4 Ramsey Prices

4.1 Motivation

Most public utilities produce more than one good or service, or sell their output in more than one market with a different price in each market. For example, power utilities often sell both gas and electricity. Those that sell only electricity nevertheless sell this good in several time periods (such as peak and off-peak periods) and to several types of customers (such as residential, commercial, industrial, and agricultural customers). Local telephone companies sell point-to-point service, often pricing on the basis of the distance of the call. Transit agencies might provide both bus and rail service. And so on. In fact, it is hard to find a public utility that actually provides only one service at one uniform price to all customer groups.

When more than one good is sold, or a good is sold in more than one market, the second-best outcome is not immediately obvious. First-best pricing, as always, is to set all prices equal to marginal cost. However, for a natural monopoly, marginal-cost pricing can result in the firm's losing money. If the firm cannot be subsidized, price must be raised above marginal costs until profit rises to zero. In a one-good situation, the requirement of zero profit is sufficient to determine the second-best price: price is necessarily equal to average cost when profits are zero. Consequently, the second-best price for one good is average cost. However, with more than one good, many different price combinations result in zero profit. For example, for a utility selling ing gas and electricity, the price of gas can be raised sufficiently for the firm to break even overall while still holding the price of electricity at its marginal cost; or, the price of electricity can be raised, holding gas price at its marginal cost; or, the prices of both can be raised somewhat above their marginal costs. There are an infinite number of possibilities. Of the various price combinations that provide zero profit, which is best from a social welfare perspective?

This question was first addressed by Ramsey (1927) in the context of optimal taxation. He developed a method for determining the tax rates for various goods that would provide the government with sufficient revenue while reducing consumer surplus as little as possible. As Baumol and Bradford (1970) have pointed out, optimal taxation rules are directly applicable for determining second-best prices for multiproduct natural monopolies. It is traditional, therefore, to refer to these second-best prices as Ramsey prices.

The following sections describe the goal that is implicit in Ramsey pricing, state the rule (or formula) that is used to calculate these prices, and demonstrate that the prices obtained by applying this rule attain the desired goal. The final section illustrates these concepts with an empirical example of pricing for urban transit.

The findings of the chapter can be summarized as follows. Of all possible price combinations for a multiproduct firm, Ramsey prices provide the greatest total surplus while allowing the firm to break even. At the Ramsey price, profits are zero, and

1. the output of each good is reduced by the same proportion relative to the outputs that would be produced when prices are at marginal cost; and

2. the amount by which price exceeds marginal cost, expressed as a percentage of price, is greater for goods with less elastic demand.

The first of these statements applies exactly only when demand is linear; otherwise, output is reduced *approximately* the same for each good. The second statement, called the "inverse elasticity rule," applies with both linear and nonlinear demand. The two statements are equivalent, but are simply described in different terms. That is, if prices are raised inversely to elasticity, outputs will be reduced by the same proportion for all goods, and vice versa.

It is important to note that Ramsey prices might not be considered equitable in certain situations. Inelastic demand can reflect a lack of options by consumers (e.g., demand for medical care, demand for bus service by low-income households without cars). Yet, under Ramsey concepts, prices for goods and services that consumers have no option but to buy would be raised *more* than prices for less essential goods. The regulator must address these equity issues in deciding whether to implement (or, more precisely, induce the firm to implement) Ramsey prices.

4.2 Description of the Ramsey Rule

If a multiproduct firm is a natural monopoly, then pricing each good at its marginal cost can result in the firm losing money. Suppose that the firm cannot be subsidized and consequently cannot continue to operate with negative profit in the long run. To remain solvent, the firm must set prices sufficiently above marginal cost to break even, that is, earn zero profit. However, there are many price combinations for the goods that will result in zero profit. The question is: Which of these price combinations is best from a social surplus perspective?

Let us first define the term "best." Because prices are raised above marginal cost, there is necessarily some loss of surplus associated with the higher prices. The amount by which total surplus decreases depends on the exact prices charged. Each price combination that provides zero profits (and hence is feasible for the firm to charge) results in a different amount of surplus loss. The "best" price combination is the one that results in the smallest loss in surplus.

Total surplus consists of consumer surplus plus firm profits (that is, producer's surplus). Because all price combinations that result in zero profit provide the same producer's surplus, the price combination that reduces total surplus the least also reduces consumer surplus the least. Therefore, the "best" price combination can also be considered to be that which results in the smallest loss of consumer surplus relative to marginal-cost pricing.

Ramsey, and others, have derived formulae for calculating the prices that result in the smallest surplus loss when prices must be raised above marginal cost in order for the firm to remain solvent. We present below a numerical example that illustrates the meaning of these formulae and indicates why they necessarily result in the prices that minimize the loss in surplus. In the following section, a more rigorous demonstration of the formulae is provided.

Consider a firm producing two goods, or selling one good in each of two markets. Generalization to cases with more than two goods is straightforward. Suppose that demand in the two markets is $P_1 = 50 - .0075Q_1$ and $P_2 = 40 - .004Q_2$, respectively. The firm incurs setup costs of \$19,800 and marginal costs of \$20 for each unit

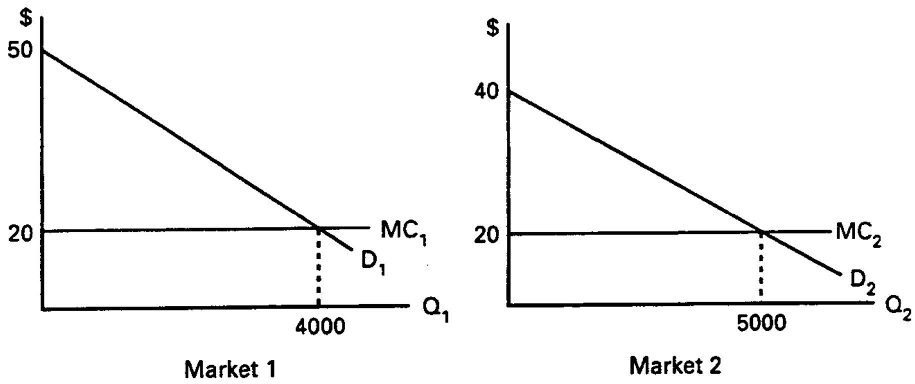

Figure 4.1
Demand and costs in numerical example

of either good produced. Its cost function is therefore $TC = 19{,}800 + 20Q_1 + 20Q_2$. The relevant curves are graphed in figure 4.1.

If the firm priced at marginal cost, it would sell 4,000 units in market one and 5,000 in market two. Its revenues would cover its variable costs, but not its fixed costs. It would therefore incur a loss of $19,800 when pricing at marginal cost. To stay in business, the firm must raise its price for one or both of the goods.

Several options are available. The firm could keep the price in market one at marginal cost and raise price sufficiently in market two to break even. With $P_2 = 25.44$, revenues in market two exceed the variable costs of production for that market by $19,800, which are its fixed costs of production. Therefore, with $P_1 = 20.00$ (that is, marginal cost) and $P_2 = 25.44$, the firm would break even.[1] Alternatively, the firm could keep P_2 at marginal cost and raise price sufficiently in market two to break even. With $P_1 = 26.25$ and $P_2 = 20.00$, the firm earns zero profit. Or, the firm could raise each price above marginal cost. If P_1 is raised $1 above marginal cost, to 21.00, then the firm would break even with P_2 raised to 23.98. With P_1 raised to 22.00, a price of 22.88 in market two is sufficient to break even. And so on. An infinite number of price combinations will result in zero profits for the firm. Some of these (for $1 increments in P_1) are listed in table 4.1.

1. At $P_2 = 25.44$, demand in market two is 3,640. Revenue in market two is $P_2 \cdot Q_2 = 92{,}600$ (rounded). Revenue in market one is $P_1 \cdot Q_1 = (20)(4{,}000) = 80{,}000$. Total cost is $19{,}800 + (20)(4{,}000) + (20)(3{,}640) = 172{,}600$. Profit is therefore $92{,}600 + 80{,}000 - 172{,}600 = 0$.

Table 4.1
Price combinations that result in zero profit

P_1	P_2	Demand in market 1 Q_1	Demand in market 2 Q_2	Revenue in market 1 P_1Q_1	Revenue in market 2 P_2Q_2	Total cost $19800 + 20Q_1 + 20Q_2$	Profit	Consumer surplus in market 1	Consumer surplus in market 2	Total consumer surplus
20.00	25.44	4,000	3,640	80,000	92,600	172,600	0	60,000	26,499	86,499
21.00	23.98	3,867	4,005	81,200	96,040	177,240	0	56,067	32,080	88,147
22.00	22.88	3,733	4,280	82,130	97,930	180,060	0	52,267	36,637	88,904
23.00	22.00	3,600	4,500	82,800	99,000	181,800	0	48,600	40,500	89,100
24.00	21.27	3,467	4,683	83,200	99,600	182,800	0	45,067	43,852	88,919
26.25	20.00	3,167	5,000	83,140	100,000	183,140	0	37,605	50,000	87,605

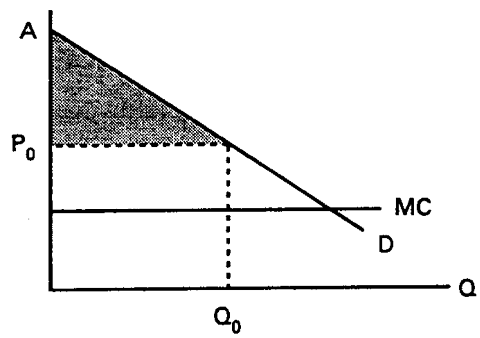

Figure 4.2
Consumer surplus

Each of these price combinations is equally acceptable to the firm. However, consumers are better off at some of these price combinations than at others. To determine which price combination is best for consumers, we calculate the consumers' surplus at each price combination.

Recall that consumer surplus in a market is the area under the demand curve and above price, the shaded area in figure 4.2. For linear demand, this area can be calculated fairly easily. It is the area of a triangle whose width is the quantity sold (Q_0) and whose height is the difference between the price and the y-intercept of the demand curve ($A - P_0$). Because the area of a triangle is one-half the width times the height, consumer surplus in this figure is $(1/2)Q_0(A - P_0)$.

Applying these ideas to the two markets in our example, we find that consumer surplus is \$86,499 when $P_1 = 20.00$ and $P_2 = 25.44$,[2] which is one of the price combinations that result in zero profit. Consumer surplus for each other price combination that provides zero profit is given in the last column of table 4.1.

Consumer surplus is greatest at $P_1 = 23$ and $P_2 = 22$. These are therefore the second-best prices: of those price combinations that provide the firm with zero profit, this price combination provides consumers with the greatest surplus.[3]

2. In market one, consumer surplus is $(1/2)(4{,}000)(50 - 20) = 60{,}000$. In market two, $(1/2)(3{,}640)(40 - 25.44) = 26{,}499$. Surplus in both markets is therefore $60{,}000 + 26{,}499 = 86{,}499$.

3. Total surplus is the sum of consumers' surplus and producers' surplus (that is, profit). Because profit is the same (zero) for all these price combinations, the price combination that provides the greatest consumer surplus also provides the greatest total surplus.

Two characteristics of these prices warrant notice; both are aspects of the Ramsey rule for second-best pricing.

1. *Output is reduced by the same proportion in each market relative to marginal-cost pricing.*

If prices are at marginal cost in both markets ($P_1 = P_2 = 20$), output in market one is 4,000, and in market two, 5,000. At the second-best prices, output is 10% lower in each market. (In market one, output decreases by 10% from 4,000 to 3,600; and in market two, output decreases from 5,000 to 4,500, for a 10% drop.) This occurrence is not a coincidence. When demand curves are linear, second-best prices always result in output being reduced by the same proportion in all markets, relative to output levels that result when prices equal marginal costs.

There is an intuitive reason for this occurrence. Marginal-cost pricing results in the first-best output level for each good. For a natural monopoly to break even, prices must be raised, meaning that output must decrease below its optimal level. If output decreases by *the same proportion* for all goods, then *relative* output levels remain at their first-best levels, even though absolute outputs change. For example, in our numerical example, first-best output is 4,000 in market one and 5,000 in market two, such that the first-best ratio of outputs is 4/5. When output is reduced by 10% in each market, the ratio of outputs remains at its optimal level of 4/5 (now 3,600/4,500). The second-best prices are those that retain the first-best ratio of outputs, even though, by necessity, the absolute output levels are not first-best.

This concept can be expressed algebraically. Let Q_1 and Q_2 be the output in markets one and two, respectively, under second-best prices. Let ΔQ_1 and ΔQ_2 be the changes in output from marginal-cost pricing to second-best pricing. (That is, ΔQ_1 is output in market one when prices are second-best minus the output that would occur under marginal cost prices; and similarly for ΔQ_2. In our example, $Q_1 = 3{,}600$, $Q_2 = 4{,}500$, $\Delta Q_1 = 400$, and $\Delta Q_2 = 500$.) At the second-best prices with linear demand, the following relation necessarily holds:

$$\Delta Q_1/Q_1 = \Delta Q_2/Q_2. \tag{4.1}$$

That is, the percentage change in output from its marginal-cost level is the same in both markets.

This relation gives us another way of thinking of the second-best prices. If a firm is charging marginal-cost prices and losing money, prices can be raised and output reduced in a number of ways to allow the firm to break even. For example, price can be raised considerably in one market and not much in another, or vice versa. Of all the possible ways of raising prices to allow the firm to break even, the price changes that keep the ratio of outputs unchanged (that is, keeps this ratio at its first-best level) are the changes that result in the least loss to consumers and hence are second best.

This fact provides a mechanism for calculating second-best prices. Start at marginal-cost prices and determine the ratio of outputs at these prices. Raise prices a little in each market in such a way that this output ratio is unchanged, that is, that output in each market is reduced by the same proportion. With these slightly higher prices, the firm will have somewhat smaller losses. Raise prices again, still keeping the output ratio constant, and the firm will incur even smaller losses. Continue raising prices in this way until the firm breaks even: these are the second-best prices.

2. *Price is raised more in the market with less elastic demand.*

Recall that the elasticity of demand is a measure of price responsiveness in a market and is defined as the percent change in output that results from a percent change in price. The elasticity is calculated as $\epsilon = (\Delta Q/Q)/(\Delta P/P)$, or, rearranging, $\epsilon = (\Delta Q/\Delta P)(P/Q) = (1/\mathrm{m})(P/Q)$, where m is the slope of the demand curve (with the demand curve giving price as a function of quantity, as in our example).

At the second-best prices in our numerical example, the elasticity of demand in market one is −.85 (calculated as (1/−.0075)(23/3,600)), and the elasticity of demand in market two is −1.2 (calculated as (1/−.004)(22/4,500)). Comparing the second-best price in each market with the elasticity in the market, we find that price is *higher* in the market with *lower* elasticity: the price in market one is higher than in market two (23 compared to 22) and the elasticity of demand is lower (−.85 compared to −1.2, where "lower" means smaller in magnitude, representing less price response).

This occurrence is not a coincidence. Second-best pricing always results in raising price farther above marginal cost in the market with a lower elasticity of demand. This characteristic of second-best prices is often called the inverse elasticity rule: prices are raised in inverse relation to the elasticity of demand in each market (raising prices more in markets with lower elasticity and less in markets with higher elasticity).

The general rule, when there are no cross-elasticities, is that, at the second-best prices

$$((P_1 - MC_1)/P_1) \cdot \epsilon_1 = ((P_2 - MC_2)/P_2) \cdot \epsilon_2, \tag{4.2}$$

where ϵ is the elasticity of demand. The term $P_1 - MC_1$ is the amount by which price in market one exceeds marginal cost for that good. Dividing this by P_1 gives the amount by which price exceeds marginal cost expressed as a proportion of price. The equation states that, at second-best prices, if the percentage by which price exceeds marginal cost in each market is multiplied by the elasticity of demand in that market, the resulting product is the same for all markets.

This equation holds in our numerical example. In market one, elasticity is $-.85$, price is 23, and marginal cost 20. Price exceeds marginal cost by 3, which is 13% (2/23) of the price. The product of the elasticity and the percent increase of price over marginal cost is $-.11$ ($= -.85 \cdot .13$). In market two, elasticity is -1.2, price is 22, and marginal cost is 20. Price exceeds marginal cost by 9% of price, which, when multiplied by elasticity, is $-.11$. In both markets, the elasticity of demand multiplied by the proportion by which price exceeds marginal cost is the same, as stated in the above equation.

Equation (4.2) is the algebraic expression of the inverse elasticity rule. For this equation to hold, price must be raised farther above marginal cost in markets with lower elasticities of demand. That is, if ϵ is smaller in one market than another, the term $(P - MC)/P$ must be higher in that first market so that the product $(P - MC)/P \cdot \epsilon$ can be the same in both markets. Thus the equation requires higher prices in markets with lower elasticities.

It is important to note that equations (4.1) and (4.2) are not two separate rules. Rather, they are two different ways of stating the same rule. Equation (4.1) states that second-best prices are attained by reducing output in each market by the same proportion. Equation (4.2) states that second-best prices are attained by increasing price in the market with the lower elasticity. However, equation (4.1) implies equation (4.2) and vice versa: if outputs are reduced by the same proportion in all markets, price necessarily rises more in markets with lower elasticity. Consider figure 4.4. First-best output is Q_F in each market, which is obtained when prices are set to marginal costs. If output is reduced by the same proportion in each market to Q_S, the price in the first market rises to P_1 and that in the second market to P_2. That is, a given proportion reduction in output in both markets

This result has intuitive meaning. Raising prices has two effects. First, it transfers money from consumers to the producer, because consumers have to pay more for the goods they purchase. Second, it reduces the quantity of goods sold, because consumers generally demand fewer of the good when its price is higher. The degree to which each of these two effects occurs depends on the elasticity of demand. If, as in panel (a) of figure 4.3, demand is highly inelastic (that is, consumers are not very responsive to price), then raising price from P_0 to P_1 transfers a considerable amount of money to the firm (its profits increase by the shaded area) and reduces the quantity sold by very little. However, when demand is more elastic (that is, consumers are more price responsive), as in panel (b), the same price increase results in a smaller transfer of money from consumers to the firm and a larger reduction in output. If the firm is losing money, a certain amount of money must be transferred to the firm for it to break even. More funds can be obtained with less disruption in consumer's consumption patterns (that is, less reduction in output) by raising price in the market with inelastic demand than in the market with elastic demand. This fact is essentially what the inverse elasticity rule is stating: raise price more in the market with a lower elasticity of demand.

The precise statement of this characteristic of second-best prices is somewhat more complex than the inverse elasticity rule might suggest. In our example, marginal cost is constant and the same for both markets. An accurate statement of the rule allows for differences in marginal cost. We give this statement below for situations in which demand in each market is independent of the price charged in the other market; that is, no cross-elasticities of demand. In a later section, we generalize to situations with cross-elasticities.

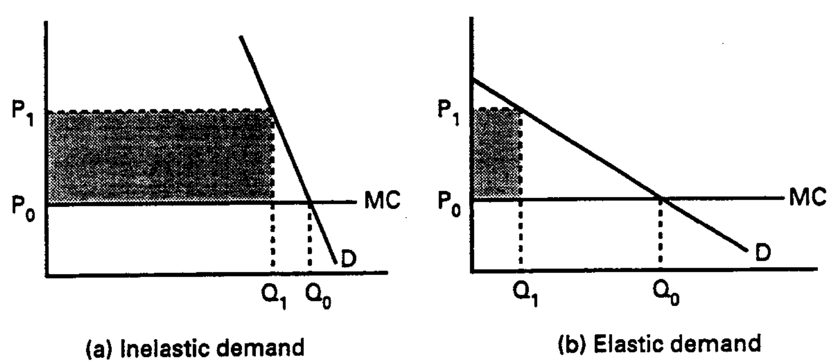

Figure 4.3
Price increase in markets with different demand elasticities

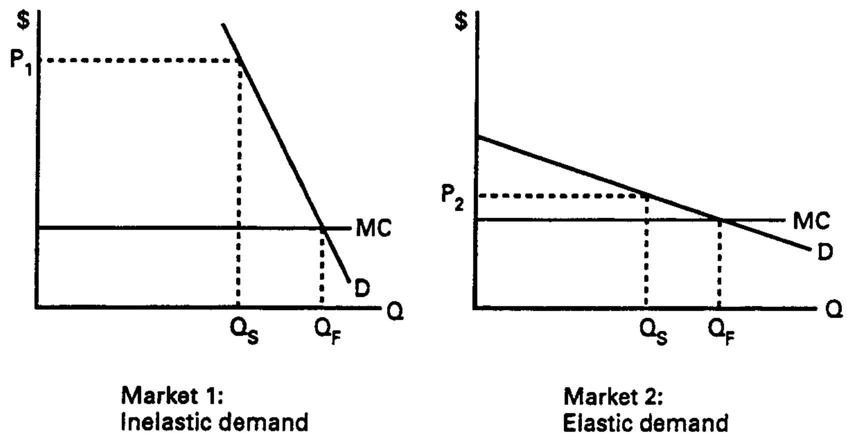

Figure 4.4
Ramsey prices

results in a higher price in market one than in market two. This is reasonable. Because there is less response to price in market one than market two, a larger price increase is needed in market one to induce a given percent reduction in quantity demanded. This illustration implies that, generally, for the output ratio to remain unchanged, price must be raised more in the market with lower elasticity.

Because the two equations are alternative ways of saying the same thing, each equation alone is, or both collectively are, called the Ramsey rule for second-best pricing. Either can be used to calculate second-best prices. That is, prices can be raised above marginal cost in such a way that output ratios remain constant, with the prices raised in this manner until the firm breaks even. Or, price in each market can be raised by an amount that is inversely related to the elasticity of demand in that market until the firm breaks even. Either method will result in the same prices.

4.3 A More Rigorous Derivation of the Ramsey Rule

In the previous section, a numerical example was used to motivate and illustrate the Ramsey rule. We asserted, without proof, that the results obtained in the numerical example occur in all such situations. We now present a more rigorous demonstration of the Ramsey rule. This demonstration is intended to be pedagogic, in that the emphasis is on understanding the meaning of the Ramsey rule and why it is generally true. The analysis avoids the use of calculus so that (1) readers who are not comfortable with calculus can obtain a clear understanding of the result, and (2) readers who know calculus will have the opportunity to think about the meaning behind the equations, which is often obscured in purely mathematical proofs. A formal analysis is provided by Baumol and Bradford (1970).

Suppose the following:

1. A firm produces two goods, labeled x and y.

2. Pricing the two goods at their marginal cost results in the firm losing money.

3. It is not possible for the firm to operate while losing money. That is, the firm cannot be subsidized.

4. Demand for the two goods is independent, in that the price of one good does not affect the demand for the other good. That is, cross-price elasticities are zero.

5. Demand for each good is linear.

The first assumption allows us to show results on two-dimensional graphs; generalization to three or more goods is straightforward. The second assumption is consistent with the firm being a natural monopoly. The third assumption reflects the way natural monopolies are generally regulated in the United States. It can be relaxed to allow the firm to lose up to a certain amount of money (the amount of its subsidy); the Ramsey rule is still applicable as long as the firm would lose more than the subsidy amount if it priced at marginal cost. The fourth and fifth assumptions are for convenience of exposition only. With these two assumptions, the Ramsey rule takes a form that is particularly intuitive. In a later section we discuss how allowing for nonlinear demand and cross-elasticities generalizes the form of the Ramsey rule.

We state the Ramsey rule first and then derive it.

Ramsey rule: Given a situation described by assumptions (1)–(5), the prices that provide the greatest surplus while also allowing the firm to break even are those at which profits are zero and

$$\frac{\Delta Q_1}{Q_1}=\frac{\Delta Q_2}{Q_2}, \tag{4.1}$$

or, alternatively,

$$\frac{(P_1-MC_1)}{P_1}\epsilon_1=\frac{(P_2-MC_2)}{P_2}\epsilon_2, \tag{4.2}$$

where ϵ is the price elasticity of demand and ΔQ is the change in output from its level when prices are at marginal cost.

We first introduce two graphical devices: the zero-profit contour for the firm and isobenefit contours for consumers. Then, by combining these two concepts, we determine the prices and outputs at which consumers obtain the greatest surplus while the firm breaks even.

Consider first the firm. The profit the firm earns is completely determined once the firm sets its prices for the two goods. Given the price of each good, the demand curve for each good determines the quantity sold. Given the quantity sold, the technology and input prices the firm faces (as embodied in the firm's cost curves) determine the minimum cost of producing the goods. The profit of the firm is simply its revenues (the product of its price and output levels) minus its costs.

The relation between profit and prices has the form of a hill, as shown in figure 4.5. When prices are very low, the firm loses money, because its prices are not high enough to cover its costs. As prices are raised, the firm's profit increases (losses decrease) and the firm starts to earn positive profit. Profit continues to increase as prices are raised. Eventually, however, prices are raised so much that demand for the goods is choked off and profit starts to drop. That is, beyond a certain point, increasing prices decreases demand sufficiently that the profit of the firm declines. Eventually, at high enough prices, profit again becomes negative.

The relevant information can be depicted in two dimensions. To

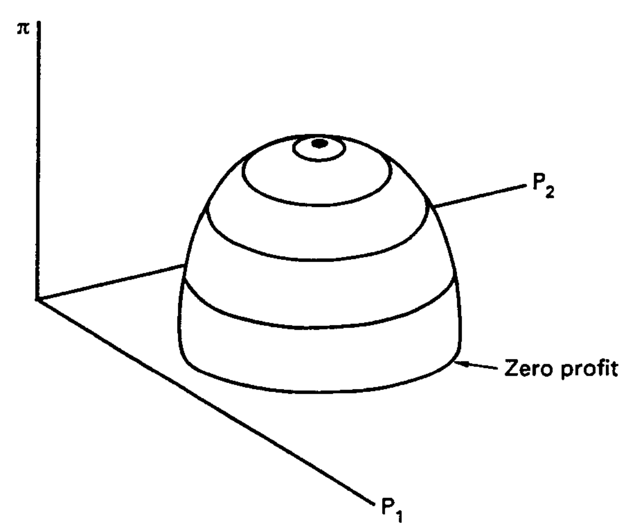

Figure 4.5
Relation of profit to prices

remain solvent, the firm must earn at least zero profit. The base of the profit hill in figure 4.5 (or, more precisely, the points at which the profit hill is cut by the $P_1 - P_2$ plane) is the set of price combinations that result in the firm earning exactly zero profits. These price combinations are depicted in two dimensions by suppressing the profit dimension, as in figure 4.6. The "zero-profit contour" in this latter figure is the set of prices that result in zero profit. Note that any price combination that is inside this zero-profit contour results in strictly positive profit, and any price combination that is outside results in negative profit. To remain solvent, the firm must charge prices that are either inside or on the zero-profit contour.

Figure 4.6 illustrates the issue that Ramsey prices address. If prices were set to marginal cost, the firm would be at point F (first-best prices), which is outside the zero-profit contour. To remain solvent, the firm must increase one or both of the prices so as to move to the zero-profit contour. The question is: Of all the price combinations that the firm could move to on the zero-profit contour, which one is best for consumers?

We now introduce a graphical method for representing consumers' surplus. At any set of prices, consumers obtain some amount of surplus. This surplus increases when the price of either good decreases; that is, consumers benefit from reduced prices. Conversely, surplus decreases as the price of either good increases. The relation between consumer surplus and prices can be represented, as in figure 4.7, as a surface that is highest when prices are zero and drops when prices rise.

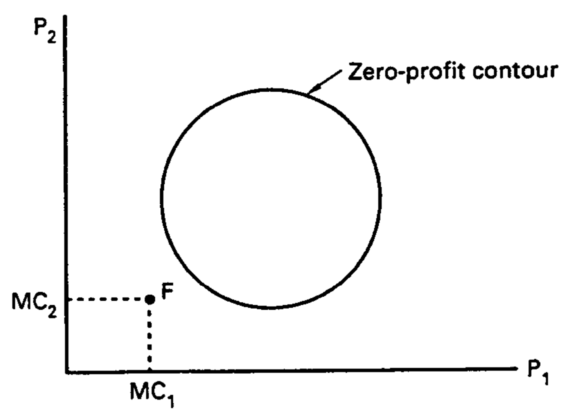

Figure 4.6
Zero-profit contour

The information in figure 4.7 can be represented in two dimensions by making a contour map of the surface (i.e., a topological map). Figure 4.8 is a contour map of the three-dimensional surface in figure 4.7. Each contour in figure 4.8 is the set of prices that result in a particular level of surplus for consumers. For example, all price combinations on the contour labeled b_1 result in $\$b_1$ of surplus for consumers. Each of these contours is called an "isobenefit contour" because it represents a set of price combinations that provide the same ("iso") level of benefits to consumers.[4]

Consumer surplus is greater on isobenefit contours that are closer to the origin (closer to zero prices.) In the figure, this means that consumer surplus level b_3 is greater that b_2, which in turn is greater than b_1. Consumers are therefore made as well off as possible by moving as far inward on the isobenefit mapping as possible.

The goal of Ramsey pricing is to make consumers as well off as possible while allowing the firm to break even. To determine which prices accomplish this goal, the isobenefit mapping for consumers is superimposed with the zero-profit contour for the firm, as in figure 4.9. The Ramsey prices are found by examining all of the price combinations on or inside the zero-profit contour (because these are the prices that allow the firm to remain solvent) and determining which of these touches the lowest isobenefit contour (because the lowest, or most inward, contour represents the highest consumer surplus). This price combination is labeled S (second-best). In particular, the Ramsey prices in this situation are P_1^S and P_2^S.

Note that at the Ramsey price combination S, the isobenefit contour is tangent to the zero-profit contour.[5] The isobenefit and zero-profit contours being tangent at S means that they have the same slope at

4. The isobenefit contours are downward sloping. This feature reflects the fact that, if one price is raised, the other price must be lowered for consumer surplus to remain unchanged. (Consumers are hurt by an increase in one price; to keep consumers' welfare unchanged—no better or worse—consumers must be helped a commensurate amount by lowering the other price.) Also, the isobenefit contours bow inward, becoming less steeply sloped as P_1 is raised and P_2 is lowered. This feature reflects the expectation that changing a price has less impact on consumer surplus when that price is relatively high (and hence consumption of the good is low) than when price is lower.

5. Stated alternatively: if the isobenefit contour and zero-profit contour are *not* tangent at a particular price combination, that price combination cannot be Ramsey. Consider point G. At this point, consumer surplus can be increased without hurting the firm by moving along the zero-profit contour to lower isobenefit contours (that is, toward S). Because consumer surplus can be higher than at G with the firm still breaking even, G cannot be the Ramsey price.

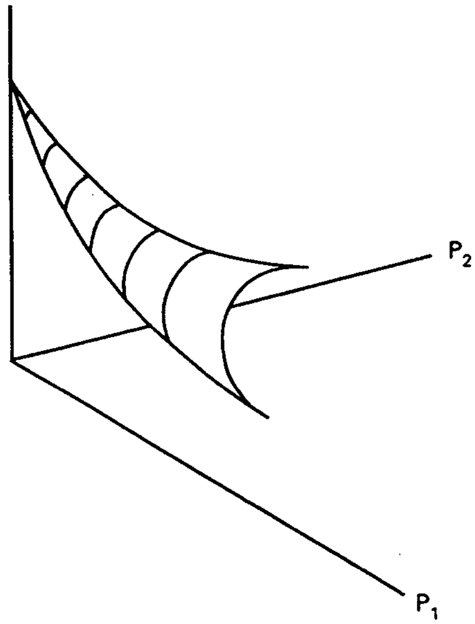

Figure 4.7
Relation of consumer surplus to prices

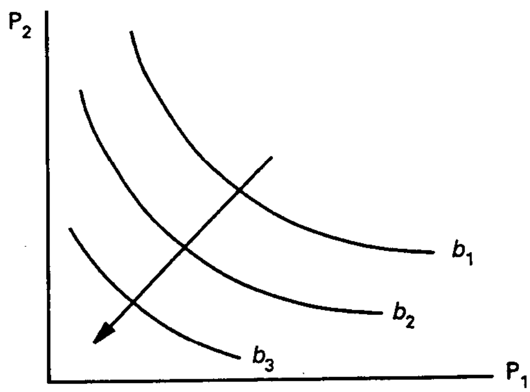

Figure 4.8
Isobenefit contours

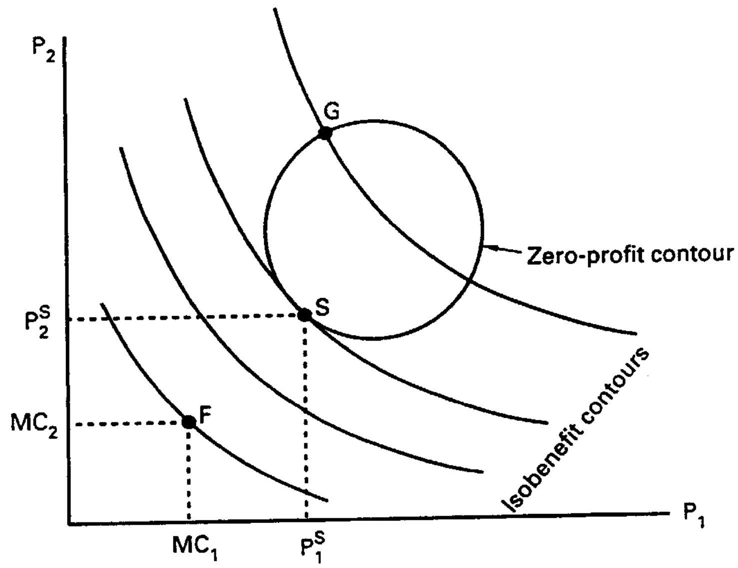

Figure 4.9
Ramsey prices

that point. The Ramsey rule (that is, equations 4.1 and 4.2) is derived from this fact.

Three steps are required to demonstrate the Ramsey rule. First, we derive a formula for the slope of the isobenefit contour. Second, we derive a formula for the slope of the zero-profit contour. Third, we set these two formulas equal to each other (because the slopes of these two contours are equal at the Ramsey prices). This equation, when rearranged, takes the form of equation (4.1) or (4.2), that is, becomes the Ramsey rule. The three steps are discussed separately below.

Step 1: The slope of the isobenefit contour at any price combination is $-Q_1/Q_2$*, where* Q_1 *and* Q_2 *are the quantities demanded at that price combination. That is, the slope of the isobenefit contour at any point is the (negative) ratio of outputs demanded at that point.*

Let us demonstrate this fact. By definition, the slope of the isobenefit contour is the amount by which P_2 must drop in order for consumer surplus to remain constant when P_1 is raised by one unit (that is, the decrease in P_2 is required to keep consumers on the same isobenefit contour when P_1 is raised one unit). Consider a person who consumes quantities Q_1 and Q_2 at given prices. Suppose P_1 increases by \$1. For the consumer to continue consuming quantity Q_1 of the good, the consumer must pay \$$Q_1$ more: \$1 more for each unit consumed.

For the person to be able to still afford Q_1 and Q_2, the price of good two must decline. Each \$1 decrease in P_2 saves $\$Q_2$; or, stated equivalently, each ($\$1/Q_2$) decrease in P_2 saves \$1. To get back $\$Q_1$, there must be a $\$Q_1(1/Q_2)$ decrease in P_2. That is, for the consumer to remain unaffected by the rise in price of good one, the price of good two must be lowered by $\$Q_1/Q_2$.

This fact can be illustrated with a concrete example. Suppose a person buys ten shirts and five pairs of jeans per year. If the price of shirts goes up by \$1, the consumer would have to pay \$10 extra for the ten shirts. To make up this \$10, the price of jeans would have to drop by \$2 such that the consumer would have to pay \$10 less for the five pairs of jeans. The \$2 is simply the quantity of shirts divided by the quantity of jeans.

The change in P_2 that allows the person to continue buying the same consumption bundle in the face of a \$1 increase in P_1 is, as we have discussed, $-Q_1/Q_2$, where the negative sign indicates that P_2 must drop. Because the person is consuming the same quantities and paying the same amount in total, the person's consumer surplus is the same. Consequently, $-Q_1/Q_2$ is the slope of the isobenefit contour, namely, it is the drop in P_2 that is necessary to keep consumer surplus constant when P_1 is raised one unit.

Perceptive readers will point out that the consumer will not choose to consume the same quantities of the two goods when their prices change, but rather will respond to the new prices by increasing consumption of good two (whose price has dropped) and less of good one (whose price has risen). This observation is correct for sufficiently large changes in prices. However, for sufficiently small changes in prices, the consumer will not change consumption levels. If a \$1 change in P_1 would induce the consumer to change consumption levels, then the units can be changed to consider, say, a 1 cent change in P_1. The analysis is the same: P_2 must drop by Q_1/Q_2¢ to compensate for a 1 cent increase in P_1. Technically, the slope of the isobenefit contour, as all slopes, is defined for infinitesimally small changes, under which consumption levels do not change.

Step 2: The slope of the zero-profit contour is

$$-\frac{Q_1+(P_1-MC_1)s_1}{Q_2+(P_2-MC_2)s_2},$$

where s_1 is the slope of the demand function for good one, and analogously for good two.

We demonstrate this fact as follows. The slope of the zero-profit contour is, by definition, the amount by which P_2 must change for profits to remain zero when P_1 is raised by one unit. If P_1 is raised by \$1, two things occur. First, the firm earns an extra dollar of revenue on each unit that it sells, such that its profits increase by \$$Q_1$. Second, the quantity demanded decreases when the price increases, and the firm loses the profits that it earned on these units. This loss is the difference between the revenues it earns per unit (P_1) and the marginal cost of each unit (MC_1) multiplied by the number of units by which demand decreases (s_1, where s_1 is the slope of the demand function, with the demand function giving quantity demanded as a function of price).[6] Summing these two effects, the change in profits that results from a \$1 increase in P_1 is $Q_1 + (P_1 - MC_1)s_1$. Label this quantity $\Delta\pi_1$. Similarly, a \$1 decrease in P_2 changes profits by $-(Q_2 + (P_2 - MC_2)s_2)$, which we label $-\Delta\pi_2$.

For profits to remain constant when P_1 rises by \$1, P_2 must drop by an amount that exactly offsets the gain in profits attributable to the rise in P_1. Profits rise by $\Delta\pi_1$ when P_1 increases by \$1. Each \$1 decrease in P_2 reduces profit by $\Delta\pi_2$, or, stated equivalently, each \$$(1/\Delta\pi_2)$ decrease in P_2 reduces profit by \$1. Therefore, to reduce profit by $\Delta\pi_1$, P_2 must be reduced by $\Delta\pi_1/\Delta\pi_2$.[7]

Substituting in the terms for $\Delta\pi_1$ and $\Delta\pi_2$, we know that the change in P_2 that is necessary to maintain constant profits when P_1 is raised \$1 is

$$-\frac{Q_1+(P_1-MC_1)s_1}{Q_2+(P_2-MC_2)s_2}.$$

This is the slope of the zero-profit contour.

Step 3: Equate the slopes and rearrange for the Ramsey rule.

At Ramsey prices, the slope of the isobenefit contour equals the slope of the zero-profit contour. Setting the expressions for these slopes equal to each other, we have

6. Demand is often represented with price being a function of quantity. For example, demand is usually graphed with quantity on the x-axis and price on the y-axis, such that the relation is price as a function of quantity. In this case, s is the inverse of the slope of the demand curve. In either case, s is the same quantity, namely, the decrease in output that results from an increase in price.

7. For example, suppose raising P_1 by \$1 increased profits by \$100 and lowering P_2 by \$1 decreased profits by \$50. It would be necessary to lower P_2 by \$2 (i.e., \$100/\$50) for profits to stay constant when P_1 is raised by \$1.

$$-\frac{Q_1+(P_1-MC_1)s_1}{Q_2+(P_2-MC_2)s_2}=-\frac{Q_1}{Q_2}.$$

Rearranging:

$$(Q_1+(P_1-MC_1)s_1)\ /\ Q_1=(Q_2+(P_2-MC_2)s_2)\ /\ Q_2$$

or

$$1+(P_1-MC_1)(s_1/Q_1)=1+(P_2-MC_2)(s_2/Q_2).$$

Subtracting one from both sides:

$$(P_1-MC_1)(s_1/Q_1)=(P_2-MC_2)(s_2/Q_2). \tag{4.3}$$

Multiplying the left-hand side by (P_1/P_1) and the right by (P_2/P_2) does not change the equation because these quantities are simply one:

$$((P_1-MC_1)/P_1)s_1\cdot(P_1/Q_1)=((P_2-MC_2)/P_2)s_2\cdot(P_2/Q_2).$$

Note that $s_1(P_1/Q_1)$ is the elasticity of demand for good one, which is labeled ϵ_1; and similarly for good two.[8] Using this fact, the equation then becomes

$$\frac{(P_1-MC_1)}{P_1}\epsilon_1=\frac{(P_2-MC_2)}{P_2}\epsilon_2,$$

which is equation (4.2), the inverse elasticity rule. Thus we have shown that at Ramsey prices, the inverse elasticity rule holds.

We now proceed to demonstrate equation (4.1), namely, that at Ramsey prices, the quantity of each good is reduced by the same proportion below its marginal-cost level. The quantity demanded of each good is lower at Ramsey prices than when each good is priced at its marginal cost; label this reduction in demand for good one as ΔQ_1, and analogously for good two. Because demand is linear, the amount by which demand for good one is reduced is equal to the amount by which price is raised above marginal cost (namely, $P_1 - MC_1$) multiplied by the slope of the demand function (s_1, which is the change in output for each one-unit change in price). That is, $\Delta Q_1 = (P_1 - MC_1)s_1$, and similarly for good two. Substitute this relation into equation (4.3) to obtain

8. The elasticity is, by definition, the percent change in quantity that results from a percent change in price: $\epsilon = (\Delta Q/Q)/(\Delta P/P)$. This term can be arranged as $\epsilon = (\Delta Q/\Delta P)(P/Q)$. The slope of the demand curve is $s = \Delta Q/\Delta P$, such that $\epsilon = s(P/Q)$, which is the term that appears in the expression above.

$\Delta Q_1/Q_1 = \Delta Q_2/Q_2,$

which is equation (4.1), stating that the percent change in output from its marginal-cost level is the same for both goods.

4.4 Finding the Ramsey Prices

The Ramsey rule describes relations that must hold at the second-best prices. For example, equation (4.2) states that, at the Ramsey prices, the elasticity of demand times the percent by which price exceeds marginal cost is the same for all goods. It is important to note, however, that the Ramsey rule can be used to *find* the Ramsey prices, as well as characterize events that occur at the prices once they are found.

To find the Ramsey prices, we start by setting price equal to marginal cost for each good. If these prices result in the firm earning zero or positive profits, we retain these prices and obtain first-best optimality. However, if marginal-cost pricing results in negative profits (as we have assumed as the motivation for this chapter), then prices must be raised to allow the firm to remain solvent.

According to equation (4.2) of the Ramsey rule, prices must be raised in such a way that the elasticity of demand times the percent deviation of price from marginal cost is the same for all goods. Note that this equation holds at the marginal-cost prices: because price equals marginal cost, the percent deviation of price from marginal cost is zero for both goods, and the product of this deviation with the elasticity is also zero for each good, independent of the size of the elasticity. In fact, the equation holds for a whole set of prices, not just the Ramsey and marginal cost prices. The Ramsey prices are unique in that they are the only prices at which equation (4.2) holds *and* the firm makes zero profits. Other price combinations that satisfy equation (4.2) result in either negative or strictly positive profits. This fact is the key to finding the Ramsey prices.

Consider figure 4.10. The upward-sloping curve denotes the set of price combinations for which equation (4.2) holds. The Ramsey prices and marginal-cost prices are necessarily on this curve. To find the Ramsey prices, prices are first set equal to marginal cost. Then prices are raised slightly and in such a way that Equation (4.2) holds; this moves the prices from F to a point, say B, somewhat further up on the curve. This change in prices increases the firm's profit (i.e., decreases its loss); however, for a small-enough change in prices, the firm will still be losing money. Prices are therefore raised again, making sure that equation (4.2) holds. This process is continued until prices are raised sufficiently that the firm breaks even. Stated succinctly: prices are raised by successively larger amounts, always in such a way that equation (4.2) holds, until prices are found that allow the firm to break even. Graphically, prices are moved up the curve in figure 4.10 until the zero-profit contour is reached; the intersection of the curve with the zero-profit contour is the Ramsey prices.

Ramsey prices can also be found using equation (4.1). This equation states that at the Ramsey prices, the output of each good is reduced by the same proportion from its level under marginal-cost pricing. Note that while this equation says that the proportionate reduction is the *same* for each good, it does not state what this proportion is. Each output could be reduced by 1%, 5%, 10%, or whatever, and equation (4.1) would hold. The percent reduction that results in zero profit is the appropriate one. This fact provides a procedure for finding the Ramsey output levels: start at the quantities demanded under marginal-cost pricing and then reduce output for each good by successively larger proportions until the firm breaks even.

Figure 4.11 depicts the situation. Note that the axes on this graph are the quantities of output for the two goods, rather than prices as in previous figures. Each point on the graph represents an output combination for the two goods. The zero-profit contour depicts the set of output combinations that result in zero profit. Any output combination that is outside of this contour results in negative profits, while

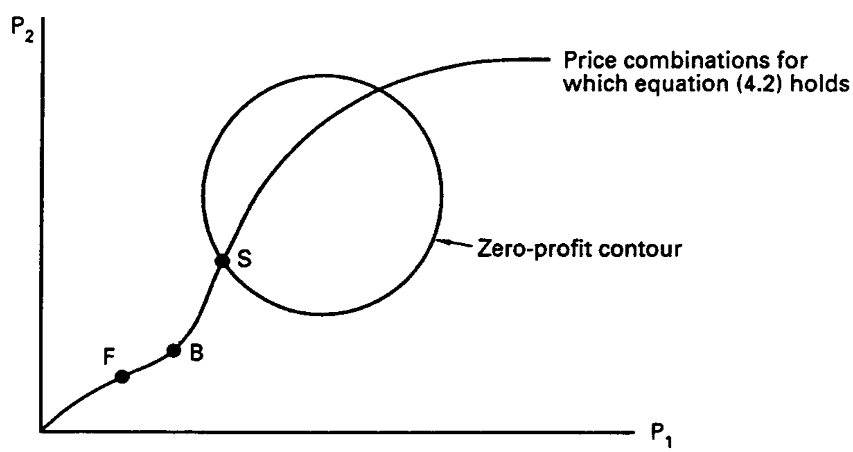

Figure 4.10
Using equation (4.2) to find Ramsey prices

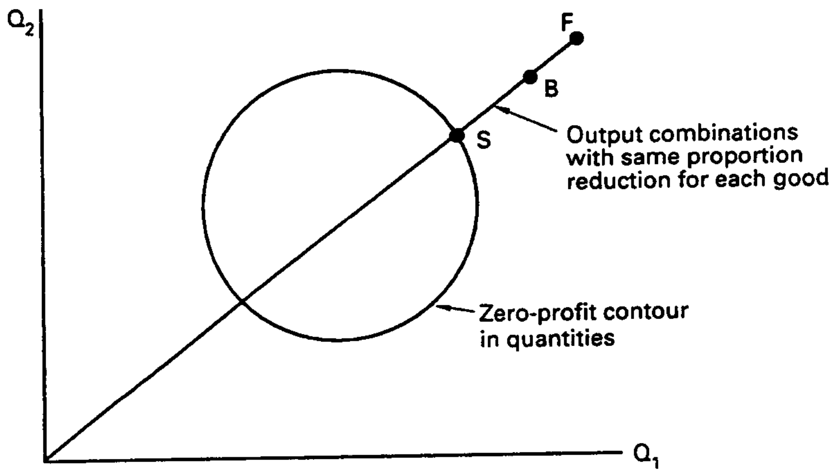

Figure 4.11
Using equation (4.1) to find Ramsey prices

a combination inside the contour provides positive profits. At marginal-cost pricing, output combination F is demanded. At this output combination, the firm is presumably losing money (and hence is located outside the zero-profit contour). For the firm to break even, output must be reduced (that is, prices must be raised). The ray from the origin through point F is the set of output combinations that are obtained by reducing the quantity of each output by the same proportion. Reducing each output by zero proportion is equivalent to no change, and so point F is on the ray. Reducing each output by 100% is equivalent to producing no output, such that the origin is on the ray. Any other point on the ray is obtained by reducing both outputs by some proportion between zero and 100%.

To find the Ramsey output levels (that is, the outputs that would result from Ramsey prices), we start at point F, the outputs obtained under marginal cost pricing, and reduce both outputs by some small proportion, adjusting the price of each good appropriately to obtain this equal-proportion reduction. This reduction moves the firm from point F inward on the ray to, say, point B. The firm loses less money as a result of this reduction in output (and the corresponding increase in price), but it is still losing money. Each output is therefore decreased again, by a larger proportion. This process is continued, reducing both outputs by an increasing large proportion until the firm breaks even.

Note that the two methods for finding Ramsey prices are equivalent, in that they result in the same prices and output. That is: if outputs are reduced proportionately until the firm breaks even, the prices that result in these output levels being demanded are the same prices that would be obtained if prices were raised in accordance with equation (4.2) until the firm breaks even; and vice versa. This correspondence is simply a reflection of the fact that equations (4.1) and (4.2) are alternative and equivalent ways of stating the same result. Equation (4.1) is expressed in terms of output, with prices being implicit; equation (4.2) is expressed in terms of price, with output being implicit.

Finally, it is important to note that these methods for finding Ramsey prices are not meant to be applied to a regulated firm in real time. That is, it is not being suggested that the regulator set prices at marginal cost, observe the firm's loss, and then slowly raise prices appropriately until the firm is observed to break even. Rather, the methods are meant to be used as a means for calculating Ramsey prices given information about the firm's costs and demand. An application of the use of the methods in a real-world setting is provided in section 4.6.

4.5 Relaxation of Assumptions

Two assumptions that have been maintained in the discussion.so far can be relaxed for a more general statement of the Ramsey rule. These assumptions are: (1) there are no cross-elasticities of demand, such that the price of one good does not affect the demand for the other, and (2) the demand curve for each good is linear. These assumptions simplify the analysis considerably and allow a clearer view of the meaning and purpose of the Ramsey rule. However, in most situations they do not hold. For example, energy utilities often sell both natural gas and electricity. Because either of these power sources can be used for heating, one would expect that if the price of electricity is raised, some consumers who use electric heaters would, over time at least, switch to gas heating. As a result, the demand for gas would increase in response to the higher price for electricity, contrary to the assumption of no cross-elasticity. Similarly, the assumption about linear demand is probably unrealistic in many if not most settings.

The Ramsey rule can be generalized to allow for situations in which these two assumptions do not hold. The more general rule is derived in a way that is analogous to our derivation in the previous section, but with more cumbersome notation. We simply state the more general rule and explain its meaning intuitively. Interested readers can work through the algebra themselves.

Consider first a situation with cross-elasticities of demand. The more general version of equation (4.2), which allows for cross-elasticities, is the following (Dreze 1964):

$$\frac{(P_1 - MC_1)}{P_1}(\epsilon_1 - \epsilon_{21}) = \frac{(P_2 - MC_2)}{P_2}(\epsilon_2 - \epsilon_{12}), \tag{4.2'}$$

where ϵ_{21} is the elasticity of demand for good two with respect to the price of good one, and analogously for ϵ_{12}.[9] This equation is essentially the same in meaning as the original version. However, in this more general version, the cross-elasticities are subtracted from the own-price elasticities. This subtraction gives, in a sense, a "net" elasticity: the effect of one good's price on the demand for that good itself *net* of the effect on the demand for the other good. Note that if the cross-elasticities are zero, then this more general statement of the Ramsey rule reduces to the original statement (that is, becomes equation 4.2).

Equation (4.2′) holds whether demand is linear or nonlinear. As such, it is a fully general statement of the Ramsey rule, applicable with zero or nonzero cross-elasticities and with linear or nonlinear demand curves.

Equation (4.1), which states that each output is reduced by the same proportion, still applies without modification if there are cross-elasticities of demand. However, if demand is nonlinear, it holds only approximately. Recall that the demonstration of equation (4.1) uses the fact that ΔQ_1 is equal to $(P_1 - MC_1)s_1$, because demand is assumed linear. That is, the amount by which output changes from its marginal-cost level is equal to the slope of the demand function, s_1, times the amount by which price is raised above marginal cost. If demand is not linear, then the slope is not constant; rather, the slope changes as one moves along the demand curve. With nonlinear demand, ΔQ_1 must be calculated using the *average* slope of the demand function between P_1 and MC_1. The slope of a nonlinear demand function at the Ramsey prices is only approximately the same as this average slope. Consequently, with nonlinear demand, ΔQ_1 is only approxi-

9. That is, ϵ_{21} is the percent change in demand for good two that results from a 1% change in the price of good one; and analogously for ϵ_{12}. To be perfectly accurate, the elasticities for this formulation are taken on the compensated demand curve rather than uncompensated demand.

mately equal to $(P_1 - MC_1)s_1$, where s_1 is the slope at the Ramsey prices. The general statement of equation (4.1), which allows for nonlinear demand, is therefore

$$\Delta Q_1/Q_1 \approx \Delta Q_2/Q_2, \tag{4.1'}$$

where an approximately equal sign replaces the equal sign.

For small deviations from marginal cost, the average slope (averaged over the part of the demand curve between marginal cost and the Ramsey prices) is nearly the same as the slope at the Ramsey prices. Consequently, the approximation (equation 4.1′) is better for smaller deviations from marginal-cost prices (that is, when the Ramsey prices are fairly close to marginal cost). Furthermore, the approximation is better for demand curves that are more nearly linear, becoming exact when demand is perfectly linear in the relevant region.

These generalizations of the Ramsey rule, while perhaps adding complications conceptually, do not introduce difficulties from a practical perspective. Equation (4.2′) is nearly as easy to apply as its more restricted version (equation 4.2). In either case, the researcher or regulator uses information on demand and costs.[10] Furthermore, equation (4.1) can often be applied as a strict equality without undue concern about the approximation. That is, while demand might not be linear throughout the entire demand curve, the part of demand between marginal cost and the Ramsey prices might be sufficiently linear, and/or marginal cost and Ramsey prices might be sufficiently close, such that reducing each output by exactly the same proportion will not result in unreasonable errors. The following section provides an application of the Ramsey rule in a real-world situation.

4.6 An Application of the Ramsey Rule: Transit Pricing

The East Bay area of the San Francisco Bay region includes Berkeley, Oakland, Walnut Creek, and numerous other cities, plus some unincorporated areas. Two forms of public transit are provided in this area: bus service by the Alameda-Contra Costa (AC) Transit Company and rail service by the Bay Area Rapid Transit (BART) system. A regional transportation agency, the Metropolitan Transportation

10. Information on cross-elasticities is required for application of equation (4.2′), whereas they are assumed to be zero for equation (4.2). However, given information on same-price and cross-elasticities, Ramsey prices are as easy to calculate from equation (4.2′) as from equation (4.2).

Commission (MTC), coordinates service among the various transit agencies in the San Francisco Bay region and exercises considerable oversight of each agency's fares.

As with most public transit providers, AC Transit and BART are natural monopolies in that their marginal cost is below their average cost over the relevant range of output. Consequently, pricing each service at its marginal cost would result in the two agencies losing money. Given the authority of MTC, the possibility of Ramsey pricing is feasible in this situation. The two transit providers can be considered one for the purpose of pricing and covering costs. Ramsey prices for the two services are those that provide the greatest surplus for travelers while allowing the combined revenues for the two agencies to cover their combined costs. MTC could administer any cross-subsidization that is required at the Ramsey prices.[11]

Ramsey prices for AC Transit and BART have been calculated by Train (1977) using demand functions estimated by McFadden (1975) and cost functions estimated by Lee (1974) for AC Transit, and Merewitz and Pozdena (1974) for BART. For the demand functions, travelers are assumed to choose among bus, rail, and auto for each of their trips and to make this choice on the basis (at least partially) of the cost and time of taking the trip by each mode. The demand for each mode therefore depends on the price for that mode as well as the price for other modes. This characteristic of the demand functions reflects the fact that, if bus fares rise, some bus patrons will switch to rail, and similarly for rail fares. Because cross-elasticities are explicitly incorporated in the demand relations, the calculation of Ramsey prices utilizes the generalized version of the inverse elasticity rule, equation (4.2').

In the current context, equation (4.2') takes the following form:

$$\frac{(P_r - MC_r)}{P_r}(\epsilon_r - \epsilon_{br}) = \frac{(P_b - MC_b)}{P_b}(\epsilon_b - \epsilon_{rb}), \tag{4.4}$$

where r denotes rail and b denotes bus. In addition to satisfying this equation, Ramsey prices allow the providers to break even. In the

11. Each provider's individual revenues will not necessarily exactly cover the costs for that service. If each provider priced separately, at average cost, each service's revenues would cover its own costs. The value of Ramsey pricing in this situation is that it allows greater surplus because it involves only one constraint on prices (namely, that combined revenues cover combined costs) rather than two constraints (namely, that each of the two provider's revenues cover its own costs).

context of AC Transit and BART, breaking even takes a slightly different meaning. BART is required by law to cover its operating costs, but its capital costs are paid through a regional sales tax. AC Transit is assumed to be required to cover all of its costs. Therefore, for the two agencies in combination to break even, bus and rail revenues must be sufficient to cover the operating costs of BART and all of AC Transit's costs.[12] This break-even constraint is expressed as

$$P_rQ_r + P_bQ_b = OC_r + TC_b, \tag{4.5}$$

where Q is quantity, OC is operating cost, and TC is total cost. Because fares for AC Transit and BART are distance-based, quantity is expressed in passenger-miles (that is, the sum over passengers of the number of miles traveled by each passenger). Price is correspondingly expressed in cents per mile of travel.

Ramsey prices for AC Transit and BART are those that satisfy both equations (4.4) and (4.5). The most straightforward way to determine the Ramsey prices in this context is to consider each possible price combination, use the demand and cost functions to calculate the terms in equations (4.4) and (4.5), and observe whether the equations hold at these prices.

Consider the break-even constraint first (that is, equation 4.5). At any price combination (that is, at any price for bus travel and price for rail travel), the demand functions determine the quantity of travel on each mode (i.e., Q_r and Q_b). Quantities times prices gives revenues. The cost function for AC Transit is then used to determine the total cost of providing Q_b passenger-miles of travel on bus, and the cost function for BART determines the operating cost of providing Q_r. Total revenues are compared with the sum of BART operating cost and AC Transit total cost to determine whether the combined transit provider, AC Transit/BART, breaks even.

The price combinations at which combined revenues equal combined costs are charted as curve A in figure 4.12. This curve is the relevant portion of the zero-profit contour. If the axes on the graph were extended (that is, if higher prices were represented on the graph), the curve would extend to form a circular contour (as in figure 4.10).

Consider now equation (4.4). This equation states that the percent by which price deviates from marginal cost, multiplied by the "net" elasticity, is the same for both rail and bus. At each price combina-

12. In actuality, AC Transit is subsidized through state funds. However, the amount of subsidy is not fixed and varies from year to year. For any given level of subsidy, the

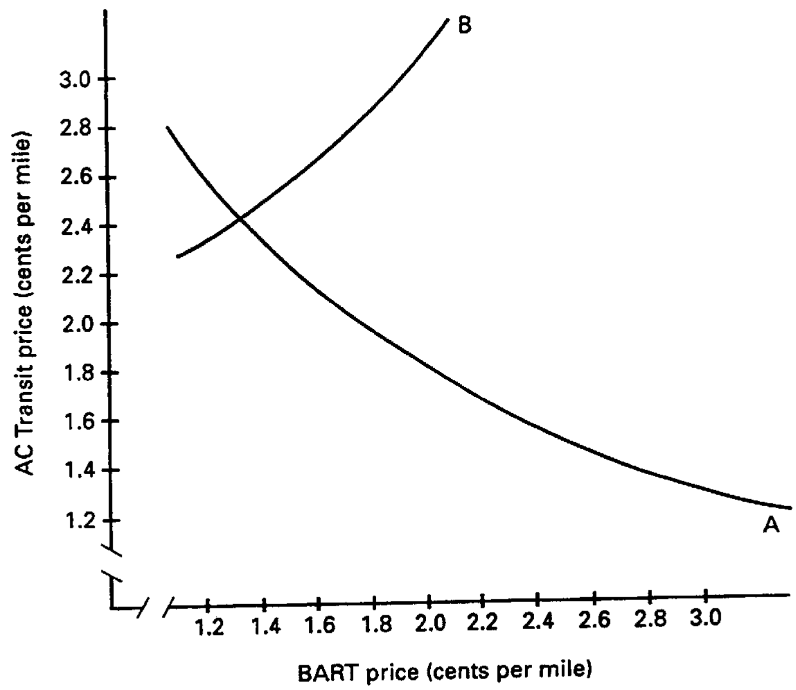

Figure 4.12
Ramsey prices for AC Transit and BART

tion, the demand functions are used to calculate elasticities. Marginal costs are determined from the cost functions. For each mode, the percent deviation of price from marginal cost is multiplied by the "net" elasticity. Curve *B* charts the price combinations at which these products are the same for bus and rail.

The intersection of curves *A* and *B* is the Ramsey price combination, because both equations (4.4) and (4.5) are satisfied at this point. The Ramsey prices are 2.42 cents per mile for bus and 1.28 cents per mile for rail. For comparison, the average cost of bus travel is 2.0 cents per passenger-mile, and the average operating cost for rail is 1.78 cents. Because the Ramsey price for bus exceeds the average cost of bus service, while the rail price is less than BART's average operating cost, bus service in this case would be subsidizing rail.[13] This subsidy is in

13. In many applications of Ramsey prices, some costs are shared jointly in the production of the goods. For example, the same generation capacity is used for the production of electricity for residential and commercial customers, even though the two groups are charged different prices. In these cases, it is not possible to calculate average cost of each service separately. However, in the case of AC Transit and BART, no costs are shared, such that the average cost of each service can be calculated.

addition to the subsidy that covers BART's capital cost (this latter being already reflected in the fact that BART's price is compared to its average *operating* cost, whereas the bus price is compared to its average of all costs).

This subsidy from bus patrons to rail patrons raises some important issues regarding the advisability of Ramsey prices. In the East Bay area, the average income of bus riders is considerably lower than that of BART patrons. This difference in income is partially the result of the routes provided by each service. For example, people who live in suburban areas can easily ride BART into the the financial and commercial centers of the area. In fact, BART is faster and in many ways easier than driving for these trips. Consequently, many of its riders are people who work at relatively high-paying jobs in the city and live in relatively high-income neighborhoods in the suburbs; they often own a car but choose to take BART to work because BART is faster. AC Transit, on the other hand, provides more short-haul trips, especially within the more central areas. A larger percentage of its patrons live in inner-city, lower-income neighborhoods and use AC Transit for travel within the city. Many of the riders do not have cars and take AC Transit not because it is faster than driving but because they do not have the option of driving.

The Ramsey prices, if implemented, would require that the lower-income riders of the bus subsidize the higher-income riders of BART. From an equity perspective, this arrangement would seem unsuitable.

The issue of equity in this application elucidates an important characteristic of Ramsey prices. By construction, Ramsey prices are those that provide the greatest *total* consumer surplus, while allowing the provider to break even. The distribution of this surplus among consumers is not considered. And, in fact, the distribution that results from Ramsey prices might very well, as in this application, seem inequitable.

If total surplus is as high as possible, then there is, theoretically at least, some way that this surplus can be redistributed such that all people are better off than at any other price combination. If the regulator can accomplish this redistribution, then the issue of equity can be resolved. However, generally the regulator cannot effectively implement a redistribution of surplus. In these cases, the regulator needs to consider the equity impacts of Ramsey prices when deciding whether to implement them.

In the current application, Ramsey prices imply that lower-income consumers would subsidize higher-income households. This result is not entirely a coincidence and in some sense is inherent in the concept of Ramsey pricing. Recall that the Ramsey rule is often called the inverse elasticity rule, because Ramsey pricing requires that price be raised further above marginal cost for goods with lower elasticities. A low elasticity of demand means that the consumers of that good are relatively insensitive to price: they will largely continue to buy the good even if its price is raised. When people do not have options and consumption of the good or service is necessary, then people will not be price-responsive: they will, of necessity, buy the good at the higher price. In the case of bus and rail, the BART riders generally own cars that they can drive to work if the cost of BART becomes too high. However, the lower-income patrons of AC Transit often do not own cars, precisely because they have less income, and consequently cannot choose to drive instead of paying a higher bus fare. Furthermore, BART is usually not a viable option for these people, because BART does not serve the inner-city residential neighborhoods as well as the bus. The primary option that these bus riders have to respond to a higher bus fare is not to travel, which for the commute to work would end up costing the person more in lost wages than the extra bus fare.

The basic point is: insofar as lower-income consumers have fewer options, their demands will tend to be less elastic. Application of the Ramsey rule will, in these cases, result in their facing higher prices relative to consumers with more options and hence higher elasticities.

A similar consequence occurs when comparing demands for different goods. Necessities, such as medical care, have very low elasticities of demand because people will largely continue to buy them even when price is raised substantially. The Ramsey rule would imply that prices be raised more on these goods than on goods with more elastic demands. However, it does not seem appropriate for people who become sick or injured to bear an even greater burden through higher prices for care.

These examples point out that the application of Ramsey pricing should be tempered with an appreciation for the distributional consequences of such pricing in any particular situation. The fact that Ramsey prices obtain the greatest total surplus does not guarantee that they are "best" or even "good" by other social criteria that the regulator might consider relevant.