) requires that the regulator knows past and present *qualities* (which is not too onerous), but she also must know the (lagged) marginal valuations of quality, which are much harder to observe than quantities.

The Lagrangian for the firm facing such a constraint is

$$\mathcal{L} = \sum_{t=1}^{\infty} \beta^{t-1} \left\{ \Pi(\mathbf{p}_t, \mathbf{x}_t) + \lambda_t \left(\sum_{i=1}^{n} (p_{t-1}^i - p_t^i) q_{t-1}^i + \sum_{j=1}^{m} \frac{\partial V_{t-1}}{\partial x_{t-1}^j} [x_t^j - x_{t-1}^j] \right) \right\}$$

and the Euler equations are:

$$\frac{\partial \Pi_t}{\partial p_t^k} - \lambda_t q_{t-1}^k + \beta \lambda_{t+1} \left(q_t^k + \sum_{i=1}^{n} (p_t^i - p_{t+1}^i) \frac{\partial q_t^i}{\partial p_t^k} + \sum_{j=1}^{m} \frac{\partial^2 V_t}{\partial x_t^j \partial p_t^k} [x_{t+1}^j - x_t^j] \right) = 0 \tag{7.5}$$

$i = 1, \ldots, n$

$$\frac{\partial \Pi_t}{\partial x_t^l} + \lambda_t \frac{\partial V_{t-1}}{\partial x_{t-1}^l} + \beta \lambda_{t+1} \left(\sum_{i=1}^{n} (p_t^i - p_{t+1}^i) \frac{\partial q_t^i}{\partial x_t^l} + \sum_{j=1}^{m} \frac{\partial^2 V_t}{\partial x_t^j \partial x_t^l} [x_{t+1}^j - x_t^j] - \frac{\partial V_t}{\partial x_t^l} \right) = 0 \tag{7.6}$$

$l = 1, \ldots, m.$

Again we focus on the steady state. This is characterized by the following conditions:

$$\frac{\partial \Pi}{\partial p^k} = (1-\beta)\lambda q^k, \text{ and}$$

$$\frac{\partial \Pi}{\partial x^j} = -(1-\beta)\lambda\frac{\partial V}{\partial x^j}.$$

Thus the necessary conditions for both Ramsey prices and Ramsey qualities hold in steady state. This quality-adjusted constraint has properties that are analogous to those of the simple tariff basket scheme described in Chapter 3.2.1. Consumer surplus rises each period and the firm's choices of prices and qualities are efficient in the long run.

The quality-adjusted tariff basket constraint for the linear example with a two-part tariff is $A_{t-1} - A_t + (p_{t-1} - p_t + x_t - x_{t-1})(x_{t-1} + 1 - p_{t-1}) \geq 0$. The constraint allows the firm to trade off increases in p_t against equal increases in x_t. Since this is exactly the rate at which consumers trade off price against quality the firm has been given the "correct" incentives to choose both price and quality. In Appendix 7.2 we confirm that steady-state pricing and quality setting are efficient, and that convergence is guaranteed as long as $\gamma > 1$. No further assumption about γ needs to be made.

If V(.) is not convex in both prices and qualities then the quality-adjusted price cap will not guarantee that consumer surplus rises each period, as prices might be raised without a sufficiently large increase in quality, and it will not necessarily have desirable properties. In this case a straightforward regulatory strategy would be to impose the simple TB price cap (without quality adjustment), in combination with a requirement that no quality level ever falls, i.e. $x_t^j \geq x_{t-1}^j$ for $j = 1, \ldots, m$. This ensures that consumer surplus does not fall each period. It suffers from the disadvantage that trade-offs between different elements of quality, and between quality as a whole and prices, are not allowed. Another case where the adjusted price cap based on (7.4) might not be feasible is when the lagged marginal valuations of quality are not available.

In this section we have shown that some simple adjustments to the standard tariff basket price cap can preserve its desirable properties, under some assumptions. These assumptions are, however, non-trivial. Quality regulation, as the British experience has shown, introduces significant complexity into price regulation.

7.3 Income Distribution and Network Externalities

In this section we analyze price cap regulation when (i) the regulator weights the surpluses of consumers differently, and (ii) there are network externalities. We focus on the case of TB regulation to avoid being unduly taxonomic and because of the general arguments in favour of the scheme. We first discuss the general structure of demand that is used in the analysis.

7.3.1 Two-Part Tariffs with a Variable Number of Consumers

We assume that the firm sets a two-part tariff that excludes some potential customers from purchasing. The demand structure specified here is standard in models of two-part tariffs (see Ng and Weisser, 1974, Leland and Meyer, 1976, Schmalensee, 1981b, Varian, 1989, Laffont and Tirole, 1993, Chapter 2). The total cost to the customer of buying q units of the product is $A + pq$, where A is the access charge and p is the marginal price. For notational ease we assume that there is only one product. A consumer of type s has indirect utility $V(p, s) - A$ if $q > 0$ and 0 if $q = 0$. The number of potential customers is normalized to unity. The type parameter s is continuously distributed on support $[\underline{s}, \bar{s}]$ with distribution function F(s) and density f(s). Customers with a higher value of s obtain more utility, i.e. $V_s > 0$ where the subscript indicates a partial derivative. The demand of a type-s consumer is $q(p, s) = -V_p$, using Roy's identity. It is assumed that customers with higher types have larger demands as well as greater utility, i.e. $-V_{ps} > 0$. There is a cut-off value of s, denoted by S, such that potential consumers with type parameters below S do not consume. S is implicitly defined by $V(p, S) = A$. Totally differentiating this gives

$$S_A = \frac{1}{V_s(p, S)} \quad \text{and } S_p = -\frac{V_P}{V_s} = \frac{q(p, S)}{V_s(p, S)}. \tag{7.7}$$

The number who consume is $N(p, A) = 1 - F(S(p, A))$. This can be thought of as the demand for access. Total demand for the product is $Q(p, A) = \int_{S(p, A)}^{\bar{s}} q(p, s)f(s)ds$. The elements of the aggregate Slutsky substitution matrix for N and Q are, using (7.7),

$$Q_p = \int_{S(p,\,A)}^{\bar{s}} q_p(p,\,s)f(s)ds - \frac{[q(p,\,S)]^2 f(S)}{V_s(p,\,S)} \tag{7.8a}$$

$$Q_A = -\frac{q(p,\,S)f(S)}{V_s(p,\,S)} \tag{7.8b}$$

$$N_p = -\frac{q(p,\,S)f(S)}{V_s(p,\,S)} \tag{7.8c}$$

$$N_A = -\frac{f(S)}{V_s(p,\,S)}. \tag{7.8d}$$

The interpretation of these partial derivatives is as follows. Equation (7.8a) gives the total effect of a rise in p on the demand for the product. Existing customers purchase less, while some marginal customers, each of whom consumed q(p, S), leave the market. The number who leave when p increases is given by (7.8c). The cross-price terms in (7.8b) and (7.8c) are equal, because utility is quasi-linear. Access and the product are complements, since the cross-price derivatives are negative. A higher access price reduces the number of customers by $f(S)/V_s$, and each of these consumers was purchasing q(p, S) of the product, so aggregate demand for the product decreases.

Aggregate consumer surplus is

$$CS(p,\,A) \equiv \int_{S(p,\,A)}^{\bar{s}} [V(p,\,s) - A]f(s)ds.$$

The derivatives are $CS_P = -Q(p,\,A)$ and $CS_A = -N(p,\,A)$. The Hessian matrix is -1 times the Slutsky matrix, whose elements are given in (7.8). The diagonal elements of the Hessian are positive. Its determinant is $Q_pN_A - Q_A^2 = -f(S)\int_S^{\bar{s}} q_p(p,\,s)f(s)ds/V_S(p,\,S) > 0$, so CS is convex in p and A. CS(p, A) acts in the same way as the consumer surplus function V($\mathbf{p}$) used in Chapters 2 and 3.

7.3.2 Income Distribution

In this subsection we examine the regulation of a firm by a regulator who is concerned about income distribution. Regulated utilities typically set two-part tariffs for their domestic customers. Setting an access charge improves efficiency by providing revenue to cover some of the fixed or customer-related costs of the network without having to raise marginal charges excessively above marginal costs, but an access charge acts like a regressive lump-sum tax for poorer customers. In Britain regulators have duties to show special concern to particular groups of customers such as the elderly and the disabled, and have used their powers when setting price caps to cap the real growth of access charges (e.g. in the cases of the telecommunications and gas industries) and to cap the growth of prices to the median domestic customer (in the telecommunications case). In theory it would be desirable for regulators to concentrate on efficiency issues and leave distribution to the tax-and-benefit system. In practice, however, regulators cannot ignore the distributional consequences of their decisions.

There is some literature on the distributional aspects of public utility pricing. Feldstein (1972a, b) examines optimal pricing when consumer surpluses have unequal weights in social welfare.[1] Auerbach and Pellechio (1978) extend Feldstein's analysis to the case where some customers are excluded. Spulber (1989) and Cabral (1990) derive Ramsey formulae for the two-part tariff case, and Cabral calibrates his model to determine optimal prices for the Portuguese telephone system. Ross (1984) shows that the implicit weights of a regulator can be inferred by inverting the first-order conditions in the weighted-Ramsey problem. Sharkey and Sibley (1993) examine the effect of different weights on different consumers' utilities when nonlinear tariffs are available.

We first consider the regulator's complete information problem, which is akin to the Ramsey problem in Chapter 2. She chooses p and A to maximize weighted consumer surplus plus (weighted) profits subject to a minimum profit constraint. The surplus of a consumer of type s is weighted by w(s). This can be thought of as the

[1] See also Atkinson and Stigltiz (1980, Chapter 12), Diamond (1975), Mirrlees (1975) and Deaton (1977) for discussion of many-person Ramsey tax rules.

social marginal utility of income, although strictly speaking s cannot be income itself because this would make the use of consumer surplus invalid. Weighted consumer surplus is

$$\int_{S(p,\,A)}^{\bar{s}} [V(p,\, s) - A]w(s)f(s)ds,$$

which is convex in p and A. The first-order conditions in the Ramsey problem for p and A are

$$-\int_{S(p,\,A)}^{\bar{s}} q(p,\, s)w(s)f(s)ds + (\alpha + \mu)\Pi_p = 0 \tag{7.9}$$

$$-\int_{S(p,\,A)}^{\bar{s}} w(s)f(s)ds + (\alpha + \mu)\Pi_A = 0 \tag{7.10}$$

where μ is the multiplier on the profit constraint and α is the weight on profits. Dividing (7.9) by (7.10) yields

$$\frac{\Pi_p}{\Pi_A} = \frac{\int_S^{\bar{s}} q(p,\, s)w(s)f(s)ds}{\int_S^{\bar{s}} w(s)f(s)ds} = \frac{Q^W(p,\, A)}{N^W(p,\, A)}. \tag{7.11}$$

Q^W denotes the weighted sum of demands and N^W is the weighted number of consumers.[2] This is a typical Ramsey result with the addition of welfare weights.

Some special sets of weights can be considered. The utilitarian welfare function has $w(s) = 1$ for all s. This generates the standard Ramsey pricing result when the regulator is not concerned with distribution, i.e. $\Pi_p/\Pi_A = Q/N$. In this case the regulator is interested in the average consumer. A regulator that is concerned exclusively with the median consumer puts all the weight on the median type, σ, defined by $F(\sigma) = 0.5$. Condition (7.11) is then $\Pi_p/\Pi_A = q(p,\, \sigma)$. A regulator with special concern for those with low utility can be characterized as having higher weights

[2]Spulber (1989, Chapter 7) derives an equivalent expression.

for lower-s types.

As the relative weight on low-demand customers increases A becomes lower and p higher. This can be shown as follows. Suppose that the regulator is initially optimizing, and has some given set of weights denoted by "W". Moving along the iso-profit curve, we have $dA = -(\Pi_p/\Pi_A)dp = -(Q^W/N^W)dp$, where the second equality follows because the regulator optimizes. If the weights now change to X, and more weight is given to lower-s types, the effect on consumer surplus (with the new weights), dCS^X, of the move along the iso-profit curve is:

$$dCS^X = -N^X.dA - Q^X dp = dp.N^X\left[\frac{Q^W}{N^W} - \frac{Q^X}{N^X}\right]$$

so if $Q^W/N^W > Q^X/N^X$ then dCS^X has same sign as dp. But the X weights give greater weight to smaller values of q(p, s), so it must be the case that $Q^W/N^W > Q^X/N^X$. Thus the greater the concern for smaller customers, the higher the marginal price should be higher, and the lower the access charge should be. This brings into focus the equity-efficiency trade-off that regulators face. In the U.K., for example, British Telecom has consistently argued that efficiency requires it should be allowed to rebalance its tariff so that the access charge is increased and the marginal price lowered, while the regulator has constrained the rate at which such rebalancing is implemented.

It is straightforward to generalize the standard tariff basket mechanism to the present case. The constraint that the firm faces in period t is given by

$$\int_{s_{t-1}}^{\bar{s}} [A_{t-1} - A_t + (p_{t-1} - p_t)q(p_{t-1}, s)]w(s)f(s)ds$$

$$= [A_{t-1} - A_t]N^W_{t-1} + (p_{t-1} - p_t)Q^W_{t-1} \geq 0.$$

At the proposed prices the *weighted* aggregate cost of last period's consumers buying last period's quantities should not rise. Given convexity the constraint ensures that *weighted* consumer surplus rises. The firm's Lagrangian is:

$$\sum_{t=1}^{\infty}\beta^{t-1}\{\Pi(p_t, A_t) + \lambda_t[(A_{t-1} - A_t)N^w_{t-1} + (p_{t-1} - p_t)Q^w_{t-1}]\} \tag{7.12}$$

The first-order conditions give:

$$\frac{\partial \Pi_t}{\partial p_t} - \lambda_t Q^w_{t-1} + \beta\lambda_{t+1}\left\{(A_t - A_{t+1})\frac{\partial N^w_t}{\partial p_t} + (p_t - p_{t+1})\frac{\partial Q^w_t}{\partial p_t} + Q^w_t\right\} = 0$$

$$\frac{\partial \Pi_t}{\partial A_t} - \lambda_t N^w_{t-1} + \beta\lambda_{t+1}\left\{N^w_t + (A_t - A_{t+1})\frac{\partial N^w_t}{\partial A_t} + (p_t - p_{t+1})\frac{\partial Q^w_t}{\partial A_t}\right\} = 0.$$

The two equations characterizing the steady state are:

$$\Pi_p = \lambda(1-\beta)Q^w \tag{7.13}$$

$$\Pi_A = \lambda(1-\beta)N^w. \tag{7.14}$$

Dividing (7.13) by (7.14) gives an equation of the form of (7.11). Thus, in steady state there are Ramsey prices, in the sense that welfare is maximized subject to the firm achieving the level of profit it actually achieves. The consumer in whom the regulator is interested (e.g. the average or median customer) is better off each period that the firm chooses to alter prices, since the cost of buying his previous bundle has not risen.

Another possible objective for the regulator is to maximize the number of customers subject to a profit constraint. A vote-maximizing government might give the regulator the duty to ensure that as many customers as possible are served. Alternatively this can be seen as a type of Rawlsian objective for reasons that will become apparent. The first-order conditions for this problem are

$$N_p + \mu\Pi_p = 0 \tag{7.15}$$

$$N_A + \mu\Pi_A = 0. \tag{7.16}$$

Dividing (7.15) by (7.16) and using (7.8c) and (7.8d) gives:

$$\frac{\Pi_p}{\Pi_A} = q(p, S(p, A)). \tag{7.17}$$

This can be thought of as a special case of (7.11) where all the weight is allocated to the marginal customers. How can this be implemented by a price cap? What is needed is a scheme that guarantees that the customer who was marginal in period $t-1$ derives at least as much utility in period t. By convexity

$$[V(p_t, S_{t-1}) - A_t] - [V(p_{t-1}, S_{t-1}) - A_{t-1}] \geq [p_{t-1} - p_t]q(p_{t-1}, S_{t-1}) + A_{t-1} - A_t$$

where the left-hand side is the change in the net surplus of the customer who was marginal in period $t-1$. The constraint that ensures that such a customer is better off in period t is

$$A_{t-1} - A_t + (p_{t-1} - p_t)q(p_{t-1}, S(p_{t-1}, A_{t-1})) \geq 0. \tag{7.18}$$

When the constraint holds $V(p_t, S_{t-1}) - A_t \geq V(p_{t-1}, S_{t-1}) - A_{t-1} = 0 = V(p_t, S_t) - A_t$, implying that $S_t \leq S_{t-1}$, and ensuring that $N_t \geq N_{t-1}$. Since the marginal customer earned zero surplus in period $t-1$ and is better off in period t, the new marginal customer is of a lower type and the number of customers must have risen.

The Lagrangian for the firm facing (7.18) is

$$\sum_{t=1}^{\infty} \beta^{t-1}\Pi(p_t, A_t) + \lambda_t[A_{t-1} - A_t + (p_{t-1} - p_t)q(p_{t-1}, S(p_{t-1}, A_{t-1})]. \tag{7.19}$$

The first-order conditions give:

$$\frac{\partial \Pi_t}{\partial p_t} - \lambda_t q_{t-1} + \beta\lambda_{t+1}\left\{(p_t - p_{t+1})\left(\frac{\partial q_t}{\partial p_t} + \frac{\partial q_t \partial S_t}{\partial S_t \partial p_t}\right) + q_t\right\} = 0$$

$$\frac{\partial \Pi_t}{\partial A_t} - \lambda_t + \beta\lambda_{t+1}\left\{1 + (p_t - p_{t+1})\frac{\partial q_t \partial S_t}{\partial S_t \partial A_t}\right\} = 0.$$

In steady state we have:

$$\Pi_p = \lambda(1-\beta)q(p,\, S(p,\, A)) \tag{7.20}$$

$$\Pi_A = \lambda(1-\beta) \tag{7.21}$$

yielding the standard Ramsey result.

This subsection has shown that the model of two-part tariffs can be used to analyze the consequences of assuming that the regulator is concerned about income distribution. The Ramsey problem can be generalized and tariff basket regulation produces the usual efficiency properties in the long run. The tariff basket scheme ensures very generally that the firm's objective of maximizing the present value of profits is aligned with the regulator's objective of maximizing welfare.

7.3.3 Network externalities.

In this subsection we discuss the effect of assuming that there are network externalities. A standard example of a network externality is in telecommunications. Having a telephone or electronic mail is only of use if others have access to the same network as well, and as the number of subscribers increases existing users gain extra utility. For a general discussion of network externalities see Katz and Shapiro (1994). We use a simple representation of network externalities that follows Littlechild (1975). The direct utility function is U(q, N, s). The utility of a customer of type s depends on the quantity of calls he pays for, q, and on the number of subscribers, N. The person who pays for the call is assumed to obtain all of the value of the externality. Since each customer is small, N is taken to be constant when individual s chooses his desired quantity, yielding a demand function q(p, N, s). The indirect utility function is $V(p, N, s) - A \equiv U(q(p, N, s), N, s) - pq(p, N, s) - A$. As usual $V_p = -q$. Since the number of subscribers is $N = 1 - F(S)$, the cut-off type, S, is given by $V(p, 1 - F(S), S) = A$. The comparative statics of S are:

$$S_A = \frac{1}{V_s(p, N, S) - V_N(p, N, S)f(S)} \tag{7.22a}$$

$$S_p = \frac{q(p, N, S)}{V_s(p, N, S) - V_N(p, N, S)f(S)}. \tag{7.22b}$$

We assume the externality is small enough that S_A, and thus S_p, are both positive. Effectively the presence of the externality makes the type of the marginal customer more sensitive to increases in prices. The elements of the aggregate Slutsky substitution matrix for Q and N are

$$Q_p = \int_{S(p,\,A)}^{\bar{s}} [q_p(p, N, s) + q_N(p, N, s)N_p]f(s)ds - q(p, S)f(S)S_p \tag{7.23a}$$

$$Q_A = \int_{S(p,\,A)}^{\bar{s}} q_N(p, N, s)N_A f(s)ds - q(p, S)f(S)S_A \tag{7.23b}$$

$$N_p = -f(S)S_p \tag{7.23c}$$

$$N_A = -f(S)S_A. \tag{7.23d}$$

Because of the externality the Slutsky matrix is no longer symmetric. The effect of a rise in p on Q can be explained as follows. The non-marginal customers buy less because of both the direct effect of the price increase and the fact that marginal customers drop out. The reduction in numbers hooking up also directly reduces demand by the final term in (7.23a). Similarly an increase in A reduces the number of calls made because of the reduction in the demands of non-marginal customers as N falls and the reduction in the number of customers.

The Lagrangian for the standard Ramsey problem is:

$$\int_{S(p,\, A)}^{\bar{s}} [V(p,\, N(p,\, A),\, s) - A]f(s)ds + (\alpha + \mu)[\Pi(p,\, A) - \bar{\Pi}].$$

The Ramsey condition is

$$\frac{\Pi_p}{\Pi_A} = \frac{Q - N_p \int_S^{\bar{s}} V_N(p,\, N,\, s)f(s)ds}{N - N_A \int_S^{\bar{s}} V_N(p,\, N,\, s)f(s)ds}.$$

The difference between this and the standard Ramsey condition, $\Pi_p/\Pi_A = Q/N$, occurs because of the consumption externality. A small rise in price, dp, lowers consumer surplus because customers must pay Qdp extra to purchase the initial quantity. But because the number of customers falls by $N_p dp$, non-marginal consumers are worse-off by $\int_S^{\bar{s}} V_N(p,\, N,\, s)f(s)ds$ for each customer that drops out. A similar result holds for the effect of a marginal change in A.

It is clear that the standard tariff basket constraint does not generate Ramsey pricing in the limit because of the presence of the externality. In steady state the firm ensures that $\Pi_p/\Pi_A = Q/N$. We would expect that this steady state is characterized by an access charge that is too high and a marginal price that is too low, given the profits of the firm. We now demonstrate this more formally. Taking small changes dp and dA along the iso-profit curve we have $dA = -(\Pi_p/\Pi_A)dp = -(Q/N)dp$. The effect of this on consumer surplus is

$$\begin{aligned} dCS &= -Qdp + N_p dp \int_S^{\bar{s}} V_N(p,\, N,\, s)f(s)ds - NdA + N_A dA \int_S^{\bar{s}} V_N(p,\, N,\, s)f(s)ds \\ &= \left\{ \int_S^{\bar{s}} V_N(p,\, N,\, s)f(s)ds \right\} [N_p dp + N_A dA] \end{aligned}$$

$$= \left\{ \int_{S}^{\bar{s}} V_N(p, N, s) f(s) ds \right\} f(S) S_A \left[-q(p, N, S) + \frac{Q}{N} \right] dp.$$

The second equality follows from the fact that $dA = -(Q/N)dp$, and in the third equality we also use (7.23c) and (7.23d). It follows that as long as the marginal customer consumes less than the average customer, i.e. $q(p, N, S) < Q/N$, the change in consumer surplus has the same sign as dp.

What alternative types of regulation could be used in the presence of externalities? As in the quality case examined in 7.2 then if CS(p, A) is still convex in p and A, as we might expect if the externalities are small, an adjusted price cap scheme might be possible. Convexity implies

$$CS_t - CS_{t-1} \geq \left[Q_{t-1} - N_p(p_{t-1}, A_{t-1}) \int_{S}^{\bar{s}} V_N(p_{t-1}, N_{t-1}, s) f(s) ds \right] (p_{t-1} - p_t)$$

$$+ \left[N_{t-1} - N_A(p_{t-1}, A_{t-1}) \int_{S}^{\bar{s}} V_N(p_{t-1}, N_{t-1}, s) f(s) ds \right] (A_{t-1} - A_t) \geq 0. \qquad (7.24)$$

The constraint would be that the right-hand side of the above inequality must be non-negative. The difficulty, of course, in implementing such a constraint is that estimating the externality effect, $\int_{S}^{\bar{s}} V_N(p_{t-1}, N_{t-1}, s) f(s) ds$, and N_p and N_A is likely to be very difficult. If the regulator is able to estimate these effects then she probably has enough information to implement Ramsey prices without the need for a price cap.

Of course it might be the case that the externality effect is so small that it is not worth attempting to adapt the standard TB formula. Baumol (1995, page 278) argues that "it is . . . quite implausible that these uninternalized externalities are large in magnitude." He claims that sensitivity tests in which Ramsey prices were reestimated using high assumed values of the externalities show little effect on the calculated Ramsey values. If it is thought that the externality effects are large, but that implementing a price cap based on (7.24) is not feasible then some alternative type of regulation must be used. The key is to note that if it faces the ordinary TB constraint the firm will tend to set A too high and p too low, thus reducing the number of subscribers below the socially efficient level. It needs to be given incentives to raise the number of subscribers. One option is to impose the standard TB price cap, but with the additional constraint that the number of subscribers cannot fall. Alternatively, the regulator could use a weighted version of the tariff basket, as described in 7.3.2. For example the Rawlsian constraint (7.18) could be used. This would ensure that, in steady state, the number of subscribers is maximized (subject to the given level of profits), and each period the number does not fall. A third possibility is to introduce a tariff that is specifically designed to encourage small users to join the network. Such a tariff would have a low access charge and a high marginal price. British Telecom has been required to offer an optional low-user tariff since 1989. Users have the option to pay half-price rental charges and obtain 30 units of free calls per quarter. Units used above the free block, however, are charged at a premium rate. This is estimated to have benefitted between 1.5 and 2 million customers, but although it might encourage network membership it is not clear that the tariff is optimal (Armstrong, Cowan and Vickers, 1994, p 227).

7.4 Growth in Demand

So far we have assumed that demand conditions are stationary. In practice many utilities face exogenous growth in demand because of growth in real incomes and in population. Tastes can also shift over time, for example if technology improves for a complementary product. What is the effect of such shifts in demand functions on the operation of the tariff basket scheme? To analyze the case of shifting demand we shall use the linear version of the two-part tariff model of Chapter 5.[3] Demand is $q_t = a_t - p_t$ where the intercept, a_t, can vary over time. The cost function is $K + cq_t$, where K is the fixed cost. It is straightforward to allow for marginal costs that increase

[3]Brennan (1991) presents a two-period model of the tariff basket in a linear pricing context and shows that inefficiencies can result.

with output, but this adds too much notation. Assuming constant marginal costs does not alter the essence of the argument. The firm is assumed to know the sequence of demand intercepts $\{a_{t+i}\}_{i=1}^{\infty}$.

We need to restrict the rate of growth of a_t to ensure that the firm's problems are well-defined. First, we assume that if $a_t = (1+g)^t a_0$, i.e. it grows at a constant proportional rate, the growth rate is below the interest rate, so $g < r \equiv 1/\beta - 1$. This helps to ensure that the present value of the firm is finite. A simple benchmark case is where the firm keeps p and A fixed at their period 0 values. The present value of the increases in profits from shifts in a_t would then be:

$$\frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}(p_0 - c)(a_t - a_{t-1}).$$

In each period profits increase by the profit margin $(p_0 - c)$ times the increase in sales, $a_t - a_{t-1}$. The assumption that a_t grows at rate g implies $a_t - a_{t-1} = a_0(1+g)^{t-1}g$, so the present value of the increase in profits is:

$$= \frac{(p_0 - c)a_0 g}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}(1+g)^{t-1} = \frac{(p_0 - c)a_0 g}{1-\beta}.\frac{1}{1-\beta(1+g)}.$$

For this to be positive and finite we need $\beta(1+g) < 1$, i.e. $g < r$. This is necessary to ensure that the transversality condition is satisfied, but sufficiency requires an even stricter assumption about the growth rate. In Appendix 7.3 we discuss transversality conditions for the standard tariff basket problem. This requires that $\sqrt{\beta}(1+g) < 1$, which implies that $g < \sqrt{1+r}\ -1$. Thus if the interest rate is 5 per cent g must be less than 2.47 per cent.

The first case we analyze is where the firm faces the standard tariff basket constraint, i.e. $A_{t-1} - A_t + (p_{t-1} - p_t)q_{t-1} \geq 0$. By a similar argument to that used in equations (5.5) to (5.7) the firm's maximand can be written as

$$\frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}\Big(A_t - A_{t-1} + (p_t - c)q_t - (p_{t-1} - c)q_{t-1}\Big).$$

When the constraint binds we have $A_t - A_{t-1} = (p_{t-1} - p_t)q_{t-1}$, which yields:

$$\frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}\Big((p_t - c)(a_t - a_{t-1} + p_{t-1} - p_t)\Big).$$

The increase in profits in period t is the margin, $p_t - c$, times the increase in sales. Increased sales can come from two sources, the shift in the demand function, $a_t - a_{t-1}$, and the fall in price, $p_{t-1} - p_t$, which is a choice variable. The Euler condition for p_t is:

$$\beta p_{t+1} - 2p_t + p_{t-1} = -[a_t - a_{t-1} + (1-\beta)c]. \tag{7.25}$$

Using the lag-operator method suggested by Sargent (1987, Chapter 9) for solving Euler equations we can write (7.25) as

$$\beta(1-\chi_1 L)(1-\chi_2 L)p_{t+1} = -[a_t - a_{t-1} + (1-\beta)c] \tag{7.25$'$}$$

where L is the lag operator $(Lp_{t+1} \equiv p_t)$ and χ_1 and χ_2 are the roots which satisfy $\chi_1\chi_2 = 1/\beta$ and $\chi_1 + \chi_2 = 2/\beta$. The values of χ_1 and χ_2 are the same as in 5.2.1, i.e.

$$\chi_1 = \frac{1-\sqrt{1-\beta}}{\beta} \quad \text{and} \quad \chi_2 = \frac{1+\sqrt{1-\beta}}{\beta}$$

and they satisfy $0 < \chi_1 < 1$ and $\chi_2 > 1/\sqrt{\beta} > 1$. Multiplying both sides of (7.25′) by $1/(1-\chi_2 L) = -(\chi_2 L)^{-1}/\big(1-(\chi_2 L)^{-1}\big)$, dividing by β, and rearranging we can solve the unstable root, χ_2, forward to obtain

$$p_{t+1} = \chi_1 p_t + \chi_1\sum_{i=0}^{\infty}\left(\frac{1}{\chi_2}\right)^i[a_{t+1+i} - a_{t+i} + (1-\beta)c] + Z(\chi_2)^{t+1} \tag{7.26}$$

For the transversality condition to hold we require that Z, the coefficient of $(\chi_2)^{t+1}$, is

0. For details of the transversality condition see Appendix 7.3. If $Z \neq 0$ then since $\chi_2 > 1/\sqrt{\beta}$ the final term will dominate as t increases and p_{t+1} will fail to be of exponential order less than $1/\sqrt{\beta}$.

Lagging (7.26) by one period gives

$$p_t = \chi_1 p_{t-1} + \chi_1 \sum_{i=0}^{\infty} \left(\frac{1}{\chi_2}\right)^i [a_{t+i} - a_{t-1+i} + (1-\beta)c]. \tag{7.27}$$

The firm uses its knowledge about future changes in the demand intercept to determine its optimal price in period t. We can take the term $(1-\beta)c$ out of the summation term to give

$$p_t = \chi_1 p_{t-1} + (1-\chi_1)c + \chi_1 \sum_{i=0}^{\infty} \left(\frac{1}{\chi_2}\right)^i [a_{t+i} - a_{t-1+i}]. \tag{7.27$'$}$$

This is proved in Appendix 7.4.

From (7.27′) we infer that when the demand intercept is growing p_t is always above the price level that would be set with no demand growth. Without demand growth the price path is determined by $p_t = \chi_1 p_{t-1} + (1-\chi_1)c$. We shall explore two special cases of (7.27′). In case (i) we suppose that a_t increases each period by a constant *absolute* amount, i.e. $a_{t+i} - a_{t-1+i} = e$. Equation (7.27′) becomes

$$p_t = \chi_1 p_{t-1} + (1-\chi_1)c + \frac{e}{\beta(\chi_2 - 1)}. \tag{7.27$''$}$$

This is a first-order difference equation with particular solution:

$$p = c + \frac{e}{\chi_1(1-\beta)},$$

using the manipulations used in Appendix 7.4. In steady state the firm sets a constant price above marginal cost. With a positive margin it benefits from future demand growth, whereas if it set price equal to c it would derive no advantage from the future upward shifts in demand. Adding the complementary solution gives the expression for p_t when the demand intercept shifts up by e each period:

$$p_t = c + \frac{e}{\chi_1(1-\beta)} + (\chi_1)^t \left[p_0 - c - \frac{e}{\chi_1(1-\beta)} \right]. \tag{7.28}$$

The difference between the price with demand growth and without is

$$\frac{e(1-(\chi_1)^t)}{\chi_1(1-\beta)} > 0.$$

Thus the price is always lower when $e = 0$.

Case (ii) has a constant *proportional* growth rate of the demand intercept, i.e. $a_t = (1+g)a_{t-1} = (1+g)^t a_0$. In this case (7.27′) becomes

$$p_t = \chi_1 p_{t-1} + (1-\chi_1)c + \frac{a_0 g(1+g)^{t-1}}{\beta[\chi_2 - (1+g)]}. \tag{7.29}$$

The assumption that $1+g < 1/\sqrt{\beta}$ entails that $\chi_2 > 1/\sqrt{\beta} > 1+g$, so the final term is positive when $g > 0$. This yields the following expression for p_t as a function of past values of a_t:

$$p_t = c + (\chi_1)^t(p_0 - c) + \frac{g}{\beta[\chi_2 - (1+g)]} \sum_{i=1}^{t} (\chi_1)^{i-1} a_{t-i}. \tag{7.30}$$

Again p_t is above the level that would obtain if $g = 0$, and this time a steady state is not reached. Equation (7.30) reveals that if $p_0 > c$ and $g > 0$ price can never equal marginal cost. As t increases the second term in (7.30), which acts to lower p_t towards c, tends to zero and the final term dominates. Demand growth leads eventually to a rising price path.

What are the welfare properties of the standard TB scheme when there is demand growth? The problem is that the standard TB constraint does not guarantee consumer surplus increases when demand grows. We know that in the long run price increases further away from the first-best level, c, and that increases in consumer surplus cannot be guaranteed under the standard TB scheme. This suggests that we should look for an adaptation of the scheme. Convexity of consumer surplus implies

$$\begin{aligned} & V(p_t - a_t) - A_t - [V(p_{t-1} - a_{t-1}) - A_{t-1}] \\ & \geq (A_{t-1} - A_t) + (p_{t-1} - p_t + a_t - a_{t-1})(a_{t-1} - p_{t-1}). \end{aligned} \tag{7.31}$$

The right-hand side of (7.31) could be required to be non-negative, thus guaranteeing that consumer surplus increases (weakly) each period. In the standard TB case the firm is only allowed to increase the access charge if it cuts the marginal price. Under the cap based on (7.31) it is only allowed to raise A if it increases *sales.* These are equivalent when the demand intercept is stationary. Implementing such a price cap requires that the regulator either knows the sequence of a_t, or knows that demand is linear and can observe current and lagged demand. This is because the final term in (7.31) is $(q_{t-1} - q_t)q_{t-1}$, whereas under TB the equivalent term is $(p_{t-1} - p_t)q_{t-1}$.

The firm's objective under such a price cap is to choose the sequence of price to maximize

$$\frac{1}{1-\beta} \sum_{t=1}^{\infty} \beta^{t-1}\Big[(p_t - c)(a_t - a_{t-1} + p_{t-1} - p_t) + (a_t - a_{t-1})(a_{t-1} - p_{t-1})\Big].$$

The Euler condition is

$$\beta p_{t+1} - 2p_t + p_{t-1} = -[a_t - a_{t-1} - \beta(a_{t+1} - a_t) + (1-\beta)c],$$

which can be solved to give a difference equation for p_t similar to that in (7.27):

$$p_t = \chi_1 p_{t-1} + (1-\chi_1)c + \chi_1 \sum_{i=0}^{\infty}\left(\frac{1}{\chi_2}\right)^i [a_{t+i} - a_{t-1+i} - \beta(a_{t+1+i} - a_{t+i})]. \qquad (7.32)$$

We should note that relative to (7.27) there is an extra term within the final term in (7.32), $-\beta(a_{t+1+i} - a_{t+i})$, that is negative when demand grows. Thus if $p_0 > c$, subsequent prices are always closer to marginal cost under the adjusted TB scheme than under the simple TB scheme, and welfare is always higher.

In case (i), where $a_t - a_{t-1} = e$, (7.32) reduces to

$$p_t = c + \frac{e}{\chi_1} + (\chi_1)^t(p_0 - c - e/\chi_1). \qquad (7.33)$$

It is straightforward to check that this induces a price that is always closer to c than the price under the standard TB scheme in (7.28), because $1-\beta < 1$. It is still the case, however, that the steady-state value of p is above c, although in this case for a high value of β the deviation is small. For example if the interest rate is 5 per cent, so $1/(1-\beta) = 21$, the steady-state deviation of p from c under the standard TB scheme is 21 times as great as under the adjusted scheme.

In case (ii) with a constant growth rates of a_t (7.32) becomes

$$p_t = \chi_1 p_{t-1} + (1-\chi_1)c + \frac{a_{t-1}g[1-\beta(1+g)]}{\beta[\chi_2 - (1+g)]} \qquad (7.34)$$

which can be solved to yield

$$p_t = c + (\chi_1)^t(p_0 - c) + \frac{g[1-\beta(1+g)]}{\beta[\chi_2 - (1+g)]}\sum_{i=1}^{t}(\chi_1)^{i-1}a_{t-i}. \qquad (7.35)$$

The difference between price under the adjusted scheme and under the standard scheme (given by (7.30)) is

$$-\frac{(1+g)g}{[\chi_2 - (1+g)]}\sum_{i=1}^{t}(\chi_1)^{i-1}a_{t-i}$$

so for $g > 0$ the price under adjusted TB is always below that under TB, and the price is closer to marginal cost. Again as t increases the final term in (7.35) becomes larger, and price will never equal c.

Thus the existence of a nonstationarity in the form of exogenous shifts in demand can be coped with by altering the standard tariff basket formula to take account of demand growth. The main problem that the regulator faces is that she must estimate the extent of the shifts in demand in order to implement the adjusted price cap.

7.5 Conclusions

In this chapter we have considered a variety of extensions of the analysis of price caps. A common theme has been that relatively simple adjustments to the standard tariff basket price cap scheme enable the regulator to preserve its desirable properties when the environment that the firm differs from that assumed in earlier chapters. In some cases, however, although the adjustments are simple the information requirements on the regulator are demanding. Price caps cannot be expected to solve all the problems of regulation, but the analysis in this chapter suggests that they can be of some help if carefully designed.

Appendix 7.1

In this appendix we analyze the case of tariff basket regulation where the quality level is not controlled. Demand is $x+1-p$, marginal production cost is c and the cost of quality provision is $\gamma x^2/2$ where $\gamma \geq 1/(1-\beta)$. Using (5.6) the present value of the firm's profits is:

$$\sum_{t=1}^{\infty} \beta^{t-1}\left\{A_t + (p_t - c)q_t - \gamma x_t^2/2\right\} =$$

$$\text{Constant} + \frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}\Big(A_t - A_{t-1} + (p_t - c)q_t - \gamma x_t^2/2 - (p_{t-1} - c)q_{t-1} + \gamma x_{t-1}^2/2\Big).$$

The constraint gives $A_t - A_{t-1} = (p_{t-1} - p_t)q_{t-1}$, which can be substituted into the summation term yielding:

$$\frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}\Big((p_t - c)(p_{t-1} - p_t + x_t - x_{t-1}) - \gamma x_t^2/2 + \gamma x_{t-1}^2/2\Big).$$

The Euler condition for p_t is:

$$x_t - x_{t-1} + \beta(p_{t+1} - c) - 2(p_t - c) + (p_{t-1} - c) = 0, \tag{A.1}$$

and the Euler condition for x_t is:

$$p_t - c - \gamma x_t - \beta\Big(p_{t+1} - c - \gamma x_t\Big) = 0. \tag{A.2}$$

Rearranging (A.2) gives x_t as a function of p_t and p_{t+1}:

$$x_t = \frac{1}{\gamma(1-\beta)}[\, p_t - c - \beta(p_{t+1} - c)] \quad \text{for } t = 1, 2, \ldots \tag{A.3}$$

Using (A.3) in (A.1) gives

$$\beta[\gamma(1-\beta) - 1](p_{t+1} - c) + [1 + \beta - 2\gamma(1-\beta)](p_t - c) + [\gamma(1-\beta) - 1](p_{t-1} - c) = 0$$
$$\text{for } t = 2, 3, \ldots \tag{A.4}$$

Note that when $\gamma(1-\beta) = 1$ this yields $p_t = c$ for $t \geq 2$ and (A.3) then implies $x_t = 0$. Equation (A.4) is a second-order difference equation in prices with roots:

$$\phi_i = \frac{2\gamma(1-\beta) - 1 - \beta \pm (1-\beta)\sqrt{4\gamma[\gamma(1-\beta) - 1] + 1}}{2\beta[\gamma(1-\beta) - 1]} \qquad i = 1, 2.$$

It can be checked that when $\gamma > 1/(1-\beta)$ the roots are real and positive, with $\phi_2 > 1$ and $\phi_1 < 1$. This condition is also necessary for $\phi_1 < 1$. The optimal sequences of prices and quality are given by

$$p_1 = c + \frac{\gamma(1-\beta)}{(2-\beta\phi_1)[\gamma(1-\beta)-1]+1}(p_0 - x_0 - c)$$

$$p_t = c + (p_1 - c){\phi_1}^{t-1} \qquad t = 1, 2, 3, \ldots$$

$$x_t = \frac{(p_1 - c)}{\gamma(1-\beta)}[1-\beta\phi_1]{\phi_1}^{t-1} \qquad t = 1, 2, \ldots$$

We assume $p_0 - x_0 - c > 0$, which is necessary and sufficient for $p_1 > c$. Price moves monotonically towards marginal production cost from period 2 onwards and the quality level moves monotonically towards zero. As t increases welfare tends to $(1-c)^2/2$.

Appendix 7.2

In this appendix we consider the quality-adjusted tariff basket scheme. As in Appendix 7.1 the present value of profits can be written as

$$\text{Constant} + \frac{1}{1-\beta}\sum_{t=1}^{\infty}\beta^{t-1}\Big(A_t - A_{t-1} + (p_t - c)q_t - \gamma x_t^2/2 - (p_{t-1} - c)q_{t-1} + \gamma x_{t-1}^2/2\Big).$$

The constraint gives $A_t - A_{t-1} = (p_{t-1} - p_t + x_t - x_{t-1})(x_{t-1} + 1 - p_{t-1})$, which can be substituted into the summation term:

$$\sum_{t=1}^{\infty}\beta^{t-1}\Big\{(x_t - x_{t-1})(x_{t-1} + 1 - p_{t-1}) + (p_t - c)(p_{t-1} - p_t + x_t - x_{t-1}) - \gamma x_t^2/2 + \gamma x_{t-1}^2/2\Big\}$$

The Euler condition for p_t is:

$$(p_{t-1} - p_t + x_t - x_{t-1}) - (p_t - c) - \beta(x_{t+1} - x_t) + \beta(p_{t+1} - c) = 0, \tag{A.5}$$

and the Euler condition for x_t is:

$$(x_{t-1} + 1 - p_{t-1}) + (p_t - c) - \gamma x_t + \beta\big(x_{t+1} - 2x_t - 1 + p_t - (p_{t+1} - c) + \gamma x_t\big) = 0. \qquad \text{(A.6)}$$

Adding (A.5) and (A.6) gives:

$$x_t = \frac{1 - p_t}{\gamma - 1}.$$

This is the level of x_t that maximizes $V(p - x) - \gamma x_t^2/2$. We can now substitute (A.6) back into the Euler condition for pricing, (A.5), to obtain

$$\beta\gamma(p_{t+1} - c) + (1 - \beta - 2\gamma)(p_t - c) + \gamma(p_{t-1} - c) = 0 \quad \text{for } t \geq 2.$$

The relevant root of the Euler equation is:

$$\xi = \frac{2\gamma - (1 - \beta) - \sqrt{(1 - \beta)^2 + 4\gamma(\gamma - 1)(1 - \beta)}}{2\beta\gamma}$$

Since $0 \leq \beta < 1$ and $\gamma > 1$, ξ is real. It can be checked that $0 < \xi < 1$. The sequences of price and quality are:

$$p_1 = c + \frac{(\gamma - 1)(p_0 - x_0 - c) + 1 - c}{\beta(1 - \gamma\xi) + 2\gamma - 1}$$

$$p_t = c + (p_1 - c)\xi^{t-1} \qquad t = 1, 2, 3, \ldots$$

$$x_t = \frac{1 - p_t}{\gamma - 1} \qquad t = 1, 2, \ldots$$

The steady state has $p = c$ and $x = (1 - c)/(\gamma - 1)$, confirming that efficient pricing and quality-setting holds.

Appendix 7.3 Transversality conditions

We use the method of Sargent (1987, Chapter 9) to derive the transversality condition. We assume the more general form of the cost function, i.e. $C = cq_t + \delta q_t^2/2$. In a finite horizon problem the maximand in the last period, T, is

$$A_{T\text{-}1} + (p_{T\text{-}1} - p_T)(a_{T\text{-}1} - p_{T\text{-}1}) + (p_T - c)(a_T - p_T) - \delta(a_T - p_T)^2/2.$$

The first-order condition for p_T is:

$$\beta^T[p_{T\text{-}1} - a_{T\text{-}1} + a_T - (2+\delta)p_T + c + \delta a_T] = 0$$

and the transversality condition for the infinite horizon problem is obtained by multiplying the left-hand side by p_T and taking the limit as T tends to infinity:

$$\lim_{T\to\infty} \beta^T p_T[p_{T\text{-}1} - a_{T\text{-}1} + a_T - (2+\delta)p_T + c + \delta a_T] = 0. \tag{A.6}$$

This is a necessary condition for optimality. Sufficient conditions for it to hold are that the sequences $\{a_{t+i}\}_{i=1}^{\infty}$ and $\{p_{t+i}\}_{i=1}^{\infty}$ are of exponential order less than $1/\sqrt{\beta}$. This implies that $a_{t+i} < K(x)^t$ for $K > 0$ and $1 \leq x < 1/\sqrt{\beta}$, and similarly $p_{t+i} < K(x)^t$. Taking the absolute value of (A.6) and using these assumptions we have

$$\begin{aligned}
&|\, \beta^T p_T[p_{T\text{-}1} - a_{T\text{-}1} + a_T - (2+\delta)p_T + c + \delta a_T] \,| \leq \\
&\beta^T \,|\, p_T p_{T\text{-}1} \,| + \beta^T \,|\, a_{T\text{-}1}p_T \,| + \beta^T \,|\, a_T p_T \,| + (2+\delta)\beta^T \,|\, p_T^2 \,| + \beta^T \,|\, cp_T \,| + \beta^T \,|\, \delta a_T p_T \,| \\
&\leq 2K^2(\beta x^2)^T x^{-1} + (3+2\delta)K^2(\beta x^2)^T + c\beta^T x^T.
\end{aligned}$$

Since $1 \leq x < 1/\sqrt{\beta}$ we have $0 < x\sqrt{\beta} < 1$ and $0 < \beta x^2 < 1$. Thus the limit of this expression as $T \to \infty$ is zero.

Appendix 7.4

We show here that

$$\chi_1 \sum_{i=0}^{\infty} \left(\frac{1}{\chi_2}\right)^i \frac{(1-\beta)}{(1-\chi_1)} = 1$$

which is then used in (7.27′). The left-hand side can be written as

$$\frac{\chi_1(1-\beta)}{1-\chi_1} \cdot \frac{1}{1-1/\chi_2} = \frac{\chi_1\chi_2(1-\beta)}{1-\chi_1} \cdot \frac{1}{\chi_2 - 1}.$$

since $1/\chi_2 < 1$. Using the fact that $\chi_1\chi_2 = 1/\beta$ we have

$$\frac{1-\beta}{\beta(1-\chi_1)(\chi_2-1)} = \frac{1-\beta}{\beta\chi_2 - \beta - 1 + \beta\chi_1} = \frac{1-\beta}{1/\chi_1 - \beta - 1 + \beta\chi_1}$$

$$= \frac{\chi_1(1-\beta)}{1-\beta\chi_1 - \chi_1 + \beta\chi_1^2} = \frac{\chi_1(1-\beta)}{1-\beta\chi_1 + \chi_1 + \beta\chi_1^2 - 2\chi_1} = 1$$

since $\beta\chi_1^2 - 2\chi_1 + 1 = 0$.

CHAPTER 8

CONCLUSIONS

In this thesis the theoretical properties of different price cap schemes that have been applied in both the UK and the USA have been assessed. The analysis was conducted in a second-best framework under the assumption that the regulator has no knowledge of absolute and relative costs. It is thus not surprising that it was not possible to provide theoretical arguments that demonstrated the superiority of one particular scheme. Nevertheless we were able to come to some broad conclusions.

The scheme based on average revenue is likely to generate inefficient pricing, and can even reduce welfare below the level obtained when the monopolist is unregulated. Yet this scheme is the basis of regulation of several important industries in the UK, including gas, electricity and airports. There have been moves in these industries to disaggregate tariffs, and the analysis of this thesis shows the dangers of such disaggregation when the price cap is based on average revenue.

We also showed that the tariff basket scheme, because it aligns the objectives of the firm and the regulator, has desirable properties and is generally robust. When the tariff basket scheme is applied the set of allowed prices contains the current prices and all prices that are *sufficient* to generate an increase in consumer surplus. Imposing such a constraint must therefore benefit consumers, though a far-sighted firm can manipulate the future structure and height of the price cap by its current price choices, and in the two-part tariff example of Chapter 5 it was shown that the fixed weights scheme can be preferable when the discount factor is high. We showed in Chapter 7 that the principle behind the tariff basket scheme, i.e. designing a constraint that ensures that consumer surplus increases (weakly) each period, can be used to develop adjusted price caps in cases where the standard assumptions of the earlier chapters do not hold. Thus endogenous quality, income distribution considerations and demand growth can be allowed for in the price cap, although the informational burden on the regulator is considerable.

Many issues in the regulation of utility prices were not addressed. In order to focus on the effects of different price caps on *price structure*, and to allow comparison with other regulatory schemes such as the Vogelsang-Finsinger (1979) mechanism, the assumption that the *level* of the price cap is never revised was made. In practice regulators revise the level of price caps at regular intervals. This affects *productive efficiency* by limiting period over which the benefits of cost reduction are translated into higher profits and, as was mentioned in Chapter 1, this effect is now well-understood. One straightforward extension of the work in this thesis would be to analyze the effect of a finite horizon on price structure and *allocative efficiency.* Sappington and Sibley (1992) analyze the average revenue (lagged) scheme for different finite horizons. Although such an analysis is certainly possible for all the price caps considered in this thesis, at least if specific functional forms are assumed, it is not clear that further insights into price structure would be obtained. It is likely that much will depend on the assumed length of the period between price reviews.

Throughout the thesis the implicit assumption has been that the regulated firm is a monopolist supplying final customers. Although this is still relevant for many industries, it is important to recognize that regulatory reform has often entailed the liberalization of markets that were previously monopolized. One issue the effect of a price cap on a dominant firm that faces competition from an unregulated competitor in a subset of its markets – see Armstrong and Vickers (1993) for a model of this case. Another issue concerns the terms of *network access.* It has been recognized that while the networks of pipelines and wires are often naturally monopolistic the provision of services across the networks is potentially competitive. Thus in Britain there is growing competition in long-distance telecommunications, electricity generation and supply, and gas supply. Regulators and theorists are paying increasing attention to the terms of access to the network that competing firms should obtain. For recent discussions of this issue see Armstrong, Cowan and Vickers (1994, Chapter 5), Laffont and Tirole (1993, Chapter 5, 1994) and Baumol (1995). The optimal access price depends on whether the network operator is also a supplier of services, the nature of downstream competition, whether the final product price is regulated or not, whether lump-sum transfers to the network operator are possible and the degree of differentiation of final products. Thus the problem is a complex one and while some progress has been made recently in the analysis of the optimal level of the access price, there has been little analysis of the structure of access charges in a multiproduct or multimarket context. This represents a future area of research.

This thesis has stressed the importance of giving regulated firms incentives to choose efficient price structures. Schemes that on first glance appear reasonable can distort incentives and generate poor welfare outcomes. Price cap schemes need careful analysis under a variety of conditions before their desirability can be assessed. Such an analysis has been attempted here, but much remains to be done.

References

Acton, J.P. and Vogelsang, I. (1989) Introduction to Symposium on Price-Cap Regulation, *Rand Journal of Economics*, 20, 369-372.

Armstrong, M., Cowan, S.G.B. and Vickers, J.S. (1994) *Regulatory Reform: Economic Analysis and British Experience.* Cambridge, MA, MIT Press.

Armstrong, M., Cowan, S.G.B. and Vickers, J.S. (1995) 'Nonlinear Pricing and Price Cap Regulation', mimeo, Institute of Economics and Statistics, University of Oxford, forthcoming in the *Journal of Public Economics.*

Armstrong, M. and Vickers, J.S. (1991) 'Welfare Effects of Price Discrimination by a Regulated Monopolist', *Rand Journal of Economics*, 23, 571-580.

Armstrong, M. and Vickers, J.S. (1993) 'Price Discrimination, Competition and Regulation', *Journal of Industrial Economics*, 41, 335-360.

Atkinson, A.B. and Stiglitz, J.E. (1980) *Lectures on Public Economics.* London, McGraw-Hill.

Auerbach, A.J. and Pellechio, A.J. (1978) 'The Two-Part Tariff and Voluntary Market Participation', *Quarterly Journal of Economics*, 92, 571- 587.

Averch, H. and Johnson, L.R. (1962) 'Behavior of the Firm Under Regulatory Constraints', *American Economic Review*, 52, 1052-1069.

Baron, D.P. (1989) 'Design of Regulatory Mechanisms and Institutions', in R. Schmalensee and R.D. Willig, eds., Vol. 2.

Baron, D.P. and Myerson, R.B. (1982) 'Regulating a monopolist with unknown costs', *Econometrica*, 50, 911-930.

Baumol, W.J. (1995) 'Modified Regulation of Telecommunications and the Public Interest Standard' in Bishop, M., Kay, J. and Mayer, C., eds.

Baumol, W.J. and Bradford, D.F. (1970) 'Optimal departures from marginal cost pricing', *American Economic Review*, 60, 265-283.

Baumol, W.J. and A.K. Klevorick (1970) 'Input Choices and Rate of Return Regulation: An Overview of the Discussion', *Bell Journal of Economics*, 1, 162-190.

Bawa, V.S. and Sibley, D.S. (1980) 'Dynamic Behavior of a Firm Subject to Stochastic Regulatory Review', *International Economic Review*, 21, 627-642.

Beesley, M.E. and Littlechild, S.C. (1989) 'The regulation of privatized monopolies in the United Kingdom', *Rand Journal of Economics*, 20, 454-472.

Besanko, D., Donnenfeld, S. and White, L.J. (1987) 'Monopoly and Quality Distortion, Effects and Remedies', *Quarterly Journal of Economics*, 102, 743-767.

Besanko, D., Donnenfeld, S. and White, L.J. (1988) 'The multiproduct firm, quality choice and regulation', *Journal of Industrial Economics*, 36, 411-429.

Bishop, M., Kay, J. and Mayer, C.M., eds., (1995) *The Regulatory Challenge.* Oxford, Oxford University Press.

Bishop, M., Rovizzi, L. and Thompson, D.J. (1991) 'Price and Quality control: regulatory reform and public utilities performance', in M. Beesley, ed., *Lectures on Regulation 1991*, Centre for Business Strategy Report. London, London Business School.

Boiteux, M. (1956) 'Sur la gestion des Monopoles Publics astreints à l'Equibre Budgétaire', *Econometrica*, 24, 22-40.

Boiteux, M. (1971) 'On the Management of Public Monopolies Subject to Budgetary Constraints', *Journal of Economic Theory*, 3, 219-240.

Bradley, I. and Price, C. (1988) 'The Economic Regulation of Private Industries by Price Constraints.' *Journal of Industrial Economics*, 37, 99-106.

Braeutigam, R.R. and J.C. Panzar (1993) 'Effects of the Change from Rate-of-Return to Price-Cap Regulation,' *American Economic Review Paper and Proceedings*, 83, 191-198.

Brennan, T. J. (1991) 'Regulating by Capping Prices', in M. A. Einhorn, ed.

Cabral, L.M.B. (1990) 'Optimal pricing of the Portuguese telephone service', *Applied Economics*, 20, 211-220.

Caillaud, B., Guesnerie, R., Rey, P. and Tirole, J. (1988) "Government Intervention in Production and Incentives Theory: A Review of Recent Contributions', *Rand Journal of Economics*, 19(1), 1-26.

Chow, G.C. (1993) 'Optimal Control without solving the Bellman equation', *Journal of Economic Dynamics and Control*, 17, 621-630.

Coase, R. (1946) 'The marginal cost controversy', *Economica*, 13, 168-189.

Cox, J.C. and Isaac, R.M. (1987) 'Mechanisms for incentive regulation: theory and experiment', *Rand Journal of Economics*, 18, 348-359.

Deaton, A. (1977) 'Equity, Efficiency and the Structure of Indirect Taxation', *Journal of Public Economics*, 8, 299-312.

Deaton, A. and Muellbauer, J. (1980) *Economics and Consumer Behaviour.* Cambridge, Cambridge University Press.

Diamond, P.A. (1975) 'A Many-Person Ramsey Tax Rule', *Journal of Public Economics*, 4, 335-342.

Diamond, P.A. and Mirrlees, J.A. (1971) 'Optimal taxation and public production I: production efficiency and II: tax rules', *American Economic Review*, 61, 8-27 and 261-278.

Dixit, A.K. and Norman, V. (1980) *Theory of International Trade.* Cambridge, Cambridge University Press.

Doyle, C. (1993) 'Regulating Firms with Monopoly Power', in Sugden, R., ed., *Industrial Economic Regulation: a framework and exploration.* London, Routledge.

Einhorn, M.A. (1991) *Price Caps and Incentive Regulation in Telecommunications*, Boston, Kluwer.

Faulhaber, G.R. and Baumol, W.J. (1988) 'Economists as Innovators: Practical Products of Theoretical Research, *Journal of Economic Literature* 26(2), 577-600.

Feldstein, M.S. (1972a) 'Distributional Equity and the Optimal Structure of Public Prices', *American Economic Review*, 62, 32-36. Corrected version of p. 33, fn. 7, *American Economic Review*, 62, 763.

Feldstein, M.S. (1972b) 'Equity and Efficiency in Public Sector Pricing: The Optimal Two-Part Tariff', *Quarterly Journal of Economics*, 86, 175-187.

Finsinger, J. and Vogelsang, I. (1981) 'Alternative Institutional Frameworks for Price Incentive Mechanisms', *Kyklos*, 34, 388-403.

Finsinger, J. and Vogelsang, I. (1982) 'Performance indices for public enterprises' in L. P. Jones, ed., *Public Enterprise in Less Developed Countries.* Cambridge, Cambridge University Press.

Finsinger, J. and Vogelsang, I. (1985) 'Strategic Management Behaviour under Reward Structures in a Planned Economy', *Quarterly Journal of Economics*, 100, 263-269.

Fudenberg, D. and Tirole, J. (1991) *Game Theory.* Cambridge, MA, MIT Press.

Graham, D.A. and Vernon, J.M. (1991) 'A Note on Decentralized Natural Monopoly Regulation', *Southern Economic Journal*, 58, 273-275.

Gravelle, H. (1983) 'Alternative Institutional Frameworks for Price Incentive Mechanisms: Some Comments', *Kyklos*, 36, 115-120.

Gravelle, H. (1985) 'Reward structures in a planned economy: some difficulties', *Quarterly Journal of Economics*, 100, 271-278.

Greenhut, J.G. and Greenhut, M.L. (1975) 'Spatial price discrimination, competition and locational effects', *Economica*, 42, 401-419.

Greenhut, M.L. and Ohta, H. (1972) 'Monopoly output under alternative spatial pricing techniques', *American Economic Review*, 62, 705-713.

Greenhut, M.L. and Ohta, H. (1976) 'Joan Robinson's criterion for deciding whether market discrimination reduces output', *Economic Journal*, 46, 96-97.

Hagerman, J. (1990) 'Regulation by price adjustment', *Rand Journal of Economics*, 21, 72-82.

Hausman, J.A. and Mackie-Mason, J.K. (1988) 'Price discrimination and patent policy', *Rand Journal of Economics*, 19, 253-265.

Hicks, J.R. (1946) *Value and Capital*, 2nd edition, Oxford, Oxford University Press.

Holahan, W.L. (1975) 'The Welfare Effects of Spatial Price Discrimination', *American Economic Review*, 65, 498-503.

Ireland, N.J. (1992) 'On the Welfare Effects of Regulating Price Discrimination', *Journal of Industrial Economics*, 40, 237-248.

Katz, M.L. and Shapiro, C. (1994) 'Systems Competition and Network Effects', *Journal of Economic Perspectives*, 8(2), 93-115.

Laffont, J-J. and Tirole, J. (1986) 'Using Cost Information to Regulate Firms', *Journal of Political Economy*, 94, 614-641.

Laffont, J-J. and Tirole, J. (1990) 'The regulation of multiproduct firms I: Theory and II: Applications to Competitive Environments and Policy Analysis', *Journal of Public Economics*, 43, 1-36 and 37-66.

Laffont, J-J. and Tirole, J. (1991) 'The politics of government decision-making: a theory of regulatory capture', *Quarterly Journal of Economics*, 106, 1089-1127.

Laffont, J-J. and Tirole, J. (1993) *A Theory of Incentives in Regulation and Procurement*. Cambridge, MA, MIT Press.

Laffont, J-J. and Tirole, J. (1994) 'Creating Competition Through Interconnection: Theory and Practice', mimeo, IDEI, Toulouse.

Law, P.J. (1993) 'Welfare Effects of Strategic Price Setting in Anticipation of Price Regulation', *Bulletin of Economic Research*, 45, 147-159.

Layson, S.K. (1994a) 'Market Opening Under Third-Degree Price Discrimination' *Journal of Industrial Economics*, 42, 335-340.

Layson, S.K. (1994b) 'Third-Degree Price Discrimination under Economies of Scale', *Southern Economic Journal*, 61, 323-327.

Leland, H.E. and Meyer, R.A. (1976) 'Monopoly pricing structures with imperfect discrimination', *Bell Journal of Economics*, 7, 449-462.

Lewis, T.R. and Sappington, D.E.M. (1988a) 'Regulating a Monopolist with Unknown Demand', *American Economic Review*, 78, 986-998.

Lewis, T. R. and Sappington, D. E. M. (1988b) 'Regulating a monopolist with unknown demand and cost functions', *Rand Journal of Economics*, 19, 438-457.

Lipman, B.L. (1985) 'Dynamic behavior of a firm subject to stochastic regulatory review: a comment', *International Economic Review*, 26, 511-516.

Littlechild, S.C. (1975) 'Two-Part Tariffs and Consumption Externalities', *Bell Journal of Economics*, 6, 661-670.

Littlechild, S.C. (1983) *Regulation of British Telecommunications' Profitability.* London, Department of Industry.

Loeb, M. and Magat, W.A. (1979) 'A Decentralized Method for Utility Regulation', *Journal of Law and Economics*, 22, 399-404.

Logan, J.W., Masson, R.T. and Reynolds, R.J. (1989) 'Efficient Pricing with Little Information: Reality in the Limit?', *International Economic Review*, 30, 851-861.

McAffee, R.P. and McMillan, J. (1988) 'Multidimensional Incentive Compatibility and Mechanism Design', *Journal of Economic Theory*, 46, 335-354.

Malueg, D.A. (1993) 'Bounding the Welfare Effects of Third-Degree Price Discrimination', *American Economic Review*, 83, 1011-1021.

Mirrlees, J.A. (1971) 'An exploration in the theory of optimal income taxation', *Review of Economic Studies*, 38, 175-208.

Mirrlees, J.A. (1975) 'Optimal Commodity Taxation in a Two-Class Economy', *Journal of Public Economics*, 4, 27-33.

Monopolies and Mergers Commission (1993), *Gas, Volume 1 of Reports under the Fair Trading Act*, Cm 2314, London, HMSO.

Nahata, B., Ostaszewski, K. and Sahoo, P.K. (1990) 'Direction of Price Changes in Third-Degree Price Discrimination', *American Economic Review*, 80, 1254-1258.

Ng, Y. and Weisser, M. (1974) 'Optimal pricing with a budget constraint - The case of the two part tariff', *Review of Economic Studies*, 41, 337-345.

Ordover, J.A. and Panzar, J.C. (1980) 'On the nonexistence of Pareto superior outlay schedules', *Bell Journal of Economics*, 11, 351-354.

Ordover, J.A. and Panzar, J.C. (1982) 'On the nonlinear pricing of inputs', *International Economic Review*, 23, 659-675.

Panzar, J.C. (1989) 'Technological Determinants of Firm and Industry Structure', in R. Schmalensee and R.D. Willig, eds., Vol 1..

Phlips, L. (1983) *The economics of price discrimination.* Cambridge, Cambridge University Press.

Ramsey, F. (1927) 'A Contribution to the Theory of Taxation', *Economic Journal* 37, 47-61.

Rees, R. and Vickers, J.S. (1995) 'RPI-X Price Cap Regulation', in M. Bishop, J. A. Kay, C. Mayer, eds.

Riordan, M.H. (1984) 'On delegating price authority to a regulated firm', *Rand Journal of Economics*, 15, 108-115.

Robinson, J. (1933) *The Economics of Imperfect Competition.* London, Macmillan.

Ross, T.W. (1984) 'Uncovering regulators' social welfare weights', *Rand Journal of Economics*, 15, 152-155.

Rovizzi, L. and D.J. Thompson (1992) 'The Regulation of Product Quality in the Public Utilities and the Citizen's Charter', *Fiscal Studies*, 13, 74-95.

Rovizzi, L. and D.J. Thompson (1995) 'The Regulation of Product Quality in the Public Utilities, in M. Bishop, J.A. Kay, C. Mayer, eds.

Sappington, D. (1980) 'Strategic Firm Behaviour under a Dynamic Regulatory Adjustment Process', *Bell Journal of Economics*, 11, 360-372.

Sappington, D.E.M. and Sibley, D.S. (1988) 'Regulating without Cost Information: The Incremental Surplus Subsidy Scheme', *International Economic Review*, 29, 297-306.

Sappington, D.E.M. and Sibley, D.S. (1990) 'Regulating without Cost Information: Further Observations', *International Economic Review*, 31, 1027-1028.

Sappington, D.E.M. and Sibley, D.S. (1992) 'Strategic Nonlinear Pricing Under Price-Cap Regulation', *Rand Journal of Economics*, 23, 1-19.

Sargent, T.J. (1987) *Macroeconomic Theory*, 2nd Edition. Orlando, Academic Press.

Schmalensee, R. (1981a) 'Output and welfare implications of monopolistic third degree price discrimination', *American Economic Review*, 71, 242-247.

Schmalensee, R. (1981b) 'Monopolistic two-part pricing arrangements', *Bell Journal of Economics*, 12, 445-466.

Schmalensee, R. and Willig, R.D. (1989) *Handbook of Industrial Organization*, Vols 1 and 2. Amsterdam, North-Holland.

Schwartz, M. (1990) 'Third-Degree Price Discrimination and Output: Generalizing a Welfare Result', *American Economic Review*, 80, 1259-1262.

Scott, M.FG. (1952) 'Criteria of Efficiency in Nationalized Industries', B.Litt Thesis, Oxford.

Scott, M.FG. (1978) 'A Profits Test for Public Monopolies', in Butler, D. and Halsey, A.H., eds., *Policy and Politics, Essays in Honour of Norman Chester.* London, Macmillan.

Sharkey, W.W. (1979) 'A decentralized method for utility regulation: A comment', *Journal of Law and Economics*, 22, 74-75.

Sharkey, W.W. and Sibley, D.S. (1993) 'Optimal non-linear pricing with regulatory preference over customer type', *Journal of Public Economics*, 50, 197-229.

Sherman, R. and Visscher, M. (1982) 'Rate-of-return regulation and two-part tariffs', *Quarterly Journal of Economics*, 97, 27-42.

Sibley, D.S. (1985) 'Response to Lipman and further results', *International Economic Review*, 26, 517-520.

Sibley, D.S. (1989) 'Asymmetric information, incentives and price-cap regulation', *Rand Journal of Economics*, 20, 392-404.

Spence, A.M. (1975) 'Monopoly, quality and regulation', *Bell Journal of Economics*, 16, 417-429.

Spulber, D.F. (1989) *Regulation and Markets.* Cambridge, MA, MIT Press.

Srinagesh, P. (1986) 'Nonlinear prices and the regulated firm', *Quarterly Journal of Economics*, 101, 51-68.

Stefos, T. (1990) 'Regulating without Cost Information: A Comment', *International Economic Review*, 31, 1021-1025.

Stokey, N.L. and Lucas, R.E. Jr., (1989) *Recursive Methods in Economic Dynamics.* Cambridge, MA, Harvard University Press.

Tam, M. (1985) 'Reward Structures in a Planned Economy: Some Further Thoughts', *Quarterly Journal of Economics*, 100, 279-290.

Tirole, J. (1988) *The Theory of Industrial Organization.* Cambridge, MA, MIT Press.

Train, K.E. (1991) *Optimal Regulation.* Cambridge, MA, MIT Press.

Varian, H.R. (1985) 'Price Discrimination and Social Welfare', *American Economic Review*, 75, 870-875.

Varian, H.R. (1989) 'Price Discrimination', in R. Schmalensee and R.D. Willig, eds., Vol 1.

Varian, H.R. (1992) *Microeconomic Analysis*, Third Edition, New York, Norton.

Varian, H.R. (1993) *Intermediate Microeconomics*, Third Edition, New York, Norton.

Vickers, J.S. and Yarrow, G. (1988). *Privatization: An Economic Analysis.* Cambridge, MA, MIT Press.

Vogelsang, I. (1983) 'Effort Rewarding Incentive Mechanisms for Public Sector Managers', *International Journal of Industrial Organization*, 1, 253-273.

Vogelsang, I. (1988) 'A Little Paradox in the Design of Regulatory Mechanisms', *International Economic Review*, 29, 467-476.

Vogelsang, I. (1989a) 'Two-part tariffs as regulatory constraints', *Journal of Public Economics*, 39, 45-66.

Vogelsang, I. (1989b) 'Price Cap Regulation of Telecommunications Services: A Long-Run Approach', in M.A. Crew, ed., *Deregulation and Diversification of Utilities*, Boston, Kluwer.

Vogelsang, I. (1991) 'A Non-Bayesian Incentive Mechanism Using Two-part tariffs', in M. A. Einhorn, ed.

Vogelsang, I. and Finsinger, J. (1979) 'A Regulatory Adjustment Process for Optimal Pricing by Multiproduct Monopoly Firms.' *Bell Journal of Economics*, 10, 157-171.

Waterson, M. (1988) *Regulation of the Firm and Natural Monopoly.* Oxford, Basil Blackwell.

Willig, R.D. (1978) 'Pareto-Superior Nonlinear Outlay Schedules', *Bell Journal of Economics*, 9, 56-69.

Wilson, R.B. (1993) *Nonlinear Pricing.* Oxford, Oxford University Press.