Can the Fed be a Payment System Innovator?

Jeffrey M. Lacker
John A. Weinberg

Jeffrey M. Lacker is a vice president and economist in the Research Department. John A. Weinberg is a research officer and economist in the Research Department. The views expressed do not necessarily reflect those of the Federal Reserve System.

We live in a time of rapid technological change, in which the arrival of new ways of conducting business has become a commonplace occurrence. One segment of the economy where these changes are having a particularly significant effect is the payment system, the web of banks and other institutions through which payments for goods and services are cleared and settled. New mechanisms such as smart cards and internet-based electronic money have captured the imagination of many payment system observers and participants. While earlier predictions of the death of paper money have proven premature, the unprecedented pace of technological advance in the last decade has given new hope to the prophets of the electronic age.

The Federal Reserve pays a prominent role in the payment system, both as a provider of payment services and a regulator. The public interest in an economically efficient payment system has been at the core of Fed payment system policy since the Fed's founding in 1914. With new electronic payment mechanisms apparently within grasp, there has been renewed attention to the role of the Fed in the innovative process. A committee headed by Federal Reserve Board Vice Chair Alice Rivlin recently completed a study of the Fed's role in the payment system, which gave special attention to how active a role the Fed should play in guiding payment system innovation.¹

Within the Federal Reserve System, electronic check presentation (ECP) is seen as a potentially promising step in
the evolution toward electronic payments. With ECP, consumers and businesses continue to make payments with paper checks, but banks and clearinghouses that clear and settle payments use electronic information “captured” from the checks shortly after they are first deposited in the banking system. (See ECP box p. 23.) While some ECP services are now available, many important aspects of full-scale implementation are still under discussion. The Fed’s role in developing and promoting ECP is clearly aimed at the public interest objective of enhancing payment system efficiency. In what follows, we ask whether the Fed can be a payment system innovator while remaining loyal to its fundamental public interest objective. In particular, how can we ensure that the Fed’s payment system leadership contributes to economic efficiency?

Our approach to this policy question is founded on the notion that the payment system is a communications industry. Such industries involve substantial common costs — costs that cannot be uniquely attributed to any one user. This cost characteristic has important implications for industry behavior. The critical issue in such industries is how common costs are allocated across users.

Markets for communication services (including payment services) tend to be heavily regulated and, in some instances, served by government-owned enterprises, such as the U.S. Postal Service. Concerns about “universal access” often motivate government intervention. Here, universal access is usually interpreted as a concern about the cost of services to a particular class of users: residential phone customers, rural postal patrons, or small and remote depository institutions. Access has been provided through price regulation, as in telecommunications, and by direct government provision, as in the U.S. Postal Service.

We show that government involvement in other communications industries offers lessons for the role of the Federal Reserve in the payment system. In both the telecommunications and postal services industries, legal barriers to competition historically have helped sustain the provision of universal access. Barriers to competition allow the shifting of common costs to be pushed to the point where some users are subsidized, in a sense that we will make precise later. Such subsidization is inconsistent with economic efficiency, and would be impossible without barriers to competition. We point out that the Federal Reserve Banks still benefit from some barriers to competition — privileged treatment under current check presentment regulations — that would allow them to subsidize should they choose to do so. Federal Reserve policy explicitly seeks to prevent subsidization, and there is no direct evidence that the Fed currently subsidizes any segment of the check collection market. At the same time, however, it is clear that available analytical methods for determining the absence of subsidies are imperfect.

Barriers to competition impede technological progress by distorting adoption choices. The contrasting experiences of the telecommunications and postal services industries illustrate this fundamental conflict. In telecommunications, the removal of barriers and a retreat from access have been accompanied by rapid technological innovation. The U.S. Postal Service has retained barriers to competition and in the view of some observers has been relatively slow to adopt innovations. Barriers provide the opportunity for cross-subsidies that distort the innovative process. Against this background, we argue that the Fed should act to preclude subsidization by removing remaining barriers to competition. As we emphasize, however, this step may require some retreat from the goal of providing universal access.
The clearing and settling of credit instruments used as means of payment intrinsically requires communication.
THE PAYMENT SYSTEM AS A Communications Industry

In the U.S. economy, roughly 220 million market transactions are made without cash daily, with a total dollar value of $1.6 trillion. These transactions all involve credit. The seller receives a financial instrument representing a claim on either the buyer or a third party. For example, a check is the liability of the check writer and his or her bank. A credit card sale results in a claim — a "sales slip" — that entitles the merchant to good funds. Similarly, a debit card transaction gives the merchant a claim to good funds.

The clearing and settling of credit instruments used as means of payment intrinsically requires communication. A vast web of bookkeeping systems record the assets and liabilities of various economic entities — bank accounts, loan balances, investment funds, and the like. Noncash payment instruments are fundamentally bookkeeping instructions to debit an account of the buyer and credit an account of the seller. Those instructions must be communicated to the relevant bookkeeping systems in order to carry out the necessary accounting entries.

The payment system bundles together communication and financial services. Arrangements governing the use of payment instruments specify the allocation of risks associated with payment failures. For example, the merchant accepting a check bears the risk that the check writer may fail to cover it, but the merchant does not bear the risk of a fraudulent credit card purchase. While these risk-sharing arrangements are an important feature of the evolution of the payment system, they do not make the fundamental function of payment arrangements inherently different from other communication services.

Every new development in communication technology brings with it a new possibility for sending payment instructions. Improvements in freight transportation increase the speed and reliability with which checks can be delivered to a buyer's bank. Improvements in computer and telecommunications technologies facilitate the sending of payment instructions in electronic form directly to and from banks. Optimism about the transition to electronic payment instruments is based on the assessment that the technologies underlying electronic communications systems are improving rapidly, while physical transportation technologies are improving only slowly at the present time.
Economists have long noted that communications industries share certain distinct characteristics that have, in turn, heavily influenced industry behavior. The most salient of these is the prevalence of common costs. The allocation of these costs among diverse users is fundamental to the operation of communications industries. Governments tend to intervene in such industries to allocate these common costs in such a way as to promote access.

In what follows we employ a few technical terms that are necessary for a clear understanding of the economics of communications industries. While these terms are defined as they are introduced, they also appear in a glossary on the facing page.

**Common Costs**

Every communication benefits two parties, the sender and the receiver. How should the costs of a message (a phone call, a letter, an e-mail) be divided between the two beneficiaries? The answer is not entirely obvious. While providers of communication services often collect fees from the sender, services are provided jointly to both parties. The costs of providing these services cannot be uniquely attributed to either of the beneficiaries. We call such costs *common costs*. Common costs also extend beyond the level of the individual message. A large part of the infrastructure costs of a communication system, such as phone lines and information processing resources, are common to all users of the system.

The significance of common costs distinguishes communications from many other industries. For most other goods and services, a large part of the costs of an individual’s consumption can be uniquely attributed to that individual. The time that a dentist, a barber or a mechanic spends serving a customer is a cost of serving that customer exclusively. Costs that can be unambiguously associated with the provision of goods or services to a particular individual are *attributable cost*.

For some costs, specifying whether they are common or attributable is not so simple. Costs that arise from a single message from a sender to a receiver (a single phone call, for instance) are attributable to that pair of users but are common between them. Similarly, the transportation costs of a single shipment of mail between two points are attributable to the group of people sending or receiving letters on that shipment, but are common among the members of that group. In communications industries, there are very few
costs that are attributable to individual users, but there are many costs that are attributable to specific groups of individuals. There are also substantial costs that are common to entire communication systems.

Common costs are often fixed costs; they do not vary with the amount of goods or services produced. An industry that has large fixed costs and relatively small variable costs will exhibit economies of scale (declining average costs) over some range of output levels. When there are costs that are common to the production of multiple goods, then production is said to exhibit economies of scope. Economies of scale and scope are important characteristics of many communications markets because many costs are common among all users of the network.

Another notable feature of communications markets is in the nature of demand for such services. The economic value to an individual of having access to a communication system depends on the individual's own demands for connection to others and on the extent of the network of individuals connected by the system. A consumer will be willing to pay more for a communication service that allows communication with a larger set of correspondents. This relationship between an individual's valuation of a communication service and the extent of the network is referred to by economists as a network effect. Note, however, that a network effect is a consequence of both the interdependence of demand for communication and the existence of common costs. The idea that an individual "belongs to a network" is only meaningful if there are common costs associated with linking people together.

The presence of common costs and network effects makes it difficult to unambiguously specify the cost of serving a particular individual or group. On the one hand, one can ask, "Given that services are already being provided to others, what would it cost to extend service to this particular group?" The answer to this question yields the incremental cost of serving a group of users. This definition excludes costs that are common to the delivery of service to this group and to others. On the other hand, one might ask, "What would it cost to provide services to this group if no one else were being served?" The answer to this question yields the group's stand-alone cost, which includes all common costs. Clearly, when common costs are substantial, incremental cost is much smaller than stand-alone cost.
When there are no network effects, a group's incremental cost is simply the attributable cost of extending service to that group. When there are network effects, the addition of the new group also has the effect of creating benefits for other users. These benefits work to reduce the net cost of adding a new group of users. Hence, we need a more general definition of incremental cost:

The *incremental cost* of extending service to a new group of users is the cost of adding that group to the network, minus the benefits for others created by that group's participation.

Some commentators have interpreted governmental concern for universal access in communications industries as a necessary response to network effects. Many believe reducing prices to some users may enhance efficiency by compensating them for the network benefits they bring to other participants. If the total benefits of an added participant, both to himself and to others, however, are greater than the costs of adding that participant, then a privately operated network will have an incentive to compensate the added participant. This would be the case even in the absence of government intervention. Network effects do not, by themselves, induce market failures.\(^7\)

**Allocating Common Costs**

In pricing a communication service, a provider must decide who should bear the common costs. There are many possibilities. One could recover all such costs from one small group of buyers, or try to spread the burden evenly among all buyers. We can evaluate alternative cost allocations according to two criteria. First, are they consistent with efficient use of the service? Second, could they arise under competitive market conditions?

While there are many dimensions to the efficiency of a communication services market, one essential consideration is that the allocation of costs must provide customers with the right incentives to participate in the network. An individual's participation is economically efficient if the resulting benefits exceed the additional costs incurred. If the prospective customer is charged less than incremental cost, his or her participation could be inefficient, creating benefits smaller than the costs incurred. Hence, a minimal requirement for a cost allocation to be consistent with efficient use of the service is that no individual or group of users should pay less than its incremental cost.

Like prices that are too low, prices that are too high can also interfere with efficient use of the service. In particular, suppose some prices were greater than stand-alone cost. There might then be users willing to pay this cost, but not willing to pay the higher cost imposed by the seller's prices. Such users would be inefficiently excluded by prices that exceed stand-alone cost. Efficiency also requires prices below stand-alone costs.

There is a natural tendency for market forces to produce prices that respect the bounds of incremental and stand-alone costs. If there are no barriers to the entry of new competitors, then the threat of such entry will serve to discipline the pricing and cost allocation practices of incumbent suppliers. Suppose, for example, that a group of customers is collectively paying more than its stand-alone costs. This market segment would be particularly vulnerable to entry by an alternative provider. The threat of such entry will limit the ability of the incumbent to charge more than stand-alone cost.
While providers of communication services often collect fees from the sender, services are provided jointly to both parties. The costs of providing these services cannot be uniquely attributed to either of the beneficiaries. We call such costs common costs.

In pricing a communication service, a provider must decide who should bear the common costs.... One could recover all such costs from one small group of buyers, or try to spread the burden evenly among all buyers.
The threat of competition, which prevents any individual or group from bearing too large a share of common costs, also prevents anyone from bearing too small a share. If a provider is to at least break even on the sale of services and tries to charge some group less than their full incremental cost, then the provider must recover from other users all of the common costs plus the deficit created by undercharging the favored group. Consequently, some set of buyers must pay more than their stand-alone cost. With potential competition, however, this allocation of costs is not sustainable. Potential competition therefore places both an upper and a lower bound on how much a customer or group of customers can bear. The lower bound is the incremental cost of serving those users, while the upper bound is stand-alone cost or the incremental cost of adding those users to a competing network. Note that these bounds are the same as those that guard against inefficient use of a service. In short, competitive pressures prevent inefficiency.

The evaluation of cost allocations on efficiency grounds is complicated by the fact that incremental cost can be difficult to measure. The categorization of costs as attributable and common is not always straightforward. Even more difficult, however, is the identification and measurement of the benefits that one individual’s or group’s participation brings to others. On the other hand, it is relatively easy to determine whether there are significant barriers to competition. If one can guard against such barriers, then market forces will tend to produce cost allocations that respect the bounds of incremental and stand-alone cost.

**Government Intervention**

In the United States and other countries, communications industries have typically been the object of substantial government intervention. Government agencies or government-owned firms have typically provided postal services and, in many countries, telecommunication services. In other cases, such as telecommunications in the United States, provision of services by private enterprises has been subject to substantial price and product regulation.

The structural characteristics of communications industries drive government intervention. There are, however, two distinct views about how these characteristics motivate intervention. These industries are conducive to relatively concentrated markets, which could give sellers the ability to exercise monopoly (or near monopoly) power over prices. One common view is that government intervention in communications industries is motivated by a desire to limit anti-competitive behavior in markets that have natural monopoly characteristics.

An alternative view states that government intervention is motivated by a desire to place the cost allocation problem inherent in the pricing of communication services under political control. In communications industries, government intervention has tended to tilt the allocation of common costs away from those buyers with high attributable costs. This group of buyers often represents individuals in remote, rural locations. For instance, postal rates are independent of the location to which mail is sent, although delivery costs are clearly higher in rural areas. Also, when there are scale economies associated with service to individual buyers, the per-unit attributable costs of serving large commercial and industrial users will be less than those of serving small residential users.
A minimal requirement

for a cost allocation to be consistent with

efficient use of the service

is that no individual or group of users

should pay less than its incremental cost.

When government intervenes to allocate service costs away from some users and toward others, it might appear that the latter are subsidizing the former. Intuitively, we might say that an individual buyer or a market segment is subsidized if it is paying less than its share of production costs. As emphasized earlier, however, common costs make it difficult to unambiguously define the share of total costs borne by an individual or group. Subsidization is less ambiguously defined with reference to incremental costs. That is, an individual or group is subsidized if it pays less than its incremental cost. If the provider must cover all costs while subsidizing a set of buyers, payments received from other buyers must be covering more than 100 percent of the common costs. In this case (and only in this case) we say that some buyers cross-subsidize others. As previously noted, competition or potential competition will limit a seller’s ability to engage in pricing that results in such subsidies.

Government intervention that respects the bounds of incremental and stand-alone costs can be consistent with the efficient provision of services. The history of public sector intervention in communications markets suggests that sometimes the beneficial treatment of groups has gone further, resulting in prices that are below incremental cost. First-class mail service to many hard-to-reach endpoints, for instance, is widely believed to be subsidized. This sort of cross-subsidization, however, is only possible if there are limits on competition. Prices in a market segment in which the seller enjoys a legally protected monopoly are not constrained to be below stand-alone cost. The seller might then be able to raise enough revenues in the protected segment to cover any losses incurred in selling to a subsidized segment.

When cost allocations are subject to political control, through either the regulation of private providers or the public provision of services, allocation choices are often justified in terms of access. Governments have tended to view themselves as guarantors of widespread access to communication systems. This interest in access has sometimes been motivated by the view that the universality of a communication network is an inherently worthy goal. In other instances, the motivation arises from the concern for the consequences of market outcomes for certain high-cost segments of users — rural postal customers, for example. In either case, interest in access may result in cost allocations in which some users subsidize others.
Communications industries, we have argued, are characterized by common costs — costs that cannot be uniquely attributed to any particular user or set of users. Government intervention in such industries is often aimed at altering the allocation of common costs across users. In the name of universal access, such intervention often reduces the portion of common costs borne by some users. Legal barriers to competition aid in cost shifting, but distort the decisions of potential competitors.

The twentieth century experience of two prominent communications industries — telecommunications and postal services — offers valuable insights. In both there are significant common costs and a tendency toward few competitors. Both were subject to significant government intervention that shifted the incidence of common costs and raised barriers to competition, although in recent decades these barriers have come under pressure. Public policy has responded very differently in each industry with divergent results, particularly with regard to technological innovation. The history of these two industries offers revealing lessons for the Federal Reserve's role as a payment system innovator.
Telecommunications

For many decades, the telecommunications industry adhered to the model of protected, regulated monopoly. The prevailing industry structure had its beginnings in the 1920s, when AT&T was allowed to amass a virtual monopoly in phone services and operate free from the threat of competition. In exchange, AT&T made large sunk investments in infrastructure to extend the national network and subjected itself to rate-of-return regulation: that sought to keep charges to any buyers from being “too high.” This deal was supported by AT&T's argument that telephone service was a natural monopoly and that it (AT&T) could provide universal access at lower cost than could a fragmented industry.

For basic local telephone services, buyer-specific fixed costs are significant and variable costs are low. Hence, attributable costs per call tend to fall with the number of calls over a wide range. Large industrial and commercial users' average attributable costs are likely to be lower than those for small business and residential users. The public interest in widespread access has typically promoted price structures that mute these cost differences by shifting common costs away from small users and toward large business users. In addition, cost allocations tended to favor local service at the expense of long distance.

Through a series of moves by market participants and regulators, the structure of the telecommunications industry has evolved from one of an integrated, regulated monopolist to one of more open competition. The Consent Decree of 1982, which settled a Justice Department antitrust case against AT&T, brought competition to long distance markets, while the regional Bell companies retained monopoly positions in local telephony.

Regulated pricing of local service continued to attempt to shift common costs away from high-cost residential and rural users in particular. Such an allocation required higher recovery from large commercial users and contributed to commercial customers' interest in alternatives to the regional Bells' local service, particularly with the proliferation of fax and data services. The long-standing status of local service providers as protected, regulated monopolies was increasingly unsustainable in the changing technological environment. The 1996 Telecommunications Act opened all markets to competition, and explicitly recognized that doing so would put pressure on the industry's ability to provide inexpensive access to such high-cost users as rural hospitals.

The dismantling of barriers to competition in telecommunications has been accompanied by rapid adoption of new technologies. While the by-pass services that hastened the arrival of competition were made possible by technological progress, competition itself has accelerated technological change by encouraging innovation. In the process, the telecommunications industry and its regulators have retreated from the goal of providing access through subsidized cost allocations.
Postal Services

The U.S. government has been involved in postal services since the founding of the nation and has long made universal access the central goal of federal postal policy. In the nineteenth century, the flow of information arising from a universally accessible and affordable postal service was seen as an important factor in the growth of a nation. The U.S. Postal Service's legal monopoly status has been seen as essential to the goal of universal access. With its protected position, the Postal Service can deliver first-class mail to all locations in the United States at a single price. Without this protection, competitors would “skim the cream” by taking low-cost local business, thereby raising the costs of serving the remaining markets. This view suggests cross-subsidized pricing, since prices that are free of subsidies would be immune to cream-skimming.

The Postal Service's legal monopoly on first-class mail appears to have affected other markets in which the monopoly does not apply. In parcel post and package delivery, for instance, private firms are allowed to compete directly with the U.S. Postal Service, although they are significantly constrained in their ability to do so. Critics have claimed that the Postal Service uses funds earned in the protected first-class market to offer new services priced below incremental cost in the more competitive package delivery business. A private, profit-seeking provider might not have an incentive to engage in such pricing of new products; unprofitable entry into a new market is not a compelling goal by itself. As a public entity, however, the Postal Service’s motivations are less well defined. While a public entity is charged with serving the public interest and may generally seek to do so, it is hard to prevent at least some decisions from being motivated by other goals. Entry into a new market, for instance, may enhance the overall size and influence of the organization.

Without the discipline of potential competition, the U.S. Postal Service's incentives to maintain and enhance the cost efficiency of its operations are muted. Some observers have noted the difficulties the Postal Service has experienced in the automation of mail processing. At the same time, potential competitors' incentives to develop innovative products and processes may well be blunted by the Postal Service's ability to subsidize its prices in competitive market segments.

In short, the postal services and telecommunications industries in the United States have followed divergent paths. While the telecommunications industry has placed increasing reliance on markets to provide pricing discipline and incentives to innovate, the U.S. Postal Service has retained a protected monopoly structure that may distort competition and can stifle technological progress. And while in telecommunications the pace of technological innovation has been quite brisk, with the U.S. Postal Service the pace has been relatively slow.
FEDERAL RESERVE
Check Clearing

Check collection and other payment services share many features with network communications industries like telecommunications and postal services. From the earliest years Reserve Banks have enjoyed legal privileges that have aided the Fed's entry into check collection and have made the shifting of common costs in the pursuit of universal access at least possible. Some competitive advantages remain today, most notably the “six-hour monopoly,” which we discuss below. These privileges make it theoretically possible for the Fed to subsidize some check clearing services, in the specific sense that term was defined above. If the Fed were engaged in subsidization, by our definition, the Fed’s presence could detract from economic efficiency. Moreover, as demonstrated by the contrasting cases of telecommunications and postal services, the capacity to subsidize would not bode well for the Fed’s ability to innovate in the public interest. The critical question regarding Fed participation in check collection, then, is whether under barriers to competition some check collection services are in fact subsidized. If so, then the Fed’s participation would not only detract from economic efficiency but could also distort the innovative process.

The Six-Hour Monopoly

The Federal Reserve Banks enjoy certain legal privileges in the check collection business. The most important is the Reserve Banks’ right to present checks to a paying bank until 2:00 p.m. and receive payment the same day; private-sector banks must present by 8:00 a.m. in order to insist on same-day funds. In practice, private-sector banks can and often do present after 8:00 a.m., but only after negotiating a voluntary agreement with the paying bank, presumably offering the paying bank compensation in the form of reciprocity or presentment fees. The Reserve Banks need not obtain prior permission. Thus, the Reserve Banks enjoy a six-hour monopoly on free par presentment for same-day funds. Other advantages also exist but they appear to be of minor significance.

The six-hour monopoly originated shortly after the founding of the Federal Reserve System. The Federal Reserve Act of 1913 authorized the Reserve Banks to offer check collection services to their member banks. An amendment enacted on June 21, 1917, extended this authorization to allow the Reserve Banks to clear checks for all banks. The amendment also prohibited charging presentment fees against Reserve Banks, but this provision only applied to banks that
voluntarily joined the Fed's collection system. The prohibition codified and expanded a stipulation the Federal Reserve Board had imposed earlier by regulatory fiat on member banks. Banks retained the right to charge presentment fees to any other banks presenting by mail, however. Only the Reserve Banks could mail checks to participating banks and demand immediate par settlement.

The Fed's par presentment privilege was by all accounts essential in the subsequent growth of the Reserve Bank check collection system. The ability to present at par to member banks gave the Reserve Banks a cost advantage over competitors. This advantage gave nonmember banks an incentive to join the Fed's collection system to obtain access to low-cost presentment at member banks. The Reserve Banks required that banks joining the system also agree to accept presentment at par. The upshot was that the more banks that joined the Fed collection system, the greater the value of joining.

From its founding in 1913, the Federal Reserve was eager to increase participation in the Reserve Banks' check collection system. For members of the Federal Reserve System, access to the system was a benefit that offset, in part, the cost of stricter Fed reserve requirements, while nonmembers gained the ability to present to participating banks at par. Despite these benefits, the Fed never completely monopolized interbank check collection. For some nonmember banks the income from presentment fees was apparently worth more than the net value of lower-cost
The six-hour monopoly could give the Reserve Banks an advantage over competitors in some market segments.

clearing services available from the Reserve Banks, so these "nonpar banks" continued to charge presentment fees, a practice that persisted for decades.  

The Monetary Control Act (MCA) of 1980 dramatically changed the nature of the Fed's check collection service. The MCA required Reserve Banks to charge fees for their payment services which must, over the long run, cover the direct and indirect costs of providing the services, including imputed costs that would be incurred if the service were provided by a private firm.  

The MCA also imposed uniform reserve requirements on all depository institutions, and granted nonmembers access to Reserve Bank payment services. Prior to the MCA, free check clearing was one of the benefits of membership. Access to Fed services was now divorced from membership and was explicitly priced.

By forcing the Reserve Banks to charge prices that cover actual and imputed costs, the MCA went a long way toward leveling the competitive playing field. The Fed retained presentment privileges nonetheless. Private collecting banks had no practical means of obtaining same-day funds.  

In response to public concerns about the remaining asymmetry, the Board sought public comment in 1988 on a proposal to extend Reserve Bank presentment rights to private-sector banks, allowing them to present until 2:00 p.m. for settlement the same day. Corporations objected to the proposal, however, because it would hamper their ability to manage their accounts within the day.  

The compromise that was finally adopted, effective January 1994, established the current regime in which all banks have the right to same-day settlement for checks presented by 8:00 a.m. The Reserve Banks retained the privilege of presenting until 2:00 p.m. for same-day funds.
The six-hour monopoly could give the Reserve Banks an advantage over competitors in some market segments. It means that the Reserve Banks can collect a given set of checks on better terms than a private provider: for example, by offering a later deposit deadline or better availability (less check float). A private-sector competitor would have to incur additional costs to clear the same checks with the same availability. In some markets, particularly for small and remote depository institutions where transportation time can be significant, this advantage has given the Reserve Banks a dominant market share. Indeed, in some locations only the Fed presents checks. In more geographically concentrated markets — large cities, for example — the six-hour monopoly provides little or no competitive advantage and the market share of the Reserve Banks is correspondingly low.

**The Allocation of Common Costs**

How do the Fed's check collection activities affect the allocation of the common costs? Since implementation of the MCA in the early 1980s, the Reserve Banks' price structure has determined the allocation of common costs. Early on, Reserve Bank pricing under the MCA was relatively uniform, although prices varied according to the destination of the check. At first, prices at various Fed offices depended only on whether the item was bound for a city or a remote location. More recently, the price structure has become increasingly complex with finer geographical differentiation.

The increasing complexity of the Reserve Banks' pricing has been a response to competitive pressures. Initially, alternatives to Fed check clearing were not well established. As private-sector clearing has grown over time, increased price differentiation has lowered margins in market segments in which alternative providers can compete effectively with the Fed. Maintaining full cost recovery then requires higher margins in market segments where customers have relatively few viable alternatives. Such markets are generally those in which the Fed's six-hour monopoly supports a dominant market share — presentment to remote banks. Accordingly, common costs have shifted away from market segments in which the six-hour monopoly yields no significant competitive advantage for the Fed — presentment to city endpoints.

The six-hour monopoly could allow the Fed to set prices below incremental costs so that subsidization results. We previously noted that in industries which have substantial common costs (like communications), competitive pressures constrain the way those costs can
be allocated across market segments; market discipline generally prevents subsidization. Governmental barriers to competition can loosen the constraints of competitive pressure, however, because they allow over-recovery of costs in protected market segments in order to fund prices below incremental costs in other market segments. The six-hour monopoly is exactly this type of barrier to competition. By raising the costs of competitors, this advantage could allow the Reserve Banks to charge more than stand-alone cost in the protected market segment (checks drawn on remote banks) in order to price below incremental cost in contested market segments (checks drawn on city banks). While these prices could further the goal of universal access, they would be detrimental to economic efficiency, since some users would face prices below incremental social cost.

Reserve Bank price setting is constrained by a specific methodology designed to prevent cross-subsidies. Per item fees must be above “floor cost,” which is defined essentially as average (attributable) variable cost. The individual check is not the only relevant increment, however. There are often significant costs that are attributable to a group of checks, but not specifically attributable to individual checks. For example, local transportation costs are attributable to the collection of checks drawn on a particular group of banks, though not to an individual customer or item. The total floor cost for a group of checks is an underestimate of incremental cost if it excludes costs that are attributable to that group of checks but not to any individual item. It is also possible that floor costs overstate incremental costs, since network effects, if they exist, reduce the true incremental cost of serving a market segment.

We need to entertain two alternative hypotheses, therefore, about the Fed’s allocation of common costs. One hypothesis is that the Reserve Banks generally do not set fees below incremental costs or above the stand-alone costs. The other is that in some market segments the Reserve Banks set some fees below incremental costs, and thus set fees above stand-alone costs elsewhere. These two hypotheses have different implications, as we will see, for how we approach questions about the Fed’s role in payment system innovation.
Access

As noted earlier, the Federal Reserve lists payment system accessibility as an important policy goal.26 The usual articulation of this goal speaks of the Fed providing payment services to all depository institutions, particularly “smaller institutions in remote locations that other providers might choose not to serve.”27 Since there is undoubtedly some price at which alternative providers would choose to serve a given location, access to the payment system must be interpreted in terms of the cost of payment system services to small and remote banks. Enhancing access to the payment system must mean lowering the cost to small and remote banks.

Does the Fed lower the costs of check clearing for small and remote banks? We have argued that the Fed’s presence tends to shift common costs toward checks drawn on remote banks. Hence, cost allocation among banks is determined by whether checks drawn on remote banks make up a smaller portion of the checks collected by remote banks than they do of checks collected by city banks. If so, the Fed’s presence tends to favor small and remote banks. Although to our knowledge no formal data is available, anecdotal evidence suggests that the difference, if there is one, is not large. The shift of common costs toward checks drawn on remote banks does not appear to alter appreciably the relative burden imposed on small and remote banks. There are, however, other dimensions of pricing along which the Federal Reserve may still be able to pursue a goal of moderating costs for small and remote banks, although direct quantitative evidence is unavailable.28

While we lack direct evidence on the extent to which the Fed shifts common costs away from small and remote banks, some indirect evidence is available. Last year Federal Reserve Board Vice Chair Alice Rivlin headed a committee that examined the role of the Federal Reserve in the payment system.29 As part of its work, the committee held a series of public forums. Many participants at these forums expressed the widely shared belief that the Fed’s exit from check clearing would raise the cost of check collection to small and remote banks. Thus according to many people intimately involved in the check collection industry, the Fed’s cost allocation does have the effect of enhancing universal access. A reasonable working hypothesis is that the Fed’s presence does shift at least some common costs away from small and remote banks.30
We have argued that the Federal Reserve's involvement in the check collection industry closely parallels government involvement in the telecommunications and postal services industries. Under this view, the Fed promotes universal access by shifting common costs in the presence of legal barriers to competition. Rapid technological change is currently creating new opportunities for innovation in payment services. As a major provider of payment services, the Federal Reserve must determine its appropriate role in pursuing and promoting innovations.

Our reading of the history of communications industries strongly suggests that barriers to competition are fundamentally incompatible with the efficient adoption of new technologies. Barriers weaken the effectiveness of an organization's innovative efforts, and they create opportunities for subsidies that can distort the choices users make with respect to new technologies. For both reasons, truly good innovations may fail to reach the market, while unworthy ones may actually take hold. Without barriers to competition, cross-subsidization would not be sustainable, and so we can have confidence that the innovative process is genuinely beneficial.

How does one resolve the conflict between cross-subsidization and innovation? One approach is to measure incremental costs rigorously in order to prevent subsidization. This approach, in essence, is the Federal Reserve's current practice. Earlier, however, we pointed out that the need to gauge incremental costs and network effects across a wide assortment of user subgroups is likely to make comprehensive measures of incremental costs difficult to obtain. Accounting data alone are not likely to convince a skeptic of the absence of cross-subsidies.

An alternative approach to the conflict between cross-subsidization and innovation as it pertains to Reserve Banks is to remove the conditions that might lead to cross-subsidization. In the absence of special legal privileges, competitive pressures will preclude cross-subsidization, as we defined it earlier. Removing the remaining barriers to competition would clearly demonstrate the Fed's commitment to efficient innovation.

These principles apply to the Fed's current efforts to implement ECP. As with any innovation, the near-term prospects of ECP are uncertain. A recent study by Joanna Stavins (1997), an economist at the Federal Reserve Bank of Boston, attempts to quantify the overall costs and benefits to society of a transition to ECP. One advantage would come from
replacing the resource cost of transporting and processing paper checks with the lower cost of sending electronic messages. On the other hand, some people prefer to get their checks back. Further, under a variety of state laws, certain check writers are either entitled or required to receive their canceled checks. While the estimates reported by Stavins favor ECP, the results are sensitive to reasonable alternative assumptions, particularly with regard to the intrinsic value of canceled checks to consumers. As with other recently proposed payment innovations, such as stored-value ("smart") cards, it is probably too early to tell whether ECP will make society better off or not.

Ideally, innovations would succeed in the marketplace if and only if they were truly beneficial to society. Accordingly, the Fed should introduce ECP in such a way that we can be assured it will succeed if and only if it improves payment system efficiency. In the absence of impediments to competition, a new product or service generally will be profitable if its value to customers, as measured by willingness to pay, exceeds the cost at which providers are willing to supply it. The usual presumption is that innovation in competitive settings yields outcomes that are beneficial to society as a whole. A necessary condition is that prices are not inefficient, that is, they do not embody cross-subsidies. Barriers to competition allow inefficient pricing. One way to ensure that the Fed’s implementation of payment system innovations contributes to payment system efficiency, therefore, is to remove artificial barriers to competition like the six-hour monopoly.

Removing barriers to competition would help avoid some of the potential pitfalls that face a public entity participating in a commercial enterprise. The Reserve Banks’ special legal status as public institutions, as opposed to private, profit-seeking businesses, could inhibit their pursuit of improvements in products and processes. The structure of Federal Reserve decision making could result in unnecessarily high costs of research and development. It is often difficult for large organizations, particularly public institutions, to respond nimbly to new technological opportunities. The difficulties experienced by the U.S. Postal Service in implementing automation illustrate the challenge of innovating at large, public-sector institutions.

An even more worrisome possibility is that an organization that is not fully subject to market discipline could make wasteful investments designed to hold on to market share. Many observers expect electronic payment instruments, such as debit cards, credit cards or smart cards, increasingly to displace checks. In this context, ECP could be viewed as an attempt to stem the expected decline in check use. By reducing the cost of paper checks, ECP could slow the transition to fully electronic payment instruments that are even more beneficial. As long as barriers shield the Fed from competitive pressures there is the potential for the Fed’s pursuit of payment system innovations to conflict with payment system efficiency.

Yet there are good reasons for the Fed to pursue ECP research and development. The Fed, the largest processor of paper checks in the economy, maintains a substantial capital stock dedicated to that activity. The Fed would need to integrate ECP investments into its current check collection infrastructure. As a result, the Fed is likely to have a comparative advantage in evaluating the technical characteristics of ECP investments. In addition, the Reserve Banks have strong incentives to pursue innovations that, if successful, would enhance the value of their existing check infrastructure. To the extent that the Fed’s decision making mimics that of a private business, the interdependence of paper and electronic check collection gives the Fed appropriate incentives regarding ECP research and development.
While many payment system innovations take the form of new payment instruments, electronic check presentment (ECP) is simply a means of bringing modern information technology to bear on the clearing and settlement of a very old payment instrument. The standard method of clearing and settlement begins when the person or firm receiving a check deposits the check in his or her bank. If the check is drawn on a different account at the same bank, the check stays there and is paid with a bookkeeping transfer. Otherwise, the check is physically transported to the bank on which it is drawn (the paying bank). After physical presentment of the check takes place, funds are sent from the paying bank to the collecting bank. Often this process is intermediated by other banks (correspondents), Federal Reserve Banks, or by private contractors (courier services, for example). A check that is not honored for some reason — because of insufficient funds in the check writers’ account, for example — are returned to the bank at which it was initially deposited.

With electronic check presentment, consumers and businesses still conduct transactions using paper checks. At some point in the process of clearing the check, the relevant payment information is transferred into electronic form and then sent on to the paying bank. The check itself may or may not continue on its path to the paying bank. If the check is not sent to the paying bank, it is called check truncation. Although truncation is not a necessary part of ECP, many proponents believe that ECP can make its greatest contribution to payment system efficiency in combination with truncation. Indeed, to the extent that there are savings associated with substituting the flow of electronic information for a paper flow, it would seem to make sense to have paper items truncated as early as possible in the clearing process. On the other hand, the occasional need to inspect the physical check suggests that it might be economical for truncation to occur at a more central point in the process in order to concentrate the storage of paper items.

All Reserve Bank offices currently offer paying banks the option of receiving electronic check presentment. Slightly less than 14 percent of the checks processed by the Fed in 1997 were presented electronically. For about another 9 percent, information was sent electronically to the paying bank, although actual presentment was made with paper checks. Reserve Bank representatives are actively involved in several collaborative efforts with industry representatives aimed at finding ways of increasing the use of ECP.
Implementing ECP

What does all this mean for the implementation of ECP? Because it is uncertain whether ECP will actually contribute to economic efficiency, the Fed should do everything possible to ensure that ECP flourishes only if genuinely warranted. If ECP truly is to enhance economic efficiency, it ought to be possible to offer it in a competitive market at prices that cover costs and attract users voluntarily. Any implicit cross-subsidy could distort outcomes by driving some prices below costs, so that users find ECP attractive even if social costs exceeded benefits. Similarly, a legal privilege that dampens competitive pressures could artificially tilt users through nonprice incentives toward an ECP service offered by the Fed.

Because the paying bank has the right to insist on presentment of the paper check, a key issue for ECP is inducing the paying bank to accept electronic presentment. Stavins' (1997) estimates indicate that while paying banks realize significant cost savings from ECP, check writers incur increased costs and lose the benefits of receiving canceled checks. Although her estimates suggest a small net gain to paying banks and their customers, there will undoubtedly be some instances in which ECP would raise the net cost to paying banks and their customers. If the total benefits of ECP exceed the total cost for payors and payees combined, then it ought to be possible for paying banks and their customers to be compensated by other participants. Such compensation could take the form of fees for checks presented electronically, or alternatively, charges to paying banks that wish to receive paper checks.

Net revenues from ECP services should cover the full incremental cost of ECP if it is to be implemented without subsidization. In the absence of barriers to competition, the Fed could not systematically violate this bound. Theoretically, the six-hour monopoly gives the Fed the capability to subsidize ECP; paying banks could be induced to adopt ECP before it is efficient to do so. Eliminating barriers to competition like the six-hour monopoly would help ensure that ECP will succeed if and only if it is truly beneficial to society.

One frequent suggestion by ECP advocates is that the Federal Reserve alter its check presentment regulations so that paying banks are required to accept presentment in electronic form as well as paper. Paying banks could no longer insist on presentment of the paper check.
This change is consistent with a competitive market approach as long as paying banks who prefer to receive paper presentment are able to compensate collecting banks. If the paying bank's willingness to pay to receive paper exceeds the cost to collecting banks of presenting paper, then the paying bank ought to be able to stay with paper. Otherwise, the paying bank will receive presentment electronically.36

Simply mandating participation by paying banks would short-circuit the competitive discipline imposed by the need to enlist voluntary cooperation. Then an ECP plan that marginally lowers the costs of collecting banks as a whole might succeed, even though it imposes greater additional costs on paying banks and their customers. Such a scheme would not be in society's interests, and yet it might be adopted if acceptance by paying banks of electronic presentment were mandated with no opt-out provision.

**What About Access?**

We have interpreted access in terms of the costs of check collection to small and remote banks. Fed participation in the check collection system is intended, in part, to make the cost to these banks lower than it otherwise would be. This interpretation is consistent with two alternative hypotheses. First, the Fed's priced services could be free of cross-subsidies, and therefore efficient, even though its allocation of common costs might favor small and remote banks. Second, the Fed's pricing could involve cross-subsidies. In order to maintain prices below incremental costs, the Fed would need to rely on market privileges such as the six-hour monopoly.

If the six-hour monopoly is essential to achieving the Fed's access goals, then its continued presence could distort the implementation of ECP or other innovations in check clearing. If current pricing involves cross-subsidies, then the revenue from customers paying more than their stand-alone costs could be used to push ECP prices below incremental cost. If the Fed's current pricing does not involve cross-subsidization, then the six-hour monopoly is not essential to the status quo price structure. In this last case it should be possible for the Fed to implement ECP efficiently without sacrificing universal access.

Which of these two hypotheses is correct? As we noted above, available data cannot discriminate between the two, and the Fed's floor cost methodology may not guarantee the absence of subsidies. Moreover, it will always be difficult to objectively verify the absence of cross-subsidies. As long as cross-subsidies are possible, there will be those who question the Fed's actions, particularly with regard to new product offerings. How can the public be confident that the Fed's innovative efforts in the payment system enhance efficiency? The simplest and most transparent measure would be to eliminate artificial barriers to competition like the six-hour monopoly.
Conclusion

We have drawn lessons for Federal Reserve payment system policy from the history of other communications industries. Government intervention in these industries has been driven largely by the desire to allocate common costs in order to enhance access for some users. We have argued that Federal Reserve Banks’ provision of check collection services fits the same pattern.

Providing access conflicts with technological progress when access is supported by subsidized prices and barriers to competition. In the telecommunication industry rapid innovation was stimulated by deregulation that required a retreat from universal access. The U.S. Postal Service provides a contrasting example in which protected markets were maintained, but at an apparent cost in foregone efficiency-enhancing innovation. The lesson for the Federal Reserve seems clear: a pursuit of access that makes use of cross-subsidization interferes with the efficient implementation of payment system innovations. Subsidies erode market discipline and distort choices among competing technologies.

Let us emphasize that it is not at all clear that the Fed currently subsidizes any segment of the check collection market. Federal Reserve policy explicitly seeks to prevent subsidization and promote payment system efficiency. With its efforts to promote ECP, the Fed seeks to establish itself as a leader in payment system innovation. The Fed is well suited to understand, evaluate, and help implement new technologies in this area. For the Fed to be an effective leader, however, the public must be confident that its choices are in the public interest. Eliminating any remaining competitive advantages would deny the Fed the capacity to subsidize, and thus would enhance the credibility of the Federal Reserve’s commitment to payment system efficiency.

The authors wish to thank Ed Green, Tim Homophrey, Ned Prescott, Marsha Shuler, John Walter, Marvin Goodfriend, Walter Venet, and Al Brunstad for helpful comments and discussions. The authors remain solely responsible for the contents of this article.
Endnotes

4. It is important to distinguish between incremental and marginal costs. Marginal cost is the added cost of the last unit of a good produced. Incremental cost is all of the additional costs that arise from extending a particular set of services to a particular set of users. This may include costs that are fixed with regard to the quantity of services provided, such as the costs of connecting a group of users to an existing network.
5. For alternative discussions of network effects as a source of market failure, see Economic (1996) and Weinberg (1997).
6. For further resources on the history of the U.S. telecommunications industry, see Brock (1986) or Bornholz and Evans (1993).
7. Sidak and Spulber (1996) give a detailed account on the restrictions facing private carriers.
9. The Reserve Banks voluntarily refrain from presenting between noon and 2:00 p.m. in more markets. The six-hour monopoly is not the only legal privilege; privilege enjoyed by the Reserve Banks. For example, private-sector banks do not have as much flexibility as Reserve Banks in choosing where to present checks to paying banks. In addition, the paying bank controls the intraday timing of payment to a private-sector presenting bank, while the Reserve Banks have the right to debit the paying bank's account within a specified time period. Because the other legal privileges appear to be of minor significance relative to the six-hour presentation monopoly, we will focus on the latter, although we have to say will apply equally well to these other privileges. See Board of Governors (1998) and General Accounting Office (1990) for more details.
10. The amendment provided that any bank could make "reasonable charges, to be determined by the Federal Reserve Board, but in no case to exceed 10 cents per $100," but that "no such charges shall be made against the Federal Reserve Banks." An opinion of the U.S. Attorney General established that this latter provision applied only to banks that voluntarily joined the Fed's clearing system. Note that a state-chartered bank did not have to become a member of the Federal Reserve System in order to participate in the Fed's check collection plan.
11. The first Reserve Bank check clearing arrangement, the so-called "voluntary plan" adopted in 1915, required that member banks join the plan accept checks at par from the Reserve Banks. The "compulsory plan" adopted in May 1916 also included the same requirement, but held the Reserve Banks covering the expense of shipping notes or lawful money from the bank to the Reserve Bank in payment. Such expenses were obviously not the only paying bank costs attributable to check collection. Note that because nonmembers had to agree voluntarily to join the Fed clearing plan, the amendment gave the Reserve Banks no real advantage over private banks, since both needed to offer inducements to obtain payee presentation rights. The amendment's effect was to codify the Reserve Bank's right to present to member banks at par, by mail, without prior permission. For discussions of the Fed's entry into check clearing see the classic account of Warren Spulber (1926), or more recently Ed Stevens (1996, 1998) and Alton Gilbert (1998).
12. Note that the effect of the size of the Fed check collection system on the value of joining did not necessarily reflect a network effect. Federal Reserve policy deliberately tied the service of collecting a bank's outgoing checks to that bank's willingness to pay part on its incoming checks. There was no technological link between the number of banks sending checks to the Fed and the number of banks to which the Fed could send checks.
14. The Federal Reserve's cost recovery requirement includes a "private sector adjustment factor" that consists of the taxes, fees, and return on capital applicable to a comparable private-sector provider.
15. The rights of private collecting banks were governed by provisions of the Uniform Commercial Code. For a description, see General Accounting Office (1989), p. 28.
16. In arrangements called "controlled disbursement," banks notify their corporate customers early in the day of the value of the customers' checks presented that day, allowing the customers to fund their accounts by selling money market securities. Later presentation makes such arrangements more difficult because money markets become progressively less liquid in the afternoon. These costly effects effectively shift the prohibition on corporate demand deposits, and are wasteful from society's point of view. Note that corporate objections to extending private presentation time to 2:00 p.m. are not directly relevant to the question of whether private and Reserve Bank presentation times should be equalized; presumably they would also object if asked whether the Reserve Banks should be able to present at 2:00 p.m. The objections might suggest that, without interest on corporate checking accounts, equalization should take place at a time earlier in the day rather than later. See Board of Governors (1991), p. 4747, for discussion of public comments on the 1988 proposal.
17. The Board of Governors has recently requested public comment on the effect of the January 1994 same-day settlement rule. In addition, the Board is considering reducing or eliminating legal disparities between Reserve Banks and private-sector collecting banks in the check collection process, including the six-hour monopoly (Board of Governors 1998).
18. Critics that have charged the Fed with unfairly subsidizing check collection have focused on whether the Fed's cost accounting methodology understates the overall cost of Fed check collection. This question is separate from the question we discuss: cross-subsidization within Fed check processing. The Board of Governors requires that the Reserve Banks annually recover the full cost of check collection services from check collection fees. Our reasonable hypothesis is that the Reserve Banks recover the full costs of check collection in the aggregate.
19. The Monetary Control Act states that prices for Federal Reserve services "shall give due regard to competitive factors and the provision of an adequate level of such services nationwide."
22. For instance, Reserve Banks' prices depend on the amount of sorting done by depositing banks prior to depositing checks with the Fed. Small, remote banks are more likely to make sorted deposits than are large, city banks. The Fed could pursue its interest in access by setting lower price-cost margins for sorted than for sorted deposits, thereby lowering the cost of check collection for small remote banks.
24. The shift of common costs away from small and remote banks might be independent of the six-hour monopoly. Some participants in the Richmond Committee Forums believe that the Federal Reserve accepts a lower rate of return than would be required by commercial providers or that the Fed does not account for the full costs of providing service.
25. We mean profitability in the sense that the expected present discounted value of net cash flows from the introduction of an innovation is positive. The Board of Governors imposes a tighter constraint on Reserve Banks; net cash flows must be positive each year in each priced service line (check collection, automated clearing house, and so on).
26. ECP was implemented quite rapidly in Switzerland under just such a scheme. Paying banks must pay a substantially higher fee to receive paper checks.

References


