Selling Federal Reserve Payment Services: One Price Fits All?

John A. Weinberg

In a large modern economy, there is a vast and constant movement of funds in the conduct of commerce and finance. The channels through which these funds move constitute the payment system, which, ultimately, forms a network connecting all participants in the economy. In dollar value, the bulk of this movement is not in cash but in the form of instructions for the crediting and debiting of accounts held with public or private financial institutions.[1] As a network for sending and receiving instructions, the payment system bears a resemblance to transportation and, especially, communication systems. Accordingly, many of the issues and questions that arise in discussions of markets for payment services have parallels in discussions of these other markets.

Markets that are characterized as networks are often thought to be driven by the existence of economies of scale. In the presence of scale economies, the average cost of providing services declines with the size of the network and the volume of traffic it carries. The belief in such economies has motivated a long history of direct government involvement and intervention in network markets, from the operation of the postal service to the regulation of telecommunications and transportation networks.

Much of the evolution of the structure of markets for payment services has been driven by the desire of participants to take advantage of the economies of network expansion. The most fundamental example is the replacement of a system in which payments are made in currency directly between individuals to one in which payments are made through accounts with financial intermediaries. Specifically, a check-based payment system opened the door to

This paper has benefited from the helpful comments of Bill Cullison, Jeffrey Lacker, Bruce Summers, John Walter, and Tom Humphrey. The views expressed herein are the author's and do not represent the views of the Federal Reserve Bank of Richmond or the Federal Reserve System.

[1] For a detailed description of the payment system, see Blommestein and Summers (1994).

network efficiencies to be gained through the centralized exchange of checks among banks in clearinghouses.[2] More recently, some payments have moved from checks into electronic forms of transmission. For instance, the use of Automated Clearinghouse (ACH) payments, for such purposes as payroll direct deposit, tripled in the number of transactions processed annually (from around 800 million to around 2.4 billion) from 1986 to 1992.[3]

In addition to technological factors, the evolving market structure in the payment system has been greatly influenced by the policy of the Federal Reserve System. Prior to 1980, many payment services were provided free of charge by the Federal Reserve to its member banks. As a result, a majority of payments cleared through the Federal Reserve, either directly or through correspondent banks. The Monetary Control Act of 1980, among its provisions, required Federal Reserve services to be made available, at a price, to all institutions. The Reserve Banks were instructed to set prices to cover all direct and indirect costs incurred in the provision of services. Since the institution of pricing, the Reserve Banks have experienced losses in market share to private providers. In check processing, for instance, renewed growth has occurred in the activities of clearinghouses on local, regional and, most recently, national levels. The resulting loss of market share by the Reserve Banks has been most significant among larger institutions.[4]

In the provision of ACH services, the Fed's position is somewhat more dominant than in check services. The Federal Reserve processed about 94 percent of all transactions in 1992 (McAndrews 1994). Private alternatives continue to develop, however. As in the case of check processing, new competition and the potential for institutions to engage in direct (nonintermediated) exchanges are focused on large-volume ACH users.

In the changing payment services environment, there have been a number of proposals for the restructuring of Fed pricing. Proposals for *market-sensitive* pricing tend to suggest advantageous pricing terms to large-volume users of services. Any such scheme amounts to some form of *price discrimination*. This term is purely descriptive: it applies to any pricing other than the setting of a single price per unit sold that is available to all buyers. The simplest example, referred to as *two-part pricing*, involves charging all buyers the same combination of a fixed fee and a per-unit price. When two-part pricing does not "discriminate enough," more complex schemes can be used. Examples include a per-unit price that varies with the quantity purchased and a schedule of combinations of fixed and per-unit charges among which buyers can choose.

[2] Goodfriend (1990) discusses this change and how banks developed institutions for enhancing payments efficiency and dealing with the resulting credit risk.

[3] McAndrews (1994) describes the growth in ACH payments.

[4] The General Accounting Office (1989) found that between 1983 and 1987 the Federal Reserve lost market share only among banks with over $750 million in assets.

Price discrimination in response to market competition raises some important questions about Federal Reserve pricing policy. For instance, are the Reserve Banks' "business interests" in conflict with their public policy responsibilities? Additional questions arise from the fact that price discrimination has the tendency to favor some institutions, particularly larger institutions, over others. Should the equal treatment of all banks be part of Fed pricing policy? Of course, at the most basic level is the question of whether the Fed should participate at all as a competitor with private providers of payment services.

This article argues that the public interest may be best served by a Federal Reserve pricing policy that is responsive to competition, within certain limits. This argument is based on the presumption that an important goal for Federal Reserve policy is the resource efficiency of the payment system. An efficiency perspective dictates that a loss of market share by the Federal Reserve is neither good nor bad per se. What matters is the overall cost efficiency of the market. If the Federal Reserve is replaced by providers with lower costs, then such a change should be accommodated. The goal of pricing policy, however, should be that only efficiency-enhancing losses are experienced.[5]

The central concept employed in this article is that of *sustainable prices*.[6] Sustainable prices are prices designed to sustain an efficient allocation of production by giving no buyer an incentive to seek to obtain the product from an alternative source. The following section briefly describes the organization and pricing in the markets for check clearing and ACH services. These two markets can be broadly characterized by a high volume of low-value transactions. As such, they make relatively intensive use of resources in transmitting payment instructions and constitute large markets for transmission services. The subsequent sections develop the notion of sustainable prices and use it to draw conclusions about Fed pricing policy. In particular, sustainable pricing can provide a guide for determining when market-sensitive pricing by the Fed is and is not in the public interest.

It is important to note that resource efficiency is not the Federal Reserve's only public policy interest in the payment services market. Indeed, the Fed's primary concern is with the overall safety and reliability of the system. This concern is expressed in the Fed's regulatory oversight of arrangements used for payment settlement. It is along the dimension of efficiency, however, that the Fed's role as a provider of many payment services should be evaluated. The Fed's participation should be determined by its ability to provide services in a cost-effective manner.

---

[5] While this article focuses on pricing, the terms of competition among alternative providers are affected by a variety of other factors. For instance, in 1994 the Board of Governors adopted a requirement of same-day settlement of checks presented by private collecting banks that put private-sector processing on a more equal footing with Fed processing.

[6] See Spulber (1989).

1. TWO PAYMENT SERVICES MARKETS IN BRIEF: CHECKS AND ACH

The concept of sustainable prices, as developed below, applies to concentrated markets.[7] Hence, it is useful to establish at the outset that markets for payment services tend to be fairly concentrated. In most of these markets, the Federal Reserve has a significant market share, while in some markets, the Fed's share is dominant. A brief description of the structure of two markets follows.

In 1992, over 72 billion noncash payments were made in the United States. Of these, 80 percent were made by check.[8] Checks are written on more than 15,000 banks and other depository institutions. In about 30 percent of all transactions made with checks, the recipient deposits the payment in an account in the bank on which the check is written. The clearing of these "on-us" items is a simple matter; the bank merely debits the account of the payor and credits the account of the payee (subject, of course, to the payor's account having sufficient funds). The remaining 70 percent of check payments must clear between banks. This clearing can proceed directly: a payee bank can send the check to a payor bank in exchange for funds. Alternatively, check clearing can make use of one or more of a number of intermediary services.[9] One such service is that provided by a clearinghouse. In a clearinghouse arrangement, a number of institutions agree to exchange checks drawn on each other at a specified place and time. Hence, a clearinghouse resembles multilateral direct exchange, except in the way that payments are cleared. With each exchange of checks, a clearinghouse member pays its net debit position or receives its net credit.

If a bank participates in a clearinghouse of any size or if it engages in direct exchange with a large number of banks, it must have the capacity to sort the checks it receives by payor bank. This task is performed by specialized equipment, reader-sorter machines. If a bank chooses not to invest in sorting capacity, it can, instead, send unsorted or incompletely sorted checks to an intermediary institution that completes the collection process. Both the Federal Reserve Banks and private collecting banks play this role. The collecting bank, private or Fed, may sort and send checks to payor banks or to subsequent collecting banks. For instance, a Federal Reserve office sends within-district checks to payor banks and out-of-district items to their respective Fed offices. In 1992, the Fed handled over 19 billion checks, about half of all checks requiring interbank clearing.

The resource costs in the check-collection process are dominated by two cost categories: the sorting and transportation of checks. Direct, bilateral

[7] For a treatment of the wide variety of theories of behavior in concentrated markets, see Tirole (1989).

[8] The data cited in this section are from the Bank for International Settlements (1993).

[9] The various paths for check clearing are reversed when a payor bank sends a "return item" (a check returned because of insufficient funds).

exchange of checks is the most costly means of clearing since it requires the payee bank to sort and ship to a large number of endpoints. Concentration of both activities can lead to cost savings. A group of banks that regularly receive checks drawn on each other can economize through a clearinghouse arrangement. Hence, the typical clearinghouse is composed of relatively large institutions within a metropolitan area. When an institution does not internalize the economies of concentration, it can instead purchase sorting and transportation services from entities that can take advantage of the cost efficiencies available.

The use of Automated Clearinghouse transactions is relatively new. In an ACH payment, the payor (or the payee with preauthorization by the payor) gives direct instructions to the payor's bank for the transfer of funds. Modern electronic information technology has made this means of transfer particularly cost-effective for recurring payments of set value. Accordingly, a growing fraction of the work force has wage and salary payments directly deposited into bank accounts by ACH. Other payments that might be made by ACH include mortgage payments and insurance premiums.

As with checks, ACH payments must clear between banks when the payor and the payee do not have accounts with the same institution. Clearing is facilitated if the payor and payee bank share an electronic connection over which instructions can be sent. Transactions can be made by direct bilateral exchange, through a private clearinghouse, or through the Fed. The first two options are likely to be used primarily by pairs or groups of banks that share a large number of payments. That is, private ACH transactions have been carried out primarily within geographic regions, while for interregional payments, the Fed has been the dominant provider. This market structure may be subject to change, however, as a private, national ACH initiative has recently begun competing with the Fed. In 1992, 94 percent of approximately 1.8 billion ACH transactions were made through the Fed.

Current pricing of Federal Reserve check and ACH services is a form of two-part pricing, a combination of a fixed fee and a per-unit price. In check services, the fixed charge is the *cash letter charge*. A cash letter is a collection of checks deposited with the Fed. The cash letter charge and the per-item fee vary with the amount of sorting that has already been done by the depositing bank and with the locations of the banks on which the deposited checks are drawn. The Federal Reserve Bank of Richmond's price structure for 1994 includes cash letter charges between $2 and $3 for most checks, while per-item fees range from less than 1¢ to 6¢.[10] These different fee combinations apply to varying amounts of sorting that may be necessary.

[10] Larger cash letter and per-item charges are assessed for some special categories of checks.

The prices for ACH services also vary with the particular services provided. The basic fee structure in the Richmond Fed's 1994 price list includes a participation fee of $20 per account per month and transaction (per-item) fees of 1¢ per intradistrict item and 1.4¢ per interdistrict item. In addition, a bank must have electronic access to the system. Access is priced with a monthly fee that ranges from $30 to $1000, depending on the type of connection maintained. While electronic access allows institutions to receive other services as well, at least part of the access fee can be considered the fixed cost of engaging in ACH transactions.

2. NATURAL MONOPOLY[11]

The main concepts to be employed can be demonstrated with a simple example of a single service that can be provided by one or more sellers. Let q_i be the quantity provided to the ith out of N buyers. Denote by q the array of quantities provided to all the buyers, $q = (q_1, q_2, \ldots q_N)$, and let Q be the sum of the q_i. The total cost incurred by a single seller in providing the service is given by

$$C(Q) = F + \sum_{1}^{N} f_i + v(Q). \tag{1}$$

The fixed cost has two components. A general cost of F, the *common* fixed cost, is incurred by any seller providing any quantity of the service (e.g., the cost of maintaining an accounting and communication system for ACH transfers). In addition, there may be a cost of f_i specific to the relationship with buyer i (the cost of an individual bank's electronic connection to the system). The variable-cost function, $v(Q)$, is increasing and convex.[12]

The basic ideas can be presented for the simple case in which only the common fixed cost, F, is present in equation (1). In this case, the relationship between total cost and output might be represented as in Figure 1. The corresponding relationship between average cost and output is shown in Figure 2. This U-shaped average-cost curve exhibits economies of scale as long as

[11] The case of natural monopoly is developed for expository purposes. The concept of sustainable pricing can be extended to any market structure. The application to concentrated, nonmonopoly markets closely parallels the case of natural monopoly. For instance, if all sellers can operate at minimum average cost, then that minimum cost is the sustainable price.

[12] It is worth pointing out that the N quantities (q_i) specified above could just as easily be interpreted as quantities of N different products. In that case, the variable-cost function $v(Q)$ might be replaced by a sum of separate cost functions, $v(q_i)$, for each of the individual products. The concepts developed here to analyze pricing of a single product in the presence of economies of scale are directly applicable to the pricing of a set of products in the presence of *economies of scope*. Economies of scope are said to exist when the costs of joint production of a set of products is less than the sum of the costs of separate production.

Figure 1 Total Cost Curve

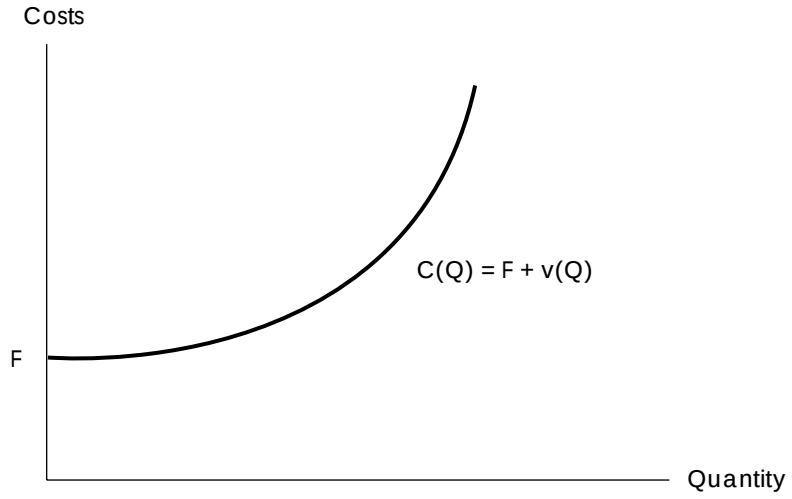
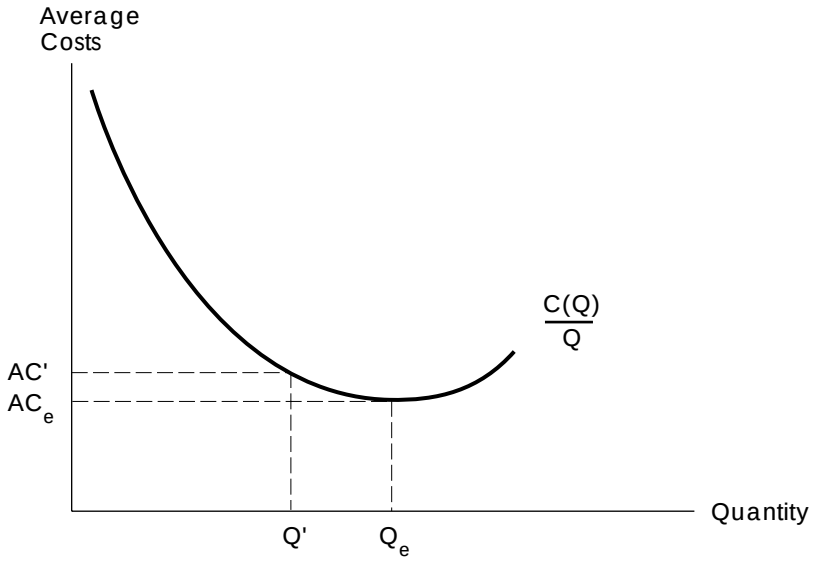


Figure 2 Average Cost Curve



Note: Figure 1 displays a total cost curve with fixed cost F and convex variable-cost function $v(Q)$. Figure 2 shows the corresponding average-cost function. The quantity Q_e is the "efficient scale" at which minimum average cost AC_e is achieved.

total output is less than the level labeled Q_e.[13] This level of output, at which average cost is minimized, is referred to as the *efficient scale* for the production of the service. Above the efficient scale, as average cost rises, there are diseconomies of scale. For now, it is assumed that all sellers and potential sellers have identical cost structures.

As in any market, pricing is affected by the structure of the market—e.g., the number and relative sizes of sellers. Market structure is, in turn, affected by the nature of the cost function for producing the service. If the total quantity demanded in this market was very large relative to Q_e, then competitive pricing and free entry among providers of the service would tend to result in a market composed of a large number of providers, each producing about Q_e. The price in this competitive market would tend toward AC_e in Figure 2. That is, when efficient scale is small relative to the size of the market, the invisible hand of competition works well; production costs are minimized and price just covers costs.

At the opposite extreme is the case in which a single seller's efficient scale (Q_e) is at least as large as the total quantity of service demanded by the market. In this case, competition among active providers cannot enhance the efficiency of production. Any division of output among sellers will only serve to raise the overall economic costs of providing the service, by duplicating the fixed costs. This is a case of *natural monopoly.* Under the belief that competition is infeasible, price regulation is often imposed on industries which are thought to operate under the conditions of natural monopoly.

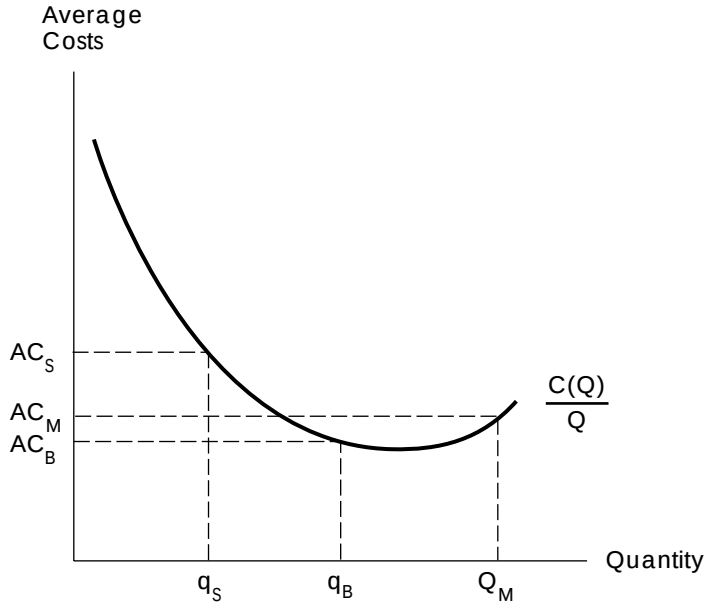
A natural focus for the pricing of the product sold by a natural monopoly, subject to the requirement that revenues just match costs, is to set a per-unit price equal to the average cost of producing the total industry output. Suppose, in Figure 2, that this output level is Q'. Suppose further that the quantity demanded is independent of price. This assumption is not essential but allows us to focus on the issue of whether or how the market quantity is divided among sellers. The price that just covers costs is AC'. Note that this price is greater than the marginal cost of production, since average cost is declining at Q'; when average cost is declining, marginal cost is less than average cost. Since price deviates from marginal cost, average-cost pricing in such cases is sometimes referred to as *second-best* pricing; "first-best" pricing would equate price to marginal cost, but would result in revenues less than costs. Second-best pricing maximizes net social benefits subject to the constraint that total revenues from the sale of the product just equal total costs.

[13] The average-cost curves should be understood as long-run average-cost curves. Although all factors of production are variable in the long-run, fixed costs are possible in the long-run if a minimum (positive) level of some input is necessary for production of any positive amount of output. For instance, to send telephone messages between two points, one must have, at least, one telephone line connecting those points. This represents a fixed cost, even in the long run.

Clearly, average-cost pricing in Figure 2 leaves no opportunity for a competitor to attract some piece of the market and at least cover its costs. Any piece of the market will involve average production costs greater than AC'. In order to win customers, however, a competitor would have to offer a price below AC'. In this case, uniform average-cost pricing (a per-unit price equal to AC' available to all buyers) is not vulnerable to the entry of competitors.

There is another case that falls into the category of natural monopoly for which pricing is more problematic. This case can be illustrated by a simple example in which there are two buyers, Big (B) and Small (S). Buyer B uses q_B units of the service, while S uses q_S. The total market quantity, then, is $q_B+q_S = Q_M$. Again, in the present example, quantity demanded is independent of price, except that each buyer seeks the lowest-cost supplier. The situation is depicted in Figure 3. Market quantity lies in the range of diseconomies of scale, and the average cost of serving the whole market is greater than the cost of serving just buyer B ($AC_M > AC_B$). Although market quantity exceeds efficient scale, the market is still a natural monopoly; any division of the market would result in higher total production costs. The average cost of serving only buyer S is greater than the average cost of serving the whole market ($AC_S > AC_M$).

Figure 3 Natural Monopoly with Quantity Greater than Efficient Scale



Note: Although market quantity, $Q_M = q_S + q_B$, is greater than efficient scale, the market is still a natural monopoly; the cost of serving the entire market is less than the combined cost of serving the market in any set of separate "pieces."

In this example, a simple price structure would set a uniform price equal to AC_M, the average cost of serving the entire market. If there are no legal barriers to entry, however, this price will induce a competitor to seek to gain a portion of the market. Specifically, a competitor can target buyer B, offering a price between AC_M and AC_B, the average cost of serving just buyer B. This strategy allows the competitor to take advantage of the economies of scale available in serving the large-volume user. Indeed, if no competitor were forthcoming and if buyer B had access to the necessary technology, then the buyer would be prompted to provide the service in-house.

A couple of comments on the competitor's pricing strategy are useful to bear in mind. First, the competitor must have reason to believe that the incumbent monopolist cannot or will not rapidly adjust prices in response to the competitor's move. Such a belief might be justified if the incumbent's pricing is subject to a cumbersome administrative procedure. Second, the competitor must be able to offer the lower price to a restricted set of buyers. If targeting a segment of the market requires making private deals with individual buyers, the competitor's task will be simpler if it is possible to identify a relatively small number of buyers with large enough volume to take substantial advantage of available economies of scale.

If the large-volume user defects to a competing source for the service, what becomes of the small-volume user? If the incumbent continues to offer the service at the price AC_M, then buyer S is just as well off as before. This price, however, no longer covers costs, which are now AC_S. Assuming the incumbent must cover costs, its price must rise. If it is resigned to serving only the remaining customer, S, then the incumbent must set its price at AC_S. Note that the end result may be an inefficient market structure. If there are two sellers operating, one serving buyer B and the other serving buyer S, then the duplication of fixed costs in serving the market constitutes social waste. The story may not end here. The incumbent may seek to win back some or all of the market share lost. This counterattack may ultimately succeed, but even temporary production by more than the efficient number of sellers is socially inefficient.

3. SUSTAINABLE PRICES

Are there pricing strategies for the incumbent that leave no room for encroachment by competitors? In the above example, the incumbent was vulnerable, because one buyer was charged a price that was greater than the cost of serving that buyer alone. The cost of serving only some subset of the buyers in a market is referred to as the *stand-alone* cost for those buyers. Accordingly, a set of buyers will be receptive to alternative sources of a service unless they face a price that is no greater than their stand-alone cost. A pricing scheme that meets this requirement for all sets of buyers is called a *sustainable* pricing

scheme. Sustainable prices leave no opportunity for a competitor with identical costs to capture any segment of the market.

How should one set prices that just cover costs and result in efficient production? In the case of natural monopoly, efficient production requires a single producer. In this case, the task is to find prices that recover costs and are sustainable. When market quantity is smaller than efficient scale, as in Figure 2, a uniform (per-unit) price equal to the average cost of producing the market quantity does the job. When market quantity is greater than efficient scale, as in Figure 3, there is no uniform (nondiscriminating) price that can satisfy both sustainability and cost recovery. On the other hand, a variety of nonuniform price structures can achieve the desired goals. One simple form for such pricing would be to give each buyer (or class of buyers) a distinct price. While this approach may not be practical in all circumstances, it is used here for illustrative purposes.

As defined above, sustainable prices generate total revenue that is necessarily no greater than total cost, since total cost is the stand-alone cost for the whole market. Sustainability, however, does not rule out prices that yield total revenue less than total cost. The Federal Reserve Banks operate under the requirement, from the Pricing Principles developed by the Board of Governors pursuant to the Monetary Control Act, that revenues be sufficient to at least cover all costs (the cost-matching requirement). Adding this condition to sustainability necessarily results in revenues that exactly match total costs.

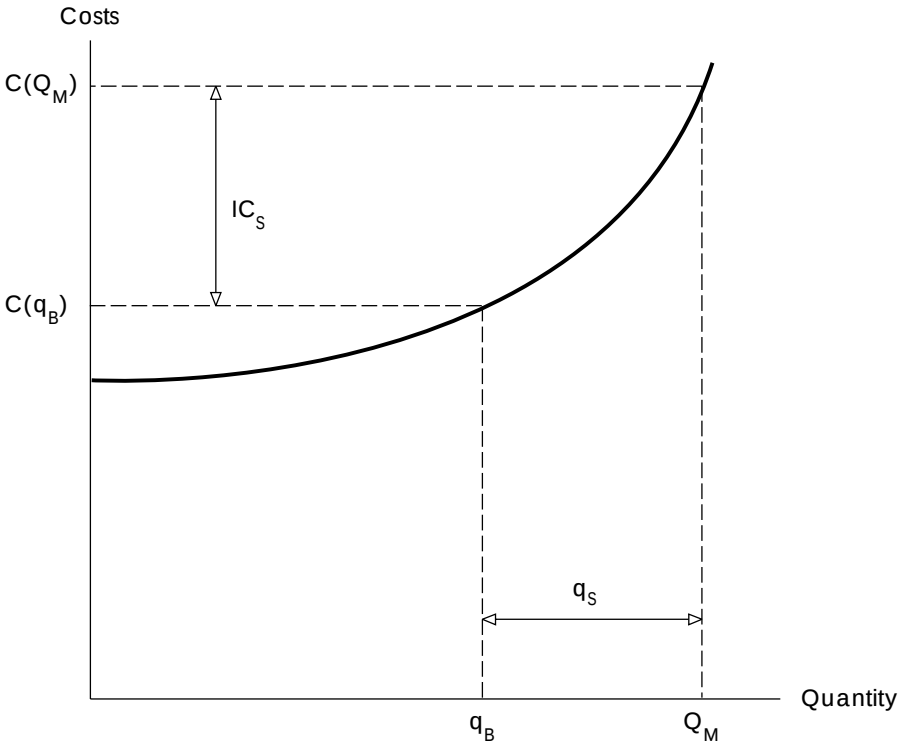
One implication of cost-matching, sustainable pricing is that at least one buyer must be given a price lower than stand-alone cost. Suppose, in the example of Figure 3, that buyer B is charged its stand-alone cost, in the form of a per-unit price of $AC_B = [F + v(q_B)]/q_B$. If both buyers are to be served, the revenue that needs to be collected from buyer S in order to just recover total costs is

$$[F + v(q_B + q_S)] - [F + v(q_B)]. \tag{2}$$

Here, the first term is the total cost of serving both customers, while the second term is the stand-alone cost of serving the large-volume customer. The difference between these two terms is referred to as the *incremental* cost of serving customer S. This cost is denoted by IC_S in Figure 4. Hence, the revenue needed from buyer S can be collected with a per-unit price equal to IC_S/q_S. If buyer S is charged anything less than this price, then in order to recover costs, the seller must charge more than AC_B. If B is charged more than AC_B, a competitor will take B's business.

It is important to note that incremental cost, as the term is used here, is not the same as marginal cost. The former, as indicated by equation (2), is the cost of providing a particular quantity to a particular buyer, given the quantity being provided to other buyers. The latter is simply the cost of providing an additional unit of the product, without regard to the identity of the recipient. It

Figure 4 Incremental Cost

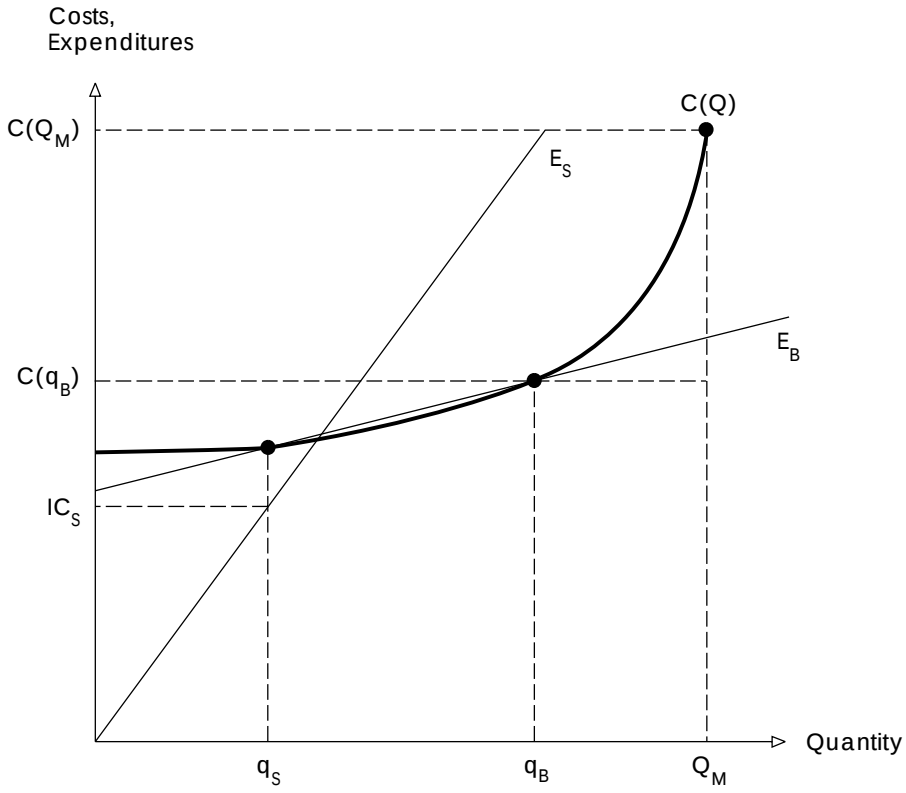


Note: $C(q_B)$ is the stand-alone cost of serving buyer B, while $C(Q_M)$ is the total cost of serving the entire market. The difference between the two, denoted IC_S, is the incremental cost of serving buyer S.

is possible to have a pricing structure in which the (marginal) price charged to some buyers is less than marginal cost while no buyer's average price is less than its average incremental cost.

The price discrimination just described requires that buyers be segmented into groups according to some observable characteristic. This task may not always be straightforward. For instance, the quantity of services used by an institution may be subject to significant change over time. In that case, setting a price to a buyer based on the buyer's previous behavior may not yield the desired results of tailoring prices to current demand conditions. Fortunately, the desired segmentation can typically be achieved by pricing schedules that allow buyers to self-select into groups. One example is "option pricing," in which buyers are given a choice between a schedule with a high fixed charge and low fee per unit and a schedule with a low fixed charge and high fee per unit. For the two-buyer example, Figure 5 illustrates the total expenditure

Figure 5 Option Pricing



Note: The lines labeled E_S and E_B give the total expenditures (as a function of quantity purchased) resulting from buying services under the two alternative options. Under one option, given by E_S, the buyer pays no fixed fee and pays a per-unit price of $P_S = IC_S/q_S$, which is the slope of the line E_S. This price generates expenditures by buyer S equal to incremental cost. The other option, given by E_B, includes a positive fixed fee and a lower per-unit price (slope). The key features of the schedule E_B are that it meets the total cost curve $C(Q)$ at the quantity q_B and that it lies below E_S at q_B. Hence, buyer B prefers the schedule E_B and has a total expenditure equal to stand-alone cost, $C(q_B)$. The fixed fee in the schedule E_B must be (and is, as drawn) high enough so that buyer S prefers the schedule E_S.

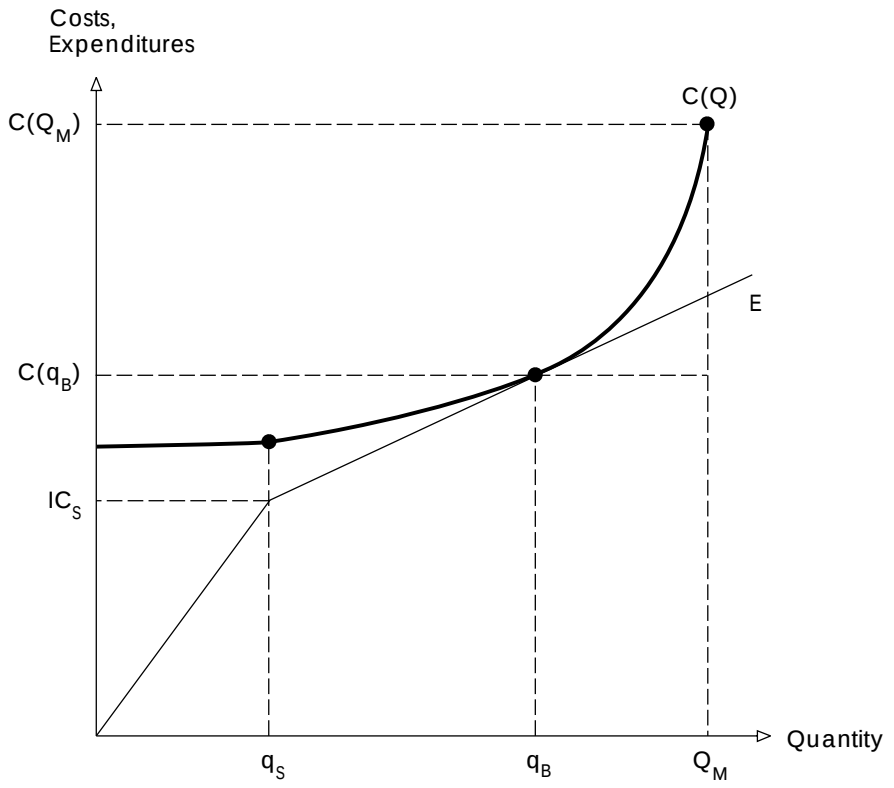
schedules generated by such pricing options. In this example, the low fixed fee is actually set at zero and combined with a per-unit price of $P_S = IC_S/q_S$. An individual buyer will choose the option for which its own total expenditures are the smallest. Hence, the low-volume buyer selects the low fixed charge and high per-unit fee, which is constructed so that total expenditures by buyer S just cover incremental costs. Buyer B selects the other schedule, resulting in expenditures equal to stand-alone cost.

The desired results also could be achieved by a pricing schedule with volume discounts, in which the per-unit fee varies with the quantity purchased.

The total expenditure schedule from one such price structure is given in Figure 6. This price schedule sets a price of IC_S/q_S for the first q_S units purchased. For each additional unit purchased beyond that threshold, the buyer pays the lower price of $[C(q_B) - IC_S]/(q_B - q_S)$. As Figure 6 indicates, this volume-discounting scheme is designed to collect exactly stand-alone cost from buyer B and incremental cost from buyer S.

An option pricing scheme, like that presented in Figure 5, has recently been adopted on a trial basis by some Federal Reserve Banks for some check-processing services. Most prices for Reserve Bank services, as discussed in Section 1 above, are a simpler form of two-part pricing, including a single fixed fee and a single per-item charge. This form of pricing is not as flexible

Figure 6 Volume Discount



Note: Like Figure 5, Figure 6 presents a pricing arrangement that results in expenditures equal to stand-alone cost for buyer B and incremental cost for buyer S. This arrangement involves no fixed fee. The per-unit price is IC_S/q_S on the first q_S units purchased. For each subsequent unit purchased, a lower price is charged. This lower price is equal to $[C(q_B) - IC_S]/(q_B - q_S)$. The resulting total expenditure, as a function of quantity purchased, is given by the kinked line, *E*.

as an option pricing schedule. Accordingly, sustainability may or may not be achievable with simple two-part pricing.

In summary, sustainability requires that no buyer (or set of buyers) faces prices greater than stand-alone costs. Adding the cost-matching condition requires that no buyer (or set of buyers) faces prices less than incremental costs. That is, each buyer must face a price between stand-alone and incremental cost. Within these restrictions, the relative pricing to different buyers can be treated in a variety of ways. In the examples above, buyer B is charged stand-alone cost and buyer S is charged incremental cost. Consequently, all of the fixed costs of production are allocated to buyer B. This allocation could be reversed, charging stand-alone cost to buyer S and incremental cost to buyer B. Price schedules that allocate some fixed costs to each buyer face both buyers with prices in between stand-alone and incremental costs. The concept of sustainability, by itself, gives no guide to the choice among these alternatives. The next section suggests that the possibility of technological differences among alternative providers of a service can help to sharpen the choice.

4. BYPASS AND TARGETED COMPETITION

The forgoing development of sustainable prices assumes that the same technology is available both to the incumbent firm (the Fed) and to any potential or actual competitors. Hence the relevant cost standard for deterring entry is the stand-alone cost of a market segment. There may be instances in which some segment of the market can be served with a technology different from that used by the incumbent seller. In such cases, the term "stand-alone cost," as defined above, is somewhat of a misnomer. This cost is the cost to a buyer, or group of buyers, of obtaining services from a source with a cost structure identical to the incumbent's. When the alternative to the incumbent's network involves a substantially different technology, the buyer's option is not so much to "stand alone" as to "bypass" the network.

An example, as described by Einhorn (1987), is in the provision of long-distance telephone services. Most long-distance calls are routed through the local telephone company, for which a charge is assessed. Large-volume callers, however, may exercise the option to bypass the local company and connect directly with their long-distance provider. The technology for bypassing the local network and, therefore, the costs associated with doing so are different from those associated with connecting through the network.

One can also think of obtaining payment services from alternative sources as bypassing the Federal Reserve network. For instance, a local check clearinghouse utilizing centralized exchange of items involves a different pattern of sorting and transportation expenditures from that arising in the use of the Fed's check-clearing services. Of course, in this regard, the most stark example of bypassing the Fed's network is direct bilateral exchange of payments.

The presence of a bypass alternative places additional constraints on the pricing choices facing the incumbent seller. Continuing with the example of Section 2, suppose that there are two buyers with quantities demanded of q_S and q_B and the sum of these quantities is denoted Q_M. Buyers B and S can bypass the incumbent and receive the service at a cost of C_B^* and C_S^*, respectively. The incumbent seller's total cost of serving the entire market is $C_M = F + v(Q_M)$, while buyer B (S) alone can be served by the incumbent for the stand-alone cost of $C_B = F + v(q_B)$ ($C_S = F + v[q_S]$). Suppose that the bypass technology is potentially attractive only to a large-volume buyer, so that $C_B^* < C_B$ but $C_S^* > C_S$. On the other hand, suppose that the market is still a natural monopoly. This is so if when buyer B bypasses the system, total costs rise, or $C_M < C_S + C_B^*$. This last statement is equivalent to saying that the incumbent's incremental cost of serving buyer B is less than the bypass cost, since $IC_B = C_M - C_S < C_B^*$.

Under the conditions just described, the incumbent is limited in how much of the common fixed costs can be allocated to buyers with a viable bypass option. Recall that allocating all of the (common) fixed costs to a single buyer amounts to charging that buyer its stand-alone cost. Here, such a price to buyer B would induce B to bypass the incumbent, to the detriment of market efficiency.

The incumbent's pricing problem is further complicated if there is some uncertainty about the viability of the bypass technology. Suppose, for instance, that the incumbent is reasonably sure that $C_S^* > C_S$ but is uncertain as to the value of C_B^*. If buyer B's bypass cost is so low that it is less than the incremental cost of serving B, then it is efficient to let buyer B bypass. Indeed, in this case, there are no prices that the incumbent can set to cover costs and guarantee against the loss of buyer B. On the other hand, it is still possible to price in such a way that buyer B will be lost only if bypass is efficient. Specifically, pricing to buyer B at incremental cost and buyer S at stand-alone cost will succeed in covering costs regardless of whether B is retained. Further, by comparing its bypass and incremental costs, B makes its choice in a way that minimizes the total (social) costs of serving the market.

To summarize, the presence of bypass options that are potentially attractive to some individual buyers or groups of buyers limits the ways in which fixed costs can be recovered from buyers. Especially when the value of the bypass option is not fully known by the incumbent, prices to segments of the market that are likely to have the most attractive bypass options should be pushed down to the incremental cost of serving those segments. Such an allocation of fixed costs is likely not to coincide with the allocations implied by standard accounting practices, and it may strike some (especially other buyers) as inequitable. It is important, however, to consider the alternative. If the incumbent seeks to recover some of the fixed cost from likely candidates for bypass, those buyers may turn to alternative sources even when it is socially inefficient to do so. If their business is lost, the incumbent will still have to recover fixed

costs from the remaining buyers. This result would be less cost-efficient and no more equitable than the result of sustainable pricing that recovers all fixed costs from those buyers with the least attractive alternatives.

5. ADDITIONAL COMMENTS ON SUSTAINABLE PRICES

While this article has presented sustainable pricing as a tool for evaluating pricing from a public policy point of view, the concept originated as a predictive notion in the theory of "contestable markets."[14] Contestable markets theory holds that in the presence of potential competition, incumbent firms will not be able to charge anything other than sustainable prices; any attempt to charge unsustainable prices would quickly prompt entry of and loss of market share to a competing seller. In other words, even in a natural monopoly, there are no economic rents earned by an incumbent monopolist. This is a strong conclusion that has not been broadly accepted without qualification. Most importantly, one cannot discuss the effects of potential entry without considering the likely response to entry by the incumbent seller. In its purest form, contestable markets theory assumes that the incumbent can alter its price in response to entry only with some lag. The incumbent's inability to respond quickly leaves an opportunity for an entrant to capture, at least temporarily, some part of the market should the incumbent's prices be unsustainable. Ultimately, the incumbent may regain the market, but the absence of sunk costs implies that even temporary profit opportunities will be exploited by entrants.

In an unregulated market populated only by private firms, there is little reason to suppose that firms do not have a great deal of flexibility in adjusting their prices to competitive conditions. The situation of a Reserve Bank, however, may come closer to that imagined by the contestable markets theory. Clearly, the process necessary to adjust pricing policy is time-consuming. Reserve Banks must set and publish prices once each year. Further, the Board of Governors' Pricing Principles, adopted pursuant to the Monetary Control Act, state that substantive changes in the structure of prices or services offered shall be made subject to public comment. Volume-based pricing for check services (on a limited basis) was approved by the Board in November 1993 and became effective in January 1994, "subject to additional staff analysis and public comment" (Board of Governors of the Federal Reserve 1994a).

When price adjustment is subject to lags, then prices that are not sustainable can attract entry, even if the market cannot efficiently support the additional seller(s) in the long run. Hence, the use of unsustainable prices can attract excessive entry when entrants can take advantage of an incumbent's administrative delays in responding to competition.

[14] See Baumol, Panzar, and Willig (1982).

It is also useful to compare sustainable prices to a pricing concept often used in discussions of regulatory price setting. In such discussions, one approach is to seek prices that maximize social welfare, subject to a zero-profit constraint for the seller. The "social welfare" to be maximized is a measure of the benefits (e.g., utility or profits) received by buyers. The resulting prices are referred to as *Ramsey* prices, because their derivation follows Ramsey's (1927) formulation of optimal taxation. Sustainability is a stronger constraint than zero profits. Hence, Ramsey prices will not, in general, coincide with sustainable prices. Accordingly, the former might be more applicable to the problem of setting prices in the public interest when an incumbent seller is protected from competition by legal barriers to entry.

Unlike Ramsey prices, the notion of sustainability used here is entirely cost-based; it does not take into account a measure of the benefits generated by the provision of payment services. A cost-based specification of sustainability is exact when demands for services are assumed to be perfectly inelastic. The more general specification would require that the net value provided to any group of buyers (benefits to buyers less payments to seller) be no less than the greatest net value those buyers could obtain from an alternate source. While this generalization is a direct extension of the basic idea, measures of benefits on the demand side of a market may be difficult to obtain. Hence, the cost-based notion of sustainability may remain useful as a practical approximation to the more general concept.

6. ARE PAYMENT SERVICES MARKETS NATURAL MONOPOLIES?

Sustainable pricing is presented above in the context of a market that is a natural monopoly. Neither of the markets discussed above, check and ACH services, is a monopoly, although the ACH market comes close. Even the market for check services, however, is fairly concentrated; in any given geographic region, the Federal Reserve serves a significant share of the market for intraregional processing. Market structure is determined in part by the degree of scale economies relative to the size of the market. Hence, a concentrated market is likely to be one in which demand and technology conditions are such that only a small number of sellers is viable. In such a market, the analysis of sustainable pricing closely parallels that of natural monopoly.

The analysis offered in this article does assume that a seller's efficient scale is at least a sizeable fraction of the size of the market. Hence the applicability of the pricing principles proposed above is partly an empirical matter. Specifically, what evidence exists on the significance of scale economies? There have been a number of studies of the Federal Reserve's check-collection services,

aimed at addressing this question.[15] These studies tend to find fairly weak scale economies in the observed range of production levels.[16] Such findings might seem at odds with the narrative description of the experience in check processing (and in ACH services), which seems to parallel the analysis of Section 2; average-cost pricing to the market as a whole led to the defection of high-volume users of the services. One possible conclusion is that the alternative means used by defecting customers do, in fact, deliver the services with lower real resource costs. That is, these users may have access to a superior bypass technology. In that case, the Fed's loss of market share would be efficiency-enhancing. On the other hand, the analysis of Figure 3 refers to a case in which the incumbent operates above efficient scale. If this case were an accurate description of the Fed priced-services environment, then one would not expect to find empirical evidence of widespread, unexploited economies of scale.

Aside from economies associated with check processing, there may be scale efficiencies in the distribution and transportation of processed checks. Fixed costs that are specific to each endpoint served may result in markets where efficient scale is a sizeable fraction of the relevant market.

There is also the possibility that economies exist less in increasing the scale of production of any given service than in the joint provision of multiple services. This is the most common use of the term "economies of scope." For instance, a single electronic connection to a Reserve Bank can allow a customer to use ACH services and other electronic services, including new electronic check-collection options. To the extent that scope economies exist, it may not make sense to talk about market structure, pricing, and cost recovery on a product-by-product basis. The concept of sustainable prices, however, can be directly extended to an environment with economies of scope. Consider the pricing of an array of services. For such pricing to be both sustainable and cost-matching, no service to any buyer or group of buyers can be priced above stand-alone cost or below incremental cost. Here, stand-alone cost is the cost of providing only a specific subset of the services to a specific subset of the buyers. Similarly, incremental cost refers to the added cost of a specific subset of services to a subset of buyers, given the services already being provided to other buyers. As before, choices among sustainable price configurations amount to choices among possible allocations of common fixed costs across buyers and services. If a particular service is targeted for competition (for instance, because of the availability of a bypass technology specific to that service), then that service's price should be set at incremental cost.

Even if the structure of cost and demand is such that these markets are not natural monopolies, the concept of sustainable prices can still provide a useful

[15] A recent example is Bauer and Hancock (1992).

[16] If, as these studies suggest, the average-cost curve is relatively flat at its minimum, then average-cost pricing should be close to sustainability.

benchmark for Reserve Bank pricing policy. The cost structure in a market might be such that the efficient number of sellers is greater than one but still small. For instance, if the market quantity sold tends to be about three times the efficient scale of production, then the efficient number of sellers is three. In such a "natural oligopoly," pricing behavior tends to be the result of a complicated dynamic game. Here, the administrative structure that governs Reserve Bank pricing can be advantageous in that it may give the Federal Reserve the ability to precommit to a pricing strategy over a long horizon.[17] When the Fed is one of several competitors, it can contribute to the efficiency of the market by adopting a clear pricing policy to which other sellers can react. Specifically, the Fed could make it known that it stands ready to sell to any market segment at no greater than stand-alone cost and no less than incremental cost. Within these bounds, it will adjust pricing to respond to competition, moving prices in more competitive segments toward incremental cost. Such a strategy makes it clear that market gains by competitors that reduce overall social costs will not be contested, while those that raise costs will not be accommodated. Under sustainable pricing, a seller cannot preserve market share that is not justified by its technological capabilities.

7. SUSTAINABLE PRICING AND THE MONETARY CONTROL ACT

The move toward market-sensitive pricing that responds to competitive conditions might raise questions about the role of the Federal Reserve in the provision of payment services. To what extent should a Reserve Bank behave like a private business? Does an attempt by a Reserve Bank to maintain its share of the market interfere with its public policy objectives with regard to the payment system? To the latter question, the discussion in this article suggests the answer, "Not necessarily." By letting its prices be guided by the notion of sustainability, the Federal Reserve establishes a benchmark for the market place. If competition targets a particular segment of the market, that segment should be served by the Federal Reserve at incremental cost. Then, any gains in market share by competitors will also be in the public interest. It is also worth noting that no market segment is being subsidized by another as long as no price is less than incremental cost.

It is important to note that the pricing behavior suggested herein is, in many cases, not the behavior one would expect from a private business. That is, the resulting pricing is not the pricing that would prevail in the market if the Fed played no operational role. Private businesses are motivated by long-run profit

[17] A treatment of the benefit of precommitment in oligopoly pricing games can be found in Tirole (1989).

maximization. This may lead to deviations from sustainable prices in a number of ways. First, an incumbent firm facing potential entry can set prices to any market segment above stand-alone cost, as long as the incumbent has adequate flexibility to adjust its prices in response to entry. In other words, revenues can more than cover costs. Second, there may be situations in which a private firm will be willing and able to set prices that fail to recover all costs in the short run. Suppose, for example, that two firms find themselves in competition in a market that has the cost and demand structure of a natural monopoly.[18] In the long run, only one of the firms can remain in the market. To determine which firm will survive, the two might engage in a "war of attrition" in which prices are below costs and losses are incurred until one firm chooses to exit. The short-run pricing would necessarily involve some prices to some market segments below incremental cost. While the Monetary Control Act does allow the Fed to have revenues that fall below costs in the short run, the Board of Governors has adopted the policy of setting prices each year with the aim of recovering all anticipated costs for that year (Board of Governors of the Federal Reserve 1994b).

Unlike the pricing behavior of a private business, market-sensitive, sustainable pricing is motivated not by profit maximization, but by an interest in the overall efficiency of the market for payment services. This motivation drives pricing as close as possible to the "first-best" result of marginal-cost pricing of all products to all buyers. The constraints that keep pricing away from that goal are the need to cover costs and the need to ensure that market share is lost only when the loss results in lowering the resource costs of serving the entire market.

Does a pricing policy that results in disparate treatment of banks conflict with the goals of Congress in writing the pricing requirement into the Monetary Control Act? The language of the Act instructs the Federal Reserve to "give due regard to . . . the adequate level of [services] nationwide." Since the sustainable pricing schemes outlined above tend to involve average and marginal prices that decline with the volume of services purchased, it appears that such pricing will favor large institutions, because small banks would pay a higher average price. Hence, disparate treatment, in the form of higher average prices, might be thought of as impeding smaller institutions' access to services. The language in the Monetary Control Act could conceivably be interpreted as prohibiting pricing that faces some institutions with a greater cost of access to services. In the presence of economies of scale or scope, pricing that achieves equal treatment of all buyers and just recovers costs is typically not sustainable. Hence, if the Monetary Control Act is interpreted strictly as mandating

[18] For instance, a market that could previously support two firms might experience a permanent decline in demand.

equal treatment, the Federal Reserve could find itself in an intractable bind; if uniform, unsustainable prices result in significant loss of business, the Reserve Banks could have difficulty covering costs without raising prices to remaining buyers. The result would be equal treatment by the Fed but disparate treatment by the market as a whole.

While the language of the Monetary Control Act may or may not be read as providing a mandate for equal treatment, it does seem to dictate a continued role for the Fed in the provision of payment services. Without such a legislated dictum, one might legitimately wonder whether there is a necessary role for the Fed in these markets. Indeed, the central result in the theory of contestable markets, as noted above, is that the force of potential competition among private businesses is sufficient to yield sustainable prices. On this point, experience in deregulated transportation and telecommunication markets has been inconclusive. These markets tend to be highly concentrated, and strategic interaction may tend to result in fluctuation between collusive and aggressively competitive behavior. In such an environment, it is conceivable that a single large provider committed to a sustainable pricing policy could provide a stabilizing influence on the market while promoting an efficient market structure.

8. CONCLUSION

This article proposes a general principle for evaluating Reserve Bank pricing strategies. The concept of sustainable pricing under conditions of scale and scope economies appears to be a useful tool. Sustainable prices that just cover total costs price all services to all customers in between their stand-alone and incremental costs. When competition from private-market providers is focused on a subset of services and customers, sustainability retains enough flexibility to respond to competitive pressures by pushing some prices down to incremental cost. This response is particularly appropriate in conditions of uncertainty about competitors' costs.

A strategy of market-sensitive sustainable pricing would result in loss of business to competitors only when such loss is efficient. Hence, this strategy provides a guideline for responding to competition in a way that respects the requirements of the Monetary Control Act while promoting efficiency in the delivery of payment services. If the Federal Reserve is going to be in the payment services business, it should use its position as a provider motivated by the public interest to guide the market in the direction of efficiency. Sustainable prices provide market participants with a benchmark for assessing the cost effectiveness of alternative modes of service delivery. Following this benchmark may or may not stem the Fed's loss of market share, but maintaining market share should not be a goal of Fed policy. The Fed's market share should be whatever is consistent with the efficient operation of the payment system.

REFERENCES

Bank for International Settlements. *Payment Systems in Eleven Developed Countries.* Bank Administration Institute, 1993.

Bauer, Paul, and Diana Hancock. "The Efficiency of the Federal Reserve Payments System," Working Paper. Washington: Board of Governors of the Federal Reserve, 1992.

Baumol, William J., John C. Panzar, and Robert D. Willig. *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt Brace Jovanovic, 1982.

Blommestein, Hans J., and Bruce J. Summers. "Banking and the Payments System," in Bruce J. Summers, ed., *The Payment System: Design, Management and Supervision*. Washington: International Monetary Fund, 1994.

Board of Governors of the Federal Reserve System. *80th Annual Report,* 1993. Washington: Board of Governors, 1994a.

_____. *Federal Reserve Regulatory Service,* Vol. III. Washington: Board of Governors, 1994b, locator number 7-134.

Einhorn, Michael A. "Optimality and Sustainability: Regulation and Intermodal Competition in Telecommunications," *Rand Journal of Economics,* vol. 18 (Winter 1987), pp. 550–63.

Goodfriend, Marvin S. "Money, Credit, Banking and Payment System Policy," in David B. Humphrey, ed., *The U.S. Payment System: Efficiency, Risk and the Role of the Federal Reserve*. Boston: Kluwer Academic Publishers, 1990, pp. 247–277.

McAndrews, James. "The Automated Clearinghouse System: Moving Toward Electronic Payment," Federal Reserve Bank of Philadelphia *Business Review,* July/August 1994, pp. 15–23.

Ramsey, Frank P. "A Contribution to the Theory of Taxation," *Economic Journal,* vol. 37 (March 1927), pp. 47–61.

Spulber, Daniel F. *Regulation and Markets*. Cambridge, Mass.: MIT Press, 1989.

Tirole, Jean. *The Theory of Industrial Organization.* Cambridge, Mass.: MIT Press, 1989.

United States General Accounting Office. "Check Collection: Competitive Fairness Is an Elusive Goal," Report to Congressional Committees, May 1989.