9 Self-Selecting Tariffs and Sibley's Mechanism

9.1 Introduction

In recent years, utilities have begun to offer their customers a choice among tariffs. For example, many electric utilities offer both time-of-use and standard (non-time-differentiated) rates and allow each customer to choose the rates under which its bill will be calculated. Local phone companies often offer both flat-rate and measured service: the customer can choose to be billed either a fixed amount per month without additional charge for local calls (flat-rate service) or a lower fixed amount per month with an additional charge for each local call (measured service). Each long distance carrier offers a variety of tariffs, including WATS and WATS-like services and plans like AT&T's "Reach Out America" program, under which the customer obtains an hour of off-peak calling each month for a fixed cost and then pays a reduced rate for additional calling.

When a customer has a choice among two or more tariffs, the tariffs are called "optional," "voluntary," or "self-selecting." The last term is probably most accurate, because it incorporates the fact that the customer necessarily pays for the service and only chooses the schedule under which its bill is calculated. The term "optional" or "voluntary" tariff seems to mistakenly suggest that paying is optional, like the voluntary admission fee of some museums and performing arts.

Self-selecting tariffs have recently become popular for a variety of reasons. It is usually easier for a regulated firm to obtain permission from its regulator to offer a new rate schedule as an option to customers, rather than as a substitute for the existing schedule. For most new tariffs, some customers would be hurt by being charged under the new schedule; for example, customers who consume a relatively large share of electricity in the peak would be hurt if their consump-

tion were charged under time-of-use rates instead of standard rates. If a new schedule is offered as an option, customers who would be hurt under the new schedule can choose to stay on the existing schedule. This possibility prevents these customers from opposing the introduction of the new schedule. It also avoids the ethical problem of needing to trade off the benefits to some customers against the harm to others in determining whether the new schedule constitutes a social improvement.[1]

Under traditional assumptions about customer behavior, self-selecting tariffs provide utilities and their regulators with a mechanism for increasing surplus. As shown in section 9.4, if prices under an existing tariff are not equal to marginal cost, then it is possible to design a new tariff that, when offered on a self-selecting basis in addition to the existing schedule, increases surplus. It is even possible to attain Pareto dominance, that is, to make every party better off without hurting anyone.

Not all self-selecting tariff offerings allow for Pareto dominance or even increase surplus. When not appropriately designed, the introduction of a new self-selecting tariff can decrease surplus. As described in section 9.4, the appropriate design of self-selecting tariffs requires information on the demand of customers, which the regulator generally does not possess. The basic question therefore arises: how can the regulator induce the firm to design and introduce self-selecting tariffs that increase surplus without the regulator knowing beforehand what tariffs are appropriate?

Sibley (1989) has proposed a regulatory mechanism that, under certain conditions, does just this. In particular, Sibley's mechanism induces the firm to offer a self-selecting tariff in each time period that increases surplus in that period; in equilibrium, first-best optimality is attained. The conditions under which this mechanism can be shown to operate effectively are fairly restrictive: (1) demand for access to the service is assumed to be fixed independent of price (at least within the range of prices considered) and (2) either all customers have the same demand, or the firm knows the demand of each individual customer and is able to offer a separate tariff to each customer. The analysis is valuable, however, even for settings in which these restrictions do

1. The offering of a new tariff on an self-selecting basis is not necessarily Pareto dominating, because it could decrease the profits of the firm if the tariff is not appropriately designed. This issue is discussed in section 9.4.

not apply and constitutes a seminal contribution on optimal regulation with self-selecting tariffs.

The chapter is organized as follows. Section 9.2 describes customers' choice among self-selecting tariffs under the traditional assumptions about customer behavior. Section 9.3 shows, under the standard assumptions, that a set of self-selecting tariffs is equivalent to one multipart tariff. This equivalence allows us in section 9.4 to show that self-selecting tariffs can be designed that increase surplus and Pareto dominate, using concepts developed in chapter 7 for multipart tariffs. In section 9.5, an application of self-selecting tariffs in the real world is examined. The evidence indicates that Pareto dominance was achieved, which shows that the theoretical results regarding self-selecting tariffs can indeed occur in the real world. Section 9.6 describes Sibley's mechanism and shows that it induces the regulated firm to design and offer self-selecting tariffs that increase surplus, reaching first-best optimality in equilibrium. The chapter concludes with a cautionary section on how the surplus implications of self-selecting tariffs may differ if, as empirical work has consistently found, customers do not behave in accordance with the standard assumptions.

The findings of the chapter can be summarized as follows:

- Under standard assumptions, each customer chooses the self-selecting tariff that provides it with the greatest surplus for its known demand.
- For any set of self-selecting tariffs, an equivalent multipart tariff can be designed. This equivalent tariff, offered without selection, results in the same consumption level and bill for each customer as would occur with the self-selecting tariffs.
- Suppose the firm offers a new tariff without changing its original tariff offerings. No customer is hurt, because each customer can choose to stay on its existing tariff. Any customer that switches to the new tariff necessarily benefits; otherwise, it would not choose the new tariff. Depending on the design of the new tariff and the distribution of customers' demands, the firm's profit could either rise or fall.
- Given any set of self-selecting tariffs with usage prices in excess of marginal cost, a new tariff can always be designed that, when offered in addition to the original tariffs, increases surplus and even Pareto dominates the offering of the original tariffs alone. Because no customer is hurt by a new tariff offering, this fact implies that, when prices exceed marginal cost under existing tariffs, a new tariff offering

can always be designed that increases (or does not decrease) the profit of the firm.

- Given a non-time-differentiated (non-TOU) price that is below marginal cost in the peak and above marginal cost in the off-peak, a time-of-use tariff can be designed that, when offered in addition to the non-TOU price, increases surplus and Pareto dominates the offering of the non-TOU price alone.

- Sibley proposes a mechanism under which the firm offers two self-selecting access/usage tariffs in each period. The firm is free to design one of the tariffs. The regulator designs the other. The regulator designs its tariff in such a way that the firm can make strictly positive profit only if the firm designs and offers a tariff that increases total surplus. In the first period after this mechanism is imposed, the firm offers a tariff with usage price equal to marginal cost. With access demand fixed (as Sibley assumes), first-best optimality is therefore attained in the first period. By the second period, profits are zero. This mechanism requires that either (1) all customers have the same demand, or (2) the firm knows each customer's separate demand and can offer separate tariffs to each customer.

- Empirical work indicates that customers do not choose tariffs in accordance with the traditional assumptions. The implications of this finding depend on how customers' behavior is interpreted. If customers are thought to make mistakes in their choice of tariff, then multipart tariffs are probably better than self-selecting tariffs. Under the traditional assumptions, a multi-part tariff can be designed that is equivalent to self-selecting tariffs, such that any function served by self-selecting tariffs can also be served by multipart tariffs. However, when the traditional assumptions do not hold, multipart tariffs protect customers from their own mistakes. It might be the case, on the other hand, that customers do not make mistakes but rather choose among tariffs on the basis of factors other than surplus under known demand, considering, for example, issues of uncertainty, risk, the effort of optimization, information costs, and so on. The implications of self-selecting tariffs for surplus have not been derived when customers' decision making is rational but more complex than traditionally assumed. It is possible that in these cases self-selecting tariffs could offer benefits that cannot be obtained with the multipart tariff.

9.2 Customer Choice among Tariffs under Traditional Assumptions

Under the traditional analysis, the customer is assumed to know, at the time of choosing among tariffs, its demand curve for the good. The customer calculates the consumer surplus that it would obtain under each tariff and chooses the tariff that provides the greatest surplus.

Consider, for example, a customer whose local phone company offers two tariffs. Under tariff I, the customer is charged \$8 per month plus 10 cents for each local call. Under tariff II, the customer is charged \$10 month and 6 cents for each call. The customer knows that its demand for local calls is that depicted in figure 9.1. Using this demand curve, the customer calculates its surplus under each tariff. Under tariff I, the consumer's surplus is the area *A* minus \$8. (Area *A* is the surplus the customer obtains from local calls given a price of 10 cents per call and ignoring the fixed charge; subtracting the \$8 fixed charge from this amount gives the actual surplus incorporating the fixed charge.) Surplus from tariff II is areas *A* and *B*, minus the \$10 fixed fee.

The customer in this example chooses tariff I if area *A* minus \$8 exceeds area *A* and *B* minus \$10; otherwise the customer chooses tariff II. Given the demand curve in figure 9.1, area *A* is \$12 and area *B*

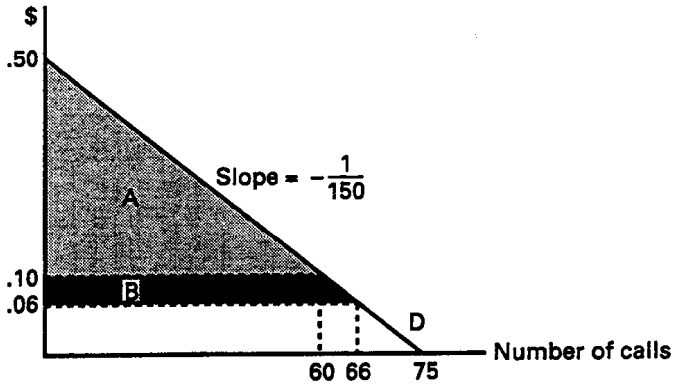

Figure 9.1
Customer choice between two access/usage tariffs

is $2.52. Surplus under tariff I is therefore $4, while surplus under tariff II is $4.52. The customer chooses tariff II.

The comparison of the two tariffs can be seen in an another way. Tariff II costs more than tariff I on a monthly basis, but levies a lower price per call. The customer's question is: are the benefits from having a lower per-call price sufficiently large to warrant the higher fixed charge? More specifically, the customer chooses tariff II over tariff I if the extra surplus the customer obtains at the lower per-call price exceeds the extra fixed charge. The extra surplus is area *B*, which is $2.52; the extra fixed charge is $2. The customer chooses tariff II.

Two other graphs are sometimes useful in examining a customer's choice among tariffs: the outlay schedule and the mapping of indifference curves against budget constraints. An outlay schedule gives the bill the customer would receive as a function of the quantity consumed, under the tariff situation the customer is facing. For the tariffs in our example, the outlay schedules are shown in figure 9.2. Under tariff I, the customer would face line I-I as its outlay schedule; and, under tariff II, line II-II. The customer has a choice between these two tariffs. For any given number of calls, the customer necessarily chooses the cheaper tariff. The outlay schedule that the customer faces in the presence of these self-selecting tariffs is, therefore, the bold, kinked line, that is, the lower of the two individual tariff schedules at each number of calls. The location of the kink, at fifty calls, can be easily calculated. If the customer makes fifty calls, its bill would be the same

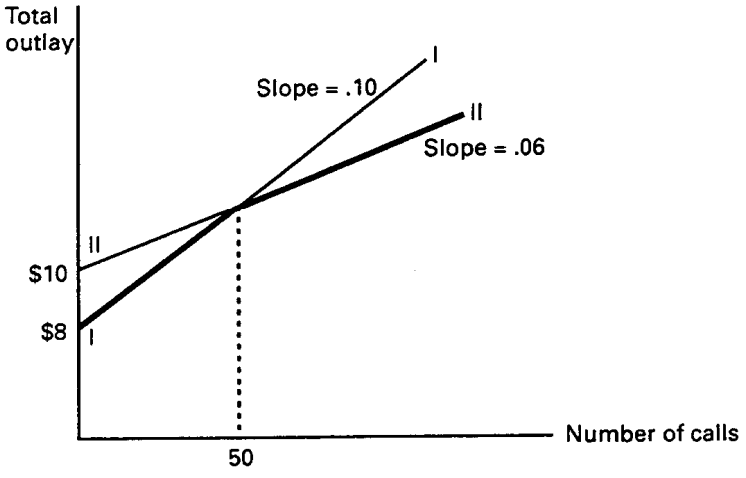

Figure 9.2
Outlay schedule

under both schedules (under tariff I: \$8 + (50) · .10 = 13; under tariff II: \$10 + (50) · .06 = 13). For fewer than fifty calls, tariff I is cheaper; whereas tariff II is cheaper for more than fifty calls.

Note that the outlay schedule provides only partial information on the customer's choice. In particular, it provides information on which tariff is cheaper and hence chosen, *given* the number of calls made. However, the number of calls the customer makes is *not* given, but is determined simultaneously with the choice of tariff. As indicated in figure 9.1, the customer would make sixty calls under tariff I and sixty-six calls under tariff II; it chooses tariff II and makes sixty-six calls. The outlay schedule can be used to determine the cost to the customer of making sixty-six calls and to verify that the customer must be choosing tariff II if it is making sixty-six calls (because tariff II is cheaper for that number of calls). However, it cannot be used (directly at least) to determine that sixty-six calls will be made.

The mapping of indifference curves and budget constraints allows the customer's choice of tariff and its choice of quantity to be presented on the same graph. For our example, consider two goods: local calls and all other goods, with the price of all other goods set at \$1 and the customer's income being \$70. The budget constraint associated with tariff I is the line I-I in figure 9.3: it starts at sixty-two because \$8 of the customer's income must be foregone in paying the fixed charge under this tariff; the slope is -0.10 because for each additional call, .10 units of other goods must be foregone. Similarly, the budget constraint for tariff II is the line II-II with y-intercept of sixty and a slope of -.06. The customer has a choice between the two tariffs. The budget constraint the customer actually faces is therefore the bold, kinked line, which is the "farther out" portions of the budget constraints for each of the two individual tariffs.[2] The customer maximizes utility subject to its budget constraint; that is, the customer moves out as far as possible on its indifference mapping while staying on the bold, kinked budget constraint. The customer chooses point

2. Consider, for example, point M. Under tariff II, the customer cannot increase its consumption of other goods without decreasing the number of calls it makes: by definition, M is on the budget constraint for tariff II. However, because the customer has a choice of tariff, the customer can switch to tariff I and consume more other goods without making fewer calls (moving from M to N). Point M is therefore *not* on the customer's budget constraint when the customer faces a choice between tariffs. At point N, the customer cannot increase consumption of other goods without making fewer calls, even if the customer were to switch tariffs. Point N *is* therefore on the customer's budget constraint for the self-selecting tariff situation.

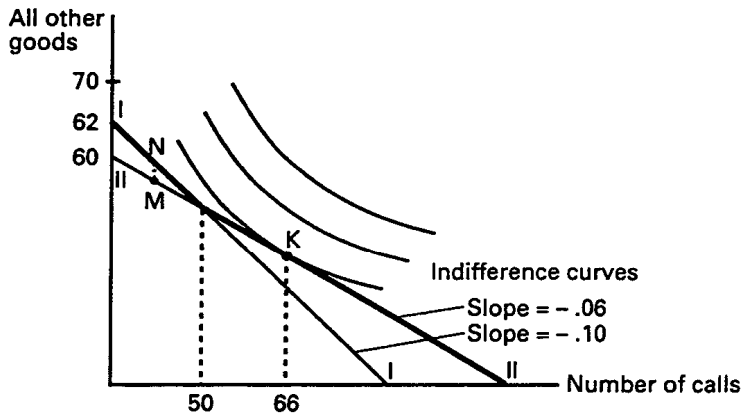

Figure 9.3
Budget constraint and indifference mapping

K, which represents sixty-six calls. Because *K* is on the line II-II, the customer in choosing point *K* necessarily chooses tariff II.

For situations with more than two tariffs, or more complex tariffs, the concepts developed in our example are applied analogously. In each case, the customer calculates its surplus under each tariff and chooses the tariff that provides the greatest surplus. The outlay schedule it faces is comprised of the lowest portions of the outlay schedules for each of the tariffs offered. The budget constraint the customer faces consists of the "farthest out" (that is, least constraining) portions of the budget constraints for the individual tariffs; the customer chooses the point that attains the highest indifference curve, thereby choosing both a tariff and a consumption level.

9.3 Equivalence to Multipart Tariffs

Given that the customer knows its demand curve for the good, a set of self-selecting tariffs is equivalent to one tariff that embodies elements of each of the self-selecting tariffs. "Equivalent" in this context means that the behavior of the customer is the same: that the customer would consume the same quantity of the good, pay the same total bill, and face the same marginal price.

Consider the example of self-selecting tariffs in the previous section: tariff I consisting of an $8 fixed charge and a 10 cents usage price, and tariff II with a $10 fixed charge and 6 cents per call. The outlay schedule and the budget constraint that the customer faces under these

self-selecting tariffs is given by the bold, kinked lines in figures 9.2 and 9.3. A three-part tariff can be constructed that has the same outlay schedule and budget constraint and hence results in the same customer behavior. Specifically, consider a tariff, called tariff M, that levies a fixed fee of \$8 plus a per-call charge of 10 cents for up to fifty calls and then a charge of 6 cents for each call over fifty. The outlay schedule for tariff M is the bold, kinked line in figure 9.2: the customer's bill is \$8 for no calls, increases by 10 cents for each call up to fifty, and then increases by 6 cents for each call over fifty. For any given number of calls, the customer would pay the same total bill if it faced tariff M without choice of tariff as it would if it faced tariffs I and II with selection.

The budget constraint is also the same, namely, the bold, kinked line in figure 9.3. With no calls, the customer under tariff M pays a fixed charge of \$8, leaving \$70 − \$8 = \$62 for other goods. The y-intercept of the budget constraint is therefore sixty-two. For each call up to fifty calls, the customer must pay 10 cents and hence forego consumption of .10 units of other goods. For each call over fifty, the customer must forego only .06 units of other goods. The budget constraint therefore has a slope of −.10 from zero to fifty calls and a slope of −.06 beyond fifty calls. Because the budget constraint is the same whether the customer faces tariff M without selection or tariffs I and II with selection, the behavior of the customer is the same: the customer makes sixty-six calls and consumes 56.04 units of other goods (\$70 minus a bill of \$13.96 for the calls, divided by the \$1 price for other goods).

The customer also faces the same marginal price. Under tariff M, the customer, at its chosen consumption level of sixty-six calls, faces a price of 6 cents for an extra call. Under tariffs I and II with self-selection, the customer, at its chosen consumption level of sixty-six calls, chooses tariff II under which a price of 6 cents is charged for an extra call.

These concepts can obviously be generalized. It is possible, for any set of self-selecting tariffs, to design one tariff that, if implemented without customers having any choice among tariffs, would result in the same behavior, total bill, and marginal price for each customer. The design of this equivalent tariff is most easily accomplished with reference to the outlay schedule for the self-selecting tariffs. Figure 9.4, for example, illustrates a situation that is increasingly common for local phone service. The phone company offers three options to

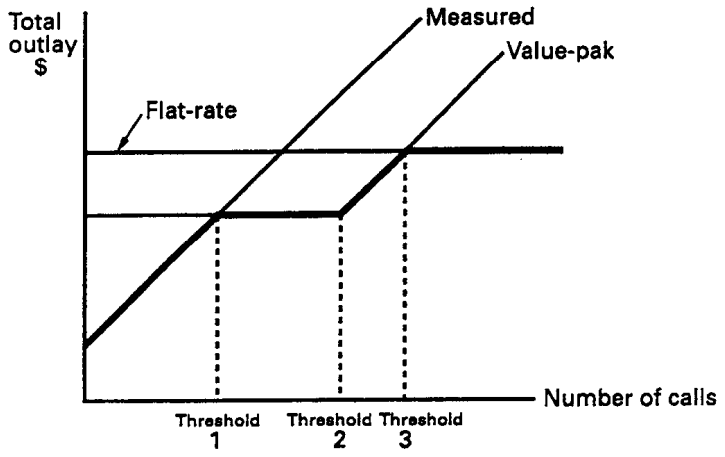

Figure 9.4
Outlay schedule for flat-rate, measured, and value-pak options, plus an equivalent tariff without selection

its customers: *flat-rate service,* which, for a fixed monthly fee, allows unlimited local calling without extra charge; *measured service,* which levies a lower monthly fee but charges for each call; and a hybrid that is often called *"value pak,"* which, for a midlevel fixed fee, allows a certain number of local calls to be made without additional charge (this is referred to as the "calling allowance") and then charges for calls beyond this allowance at the same rate as under measured service. The non-self-selecting tariff that is equivalent to this set of options consists of a fixed charge and four per-call prices, each charged in a different range of calls. The fixed charge is the same as that for measured service. For low levels of calling, the per-call charge is the same as under measured service. After the first threshold (which occurs at the number of calls at which measured service and value pak would cost the same), the per-call charge becomes zero. After the second threshold (the calling allowance under value pak), the per-call charge becomes that charged under measured service (because calls under value pak are charged the same as under measured service in this range of calls). Finally, after the third threshold (at which value-pak and flat-rate service cost the same), the usage charge becomes zero again. For customers who know their demand curves, being charged under this one tariff without any choice would be the same as being offered a choice among flat-rate, measured, and value-pak services. And because the number of calls made by each customer

and the bill paid by each customer would be the same, the costs, revenues, and profits of the phone company would also be the same.

9.4 Welfare Implications of Self-Selecting Tariffs

In chapter 7 it was shown that, in many circumstances, surplus can be improved and Pareto dominance even achieved through the judicious use of multipart tariffs. We have just shown that self-selecting tariffs are equivalent to appropriately designed multipart tariffs without selection. It follows therefore that, in many situations, surplus can be improved and Pareto dominance achieved by the judicious use of self-selecting tariffs. This fact is amplified below, using methods that are analogous to those in chapter 7.

Self-selecting tariffs possess a unique feature that is important in itself and also facilitates the analysis of surplus. Specifically, the introduction of a new self-selecting tariff in addition to existing tariffs can only benefit customers. Consider a firm that is initially offering a set of tariffs. If the firm, at some point in time, adds a new tariff without changing the original ones, then no customer is made worse off by this addition and some customers might be made better off. No customer is hurt because each customer can choose to remain on its original tariff, with no change in its behavior or surplus. Any customer that would obtain more surplus from the new tariff than it does from its originally chosen tariff can, and will, switch to the new tariff, becoming better off. Consequently, some customers—namely those who switch to the new tariff—benefit, and no customer is hurt.

Although no customer is hurt, the firm might be. Depending on the design of the tariff, the firm's profits could either increase or decrease from the offering of a new tariff in addition to existing tariffs. Several examples will suffice to illustrate the fact. Consider an energy utility that originally provides electricity under standard, non-time-differentiated rates. If the utility offers time-of-use (TOU) rates as an option to its customers in addition to the standard rate, then customers only stand to gain: those who would benefit from the TOU rates switch to them while those who would not benefit remain on the standard rates. However, depending on the extent to which customers who switch to TOU rates change their consumption patterns, the firm could either gain or lose money. At one extreme, suppose that the customers are totally non-price-responsive in their TOU con-

sumption of electricity. Those customers with sufficiently low peak-period consumption would receive a lower bill for the same consumption under the TOU rates compared to the standard rate. These customers would switch to the TOU rates; but because their consumption is not price responsive, they would consume the same amount of electricity in each period as under the standard rates. Because consumption levels are the same, the costs of the firm would be the same. However, the customers who switched to TOU rates would be paying lower bills, such that the revenues of the firm would be lower. With lower revenues and the same costs, the firm's profits decrease.

At the other extreme, customers might be highly price responsive. Suppose the customers who switch to TOU rates shift all of their peak-period consumption to the off-peak. The costs of the firm decrease substantially, because off-peak production is cheaper than peak production. Although the customer's bills, and the firm's revenues, are also lower, the reduced costs can easily dominate the reduced revenues, such that the firm's profits increase.

Another example arises in the context of local phone service. If a phone company that has traditionally provided only flat-rate service starts to offer measured service as an option, its profits could either increase or decrease. If customers are not price responsive and choose the service that offers the lower bill for their fixed number of calls, the phone company will lose money. However, if customers that switch to measured service reduce their calling in response to the higher price, the costs of the firm could decrease more than its revenues.

To summarize: if a new tariff is offered as an option in addition to existing tariffs: (1) no customer is hurt, and any customer that switches to the new tariff is benefited, and (2) the firm might obtain either more or less profit.

Although it is true that customers are not hurt when a new tariff is introduced as an option in addition to existing tariffs, the fact that the firm might be hurt can eventually have detrimental consequences for customers. In particular, if the firm loses money, it might be necessary for the firm to change the original tariffs, raising rates under these schedules. For example, if an energy utility introduces TOU rates and loses money from doing so, it might raise the standard rate in an effort to recoup the loss. In this case, the new tariff is not being offered in addition to the existing tariffs; rather, the existing tariffs are being changed. With higher rates under the original tariffs, customers who remain on their original tariffs would be hurt. This scenario—of

the firm losing money and raising rates under the original tariffs—is especially likely if the form of regulation assures that the firm breaks even. Consequently, the issue of whether the firm losses or gains from the offering of a new tariff option is of concern to customers as well as shareholders.

If a tariff can be designed that, when offered in addition to existing tariffs, increases the firm's profits, then Pareto dominance can clearly be achieved. No customer is hurt, and the firm is better off. In fact, the extra profit can be returned to customers in the form of lower rates under all tariffs. Two results establish some conditions under which it is possible to design such a tariff.

Result 1: Given any set of N self-selecting tariffs with usage prices that exceed marginal cost and customers that are at least somewhat price responsive, a set of $N+1$ self-selecting tariffs can be designed the offering of which Pareto dominates the offering of the original N tariffs.

If the new set of $N+1$ tariffs consists of the original N plus an additional tariff, we know that no customer will be hurt by the addition and some customers might benefit. The task in demonstrating the result is therefore to show that a new tariff can be designed that, when added to the original set of tariffs, increases the profits of the firm (or, alternatively, benefits at least one customer without decreasing the profits of the firm).

Consider the customer with the highest level of consumption under the original N tariffs. This customer has the largest bill (given, as will always be the case in practice, that the outlay schedule under these tariffs increases with consumption). Call this customer "the largest customer" and denote its consumption as Q_L, its bill as B_L, and the marginal price the customer faces as P_L, which, by assumption, exceeds marginal cost. The demand of this customer is depicted in figure 9.5.[3]

We can design a new tariff offering that this customer will choose and that will generate additional profit for the firm. Let the new tariff consist of: (1) an access fee of B_L (that is, an access fee equal to the largest customer's original bill), (2) a price of zero for the first Q_L units

3. The customer faces a price of P_L for units of consumption near Q_L; the line at P_L is dotted for lower units of consumption because a price other than P_L might be charged for other levels of consumption (that is, the inframarginal price might not equal the marginal price). All that is required for our analysis is a designation of the marginal price.

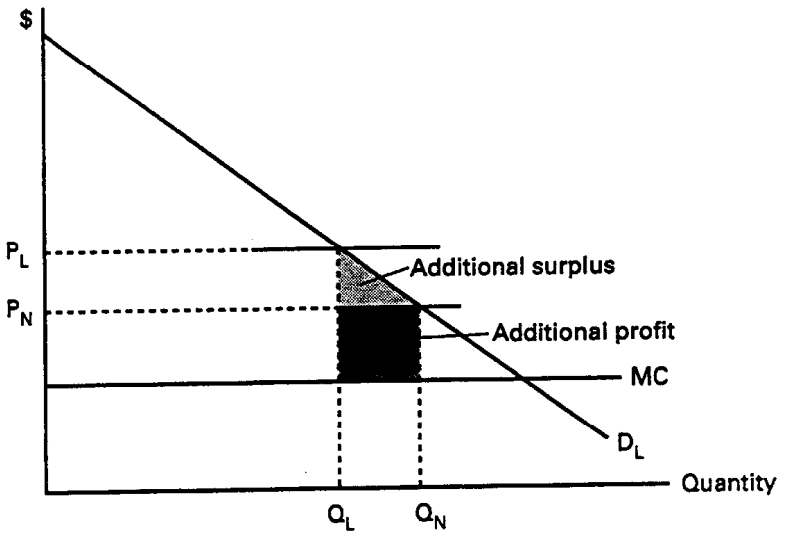

Figure 9.5
Pareto-dominating tariff

of consumption, and (3) a price of P_N for each unit of consumption over Q_L, where P_N is below P_L and above marginal cost. For its original level of consumption, Q_L, the customer is just as well off under the new tariff as under its original tariff: in either case it pays B_L. However, under the new tariff, the customer faces a lower marginal price for consumption beyond Q_L, which allows it to obtain greater surplus. In particular, the customer, if it chooses the new tariff, will increase consumption from Q_L to Q_N and obtain additional surplus equal to the shaded area in figure 9.5. Because the customer obtains greater surplus under the new tariff than under its original tariff, the customer will choose the new tariff.

Consider now the firm. Because the new marginal price is above marginal cost, the firm earns additional profit when the customer increases its consumption. For the original Q_L units of consumption, the firm obtains the same revenue from the customer as under the customer's originally chosen tariff and incurs the same costs; however, for the additional units of consumption (from Q_L to Q_N), the firm collects extra revenue in excess of its extra costs, such that its profits increase.

Consider, finally, all other customers. The access fee under the new tariff is higher than the bill that each other customer paid under the original tariffs (recall that the access fee is equal to the total bill of the

customer with the highest bill under the original tariffs). Consequently, the bill that each of these customers would receive for its original level of consumption would be higher under the new tariff than under the original tariffs. It might be the case that none of these customers would choose the new tariff because of this. In this case, the new tariff benefits the largest customer and the firm, and does not affect any other customers: Pareto dominance is achieved.

It is possible, however, that some customers (in addition to the largest) will choose the new tariff. Recall that the new tariff offers a lower marginal cost. The extra surplus that a customer can obtain due to the lower marginal price may exceed the extra bill that it is charged for its original level of consumption. If this is the case, the customer will choose the new tariff, obtaining a net increase in surplus. The firm also benefits from any customer that chooses the new tariff. Each customer pays a higher bill for its original level of consumption such that the firm obtains more revenues, and any increase in consumption that the lower marginal price induces generates even more profit for the firm, because the revenue of these extra units exceeds the cost of producing them. Again, Pareto dominance is achieved.

The result can made even stronger. A portion of the extra profit that the firm obtains can be refunded to all customers in the form of lower prices under all tariffs. If all of the extra profits are refunded (as would be required if the firm were allowed to make no more than zero profit), all customers would benefit and the firm would not be hurt by the offering of the new tariff. If the firm were allowed to retain some of the additional profits, customers and shareholders would all benefit.

Note that, in the above argument, the marginal price under the new tariff is set above marginal cost. This feature of the new tariff is what allows the firm to increase its profits by offering the new tariff, which in turn allows the opportunity of reduced rates under all tariffs. If the marginal price under the new tariff is set *equal to* marginal cost, rather than above it, then the largest customer, and any other customer that chooses the new tariff, would benefit from the new tariff. The firm would not earn additional profit from additional units of consumption. However, if any customer aside from the largest chose the tariff, the firm would obtain extra revenue from these customers' original consumption. Pareto dominance is still achieved.

A corollary to result 1 is that, under the same conditions as stated in the result, a new tariff can always be designed that increases sur-

plus when offered in addition to existing tariffs. Result 1 shows that Pareto dominance can be achieved. The offering of the new tariff that attains Pareto dominance necessarily increases surplus, because the surplus of some parties increases and no one's surplus decreases. Other tariffs might obtain an even greater increase in surplus.

The tariff that we designed to achieve Pareto dominance might provide only a small increase in surplus. It is possible that only the largest customer will choose the new tariff, such that the surplus gain is only the increased surplus to this one customer and the extra profits generated from this one customer. When averaged over a large number of customers, the benefits might be so small as to be negligible. The point of result 1 is, however, simply to show that a surplus improvement and Pareto dominance is possible. In particular situations, it may be possible to identify tariffs that increase surplus considerably. These may, and probably will, be quite different from the one designed for the purposes of the proof. The importance of the result is to indicate that a search for new tariff offerings that increase surplus is worth pursuing because we know that at least one such tariff can necessarily be found.

For result 1, we assumed that usage prices exceed marginal cost. When demand and/or costs vary over times of day or periods of the week, marginal cost may at certain times (e.g., the "rush hour" or "peak") exceed the price charged under existing tariffs. The next result indicates that Pareto-dominating tariff offerings can be designed for these situations as well (Train 1990).

Result 2: Given a non-time-differentiated rate that is below marginal cost in the peak and above marginal cost in the off-peak, and customers that are at least somewhat price responsive, a time-of-use tariff can be designed that, when offered as an option to customers in addition to the original non-time-differentiated rate, Pareto dominates the use of the non-time-differentiated rate alone.

Suppose there are two time periods, called "on-peak" and "off-peak," with the same usage rate, r, charged in both periods. Marginal cost in the on-peak, MC_{on}, exceeds r, while MC_{off} is below r. To construct a time-of-use (TOU) tariff, identify the customer with the largest ratio of off-peak to on-peak demand at the original non-TOU rate. Call this customer the "target customer" and its original consumption levels Q^0_{off} and Q^0_{on}. The TOU tariff is designed by raising price in the on-peak *toward*, but not beyond, the marginal cost in the on-peak, and

by lowering the price in the off-peak toward, but not beyond, the marginal cost in the off-peak. Furthermore, the TOU prices are set such that the target customer's bill for its original consumption is the same under the TOU rates as the original non-TOU rate.

To be concrete, let us lower the price in the off-peak by Δp_{off}; that is, set the off-peak price at $p_{off} = r - \Delta p_{off}$. In the on-peak, raise price by exactly the target customer's ratio of off-peak to on-peak consumption times the amount of the price drop in the off-peak; that is, set the on-peak price at $p_{on} = r + (Q^0_{off}/Q^0_{on})\Delta p_{off}$. In making these changes, be sure that Δp_{off} is small enough that p_{off} *is above* MC_{off} and p_{on} *is below* MC_{on}. At these TOU rates, the target customer would receive the same bill for its original consumption as it does under the non-TOU rate:

$$
\begin{aligned}
\text{Bill under TOU rates} &= p_{off}Q^0_{off} + p_{on}Q^0_{on} \\
&= (r - \Delta p_{off})Q^0_{off} \\
&\quad + (r + (Q^0_{off}/Q^0_{on})\Delta p_{off})Q^0_{on} \\
&= rQ^0_{off} + rQ^0_{on} \\
&\quad - \Delta p_{off}Q^0_{off} \\
&\quad + ((Q^0_{off}/Q^0_{on})\Delta p_{off})Q^0_{on} \\
&= rQ^0_{off} + rQ^0_{on} \\
&\quad - \Delta p_{off}Q^0_{off} + Q^0_{off}\Delta p_{off} \\
&= rQ^0_{off} + rQ^0_{on} \\
&= \text{bill under non-TOU rate.}
\end{aligned}
$$

For all other customers, who have a lower ratio of off-peak to on-peak consumption, the bill for their original levels of consumption would be higher under these TOU rates than under the non-TOU rate.

The target customer will choose this TOU tariff. Figure 9.6 provides the relevant information. The shaded areas in the on-peak and off-peak graphs are equal in size, because the TOU rates are designed such that the increase in the target customer's bill for its original on-peak consumption is exactly offset by the reduction in its bill for its original off-peak consumption (such that its total bill is unchanged). Under the TOU rates, the target customer will adjust consumption in each period and obtain additional surplus. In the on-peak, the customer will decrease consumption from Q^0_{on} to Q^1_{on}, foregoing the units of consumption for which marginal benefit is below the new price. This reduction in consumption at the new higher price provides the customer with extra surplus equal to the area *ABC*. Similarly, in the off-peak, the customer increases consumption in response to the new,

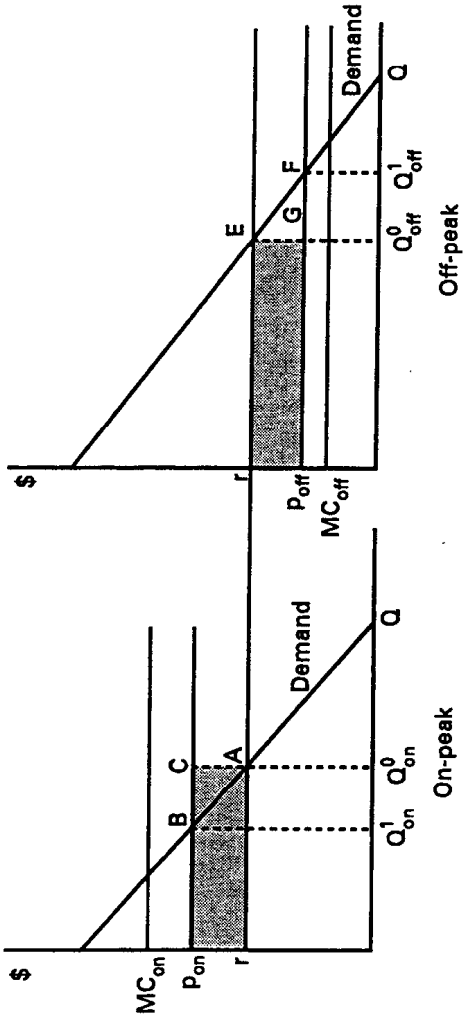

Figure 9.6
Pareto-dominating TOU rate option: target customer

lower price, obtaining additional surplus in the amount of area *EFG*.[4] If there is the possibility of shifting consumption across periods, the customer will do this also, obtaining further increases in surplus. Because the target customer obtains greater surplus under the TOU rates than under the original non-TOU rate, it will choose the TOU rates.

The firm will earn additional profit from the target customer under the TOU rates. Because $MC_{on} > p_{on}$, profits rise when the customer decreases its consumption in the on-peak: the lost revenues are more than made up by the cost savings to the firm of not producing those foregone units. And because $MC_{off} < p_{off}$, the extra consumption in the off-peak generates additional profit for the firm.

Consider now all the other customers. The bill that each of these customers would receive for their original consumption would increase under the TOU rates. Nevertheless, some of these customers may choose the TOU rates if they are able to respond sufficiently to the TOU rates such that their surplus increases. The firm will earn additional profits from these customers, because their bills for their original consumption are higher and their response to the TOU rates consists of decreasing on-peak consumption and increasing off-peak consumption.

The peripheral statements that were made about result 1 also apply to result 2. If profits are refunded to customers in the form of lower rates under both the non-TOU and TOU tariffs, all customers can benefit. The TOU tariff designed for the purpose of the result may increase total surplus by only a small amount. Other tariffs may produce a far greater surplus gain. The result simply indicates that, under the specified conditions, the search for a surplus-increasing, and even a Pareto-dominating, optional TOU tariff will necessarily be successful.

---

4. It might seem peculiar that the customer obtains additional surplus by decreasing consumption in the on-peak and increasing consumption in the off-peak: it seems that the customer gets more surplus no matter what it does. Actually, the customer obtains extra surplus only when it adjusts its consumption in the opposite direction of the change in price. In the on-peak, the price rises under the TOU rates compared to the non-TOU rate, such that the customer obtains extra surplus by decreasing consumption. In the off-peak, price drops, and the customer obtains more surplus by increasing consumption. In a sense, both adjustments are in the same direction, namely, opposite that of the price change.

9.5 An Application of Self-Selecting Tariffs

This section presents the results of an empirical investigation of a real-world application of self-selecting tariffs. Train and Toyama (1989) examined the optional time-of-use (TOU) rates for electricity that were offered to agricultural customers of the energy utility in northern California. They found that the offering of TOU rates in addition to the standard, non-TOU rate under which the customers had traditionally been charged Pareto dominated the offering of the non-TOU rate alone. Specifically, many customers chose the TOU rates and the firm's profits were estimated to have risen due to the TOU tariff offering. This result indicates that surplus-increasing and even Pareto-dominating self-selecting tariffs are possible in practice as well as in theory.[5]

Farms and other agricultural firms (such as nurseries) use electricity to pump water for irrigation. Under mandate from the California Public Utilities Commission, the northern California energy utility introduced optional TOU rates in the agricultural sector, first in 1981 for selected firms and starting in 1983 for all eligible firms. The tariff was available for any agricultural pumping installation that draws a maximum load of a least thirty-five kilowatts and is used at least 300 hours a year (i.e., all but the smallest installations). Farms with several pumping installations had the option of TOU rates for each installation. Commercial and industrial firms were also eligible for the rates for their agricultural pumping installations (e.g., for watering lawns) though not for electricity used in their commercial and industrial activities.

Under the standard, non-TOU tariff, installations were billed at a rate of $0.07877 per kilowatt-hour (kWh) as of January 1, 1985. Under the TOU rates, the kWh charge was differentiated by three times of day (called on-peak, partial-peak, and off-peak) in each of two times of year (called summer and winter). The definitions of these periods and the rates charged in each period are given in table 9.1.

Of the 33,382 eligible pumping installations, 7,675 (23%) chose the TOU rate. Using the traditional concepts, Train and Toyama assumed that these customers benefited from the TOU rates, because otherwise they would have remained on the non-TOU rate. The customers

5. As further support, Heyman, Lazorchak, Sibley, and Taylor (1987) have designed self-selecting tariffs for NYNEX's telecommunications services and shown that these tariffs Pareto dominate the existing offering.

Table 9.1
TOU rate schedule

	Times	
	Summer (May 1–Sept. 30)	Winter (Oct. 1–April 30)
On-peak	12:30 p.m.–6:30 p.m. Mon.–Fri.	4:30 p.m.–8:30 p.m. Mon.–Fri.
Partial-peak	8:30 a.m.–12:30 p.m. and 6:30 p.m.–10:30 p.m. Mon.–Fri.	8:30 a.m..–4:30 p.m. and 8:30 p.m.–10:30 p.m. Mon.–Fri.
Off-peak	10:30 p.m.–8:30 a.m. Mon.–Fri. All day Sat. and Sun.	10:30 p.m.–8:30 a.m. Mon.–Fri. All day Sat. and Sun.
	kWh Rates as of January 1, 1985	
	Summer	Winter
On-peak	0.13114	0.10689
Partial-peak	0.07945	0.08222
Off-peak	0.06780	0.06780

who remained on the non-TOU rate were not affected by the offering of TOU rates, because the non-TOU rate was not changed. The only issue therefore was whether the firm's profit rose or fell.

Using data on the energy consumption of eligible installations, Train and Toyama estimated a model of the demand for electricity by time of use. This model predicts the total electricity consumption of the customer and the share of consumption in the on-peak, partial-peak, and off-peak, respectively, taking as input the price for electricity in each period.

The demand model was used, along with cost information, to approximate the impact of the TOU tariff offering on the profits of the utility. In particular, the profits of the firm were determined twice. First, the profits the utility earned under the self-selecting tariffs were calculated. The demand model was used to estimate each customer's consumption under the tariff they actually chose. The revenues from this estimated consumption were calculated. Using the cost information, the cost of meeting this demand was estimated. Subtracting the costs from the revenues provided an estimate of the utility's profit under the TOU rate offering.

Second, the profit the utility *would have* earned had it not offered the TOU rates was calculated. The demand model was used to estimate each customer's consumption in each period under the stan-

dard, non-TOU rates. Revenues, costs, and profit from this consumption were then calculated.

Train and Toyama found that the customers who chose the TOU rates reduced their consumption in the on-peak, increased in the off-peak, and remained about the same in the partial-peak. Furthermore, the magnitudes of these shifts were such that the utility's profits *increased* as a result of the TOU tariff offering. The extra profit was small (about $360,000, a tiny percentage of the utility's total profit). The firm cannot therefore be considered to have obtained a windfall. However, because profits did not fall, there was no need for the non-TOU rates to be raised to recoup any losses.

In summary, the TOU tariff offering benefited 23% of the eligible agricultural customers (as evidenced by the fact that they chose it), did not hurt any other customers, and provided a small amount of extra profit for the firm. In this case at least, surplus was raised and Pareto dominance achieved by the judicious use of self-selecting tariffs.

9.6 Sibley's Mechanism

Sibley (1989) has proposed a mechanism under which the regulator uses self-selecting tariffs, given certain conditions, to induce the firm to move to first-best optimality. In fact, first-best optimality is achieved in the first period after the regulatory mechanism is imposed. To implement the mechanism, the regulator does not need to know the demand curve or the cost function of the firm. The regulator need only observe for each period the price the firm charges and the profit the firm earns.

The conditions under which the mechanism can be shown to operate effectively are fairly restrictive. The number of customers is assumed to be fixed; that is, access demand is assumed not to be price responsive, at least within the relevant range of prices. Regarding usage demand, either of two conditions is required: (1) each customer has the same demand, or (2) the firm knows the demand curves of each individual customer and can design a different tariff for each customer. While these conditions are unlikely to occur in the real world, the mechanism provides insight that is valuable in more general situations. In particular, the mechanism shows that by requiring the firm to offer one or more self-selecting tariffs that in some sense maintain the status quo, the firm is only able to increase its own profits

over time by introducing new tariff options that increase consumer surplus. Compatibility between the firm's profit drive and surplus maximization is thereby established.[6]

9.6.1 Identical Customers

Consider first the situation in which each customer has the same demand. The mechanism operates as follows. In each period, the firm is required to offer two self-selecting tariffs to its customers. Each of the tariffs is an access/usage tariff, consisting of a fixed charge and a usage price. The rates under the two tariffs are specified for each period. The regulator designs one of the tariffs for each period on the basis of the price and operating profits of the firm in the previous period. The firm is free to design the other tariff to consist of whatever access and usage charge it chooses. For convenience, we call the tariff the regulator designs "the regulator's tariff" and that designed by the firm "the firm's tariff," remembering of course that both tariffs are offered by the firm.

The regulator's tariff in each period is designed on the basis of the firm's tariff in the previous period. Let P_t be the usage price under the firm's tariff in period t. (How this price is established by the firm is discussed below; for the purposes of the regulator's tariff, it is simply taken as given.) At this price, the firm collects revenues Q_tP_t and incurs costs C_t, earning an operating profit per customer of $R_t = (Q_tP_t - C_t)/N$, where N is the number of customers. Note that operating profit is the profit the firm earns from usage independent of the access fee. The operating profit may be either positive or negative; for example, if P_t is set at marginal cost, then, given that the firm is a natural monopolist, R_t is negative.

The regulator uses this information to design the tariff it will require the firm to offer in the next period. In particular, the regulator's tariff in period $t + 1$ has a usage price of P_t (the usage price under the firm's tariff in the previous period) and a fixed fee of $-R_t$ (the

6. In results 1 and 2, the same concept operates as in Sibley's mechanism, but the status quo maintained by the self-selecting tariffs is somewhat different. Under Sibley's mechanism, the regulator requires that a self-selecting tariff always be offered that maintains *total* surplus; while in the demonstration of results 1 and 2, the offering of the original tariffs without change maintains the surplus of *each individual customer*. This difference is appropriate, because Sibley's mechanism is concerned with attaining first-best optimality, which relates to total surplus, whereas results 1 and 2 are concerned with attaining Pareto optimality, which requires maintenance of each individual's surplus.

negative of the firm's operating profit per customer in the previous period). If the firm earned a positive operating profit in the previous period, each customer is provided with a *negative* fixed charge—that is, a refund—equal to the amount of profit per customer. If the firm's operating profit was negative (e.g., if it charged marginal-cost prices), the fixed charge under the regulator's tariff is positive. Essentially, the regulator's tariff allows customers to face the same usage price that the firm offered in the previous period under its own tariff while refunding the excess profit or subsidizing the losses from the previous period.

The firm is required to offer the regulator's tariff. However, the firm is also allowed to offer a tariff of its own design. Note, however, that for the firm to get any customers to choose its own tariff, its tariff must provide greater surplus to customers than the regulator's tariff. That is: if the firm wants to make more profit than it would under the regulator's tariff, it must design a tariff that benefits customers as well as itself. This, as we see below, is the crux of why Sibley's mechanism works.

Let us start the analysis in a period, called period 0, before this form of regulation is imposed. Either the firm is unregulated in this original period, or some other form of regulation is operating.[7] The firm charges a price of P_0 and earns an operating profit of R_0 per customer. This profit may be either negative or positive, depending on whether the firm was losing or earning money before Sibley's mechanism is imposed.

In the first period after the mechanism is imposed, the regulator requires that the firm offer a tariff that consists of a fixed charge of $-R_0$ and a usage price of P_0. That is, the regulator designs a tariff that consists of the same usage price as the firm had been charging and a fixed charge that refunds to customers the profit the firm had been earning (or, if the firm had been losing money, charges customers an amount sufficient for the firm to recoup its losses). The regulator requires that the firm offer this tariff. However, the regulator also allows the firm to offer another tariff with whatever fixed charge and usage price the firm chooses. Customers will have a choice between the tariff that the regulator designs and the one designed by the firm.

The question is: what fixed charge and usage price will the firm

7. To avoid the possibility of strategic behavior, the firm is assumed not to know in period 0 that Sibley's mechanism will be imposed in the upcoming period.

choose for the tariff it is allowed to design? Because all customers have the same demand, all customers will choose either the tariff designed by the regulator or that designed by the firm, whichever provides the greater surplus. If all customers choose the tariff designed by the regulator, the firm will make zero profit. This fact is shown as follows.

Consider first the usage fee only. At P_0, the firm earns profit on usage of R_0 per customer, the same operating profit the firm earned before regulation was imposed. However, the regulator's tariff imposes a fixed charge of $-R_0$. Therefore, considering both the fixed charge and profit from usage, the firm ends up earning zero profit: $R_0 - R_0$.

If the firm wants to make positive profit, it must design a tariff that attracts customers away from the tariff the regulator designed. Suppose the firm sets the usage price under the tariff it designs at P_1. We can show that the maximum profit the firm can earn from offering a tariff with a usage price of P_1 is the change in total surplus (consumer surplus plus profit) that this price generates compared to the price P_0 under the regulator's tariff.

First, we must determine: What is the highest fixed fee that the firm can charge under its tariff and still have customers choose its tariff over the regulator's tariff? With this information, we can then determine the profit the firm would earn with its tariff. Customers would obtain a different level of surplus at price P_1 than at the P_0 they would face under the regulator's tariff. Label this difference in surplus (surplus at P_1 minus surplus at P_0) as ΔCS_1. For example, in figure 9.7, P_1 is below P_0 (the firm designs a tariff with a lower price), and consumer surplus is greater by the shaded area. In the discussion below, we assume for convenience that P_1 is below P_0; however, the arguments and results are the same if P_1 is above P_0 such that ΔCS_1 is negative.

Because the usage price is lower, the firm can charge a higher fixed charge on its own tariff than under the regulator's tariff and still have customers choose its tariff. In particular, the firm can charge a fixed fee that is higher by just about as much as the increase in surplus that the lower usage fee provides. Because the fixed fee under the regulator's tariffs is $-R_0$, the firm can charge a fixed fee of *nearly* $-R_0 + \Delta CS_1$ and still induce all customers to choose its own tariff. (If the firm sets the fee at exactly this amount, customers will be indifferent between the two tariffs, because both would provide the same total

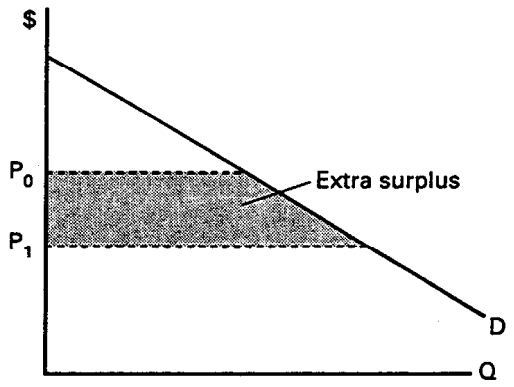

Figure 9.7
Change in consumer surplus for a price decrease

surplus. The firm can raise the fee to just slightly less than this amount and have all customers choose it.) The firm will therefore offer a tariff with a usage fee of P_1 and a fixed fee of (nearly) $-R_0 + \Delta CS_1$. All customers will choose this tariff over the one designed by the regulator. Under this tariff, the operating profits of the firm in the first period of regulation, labeled R_1, are $(P_1Q_1 - C_1)/N$, where Q_1 is the quantity demanded at P_1, and C_1 is the cost of producing Q_1. The total profits of the firm, including the customers' fixed fee, are these operating profits plus the fixed fee of (nearly) $-R_0 + \Delta CS_1$:

$$\pi_1 \simeq R_1 - R_0 + \Delta CS_1.$$

This profit is simply the change in total surplus:

$$\pi_1 \simeq \Delta R_1 + \Delta CS_1 = \Delta TS_1,$$

where $\Delta R_1 = R_1 - R_0$ is the change in operating profits from before regulation to the first period of regulation, and ΔTS_1 is the change in total surplus (profits plus consumer surplus) over the same periods. That is, the firm earns profits in the first period that are (nearly) equal to the change in total surplus that its own tariff generates.

So far, we have said that the firm charges a usage price of P_1, but we have not determined the level of this price. We have shown that the firm will earn profit equal to the change in surplus generated by the tariff the firm designs. The firm therefore obtains the greatest profit by designing a tariff that increases total surplus as much as possible. This is accomplished by inducing all customers to face a price equal to marginal cost, since total surplus is highest when price equals mar-

ginal cost. The firm will therefore, in the first period under this mechanism, offer a tariff whose usage price equals marginal cost. The firm will set the fixed fee under this tariff at a level that induces all customers to choose it over the tariff designed by the regulator, generating profits for itself by doing so. Because all customers will choose the firm's tariff, all customers will face a usage price equal to marginal cost.

As well as attaining the first-best output level, the firm uses the cost-minimizing inputs to produce this output. The firm is allowed to keep whatever profit it earns in the first period under the self-selecting tariffs. By minimizing costs, the firm's profits are larger.

In this first period after the mechanism is imposed, the firm makes positive profit. In fact, as stated, its profit is equal to the increase in total surplus that its move to first-best optimality generates, which can be quite large. While this excess profit is not inconsistent with first-best optimality,[8] equity considerations might suggest that profit should be zero. It turns out that Sibley's mechanism reduces profits to zero in the second period, while maintaining marginal-cost pricing and cost minimization. In the second period, the regulator requires that the firm offer a tariff that has a fixed charge of $-R_1$ (the operating profit/loss of the firm, per customer, in the first period) and a usage price of P_1 (the usage price in the first period). The firm is allowed to offer another tariff with whatever fixed and usage charges it chooses. Using the logic above, the firm earns zero profit if customers choose the tariff the regulator designed. If the firm designs a tariff that customers choose, the most profit the firm can earn is the change in total surplus generated by that tariff. However, because the firm is already at first-best optimality, total surplus cannot increase. Therefore the firm can do no better than to allow all customers to consume under the tariff designed by the regulator. Under this tariff, customers face a usage price equal to marginal cost (because P_1 was set in the first period to equal marginal cost), and the firm obtains through the fixed fee exactly enough revenue to break even (because R_1 is the operating profit/loss per customer of the firm when charging P_1, and $-R_1$ is the fixed charge). Note that because P_1 equals marginal cost, operating profits are negative for a natural monopolist; the fixed charge is therefore positive, equal to exactly the loss per customer that the firm in-

8. Because the tariff includes a fixed charge and the number of customers is not affected by this charge, the profits represent only a transfer from customers to the firm, with no consumption implications.

curs under marginal-cost pricing. In all subsequent periods, the situation remains the same: first-best optimality with zero profit for the firm.

9.6.2 Heterogeneous Customers

In the analysis above, the firm designs a tariff in the first period that extracts as profit all the potential surplus gain. The firm, by maximizing profit, thereby maximizes surplus. After the first period, there are no further potential surplus gains available, because first-best optimality has been achieved, and so the firm's profits become zero.

If different customers have different demands, the firm cannot design *one* tariff that captures the increase in surplus for all customers. If the firm offers a tariff with a lower usage price than the regulator's tariff, it can attract all customers to this tariff only if it raises the fixed fee by no more than the *smallest* increase in surplus that any customer obtains from the lower usage price. In this case, the firm is not extracting the entire gain in surplus for customers whose surplus gain is higher. If the firm instead charges a higher fixed fee, some customers will not choose the firms's tariff. Again, the firm is not extracting the potential surplus gain for all customers.

To enable the firm to extract all the potential surplus gains, Sibley proposes that the mechanism be applied separately to each customer in situations in which customers' demands differ. In each period the firm is allowed to design a tariff for each customer and offer that tariff only to that customer. The regulator also designs a tariff for each customer that the firm is required to offer the customer. The regulator's tariff consists of the usage charge that the firm charged under its own tariff for the customer in the previous period; the fixed charge is the average operating profit per customer that the firm made in the previous period.

The mechanism operates the same as described above, only on a customer-by-customer basis. The logic is exactly the same for each customer as when all customers have the same demand. In the first period, the regulator requires that the firm offer each customer a tariff consisting of a usage price that is the same as charged by the firm prior to the implementation of the mechanism, and a fixed fee that is the negative of the operating profit per customer that the firm had been earning. The firm designs for each customer a tariff that consists of a usage price equal to marginal cost and a fixed fee that is different

from the fixed fee under the regulator's tariff by (nearly) the amount of change in surplus that the customer obtains by moving to marginal-cost price. Because this change in surplus is different for customers with different demands, the fixed charge is different for customers with different demands. First-best optimality is attained in the first period, with the firm earning as profit the entire surplus gain of each customer. In the second period, profit is zero, because no further surplus gains are possible.

Essentially, Sibley's mechanism for customers with different demands works by combining the advantages of price discrimination with those of self-selecting tariffs. By itself, primary price discrimination allows the firm to extract all surplus; first-best optimality is achieved because the profits of the firm are the same as total surplus. However, the firm earns positive profits indefinitely. Self-selecting tariffs, judiciously used by the regulator, can reduce the firm's profits to be the *increase* in surplus from one period to the next, rather than the total surplus in any period. In particular, if the regulator requires the firm to offer tariff options that maintain the current level of total surplus and provide zero profit, the firm can earn more than zero profit only if it introduces other tariff options that benefit customers as well as itself, that is, that increase total surplus. When no further surplus gains are possible, the firm does not offer any new tariffs and earns zero profit under the tariffs required by the regulator. With self-selecting tariffs and price discrimination combined, first-best optimality is achieved along with zero profit.

9.7 Welfare Implications When Standard Assumptions Are Inappropriate

The analysis in this chapter has proceeded under the standard assumptions regarding customer behavior. In particular, customers are assumed to know their demand for the good at the time of choosing among tariffs and to choose the tariff that provides the most surplus. Empirical research repeatedly indicates, however, that customers do not choose in this fashion.[9] In particular, a significant share of customers are found to have chosen tariffs that do not provide the lowest bill for their observed level of consumption. In the context of phone

9. See for example, Kling 1985, Hobson and Spady 1987, Train, McFadden, and Ben-Akiva 1987, Kling and van der Ploeg 1989, and Train 1989.

service, for example, many customers choose flat-rate service over measured service, even though their bill would be lower under measured. Customers often do not enlist in calling plans offered by their long-distance carriers, even though the plans would reduce their bills for their current levels of calling; and, conversely, other customers are observed to join plans that actually cost them more money than under the carrier's standard rates for the same number of calls. Optional time-of-use (TOU) rates have been offered by energy utilities in many areas, and yet many customers do not switch to these rates even though they would receive a smaller bill for their current consumption levels under the TOU rates.

The implications of self-selecting tariffs are different when customers behave differently than assumed. Depending on how the observed behavior of customers is interpreted, the implications for surplus maximization can be either diminished or strengthened. Two interpretations with opposite implications are possible.

First, we might think that customers who choose tariffs that are not least costly for their observed consumption levels are simply making mistakes. If this is the case, one of the primary advantages of self-selecting tariffs is lost. Recall that under the standard assumptions, no customer is hurt if a firm offers a new tariff option without changing the original options. If, however, customers make mistakes in their choice of tariff, customers *could* be hurt in such a situation. In particular, a customer might mistakenly choose the new tariff, even though the new tariff actually decreases the customer's surplus.

The results on Pareto dominance (i.e., results 1 and 2) rely on the concept that offering a new tariff without changing the original ones does not hurt any customer. These results do not hold if customers make mistakes in their choice of tariff. It is not necessarily possible to design a tariff option that Pareto dominates, or even increases surplus, relative to an original set of options, if some customers whose surplus would decrease under the new tariff would mistakenly choose it.

A regulator who thinks that a significant portion of customers might make mistakes in their choices among tariffs is well advised to mandate multipart tariffs rather than self-selecting tariffs. As shown in section 9.3, a set of self-selecting tariffs can be represented equivalently as one multipart tariff when customers behave in accordance with the standard assumptions. Therefore, multipart tariffs serve the regulator as well as self-selecting tariffs when customers behave as assumed: any objective that self-selecting tariffs can achieve can also

be achieved by multipart tariffs. Unlike self-selecting tariffs, however, multipart tariffs protect customers from making mistakes. Consider a multipart tariff that has the same outlay schedule as a set of self-selecting tariffs. If all customers choose among the tariffs correctly, their bill would be the same under the self-selecting tariffs as under the multipart tariff. However, if a customer mistakenly chooses the wrong tariff, its bill under the self-selecting tariffs would be higher than under the multipart tariff. For example, in figure 9.4, a customer who chooses flat-rate service and then makes fewer calls than the threshold at which flat-rate becomes advantageous ("threshold 3" in the figure) would pay more under the self-selecting tariffs than under the "equivalent" multipart tariff. Because multipart tariffs without selection can perform all the functions of self-selecting tariffs when customers behave as traditionally assumed, and can also protect customers from making mistakes, any regulator who thinks that customers are likely to make mistakes would be better served by multipart tariffs.

This argument would suggest that the recent proliferation of self-selecting tariffs is misguided. If, however, the behavior of customers is seen in a different light, the use of self-selecting tariffs is perhaps justified. In particular, customers may choose tariffs that are not least costly for their consumption—that is, choose the seemingly "wrong" tariff—because there is some other feature of the tariff that appeals to them, beyond the surplus obtained under observed demand.

For example, customers may realize that their demand fluctuates over time, or customers may be uncertain of their future demand at the time of choosing a tariff. In the face of fluctuating or uncertain demand, different tariffs subject the customer to different levels of risk. Under flat-rate service, the customer's bill is fixed, with no variation or uncertainty. Under measured service, the customer's bill fluctuates from month to month, depending on the demand that the customer ends up having in the month. Which tariff is better for the customer depends on the customer's risk preferences. In particular, it is quite possible that a rational customer would choose flat-rate over measured service even though, for the number of calls that end up being made, the customer's bill is higher under flat-rate than measured service. That is, what might appear as a mistake under the (false) assumption that customer's demand is known and fixed can actually be the outcome of a more complex, but rational, choice process that incorporates issues of risk.

It is important to note that risk aversion does not necessarily induce

customers to prefer flat-rate service. Depending on the nature of the risk faced by the customer, a risk-averse customer might prefer measured service over flat-rate, all else equal. Consider for example a customer of a local phone service who does not know the amount of income he/she will receive in the upcoming months. With flat-rate service, the customer's phone bill is fixed, and the amount of money remaining for consumption of other goods varies as the customer's income varies. With measured service, the customer has the option to adjust its phone bill as its income fluctuates. In particular, when the customer's income is low (e.g., when between jobs), the customer can make fewer calls, saving money on its phone bill that can be used for consumption of other goods, such as food. Flat-rate service does not offer this type of protection against the possibility of low income. Risk-averse customers whose risk centers on income uncertainty might rationally choose measured service in this case, even if the customer's bill in months of normal income is higher.

The effort of optimization can also be a concern of customers. Under flat-rate service, for example, the customer does not have to worry about whether an extra unit of consumption is worth the price that must be paid for it. The convenience of not having to make such evaluations might in itself be worth something to the customer, such that a customer would be willing to pay more for flat-rate than measured service. Similarly, in choosing between time-of-use and non-time-differentiated rates, customers might compare the informational requirements of optimization under the two tariffs. To optimize under TOU rates, the customer must know the timing of its consumption, while this information is not necessary under non-time-differentiated rates.

In short: different tariffs are not simply different billing algorithms under known demand. They possess other features, such as the degree of risk, or the effort required for optimization. Customers have preferences over these features, and these preferences affect customers' decisions.

The implications of self-selecting tariffs (e.g., when they increase surplus, whether Pareto dominance is possible) have not been derived under behavioral assumptions that include these various factors. As a result, it is currently unknown whether self-selecting tariffs offer advantages that cannot be attained with multipart tariffs. It seems likely that they do. If tariffs are considered to possess features other than the surplus under known demand (e.g., degree of risk, effort of

optimization) and different customers have different preferences over these features, then offering various tariffs allows customers' preferences to be matched more closely than with only one tariff. In a sense, self-selecting tariffs might possess the advantages of product differentiation, by which surplus is increased by providing different kinds of products (in our case, tariffs) to customers with different tastes. This analogy is by no means clear, however. If a multipart tariff is designed that combines the least-cost components of a set of self-selecting tariffs, can customers ever be strictly better off with the self-selecting tariffs than the multipart tariff? Answering this question is an important task for future research.

We can now summarize the concepts relating to the use of self-selecting tariffs in situations in which customers do not behave in accordance with the standard assumptions. If customers are thought to make mistakes in their choice among tariffs, self-selecting tariffs are probably inadvisable. The regulator's goals could be better served with multipart tariffs. If, on the other hand, customers are thought to choose tariffs that best satisfy their preferences—in a context that includes risk, uncertainty, information costs, the effort of optimization, and other issues that are not included in the standard assumptions—then self-selecting tariffs might offer advantages over multipart tariffs.