WPS1833

POLICY RESEARCH WORKING PAPER 1833

Cost Recovery and Pricing of Payment Services

David B. Humphrey
Robert H. Keppler
Fernando Montes-Negret

The cost of providing payment services is substantial — about 3 percent of GDP. Cost reduction requires the appropriate pricing of those services.

The World Bank
Financial Sector Development Department
October 1997

Summary findings

A modern payment system is essential for promoting domestic and international trade and exchange as well as developing financial markets. Payment users will be directed toward the most efficient payment methods when the costs of producing those services are reflected in the prices paid.

Resources are being wasted in the United States because consumers see no important difference in transaction prices or bank costs between using a check or using electronic direct debit in paying a bill, even though the social costs of these two instruments are different. Electronic payments cost only a third to half as much as paper-based payments. An estimated $100 billion (or 1.5 percent of GDP) is being lost by the continued use of paper-based checks.

When payment instruments are not appropriately priced, the costs must be covered elsewhere. One common solution is to let loan revenues cover part of payment expenses (keeping loan rates higher to compensate). When prices reflect the full cost of producing the service, users demand the services that use the fewest real resources.

Humphrey, Keppler, and Montes-Negret give examples of payment prices and price schedules and show how underlying cost data are used to "build up" to a price. They outline how payment services may best be structured to:

- Appropriately reflect economies of scale or scope in the production of payment services.
- Adjust cost recovery percentages to accommodate how much demand conditions associated with start-up differ from those associated with mature operation. (During a new system's early years of operation, the transaction volume may be low and some form of underrecovery of costs may be required to encourage use of the system. But any such underrecovery must be built into future pricing arrangements once the systems are established and traffic volumes are at a level where full cost recovery is practical. To ensure fairness, the pricing structure must also guarantee that latecomers to the system do not get more favorable treatment than the initial user group.)
- Induce efficient use of scarce resources.

They note the economic principles that recommend certain pricing methods over others and apply equally to payment services provided by the private sector or through a government agency. They show why costs should be recovered through user transaction fees.

This paper — a product of the Financial Sector Development Department — is part of a larger effort in the department to promote the development of financial sector infrastructure to support banking and capital market activities. Copies of the paper are available free from the World Bank, 1818 H Street NW, Washington, DC 20433. Please contact Tomoko Ishibe, room F6K-280, telephone 202-473-8968, fax 202-522-3199, Internet address tishibe@worldbank.org. October 1997. (41 pages)

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent.

Produced by the Policy Research Dissemination Center

Cost Recovery and Pricing of Payment Services:

Theory, Methods, and Experience

David B. Humphrey

Robert Keppler

Fernando Montes-Negret

This paper carries the names of the authors and should be used and cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors' own and should not be attributed to the World Bank, its Executive Board of Directors, or any of its member countries.

Authors' profile

David B. Humphrey *is a professor of finance and Smith Eminent Scholar in Banking in the College of Business, Florida State University.*

Robert H. Keppler *is Principal Payment Systems Specialist, Financial Sector Development Department of the World Bank*

Fernando Montes-Negret *is Principal Economist, Financial Sector Development Department of the World Bank.*

CONTENTS

1. Introduction

A modern payment system is a necessary prerequisite for promoting domestic and international trade and exchange as well as developing financial markets. An efficient payment system, one which minimizes the expense of making payments relative to benefits received, lowers the resource cost of achieving these goals. Payment users will be directed toward the most efficient payment methods when the underlying costs of producing these services are reflected in the prices paid.

Efficiency considerations are increasingly important since the introduction of computer based national payments systems requires significant investment during the design, development, testing and operational phases. At some point during the design and development process, decisions must be made as to the fees to be charged to initial and future users of the new systems. Towards this end Part I of this document will:

- Outline the cost recovery principles and objectives, showing why costs need to be recovered through user transaction fees;
- Note the level and sources of payment costs;
- Indicate how sensitive users are to changes in payment prices;
- Discuss the various methods used to determine prices;
- Illustrate how some countries have chosen to price their payment services; and
- Note the data needed for pricing payment services in emerging market economies

Part II of this document contains a discussion of the practical issues of setting prices for the payment systems provided from the perspective of all payment systems participants including; the central bank as a service provider, commercial banks as users of the central bank provided services and the customers of the commercial banks. The discussion also covers the essentials for developing a two-part pricing mechanism based on the calculation of fixed and variable cost components as a basis for determining unit transaction fees.

PART I

2. Cost recovery: principles and objectives

Should payment costs be recovered? In general, prices for bank and central bank payment services should be set so as to recover the resource costs involved in supplying them. In deciding which payment service to use, payers essentially compare the costs and benefits of each payment instrument, choosing the one with the largest net benefit. If prices are less than costs then net benefits will be overstated and instruments will be overused. This wastes real resources.

Practical cost recovery strategies can be designed and accepted for implementation during the system development life-cycle within the following framework:

- partial cost recovery, where the provider aims to recover only a proportion of its development and operational costs;
- full cost recovery, where the prices are set at a level which are designed to recover the full development and operating costs;
- planned growth cost recovery, where the price is set to recover the full costs and provide funds for the next purchase of capital equipment; and
- profit, where the prices are set to recover total costs, provide for future enhancements and provide a suitable return to the owners of the system for the capital invested.

Careful consideration should be given to pricing policies when the initial services are provided by a government entity such as a central bank. In many cases, the central bank will be the only viable initial provider of services. However, the pricing policy adopted should not inhibit the introduction of competitive services by alternate suppliers at a later date or prevent the central bank from replacing its ownership with the eventual ownership of the users. It will be difficult, if not impossible, for private suppliers to compete with services that are priced at a level that are less than the cost of production or are explicitly subsidized, for example, by the non allocation of taxes that would be incurred by a private supplier.

A specific cost recovery issue arises during the early years of operation of a new system in a developing economy. The initial transaction volume may be low. It is likely that some form of under recovery of costs will be required to encourage use of the system. However, any such under recovery will need to be built into future pricing arrangements once the systems are established and traffic volumes are at a level where full cost recovery is practicable.

An additional problem must also be addressed at the outset and concerns the entry costs to be levied on future participants that, for example, join the system after the initial design and development costs have been recovered. To ensure fairness, the overall pricing structure must guarantee that new users do not get more favorable treatment than the initial user group.

Resources are being wasted today in the U.S. because consumers see no important difference in per transaction prices or bank costs between using a check or ACH direct debit in paying a bill, even though the social cost of these two instruments are quite different (Table A1).[1] The same applies to consumer choice between check, credit card, debit card, or cash use at the point of sale. Each business payer, however, initiates a much higher volume of payments than does each consumer. As a result, business payers typically face per transaction fees from banks which may be paid for directly or by holding a compensating deposit balance which varies with the value of bank payment services used.

In effect, banks typically charge consumers an average fee (minimum balance) which recovers the bank expense of payment services for the average consumer. Individual consumers may initiate many or few payments each month and still hold the same balance, so their marginal cost of an additional payment is zero. Although businesses do experience positive marginal costs from a bank, and are thus more sensitive to differences in payment instrument expenses, they initiate only 41% of U.S. non-cash payments. Consumers as a group initiate 56% of payments but do not generally face either a per transaction fee nor payment prices that reflect the lower costs of initiating an electronic payment. As an upper bound estimate, $100 billion or 1.5% of GDP is being lost by the continued use of paper-based checks rather than electronic payments.[2] This loss is projected to be reduced by 3% (the government's share of check payments) by 1999 since new legislation mandates that all federal government payments be made electronically by this date. State governments are also mandating that more payments be made electronically.

Fairness versus cross-subsidization. An additional argument for recovering payment costs in the prices paid by users concerns equity or fairness. When payment instruments are not appropriately priced, their costs will be covered elsewhere (even if a bank is not sure where exactly this is). This involves cross-subsidization, where some other party will bear the costs but not the benefits. A common type of cross-subsidization has been to use loan revenues to cover a portion of payment expenses. In this instance, borrowers end up paying a higher loan rate, and depositors a lower payment price, than otherwise would occur if both banking services were properly priced. But unfairness is not the only effect; cross-subsidization distorts decisions made by borrowers and depositors. Borrowers will respond by borrowing less, reducing domestic spending and investment, while depositors will respond by overusing payment instruments or using expensive instruments, wasting resources.[3]

[1] *To cover payment costs, U.S. banks give consumers a choice of paying a per transaction fee or holding a minimum balance. As most choose a minimum balance, the extra expense to consumers of an additional transaction for different payment instruments is zero.*

[2] *The difference between the social cost of paper-based and electronic payments in Table A1 is $1.62 (= $2.93 - $1.31). Times the 62 billion checks currently written, this gives $100 billion in resource costs that potentially could be saved if all payors switched from checks to electronic payments. During the transition from paper to electronic payments it is necessary to run a dual system. As well, the average cost of checks (electronic payments) estimated here would rise (fall) since, with scale economies, volume would be falling (rising).*

[3] *Although resources are still wasted, banks can recoup their higher payment costs associated with payment service mispricing by paying depositors a lower interest rate.*

Loss of seigniorage revenues. Many countries have sought to expand the use of non-cash payment instruments, both to facilitate the emergence and growth of financial markets and to improve the ability of firms to engage in trade and exchange. There is a hidden cost to this effort, albeit one that governments seem willing to incur. This cost is the loss of seigniorage revenues from the issuing and use of cash in domestic transactions. If the price of non-cash payment instruments do not reflect their full cost, then the loss in seigniorage revenues will be larger than otherwise would occur.

Reasons why payment costs may not be recovered. Some valid reasons why payment costs may not be fully priced to users, at least initially, concern the desire to realize internal scale economies and/or network externalities. Non-cash instruments are often initially underpriced to encourage use and generate volumes sufficient to achieve lower unit costs through scale economies. While this is valid initially, such underpricing should be eliminated at a later date once the desired scale economies have been achieved and unit costs have been sufficiently reduced.

A similar argument can be made, again only initially, when the benefits of participating in a payment network expand as more and more users are attracted to it. For example, a wire transfer network is more valuable as more banks belong to it, since all existing users now can send and receive payments to more endpoints. Indeed, the use and acceptability of a particular payment instrument expands as the number of network participants grows. This has clearly been the case for both paper-based and electronic networks. Point-of-sale and bill payment payees, in particular, are very reluctant to make investments necessary to be able to accept a new payment instrument unless they believe that the volume they will receive is sufficiently large to justify the expense. Correspondingly, payers are reluctant to use a new instrument until it is accepted by a sufficient number of payees. Banks are caught in the middle of this conflict. While underpricing a new payment instrument can occur in order to expand benefits from network participation, there comes a point where the added benefits are small and the costs of providing payments over the network should be fully recouped in their price.

Although valid reasons for not ever properly pricing a service or output do exist, they have little application to payment instruments or services. For example, if the main benefits of using a particular payment instrument were not captured by payment originators or receivers but instead benefited some other group, then charging payment users would merely "tax" users and subsidize this other group. This is an example of the so-called free rider problem. Having payment users bear the full cost of an instrument for which they obtain little of the benefits would result in under-utilization of the instrument. Fortunately, payment users–both payers and payees–are the main beneficiaries in a payment transaction and this argument for not pricing payment instrument use does not apply.

Another argument made for not pricing non-cash instruments is that substitute cash payments are not explicitly priced. If cash payments are not priced, while non-cash transactions are, then payers will presumably have an incentive to rely on cash.[4] Although this can boost seigniorage revenues, the argument is invalid because there are considerable implicit costs to using cash in precisely those situations where non-cash payments can be most beneficial. This is where large value payments are made among firms, both domestically and internationally. While cash use is often favored for retail payments when the environment is safe (as in Japan, less so in Europe, but not the U.S.), large value cash transactions have a high opportunity cost and are just too difficult to handle to be the instrument of choice for business payments even when non-cash payments are directly priced. This is particularly true when the parties to a transaction are distant from one another, as in business transactions and consumer bill payments. Overall, the only valid reason for not pricing payment instrument use involves only a temporary suspension of pricing, until scale and network economies can be largely realized. Once realized, payment costs should be reflected in the prices faced by users.[5]

3. The cost of making a payment

Payment expenses are incurred by all parties to a payment transaction. This includes payers (individuals and businesses) who initiate payments and payees who receive them. For non-cash payments, it also includes the financial institutions (hereafter banks) acting as agents for payers and payees who process the payment information and actually transfer the funds. If more than one bank is involved, as when the payer and payee do not have the same payment agent, then inter-bank settlement of the payment is required. This usually involves the central bank.

While most payment expenses can be directly measured, very little information is publicly available. This is especially true for payer and payee expenses, although less so for bank processing costs and central bank settlement expenses.

Two country examples for the US and Norway illustrate (see Annex 1) the cost per transaction of check and electronic payments, demonstrating the much lower cost of electronic (ACH) payments in both countries.

Changes in payment costs: economies of scale and scope. The larger is the share of fixed expenses in total payment cost, the greater will be the economies of scale. Payment scale economies arise because fixed expenses can be spread over a larger volume of payments, lowering

[4] While consumers incur an opportunity cost of holding cash equal to one plus the rate of inflation times the interest rate, unless inflation is especially large most consumers do not actively respond to this "price".

[5] A seemingly negative aspect to pricing payment use according to its cost is that low income groups may decide not to use non-cash payment instruments because they are "too expensive". However, any effort to lower the price below cost to specific user groups will involve cross-subsidization and lead to inefficiencies in overall use, wasting resources in the process.

the average or unit cost of a payment. Up to some point, the larger is payment volume, the lower is the average cost of processing each payment.

Similarly, if the extra or marginal cost of processing one type of payment instrument is reduced as the volume of another payment instrument rises, then there will be economies of scope. When credit card payments over a network use the same terminals, communication links, or computers as debit card payments, the marginal cost of a credit card payment can fall as debit card volume rises (or vice versa). Similar economies of scope arise when ATM terminals use the same communication links as debit card transactions (as is more common). Scope economies exist when the total cost of processing two or more payment instruments together (jointly) is lower than the sum of the costs of processing the same volume of each instrument separately.[6] (For a discussion of some empirical evidence of scope and scale economies see Annex 2).

4. Demand for payment services

The demand for a payment instrument is influenced by economic variables such as own price, the price and availability of substitutes, and user income. It is also influenced by such difficult to quantify influences as user convenience, acceptability by payees, and safety in use. For this and other reasons, the effect of price on the demand for payment instruments has been difficult to measure, although inferences can be made.

Substitution among payment instruments. Due to a lack of extended time-series data within any one country, it has been necessary to infer the demand for and substitution among non-cash payment instruments from pooled time-series cross-country analysis using data developed by the Bank for International Settlements. The relationships among the per person use of five non-cash payment instruments plus cash is illustrated in Table 1. Of the 15 pair-wise correlation coefficients, 11 are negative suggesting substitution.[7] However, these substitution effects are weak since the largest negative correlation coefficient (or r) is -.50.[8]

[6] *Scope economies can have two sources. First, if excess capacity exists, the fixed costs associated with processing one type of payment instrument can be spread over processing other instruments as well. For example, if buildings and large computers have unused productive capacity, other types of payment instruments may be processed using these same facilities at low additional expense. Second, joint production costs can be lower when certain variable expenses can also be shared, as when credit and debit card payments use the same card terminals, communication links, or personnel.*

[7] *The correlations relate the annual number of transactions per person across 14 developed countries in each year over 1987-93. Cash use reflects the real value of cash holdings per person, also across countries and for each year.*

[8] *The strongest relationship is between checks and credit cards which, with an r of .81, suggests strong complementarity. When the U.S. (which uses the most checks and credit cards) is excluded, the complementarity between checks and credit cards is reduced but not eliminated (r = .52). The complementarity of debit cards with both paper and electronic giro transactions reflects their positive cross-country and time-series association in various European countries while the complementarity between electronic giro use and cash reflects a positive association in Europe and Japan.*

The simple elasticity between the annual number of non-cash payments per person and real cash holdings per person is -.68, so a 10% reduction in cash holdings is associated with a 6.8% rise in non-cash transactions. A deeper analysis of these cross-country payment effects within a standard demand function framework suggested that while cash and non-cash payment instrument use are negatively related, the implied substitution between them is due more to differences in use across countries than it is to changes in use over the 1987-93 time period, a result that also applies to the substitution among the five non-cash instruments as well.

Table 1: Correlation matrix for per person use of payment instruments
(14 developed countries. 1987-93)

	Paper Giro	*Electronic Giro*	*Credit card*	*Debit card*	*Cash*
	(1)	*(2)*	*(3)*	*(4)*	*(5)*
Check	-.37	-.39	.81	-.11	-.42
Paper Giro		-.09	-.23	.30	-.14
Electronic Giro			-.50	.12	.19
Credit card				-.17	-.38
Debit card					-.39

Source: Computed from time-series, cross-section data from Bank for International Settlements (May 1989, December 1993, July 1994, December 1994, and May 1995)

Payment price elasticities. Own prices, which are small in magnitude, vary little over time, and may not depend on incremental use, seemingly have exerted little influence on the choice or use of payment methods. Mean own price elasticities for paper giro, electronic giro, and credit card use ranged from -.09 to -.26 and, while significant, appear to be quite inelastic. Price elasticities for check and debit card use were slightly positive but insignificant. In contrast, the influence on payment instrument use from cultural and institutional factors, such as crime rates, bank concentration, and the availability of alternative payment methods, has been strong (Humphrey, Pulley, and Vesala, 1996).

Payment instrument price information is more suitable in Scandinavia as countries there are among those that actually charge all users a price per transaction for different non-cash payments and have also been increasing the relative price of paper-based instruments to reflect better the lower supply costs of most electronic payment alternatives. Based on survey information of bank prices for check and point-of-sale electronic payments in Norway between 1989 and 1993, the implied elasticity of substitution between paper and electronic payments is $s = 4.13$ (computed from Bank for International Settlements, May 1995). This implies that payment users are in fact very sensitive to relative payment prices since a 10% rise in the relative user price of checks seems to be associated with a 41% reduction in their relative use. Although additional (and verifying) information is not available, the experience of Norway would likely be duplicated elsewhere as more countries institute explicit transaction pricing of their payment services and had them reflect the lower unit cost of electronic payments.

5. Methods for pricing payment services

Indirect pricing methods: float and minimum balances. Payment float is a natural consequence of debit transfer paper-based payments (e.g., checks) since these instruments typically require some form of physical transport or delivery. Paper-based credit transfers (e.g., paper giro) also generate float if these transactions are not value dated (where the transfer request is submitted prior to the date the transfer is to occur). Electronic debit or credit transfer payments, of course, need not incur float at all since processing is more rapid and physical transport is not needed. One indirect method used by banks to recoup payment costs has been to create even more float by debiting payers earlier than needed and/or delaying the crediting of payees. Revenues earned on the float created are used to cover bank payment expenses, rather than directly assessing a fee on users.[9] A second indirect method banks use to cover payment expenses concerns a required minimum deposit balance. Here revenues are earned from the spread between interest paid (if any) on minimum balances and the interest received when these balances finance loans or other assets.

One reason indirect (or non-price) methods are used for recovering payment expenses is that the alternative of paying a market rate of interest on deposits and charging a price for payment use generates a tax liability for depositors, as interest income is often taxable. A related reason, which applies to both float and minimum balances, is that the true cost of payment services to users–especially to consumer users–is less obvious than would be a direct fee per transaction. This, of course, is precisely the problem from a resource allocation and payment efficiency standpoint.

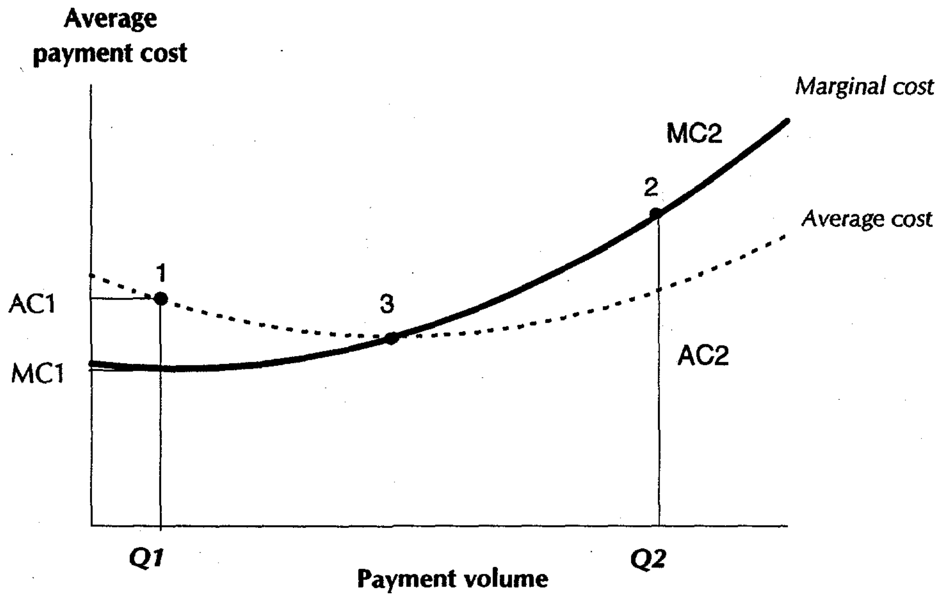
Neither float nor minimum balance methods of covering payment costs will have much effect in inducing payment users to use the lowest cost payment instrument. This is because users will not "see" how the cost of different instruments vary and will have no incentive to choose the instrument that meets their needs while having the lowest resource cost. To do this, some sort of direct fee or explicit price is needed. In what follows, a number different procedures for pricing payment services are outlined, along with their benefits and problems in implementation.

[9] *Some emerging market economies generate even more payment float. This occurs when banks are required to hold reserves with the central bank sufficient to cover the sum of each day's gross debits, which are posted prior to that day's gross credits, to clear interbank payments. In Russia, it has been estimated that close to one-third of the expansion of the money supply was tied up in this manner at various offices of the central bank. Such a settlement arrangement was a holdover from a planned economy where the time value of money was zero, as payments were merely made from one government enterprise to another. With improvements in payment system efficiency, these balances will fall, effectively expanding the useful money supply (as opposed to the measured money supply, which includes payment float).*

Marginal cost pricing. Marginal cost pricing is were the extra cost of producing an additional payment transaction is fully reflected in its price. This enables payers to demand payment services up to the point where the extra benefits equal the extra resource costs. However, two practical considerations intervene. First, it is difficult to accurately measure marginal cost in practice. Second, since scale economies exist for payment services (and scale diseconomies are rare), marginal cost pricing would consistently under-recover the full costs of production.

The cost recovery problem is illustrated in Figure 1. Scale economies exist at payment volume Q1. Setting price (P) equal to marginal cost at MC1 would thus give P = MC1 < AC1, so the difference between average and marginal cost, times payment volume, would be the value of costs not recovered (i.e., unrecovered costs = (AC1 - MC1)*Q1). Over-recovery of costs would occur in the rare event of scale diseconomies (rising average cost) since, at volume Q2, P = MC2 > AC2. Only in the special case of no scale economies or diseconomies, where marginal and average costs are equal (so price also equals average cost at point 3), would cost recovery not be a problem.[10]

Figure 1: Scale economies and pricing



[10] *It may seem that a natural monopoly exists in supplying payment services if scale economies at a single processing facility would not be exhausted even if it supplied the entire market. However, unless a country is very small, the extra transportation expense (and delayed availability of funds) associated with transporting paper-based payments to and from a single processing facility will typically offset the savings from scale economies at a single centralized facility. A different result applies to electronic payments today since communication costs to and from a single facility have been falling while scale economies remain unchanged. Thus electronic payments are close to being a natural monopoly while paper payments are not (unless the country is small).*

Optimal departure from marginal cost pricing. With scale economies in payment operations, a departure from marginal cost pricing that comes close to delivering the same information for efficient resource allocation is to relate the degree to which price exceeds marginal cost to the size of the own price elasticity of demand for payments (h). Here price is determined from (P - MC)/P = - 1/h.[11] When demand is inelastic,[12] price can exceed marginal cost and recover a larger portion of total costs (or even over-recover total costs), but not have much effect on reducing quantity demanded. This preserves the cost reducing effects from scale economies and supplies revenues to cross-subsidize other services.

Many banks implement pricing in this manner, using their best guess as to the price elasticity of demand (equivalent to a judgment regarding the likely sensitivity of payment users to changes in prices). The U.S. central bank, an institution that processes a relatively large volume of retail payments in competition with banks, prices some of its payment services using such judgments and with the restriction (by law) that all costs are recovered, even those that would have been earned as profits by a private sector supplier.

To recover all payment costs, however, requires some cross-subsidization within or among payment services. Those separately priced services with scale economies and elastic demands will under-recover their expenses, requiring other services with less or no scale economies to over-recover their costs in order for there to be an overall cost-revenue match for payment services. The need to over-recover in some areas, however, provides an opportunity for "cream skimming" on the part of other payment suppliers. Competitors who merely price according to their average cost will both recoup all their expenses and have a competitive advantage over the supplier who cross-subsidizes. As payment volume shifts to these lower price competitors, the ability of the cross-subsidizing supplier to obtain an overall cost-revenue match is impaired. Ultimately, this can force the cross-subsidizing supplier to lower or eliminate the cross-subsidy or see its market share reduced.

The difficulty in accurately determining payment demand elasticities prior to implementing a pricing structure, the need for cross-subsidization when scale economies exist (as is common in many payment services), and the possibility of cream-skimming by competitors are the three main problems with this pricing method. These problems do not arise with two-part average cost pricing.

Average cost, two-part, and benefit-flow pricing. Average cost pricing is easy to implement: average costs are not difficult to determine and all costs can be recovered without

[11] h is measured as the percent change in the volume of payment service demanded divided by the percent change in its price. h is negative because quantity demanded tends to fall as price rises, so -1/h is a positive value. The equation shown is also known as the inverse elasticity rule or market sensitive pricing.

[12] Demand is inelastic when h < 1.0 in absolute value, so - 1/h is larger the more inelastic is demand (recall that the value of h is negative, so -1/h becomes positive).

cross-subsidization. If adopted, average cost pricing is most appropriate for consumer payers since their payment volume is relatively small but the number of users is large (making a simple price structure a virtue).

When payment volumes are large but the number of users is small, as is the case for business, the average cost of a payment service is best divided up into two separately priced components. In ***two-part pricing***, one price reflects the average fixed cost of providing the payment service while a second price reflects the average variable cost. Two-part pricing, by distinguishing costs and prices that may fall with volume from those that do not, is the most accurate way for costs to be reflected in prices. This is equivalent to having a fixed charge (per account serviced, per batch of payments from a single payer, etc.) plus a volume-related fee (reflecting average variable costs) for each payment processed. The net effect is that a different "price" will apply to each payer who initiates a different volume of payments, a "price" that will reflect the scale economies realized in production.[13] In this way, the lower costs associated with high volume use are appropriately passed on to users with high volumes (and vice versa for low volume users) so that cross-subsidization among different volume users is minimized. In such an environment, prices are said to be sustainable and opportunities for cream skimming on the part of other payment suppliers is minimized.

In summary, two-part pricing is used to discriminate among payment service users according to the volume of services they demand and thereby accurately matches the cost of processing their different payment volumes with the price they are charged. As volume of use is the most important distinguishing characteristic among payment users when scale economies exist, two-part pricing is usually all that is needed for prices to be sustainable (Weinberg, 1994).

To the extent that other important characteristics among users can be identified and measured, such as demand elasticity or a splitting of the benefit of making a payment between payer and payee, two-part prices can be adjusted to assure sustainability. The latter case has led to **benefit-flow pricing** where two-part prices are split between payer and payee according to a judgment of how much each party gains from a transaction. This not only is equitable, but reduces a free-rider problem if a party who clearly benefits from a transaction does not bear any costs (and so has an incentive to overuse payment resources). As is described below, benefit-flow pricing is used by the U.S. central bank to price its wire transfer and ACH payment services.

Other pricing methods. Electric utilities use ***peak-load*** or ***time-of-day pricing*** to cover the additional investment needed to build excess generating capacity used only when demand is at its peak. This pricing method allocates more fixed cost to times of peak demand, so price is higher at the peak but lower when demand falls and fewer and less expensive generators can supply all the

[13] *In effect, the fixed fee is spread over a number of individual payments and, when combined with the (likely constant) average variable cost per payment, results in an overall "price" per payment that falls as payment volume rises.*

needed output. Peak-load pricing tends to stabilize the fluctuation in demand over the day and eliminates the cross-subsidy which would otherwise occur if a single price were charged regardless of the level of demand. Peak-load pricing could be used in situations were the volume of payment instrument use varies considerably over time (over 24 hours or day of the month). If substantial excess payment processing capacity has to be maintained, for example, to handle monthly bill payments over and above the capacity needed for daily point-of-sale transactions, then a peak-load pricing approach could be used to more fairly apportion the costs between these two types of transactions.

An alternative pricing method, one that is not recommended, is ***par value pricing.*** Here price is tied to the value of the payment being made rather than to the actual resource cost of making the transaction. The only time par value pricing is justified is when the risk of monetary loss associated with making a payment is proportional to the value of the payment made. While this can occur if a payment supplier is providing (daylight or overnight) credit

to a payer in addition to processing a payment, this usually occurs only on some large value payment networks (e.g., Fedwire in the U.S.). The more usual case is to cover this risk of loss through alternative means (collateral on CHIPS and CHAPS, intraday borrowings on BOJ-NET) while still charging a per transaction fee based on the processing cost incurred.

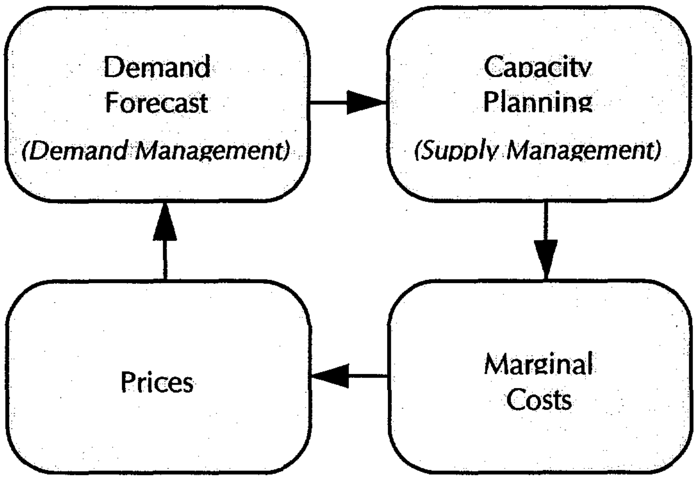
On small value retail payment networks, the risk of loss is correspondingly lower. Here legislation, case law, and payment rule-making have clearly set out the rights and liabilities of payers, payees, and their agents in a payment transaction. These have usually been sufficient to minimize losses from settlement risk for retail payments so no value-related fee is warranted.

6. From marginal costs to prices

The regulation of the supply and demand for payment services implies that processing and operating centers and telecommunication networks are constructed and located in optimal locations to meet the changing demand. Like it happens with other public utilities (particularly power generation and distribution), the demand for payment services will in part be shaped by the prices charged to users, while such prices should reflect the costs of providing the services. The system forms a consistent whole, ideally a self regulating loop.

As indicated above, each component of the services offered should ideally be sold at a price reflecting its marginal cost. However, due to difficulties in predicting accurately the need for processing capacity and customer demand an optimum system could be achieved through the posting of prices for alternative scenarios, with the aim of equalizing the prices charged to the relevant marginal costs. One particularly important very short-term application includes the intraday pricing of payment services. A pricing strategy that influences user behavior to evenly spread usage of the system throughout the operating day and thus reduce the required peak hour capacity of the system with its associated lower capital requirement may be necessary and should be recognized from the outset.

Figure 2: Regulating supply and demand for payment services



Part II of the document deals with the more practical question of moving from costs to prices, providing a discussion of the informational requirements, implementation problems, and some experiences and examples in pricing payment services.

PART II

7. Information needed to price payments and problems in implementation

Information needed. Cost and payment volume information are needed to properly price payment services. Importantly, prices should generally be set to recover all costs since this results in proper use of payment services by consumers, businesses, and banks. One of the best pricing methodologies is ***two-part pricing*** as this pricing structure will (a) reflect the likely scale economies in payment processing and (b) pass the benefits of high volume operation on to those high volume users who generate the economies of scale.

The implementation of two-part pricing is illustrated in Table A3, which follows procedures on how the U.S. central bank develops its payment pricing structure. The first step is to obtain estimates or projections of total variable cost (TVC_i) and total fixed cost (TFC_i) by each of the i payment service categories to be offered. TVC_i reflects the direct costs (wages, fringe benefits, supplies, transportation, etc.) incurred in producing the ith payment service. TFC_i represents (1) the annual cost of leased/rented facilities or equipment plus (2) the depreciated value of any whollyowned capital equipment or facilities (data processing equipment, computers, furniture, and buildings) plus (3) any rise in replacement over historical cost of the owned equipment or facilities (approximated by the inflation rate times the total value of the purchased capital goods).

When one payment service jointly uses or shares the same facilities or equipment with another service, the costs have to be properly apportioned between them. A standard approach to dividing up jointly allocated costs is to apportion the shared costs according to the percent of floorspace used for each service when dealing with occupancy expenses or the percent of time each service uses a shared computer for equipment expenses.

The second step in determining two-part prices is to estimate the volume of payments to be processed for the i^{th} payment service ($V_{i,processed}$), the number of files submitted to be processed as batched payments ($V_{i,file}$), and the number of payment accounts serviced in real time ($V_{i,accounts}$). From this, it is possible to approximate the two-part prices for smaller value retail payments which are normally processed in batch mode: $P_i = TVC_i/V_{i,processed}$ for the price which is to cover variable expenses and $P_i = TFC_i/V_{i,file}$ for the price which covers fixed expenses. Similarly, the approximate two-part prices for larger value business and financial market payments that are usually made in real-time (or at least on an expedited basis) would be: $P_j = TVC_j/V_{j,processed}$ for the price covering variable expenses and $P_j = TFC_j/V_{j,account}$ for the price covering fixed expenses. In effect, this is how the two-part prices shown in Table 4 were derived, except that TVC and TFC also included a reasonable return on invested capital (or equity) and taxes which would have been paid on profits.

Problems in implementation. The main difficulty will be in obtaining accurate payment cost and volume data. Initially, this will involve some educated guesses. However, once a proper cost accounting system is in place and experience is gained in projecting payment volume growth, the process will stabilize and more accurate estimates of costs and volumes–and hence prices–will be obtained. For operational management purposes, it is recommended that procedures be developed and implemented that will measure and record payment volumes and values processed, capacity utilization, and the extent to which quality problems are encountered (e.g., delayed payments, payments to the wrong account, and other payment errors).

8. Cost management and control

A justifiable pricing structure based on acceptable cost recovery principles requires that all relevant capital/set-up costs and operating costs are known and are managed on a rigorous basis throughout the system life-cycle.

The capital/set-up costs, which typically may be depreciated over a five year period are those costs incurred by the service provider prior to full system operation. These costs can be significant and should be carefully managed and controlled. Such costs will include:

- Site Construction: facilities, security equipment, telecommunication equipment (including taxes and import duties)

- Computer Equipment: including taxes and import duties
- Application System(s): design, development, and testing
- Management and other staff costs: directly attributable to system set-up.

Annual operating costs will include all operating, maintenance, customer support and administrative and other management costs incurred in providing the services on an on-going basis. Such costs include:

- Lease of space
- Equipment (rent and maintenance)
- Depreciation
- Staff costs
- Administrative costs (direct and indirect)
- Management costs (direct and indirect)
- Corporate and other taxes

Direct administrative and management costs are those that are directly within the control of the service provider organization. Indirect costs, are those associated with services provided, say by the parent central bank organization, such as internal auditing and accounting, when the service provider is the central bank. In all cases, a well designed budgeting, cost accounting and management information system will be required to both manage costs and provide the basic data for building the service pricing structure.

The specific treatment of capital investment depreciation will depend on the prevailing accounting rules and are not discussed in this paper. As a general principle, capital investments should be depreciated over the useful life of the asset. Useful life is the period over which the asset is to be effectively used by the service provider and may therefore be shorter than its physical life. Typically, the depreciation period for computers and other electronic equipment is usually between five and ten years. Factors which must be considered in determining useful life include obsolescence arising from technological changes or improvements in production techniques. If the equipment becomes obsolete due to technological changes during the period determined to be the useful life at the beginning of a period, the book value of the equipment should be reduced and the depreciation accelerated over the remainder of the current assessment of its useful life. This concept is particularly important when considering the pace of change in technology and its impact on determining realistic payment system costs.

After the initial capitalization, the payments systems will need to be continually updated and enhanced. Any additional capital expenditure should be treated in a similar manner to that described above.

The distinction between direct and indirect operating costs mentioned above is important as direct costs can usually be controlled by individuals within the payments systems area, while *indirect* costs are managed and controlled outside the immediate influence of the area. Responsibility for controlling direct costs, both fixed and variable as discussed in Part I of this paper, should be individual accountabilities assigned to nominated operating managers. They should be required to manage costs within budget and challenged over time to reduce the cost per item processed to ensure that user fees are optimized.

Key features of an acceptable pricing/price structure. Discussion with commercial bank users of payment systems in several countries strongly indicates that one of the most important features of any charging system is that the final price structure should be simple. Commercial banks and other users should be able to determine the impact of the charges on their profits and should be able to monitor costs as they arise during any period. The prices should therefore be established prior to the beginning of appropriately defined calendar periods and should remain in force throughout the period. Any under- or over-recovery of costs may, if full recovery is the objective, be used in determining the prices for the next period.

In a typical situation in a developing economy more than one payment service is planned for introduction. For example, the on-going World Bank supported reform initiatives in China, Viet Nam and Mauritius will result in the introduction of discrete computer based systems to process both large value/time critical transactions using a gross settlement method as well as batched files of low value payments using a net settlement method. Situations also arise in which the planned data communications services will support both payment instruction processing as well as other types of transactions, such as, securities trading instructions and credit card authorization requests. Although several aspects of the services are quite discrete, they will utilize specific infrastructural items on a shared cost basis. Each of the services have different commercial features, and in some cases will be used by different customers. To encourage the customers to use each of the services in a cost effective manner, a system of discretionary pricing may be required. Different features have different associated costs and therefore should be priced differently. The primary objective is to establish an overall price structure that is fair to all participants. The major components are discussed in the following paragraphs.

- ***Entry costs:*** The most common form of discretionary pricing is the use of an entry cost or initial membership fee. In systems which are owned and operated by the central bank the entry cost is typically by way of a fixed fee, while where the systems are owned and operated by the participants entry fees are frequently based on a percentage of equity determined using expected transaction volumes. The entry cost can also be an effective mechanism for passing some of the costs of development to the users and, where appropriate, may also be used to discourage low volume users from having direct access to the service. Entry pricing is usually used in addition to per item charges.

- ***Service level costs:*** In the example of China mentioned above, three primary payment systems – bulk paper; bulk electronic, and high value items – will be made available by the People's Bank of China (PBOC) for use by the commercial banks. The technological design anticipates using the same data communications network fōr the latter two systems. However, the base product functionality of the two electronic systems differs substantially according to the required response times, security and other features. These differences should be reflected in the pricing structure in line with the costs of providing these different features. In addition to the different basic features of each system it is also clear that the two systems have differing levels of service, for instance the high value system allows for direct electronic interface, voice interface and batched electronic or paper based interface via the branches of the PBOC. Each of these different features should be separately priced and perhaps added to the unit price of the basic service.

- ***Incentives:*** Various forms of incentive pricing schemes can be used to influence the behavior of the user.

  High volume discounts: The pricing structure could include high volume discounts to encourage the participating banks to use the system. The discounts might be activated once a previously agreed minimum transaction volume for each participant bank or bank branch has been achieved. For example, it is not untypical to base overall costs on some percent utilization of overall system capacity, say sixty percent. So, if the actual system utilization is in excess of sixty percent the service provider will be able to give discounts and still recover costs.

  Minimum daily usage charge: A minimum daily charge might be introduced to encourage subscribers to use the system. However the fixed charge should not be set at a level which is so high as to discourage the commercial banks from using the system. This minimum charge could be presented as a fixed annual subscription fee to all participants.

  Subsidized price structures: Subsidized prices may be required during the initial implementation period as a basis for positively encouraging use of newly introduced systems. The subsidy can be applied using a variety of approaches such as no charges at all, discounts based on volumes (principally based upon long term (high volume) per item costs), or full cost recovery with annual rebates to high volume users. The level of the subsidy should be considered carefully. In essence, it should be pitched at a level that encourages use, is affordable, yet does not institutionalize the concept of subsidization. Some form of gradually reducing subsidy might be appropriate to overcome a low transaction volume start-up situation.

Based on the above, it would seem appropriate that an initial price structure is developed around one or more of the following four core prices.

- ***Entry fees:*** All new subscribers to a particular payment service could be charged an entry fee. The fee would be set at a price which is designed to recover a part of previous capital expenditure. The price should be motivational and encourage users to be selective in the services they use.

- ***Annual fees:*** All subscribers could pay an annual fee which is set according to the services utilized. The purpose of this fee is to encourage the users to be selective in the service they require and/or provide lower per item fees to encourage marginal usage. These fees can be used to off-set some of the capital expenditure.

- ***Per item prices:*** This is the principal mechanism for cost recovery from the payments system. Prices are set by transaction volume according to the nature of the transaction and the features of the service. The system selected for use may have a number of optional features such as security, guaranteed finality, ability to retrieve copies of past transactions and ability to generate specific management reports. It is therefore appropriate to have a range of prices for different basic services. In addition there may be volume discounts to encourage high volumes of traffic. The two part pricing concepts discussed in Part I and illustrated in Table 4 should be considered as it represents a practical and fair method of recovering fixed and variable processing costs.

- ***Ad-hoc fees and charges:*** There may be some service that the participating bank requires such as software for upgrading his interface with the payments system. Any such assistance or service should be outside the price structure and should be priced with the participants' agreement.

Indicative cost recovery framework: Pricing and cost recovery related issues are not simple or straight forward. The decisions made must be tailored to the specific situation and must reflect both funding and operational realities. In particular, as many payment system improvement initiatives are funded initially through re-payable loans or credits, attention should be given to ways in which these loans and any associated interest charges might be re-paid from user fees. A situation may also arise, for example, in which a central bank may have full ownership of the payments system at inception, but might choose to replace its ownership with the eventual ownership of the users. The situation is further complicated in that new users of the systems will be admitted over time, thus in the interests of fairness, the pricing arrangements must ensure that these new users do not get more favorable treatment than the initial users.

The purpose of the following discussion is to illustrate how some of these issues might be considered and addressed. The described framework is not put forward as a rigorous treatment of the associated issues but is put forward to demonstrate how the basis for pricing decisions might be calculated. The primary objective of the discussion is to create awareness and stimulate the detailed analysis that must take place on a country by country, system by system basis.

For purposes of illustration, it is assumed that the central bank owner of the new systems has decided to pursue a ***planned growth cost recovery strategy*** in which, the price will be set to recover operating costs and provide funds for future purchases of essential capital equipment. For ***illustration*** purposes, we have assumed that several payment applications will be covered by the new arrangements, as is the case in China, Viet Nam, Mauritius and in many other developing economies. Each of these applications will require a different price as their cost structures will be different. Despite these differences, the formula to be used to determine the price will be the same. The differences will be reflected in the associated specific cost and volume data. This formula will be based on the following generic elements:

Element	*Average cost from the 3-year forecast*
Direct operating costs	A
Depreciation of equipment and other assets used solely in operating the system	B
Proportion of shared operating costs	C
Proportion of depreciation of equipment and other assets used jointly in operating the systems	D
Operating costs	**E = A + B + C + D**
Start-up loan repayment provision	F
Planned growth element	G
Total costs to recover	**H = E + F + G**
Average number of transactions forecast over a three-year period	X
Per item tariff for the year	**H/X**

The entries referenced in the above table are explained in the following paragraphs.

A. Direct operating costs

Direct operating costs are the costs which can be identified as being incurred exclusively to provide the service being costed. For the purpose of this generalized formula, depreciation has been excluded from the direct costs and separately analyzed (*Depreciation of Equipment*). Therefore, where loans are made to fund specific projects which can be identified, the interest should be attributed to the payments system using the loan.

B. Depreciation of equipment and other assets used solely in operating the system

Certain equipment and other assets will be used exclusively for the purpose of providing the services costed. All assets will be depreciated over their useful lives and the related depreciation cost will be treated as an expense each year. Where assets can be identified for the exclusive use in any one particular system, the depreciation will be separately recorded and included at this point in the formula.

C. Proportion of shared operating costs

In all operations of this nature there are several costs which can not be easily identified as solely arising from one area of the operations, but are the result of the combined operations. An example of this type of cost would be the central processing unit and some telecommunications links. These shared costs should be allocated between the systems in an equitable (or other manner determined by management). Activity based costing principles could be used to determine the costs attributable to each service.

An important element of these shared costs will be interest to be paid on any loans which are used to fund joint projects. As noted above, where loans have been obtained to fund specific projects the interest should be considered a direct cost of that project while in all other cases it will be a shared cost to be allocated using the selected methodology.

D. Proportion of depreciation of equipment and other assets used jointly in operating the systems

As noted above there will inevitably be some equipment and other assets which are used jointly by all systems. The basis of allocation of costs should be determined using the same methodology, as selected under *Proportion of Shared Operating Costs*, to allocated costs.

E. Start-up loan repayment provision

To develop the payment system the project requires substantial funding which, we have assumed, will initially be provided by the central bank and has for the purpose of this generalized formula been treated as a repayable loan. We have further assumed that the loan should be repaid by the system users over a period of time, yet to be determined. The following formula can be used to determine the element in the price which is related to the loan:

$$\mathrm{Pmt} = \frac{(P * i)}{[(1 + i)^{n} - 1]}$$

Where:

P	= The principal to be repaid
Pmt	= The amount of the repayment to be included in the main cost formula
n	= The number of years remaining until repayment date (payment at the end of each period)
i	= The annual rate of interest which will be earned over the period to repayment

The above assumes that the element of each payment which relates to the loan will be invested in a secure account, at the end of each period, such as guaranteed (insured) government securities, and is fully auditable. A method used in some countries is to treat loan repayment as equity in the operating cooperative and any new entrant will be required to purchase such equity from the existing members based on their anticipated traffic volumes. In this way, the initial users will not be unduly penalized by having to repay the loan during the start-up period while new entrants make no contribution to past repayments. The size of each participants equity stake could be adjusted annually based on the traffic volumes of the previous year.

F. Planned growth element

It is assumed that the cost recovery system should provide funds for future investments to expand or enhance the systems. The following formula can be used to determine the element in the price related to future investments:

$$\mathrm{Pmt} = \frac{(\mathrm{FI} * i)}{[(1+i)^{n} - 1]}$$

where:

FI = The anticipated cost of future investments

Pmt = The amount to be included in the cost formula

n = The number of years remaining until investment date

i = The annual rate of interest which will be earned over the period until the investment is required

The above assumes that the element of each payment, made at the end of each period, which relates to the future investments will be invested in a secure account such as a guaranteed (insured) government securities, and is fully auditable. It is probable that over time several investment projects will be identified and the above formula should be used for each defined project. In addition the formula has assumed that any inflationary effects have been accounted for in the cost of the Future Investment (FI).

G. Average number of transactions over, say, a three year period

The number of transactions from which the costs are recovered needs to be determined in advance of the year so all the system users are aware of the charges for the following year.

Therefore, the forecast system usage should be estimated through discussions with users before the year begins. The above formula assumes a three year average to ensure that the costs are reasonable and can be maintained at broadly the same level over the period.

H. Per item price for the year

The per item price for the year is determined by dividing the total costs by the anticipated total traffic volume.

As mentioned at the outset, the framework described above is not intended to be a comprehensive treatment of pricing issues from an accounting perspective. Neither does it cover all issues that might be relevant in a specific case. Clearly, a number of specific management decisions will have to be made in regards to the operation and modification of the above formula for price estimation purposes. Examples of such issues are as follows:

During the start-up years, it may be necessary to under recover the actual costs to encourage usage of the system. Any such under recovery will need to be built into future pricing formulae once the systems are established and traffic volumes are at a level where full recovery is practicable;

- Rigorous management accounting systems will be required to identify, capture and record the costs actually incurred in operating the systems. In the initial implementation period, these costs will have to be estimated based probably on the experience of operating similar systems in other countries adjusted as appropriate to reflect local conditions;

- There will be a need to establish a system of entry fee calculations based on projected system usage to ensure that new entrants contribute to the loan repayment fund and the investment in existing and future equipment and assets. A possible approach is to use equity as a means of guaranteeing user rights. The cost of this equity could be based on a valuation of the assets at the date of entry which would include both the investment reserve and the value of the loans less any repayments. The importance of this element in the equation must not be underestimated and as the exchange of loan capital for equity has implications on the ownership of the future "operating cooperative" it should be resolved before finalizing any decisions on the specifics of the pricing formula to be used. Professional accounting advice must be sought when addressing this issue;

- Rebates may need to be applied for over-recovery of costs in a period. However as costs, in the above example, are calculated using a three year average, the rebates could be calculated in arrears. The exact basis for calculating rebates should be defined in the user agreement documents;

- The formula does not take into account any taxation issues which might arise on either the expenses or interest earned on investment / loan accounts. In addition the formula assumes that no taxation will be payable on the receipts from users; and finally

- The formula takes no account of the potential to use differential pricing for different levels of service nor does it consider the use of differential pricing to influence behavior amongst the users.

9. Experience in pricing payment services: banks, payees, and central banks

Bank payment costs and prices. Bank costs per transaction for different types of U.S. payment instruments are shown in Table 2. These range from $.15 for an electronic (ACH) payment to a weighted average of $.41 per check. Credit card costs are higher still ($2.45) but include the extra service of extending credit as well as being a payment, and so is not directly comparable with the other instruments shown. Deposit account maintenance expenses are shown separately since they would apply regardless of which payment instrument is used. As noted earlier, consumer payors do not face these differential bank payment costs as explicit fees when choosing an instrument (although business payors do because of their larger payment volume).

Table 2: Average bank cost of payment services (United States, 1994)

Payment instrument	*Average cost*	*Percent of check operating expense*
	(In US$)	*(%)*
Account maintenance (monthly)	$7.42	36%
Paper-based (check):		
Payor bank activity		
On-us debit	0.27	28
Check cashing	0.48	10
Issue official check	0.80	1
Payee bank activity		
Deposit	0.55	11
Transit check deposit	0.15	14
Volume weighted average	$0.41	1
Electronic:		
Credit card [(1)]	2.45	na
ACH	0.15	na
ATM withdrawal	0.43	na

Source: Federal Reserve System (1994), demand deposit and other functions. Data are for the average medium sized bank with deposits of $200 million to $1 billion.

(1) Credit card expenses include both payment processing and a loan component associated with outstanding balances (which incorporates loan losses and the opportunity cost of funds needed to finance unpaid balances).

Norway has gone further and implemented policies designed to shift consumer as well as business use of paper-based (giro and check) payments to lower cost electronic transactions (giro and point-of-sale). Norwegian banks have been encouraged by the central bank to institute per transaction fees that increasingly reflect the differential bank costs of producing different types of payments. Currently, 48% of bank payment instrument expenses are covered through transaction fees whereas only 13% was so covered in 1988 (Flatraaker and Robinson, 1995). Previously, bank payment expenses were largely recouped through earnings on float by debiting a payor's account one or more days earlier than necessary to make a payment, and earning interest on these funds before a payee's account was credited.

Average prices for different bank payment services in Norway are shown in Table 3 along with their average cost and the percent of total cost recovered in the price. In the future, float revenues are to be reduced further and an even larger share of bank costs are to be recovered through explicit fees. The goal is to reflect better the true resource cost of using the different payment instruments. This enables users, who demand payment services, to make more informed decisions on which instrument has the greatest net benefit, and hence which instrument is best used for point-of-sale or bill payments. It also enables payment suppliers to guide users toward those instruments which have the lowest resource costs, reducing bank expenses. The clear implication in Table 3 is that electronic payments have both a lower cost and a lower price than paper-based payments.

Table 3: Average bank prices and costs of paper-based and electronic payments (Norway, 1994)

				Price as a
Payment instrument		*Average price*	*Average bank cost*	*% of cost*
		(In US$)	*(In US$)*	%
Paper-based:				
Mail giro		$0.49	$1.06	46%
Giro collection box		0.71	1.55	46
Giro at the counter		1.20	2.26	53
Giro cash payment		1.27	2.54	50
Check		0.92	2.15	43
Electronic				
Direct debit:	No notification	0.14	0.49	29
	With notification	0.42	0.92	46
Direct deposit:	No notification	0.14	0.21	67
	With notification	0.35	0.56	63
EFTPOS (debit card at point-of-sale)		0.46	0.63	73
ATM withdrawal		0.25	0.49	51

Source: Flatraaker and Robinson (1995), table 3: exchange rate is NOK709=$1.

Explicit payment pricing in other countries. The clear trend in recovering bank payment costs in Belgium, Finland, Germany, Ireland, Italy, Luxembourg, Netherlands, South Africa, Switzerland, and Sweden has been to implement explicit pricing for payment services (Llewellyn and Drake, 1993).[14] Two-part pricing is often used, where a transaction fee is combined with a

[14] *The U.K. is an exception to this trend, as is the U.S.*

monthly or quarterly fixed fee (covering account maintenance expenses and/or providing for a fixed number of "free" payments to be made before transaction fees are assessed). Payment prices are also often differentiated and reflect the lower cost of electronic payments.

Bank pricing of consumer versus business payments. In marketing bank payment services, the typical arrangement for consumer payments has been to offer different payment service "packages". For consumers, these packages are usually differentiated according to the payment volume initiated, the interest paid (if any) on a transaction account, and the choice of using bank personnel and paper-based instruments rather than electronics as the main point of service contact for cash withdrawal, bill payments, and point-of-sale transactions.

In contrast, business payments–due to their higher volume for each payor–are typically separately priced per transaction, along with a fixed monthly fee for account maintenance. Businesses usually have a choice of paying these fees directly or holding a non-interest earning compensating balance which generates the same cash flow to the bank. Since a compensating balance rises if the payor initiates more payments, it functions exactly like a per transaction fee and provides the same incentive structure.

Payee restrictions on payment instruments. As noted, payees incur sometimes large differences in costs when accepting different payment instruments (see Table A2). Even so, payees may not vary the price of the good or service they provide according to the payment instrument used.[15] Differences in payee payment costs are often absorbed, on average, in the sales price charged. Payees, however, do at times place restrictions on which payment instruments they will accept. For example, credit cards are often not accepted for payment because of their relatively high cost to payees or, alternatively, a sale has to exceed some minimum value before a credit card can be used. In other situations, cash payments above a certain value are refused due to safety and fraud reasons.

The payee response to the high cost of accepting a credit card payment is largely the result of restrictions imposed by credit card providers. Contracts between credit card providers (usually banks) and retail payees typically expressly prohibit payees from imposing a surcharge for payor use a credit card, which some payees would like to do because the cost to payees is so much higher than accepting other payment instruments. Although these contracts allow payees to provide a discount for (say) the use of cash instead of a credit card at the point-of-sale, this would require the payee to raise its price so that after the discount it would still earn the same revenues. While a discount for the use of certain payment instruments does occur, it is not common.

[15] *An exception to this arrangement occurs when especially large value payments are to be made. In this instance, especially if an instrument contains a good deal of float benefit for the payor (as can a check payment), it is not uncommon to negotiate both the payment method and the sales price together.*

Central bank payment costs and prices. A selection of prices for the main payment services offered by the U.S. central bank are shown in Table 4. As seen, a ***two-part pricing*** structure is common. The first part of the price structure reflects the average variable cost of providing the service and is covered through a per item or per transaction fee. The second part of the price structure effectively covers the average fixed cost of the service or payment activity indicated. The fixed cost component of price can contain labor (normally thought of as a variable input), materials, or an allocation of physical capital -- the key point is that it reflects the fixed elements in providing the service. The fixed cost component of price is assessed either each time the service is used or through a monthly or recurring fee, as is evident in Table 4. In addition, for ACH, funds, and securities transfers, ***benefit-flow pricing*** is also used. It was determined that the originator or sender of funds benefited by being able to initiate and complete a transaction but also that the receiver of funds benefited by having its account credited. Therefore, many of the prices shown in Table 4 have split the item fee evenly between the two parties to a transaction. A similar logic applies to checks but, due to the historical precedent of placing all the cost on the receiver (payee) of a check, was not adopted.

The central bank prices shown recover the total cost of each of the services separately over time, so there is no cross-subsidization among payment services.[16] That is, the check service does not consistently run a surplus or deficit to subsidize or tax one of the other services. However, within a particular service, there may be instances were a subset of payment products offered over- or under-recovers their directly allocated costs. Thus there can be some cross-subsidization within a particular service line (e.g., checks). Since the implementation of central bank pricing in the 1980s, the extent of cross-subsidization has been markedly reduced due to cream skimming by commercial banks who also offer payment services and thus force the central bank to price closer to actual cost for each product or lose volume.[17]

Central bank revenues from providing priced payment services totaled over $800 million in 1996. Because the prices charged include the imputed expense of taxes and return on invested capital equivalent to that of a private firm, the central bank has transferred almost $900 million in "profit" to the Treasury over the last 10 years, reducing government debt by the same amount. As seen in Table 5, fully three-fourths of payment service revenues are associated with checks, reflecting the predominant share of this payment instrument in total transactions.

[16] *Central bank prices are required, by legislation, to be set so that all direct and indirect costs are recovered over time, including the imputed cost of taxes and return on invested capital that would have been incurred if the payment services had been provided by a private business firm. Realized tax rates and returns on capital for a set of representative large banks are used in determining these imputed costs.*

[17] *The central bank originally adopted a pricing structure similar to what was described earlier as an optimal departure from marginal cost pricing. This arrangement, called "market-sensitive" pricing, contained some cross-subsidization and relied on informal estimates of the price elasticity of demand. As noted above, cross-subsidization is not viable in a competitive market since competitors can always cover their full cost and still underprice the payment product in which revenues exceed costs, the excess revenues of which are used to subsidize another product where costs exceed revenues.*

Table 4: Central bank prices for payment services (United States, 1996)

Payment service	*Price range per item (reflects average variable cost)*		*Price per batch of payments (reflects average fixed cost)*	
	(In US$)		(In US$)	
Check				
Unsorted checks	.003 - .080		1.50 - 9.00	
Presorted checks	.003 - .012		2.50 - 11.00	
Returned checks	.100 - 1.110		1.50 - 8.00	
Payor bank services:				
MICR information	.001 - .005		5.00 - 30.00	minimum
Electronic presentment	.001 - .0045		3.00 - 14.00	minimum
Check truncation	.010 - .017		3.00 - 25.00	minimum
ACH	***Origination***	***Receipt***		
Unsorted deposit	.01	.01	1.75 (per input file submitted)	
Presorted deposit	.009	.01		
Addenda record	.003	.003	25.00 (monthly account servicing fee)	
Returned payment	.04	.04		
Funds transfer				
Wire transfer	.5	.5		
Net settlement	1.00			
Off-line surcharge	10.00			
Telephone advice		10.00		
Book-entry securities transfer				
On-line transfer	2.25		15.00 (monthly account servicing fee)	
Off-line transfer	10.00	10.00		
Electronic connection fees				
Telephone dial up			30.00 - 450.00 per month	
Dedicated leased line			750.00 - 2,000.00 per month	
Encryption certification			0 - 8,000.00	

Source: *Board of Governors of the Federal Reserve System (1997), various tables. Two additional priced services are: the collection of definitive securities (bond coupon collection, etc.) and special cash services (provision of wrapped coin, special packaging of currency, and cash deposits/withdrawals above a standard number per month).*

In order for revenues to cover all costs, the central bank needs to forecast the growth in the demand for its various services and set prices accordingly. First, cost projections (including imputed expenses) are made for total variable cost (TVC) and total fixed cost (TFC) for the coming year. Second, the projected TVC for each product or service is divided by an estimate of the volume of payment items expected to be processed and cleared, giving an estimate of what the per item fee must be to have revenues cover all variable costs. The volume changes experienced for each payment service, which differ somewhat from projections made earlier, are shown in Table 5. Third, the projected TFC is divided by an estimate of the number of batches of payments expected or the number of accounts serviced, yielding an estimate of the fixed fee to be charged to have revenues cover fixed expenses. The end result is a set of prices which comprise the two-part pricing schedule of Table 4.

Table 5: Central bank revenues and payment volumes (United States, 1996)

Payment service	*Revenues*	*Revenue composition*	*Volume growth*
	(US$ millions)	*(%)*	*(%)*
Check	603	74	-.4 [1]
ACH	79	10	16.1
Funds transfer	97	12	8.3
Book entry securities transfer	17	2	9.7
Non-cash collection	7	1	23.2
Special cash services [2]	7	1	n.a.

Source: Board of Governors of the Federal Reserve System (1997), various tables.

(1) Unsorted check volume increased by 1.6%, presorted checks fell by 9.1%, and returned checks rose by 2.9% (giving the -.4% weighted average shown).

(2) The provision of currency and coin to the general public is deemed a public service and therefore is provided at no charge to banks. Special cash services account for only about 2% of the total cost of the (priced and nonpriced) cash service.

10. Example of a pricing system - S.W.I.F.T.

The principles discussed above can be well illustrated by reviewing the structure of the Membership Pricing and Cost Recovery philosophy of S.W.I.F.T. The following information is based on the content of the November, 1996 S.W.I.F.T. User Handbook and is referenced with the approval of S.W.I.F.T.

Membership pricing: The S.W.I.F.T. Membership Pricing Structure consists of the following elements:

- Joining fees;
- Annual support charges; and
- Annual charges for additional services.

Joining fees, which are payable by all new customers joining the S.W.I.F.T. network, consist of the entry fee and one-time charges related to connection.

All financial institutions joining S.W.I.F.T. pay an entry fee. The entry fee, to some extent, is designed and levied to ensure that new members are not being disproportionately subsidized by existing members. Financial institutions eligible for full membership (as contrasted with other categories of membership) may also purchase shares at transfer value (not nominal value) and thus qualify to participate in the shareholder decision making process. The allocation of shares to new members is made in accordance with the S.W.I.F.T. policy prevailing at the time of membership. There are also one-time charges related to connection for new S.W.I.F.T. customers. The connection charges relate to S.W.I.F.T. products and services including port access and documentation, with each product and service charged individually. For reasons of compatibility, some products and services are mandatory. The products and services included in this scheme include: ISO registration of 4-character bank code (mandatory), additional 4-character bank code registration, registration of S.W.I.F.T. address (mandatory), registration of logical terminal (mandatory), dedicated port connection, shared port connection, dedicated cross border emergency connection, and S.W.I.F.T. User Handbook (one set being mandatory).

Annual support charges are levied based on the category of membership. The annual support charges gives the customer certain entitlements covering: registration of BIC, permitted number of logical terminals, user handbook updates, BIC directory, and addresses in the BIC directory. The actual shareholding of a member provides additional address registrations entitlements. When a customer cuts over during the budget year, the annual support charge is charged pro rata for the remainder of the budget year.

Annual charges for additional services, in the main, are payable in full even if the service is only provided for part of the budget year. Annual charges are made for:

- each logical terminal additional to the basic entitlement. This charge is also payable for permanent training logical terminals;
- each branch code registered above the basic entitlement; and
- for updates to additional copies of the User Handbook.

Cost recovery: In April 1984, the Board of Directors adopted the principle that the user community of each country or independent constitutional entity accessing the S.W.I.F.T. network must guarantee recovery, of at least, the direct operating costs incurred by S.W.I.F.T.

The operating costs incurred by S.W.I.F.T. are compared to the revenues generated by the users of the country concerned. If these revenues do not cover the operating costs, the shortfall will be invoiced separately.

The costs incurred by S.W.I.F.T. under the cost recovery system fall into two categories: set-up (one-time) costs and annual operating costs.

- ***Set-up costs,*** which are depreciated over five years, are those incurred by S.W.I.F.T. prior to cutover and include: site construction (facilities, security equipment, and telecommunication equipment including taxes and duties), computer equipment (if purchased), S.W.I.F.T. staff costs directly attributable to the country charged at a standard rate, and operating costs (international telecommunication circuits, lease of space and regional administration).

- ***Annual operating costs*** include: lease of space, equipment (rent and maintenance), S.W.I.F.T. staff costs charged at a standard rate, regional administration, local corporate taxes (where applicable), and international telecommunications circuits.

The following calculation applies:

a) Costs

One-time costs / 5 (five years depreciation) = Annual portion of one-time costs
+ Annual operating costs

= Total annual operating costs to be recovered.

b) Revenues and subsidy

Joining fees (one time fees already paid in full) / 5 = Annual portion of joining fees

(for purpose of this calculation only) + Revenues from traffic sent + Revenues from traffic received + Subsidy on traffic received

= Total annual revenues and subsidy.

c) Shortfall

Shortfall (if result is positive)

= Total annual operating costs - Total annual revenues and subsidy.

The specific meaning of the terms used in the above revenues and subsidy calculations are as follows:

- ***Joining fees*** refer to those paid before cutover depreciated over five years.

- ***Revenue from traffic sent*** means the number of messages sent over the S.W.I.F.T. system and billed to the customer. "Revenue from traffic received" means the number of messages received multiplied by the basic message price. An additional subsidy per message received is allocated under certain circumstances.

- It should be noted that extra costs, for example those exceeding the standard costs for international circuits, are deducted from the subsidy on traffic received.

- ***New joining fees*** are the joining fees of new members or sub-members joining S.W.I.F.T. after the country cutover, and are added to the annual revenues during the year following the user's cutover.

Invoicing method: Before cutover, new customers are asked to provide S.W.I.F.T. with the results of a traffic survey. On the basis of this survey and S.W.I.F.T.'s estimated costs related to the customer's country, S.W.I.F.T. informs the customers whether or not the traffic is expected to cover costs, and if not, gives an estimate of the amount which will have to recovered.

After cutover, the costs are reviewed in the light of, for example, exchange rate, contract prices, local taxes, and the amended annual amount is divided by four to give the direct quarterly costs. The costs fixed at this point remain in force for the rest of the budget year.

At the end of each quarter, the following calculation is performed by S.W.I.F.T.:

$$A - B - C - D - E - F = R$$

where	A	=	direct quarterly costs
	B	=	joining fees (divided by 5)
	C	=	quarterly revenue from traffic sent
	D	=	quarterly revenue from traffic received
	E	=	new joining fees, if any (divided by 4)
	F	=	subsidy for traffic received (if any)
and	R	=	the amount (if any) to be recovered.

If the direct costs are lower than the revenues (that is, R is negative), no reimbursement will be made to the user.

If the direct costs are higher than the revenues (that is, R is positive), an additional invoice will be sent to each user to recover his share, calculated according to the allocation formula defined by the User Group Chairperson.

At the beginning of each subsequent year, the annual costs are reviewed in the light of the actual costs.

S.W.I.F.T. connection and supplied equipment costs: If it is necessary for S.W.I.F.T. to establish a connection between a S.W.I.F.T. Access Point (SAP) and a user by means of either a PSTN connection or a PDN connection, the cost of doing so is charged to the receiving customers. Invoicing frequency depends on usage; it may be yearly, half yearly or quarterly.

The price for PSTN connections are identical for all customers and will be based on a per minute price that includes all charges incurred by S.W.I.F.T.

PDN prices can either be based on a flat rate or can be usage based. A flat rate implies a fixed periodic subscription fee regardless of the number of messages transmitted. A usage based price as applied by S.W.I.F.T. can include the following:

- number of calls made, multiplied by a unit price;
- total duration of the calls made, multiplied by a unit price; or
- total volume of data transmitted, multiplied by a unit price.

Any equipment or service, for example, modems, supplied by S.W.I.F.T. at the request of a customer, is charged at cost, plus a 6% administrative charge.

S.W.I.F.T. message pricing: A clearly defined price structure exists for the many different types of S.W.I.F.T. messages. In the main, the message pricing structure, depends on a combination of message type, message length, message volume, message routing category, message priority and message delivery monitoring options.

In essence, the message price structure, is developed based on the proportion of S.W.I.F.T. costs that are attributable to the provision of the specific service.

Other S.W.I.F.T. services pricing: S.W.I.F.T. also supplies a variety of other services to its customers including security products, such as, card readers, integrated circuit cards, and secure X25L service; BIC products, education services and documentation. A clearly defined pricing structure for each of these services is made available to customers and, in the main, is designed to facilitate full cost recovery.

Conclusion: The above summary of key aspects of the S.W.I.F.T. pricing arrangements has been included to demonstrate that with careful attention to detailed cost monitoring, a comprehensive price structure can be developed that is both fair to all system participants and consistent with a management philosophy of full cost recovery. It is worth noting that an additional benefit of the rigorous approach to cost monitoring and control applied by S.W.I.F.T. has resulted in the realization and maintenance of lower customer costs from those prevailing a few years ago.

11. Summary and conclusions

The cost of providing payment services in a country are substantial, on the order of 3% of GDP. There are two ways these costs may be reduced and both involve appropriately pricing payment services. This applies to commercial banks supplying payment services to the general public as well as central banks supplying a more narrow set of payment services to the banking system. First, pricing payment services will induce users to choose those payment instruments which minimize costs relative to the benefits received. Second, when prices closely reflect the full cost of producing each service, users will demand those services which use the fewest real resources. Since the available data indicate that electronic payments generally cost only from one-third to one-half as much as paper-based payments, the current cost of a country's payment system could be substantially reduced if payments are properly priced. Indeed, such a goal has been a matter of public policy in countries in Scandinavia, especially Norway.

Scale economies exist in making and processing payments. As a result, the primary discriminating characteristic among payment users (consumer, business, and government payors and payees, plus intermediary banks) is the volume of payments associated with each participant or participant group. A pricing methodology which recovers all costs but yet properly discriminates among users according to their payment volume is two-part pricing. Two-part pricing contains a price covering the average fixed cost of serving each participant and a second price covering the average variable cost. Two-part pricing is sustainable in the sense that a competitor could not supply the same service without incurring a loss, unless its costs were truly lower. If a competitor's costs were indeed lower, then it–and not the existing supplier–should be the entity providing the payment service, otherwise resources are being wasted. Other, less important, discriminating characteristics among payment service users would include possible differences in the elasticity of demand among payment users and the possibility that both payors and payees benefit from making a payment. There are ways of accommodating these additional discriminating characteristics through so-called market sensitive pricing and benefit-flow pricing, and both are noted in the text. However, most of the goals of pricing can be achieved by implementing two-part pricing. As experience with pricing is gained over time, the pricing components may be modified to accommodate additional discriminating characteristics (if they are deemed to be important), potentially improving the sustainability of the pricing structure.

Two-part pricing, fortunately, is relatively simple to implement. It only requires that: (1) payment service costs (or cost estimates) be decomposed into their fixed and variable cost components; and (2), an estimate be made regarding the expected volume of payments that will be demanded. The first requirement is met through application of an elementary cost accounting system while the second relies on having collected some current or historical volume data. Many banks and central banks in developed countries have implemented two-part pricing, as well as some of its variants, for the payment services they offer. The fact that this pricing structure has remained in place is testament to its usefulness in attaining the payment goals outlined above.

The Bank's work in the payments system reform environment focuses, in the main, on the initial development and implementation of payment mechanisms by the central bank or by the central bank in conjunction with some or all commercial banks in a specific country. In most cases, new inter-bank payment mechanisms, are established under the leadership of the central bank and are initially owned and operated by the central bank. However, in some cases, the Bank may become involved with the establishment of private clearing houses. With this in mind, the primary purpose of this paper is to further consider the specifics of cost recovery and pricing policy from the view of the publicly or privately owned system provider in both the initial and longer term more stable operating phases of a payments systems reform initiative in a developing economy. As such initiatives are frequently funded through re-payable loans some suggestions are made as to how the cost recovery policy might relate to loan re-payment.

Annex 1

Transaction Costs of Paper-based and Electronic Payments in the US and Norway

Total payment costs in the U.S. Table A1 illustrates the (estimated) per transaction cost of a paper-based (check) and electronic (ACH) payment in the U.S. An electronic payment costs $1.31, which is only 45% of the $2.93 total expense of a paper-based payment.[18] Weighted by the shares of paper-based check and electronic (credit card, debit card, ACH) payments, the average U.S. payment transaction costs $2.60 and totals $204 billion a year. This represents 3% of GDP. On a per person basis, each adult directly or indirectly pays $1,050 annually just to make payments. Since a consumer payment averages around $50, transaction costs make up 5% of the value of a typical consumer payment. Thus the total payor, payee, and bank cost of initiating, processing, and settling a payment is not small. Indeed, it is larger than most would have expected.

Bank payment costs in Norway. Although data on the total cost of payments in other countries are not available, survey data from Norway provides an estimate of the bank cost of processing paper-based (giro) and electronic (giro) payments. The payor plus payee bank cost of a paper-based giro payment is $1.34 while an electronic giro payment–an average of a direct debit and a direct deposit–is $.35.[19]

At the bank level, an electronic payment in Norway costs only 26% of a paper-based payment. While the estimated bank costs of electronic and paper-based payments in the U.S. appear to be equal (Table A1), the overall cost of a U.S. electronic payment is 45% of the cost of a paper-based payment. Although the source and magnitude of the cost advantage for electronic payments differs between these two countries, the clear implication is that electronic payments have a lower cost, and therefore should also have a lower price, than their paper-based alternatives. Resources can usually be saved when electronic payments replace paper payments.

Costs faced by payors. There are three primary payment cost elements in Table A1 but payors may not directly face all three when making their decision on which payment instrument to use. Consumer payors typically do not face directly either bank or payee costs. While consumers will directly face their own "expenses" (time taken to initiate a payment, the maintenance of a adequate supply of cash/non-cash balances, etc.), most of these will not be in explicit money terms per transaction but instead will be evaluated in terms of differences in the convenience, acceptability, and safety of using different payment instruments. Mailing expenses associated with paper-based bill payments are the exception.

[18] *The notes to Table 1 summarize the main elements underlying these cost estimates. Even greater detail is in the Data Appendix to Wells (1996).*

[19] *Robinson and Flatraaker (1995), Table 1, provide this information in Norwegian krona which is translated into dollars at NOK 7.09 = $1.*

Table A1: Payor, payee, and bank cost for paper-based and electronic payments (U.S., 1993)

Per transaction expenses for	*Paper-based payment (check)*	*Electronic payment (ACH)*
	(In US$)	*(In US$)*
Payor [(1)]	1.39	0.80
Payee [(2)]	1.25	0.23
Bank [(3)]	0.29	0.28
	2.93	1.31

Source: Wells (1996), using averages of ranges reported for the various components. The reported figures represent weighted averages for consumer, business, and government payments. Float costs ($.09 for checks) have been excluded since float is a transfer payment.

(1) *Payor costs of check use are composed of check printing and distribution costs ($.0345), postage cost ($.18), and business cost of issuing checks ($1.18). ACH payor costs only include the business cost of initiating a preauthorized direct debit or a direct deposit of payroll ($.80).*

(2) *Payee costs are composed of the cost of accepting a check ($1.25) at the point-of-sale or for bill payments or the cost (including accounting expenses) of accepting a preauthorized direct ACH debit ($.23).*

(3) *Bank costs of processing checks and ACH payments ($.29 and $.28, respectively) include fraud costs and central bank processing and settlement expenses.*

Business payors, in contrast, will face directly a large percentage of their total payment costs. First, business expenses are routinely identified and reported to management, so payment costs are typically quantified. Second, each business payor initiates a larger volume of payments (for employee payroll and purchases from other firms) than does each consumer payor, and these expenses are usually too large to be ignored. Thus businesses are better able to put an explicit monetary value on their own internal payment expenses with those directly paid to banks. Since businesses will directly face a larger portion of the payment expenses they incur, they will be more sensitive to differences in payment instrument costs.

Costs faced by payees. Payees receiving consumer point-of-sale and bill payments will experience different expenses depending on the payment instrument used. In some countries, these are made explicit to consumer payors but in other countries they are not. This aside, Table A2 illustrates how the average cost of accepting different payment instruments can vary whether expressed on a per transaction basis or for each $100 of sales. Despite this difference, retailers in the U.S. rarely charge different fees or give discounts to encourage the use of particular instruments at the point of sale. More commonly, differential payee payment expenses are simply folded into the overall price of the goods being sold or the bill for services rendered.[20]

Table A2: Supermarket payment costs for different payment instruments (United States, 1994)

Payment instrument	*Cost per transaction*	*Cost per $100 of sales*
	(In US$)	*(In US$)*
Cash	0.07	0.52
Check	0.43	1.20
Credit card	0.81	2.27
Debit card	0.30	0.94

Source: Food Marketing Institute (1994)

Costs incurred by banks. Bank payment costs for the demand deposit activity can be decomposed into variable and fixed components. Variable costs (labor, supplies, transportation, etc.) rise with significant (say 5% to 10%) increases in payment volume, even though their cost per unit of payment volume processed may remain stable. Fixed costs (in this case buildings, computers, etc.) remain stable with significant changes in payment volume, although these too would rise if volume increased by an especially large amount (say more than 20% to 30%).

Table A3 illustrates the major components of bank payment costs associated with demand deposit activity. For the average medium sized U.S. bank, the largest allocated expense is for labor (48%), followed by other variable expenses (27%) and capital or fixed expenses (25%). The distinction between variable and fixed costs is important since this can indicate the potential for scale economies.

[20] In Table A2, the payee cost of accepting a credit card is almost twice that of other instruments. Payees, not the payor, pay a transaction fee of from 1% to 3% of the value of the sale when a credit card is used. When business payees do not vary their output prices depending on the payment instrument used, credit card users will be cross-subsidized by consumers who choose to use instruments (cash, check, electronic debits) that have a lower cost to the payee.

Table A3: Components of payment costs for the average bank (United States, 1994)

Cost component	*Operating cost*	*Percent composition*
	(In US$)	*(%)*
Labor expenses:		
Salaries	1,978,072	38
Fringe benefits	525,473	10
Other expenses:		
Supplies	142,282	3
Transportation	172,278	3
Other	1,084,771	21
Total variable costs	**3,902,876**	**75**
Capital expenses:		
Data processing	480,376	9
Furniture and equipment	293,511	6
Building occupancy	537,700	10
Total fixed costs	**1,311,587**	**25**

Source: Federal Reserve System (1994), demand deposit function. Data are for the average medium sized bank with deposits of $200 million to $1 billion.

Annex 2

Empirical Evidence of Scope and Scale Economies

Empirically significant scope economies do not appear to exist among broad classes of payments such as checks, ACH, and wire transfers (Bauer and Ferrier, 1996). Nor do scope economies apply to the joint processing of cash and non-cash payments. This is because these particular payment instruments do not appear to share a significant portion of their costs.

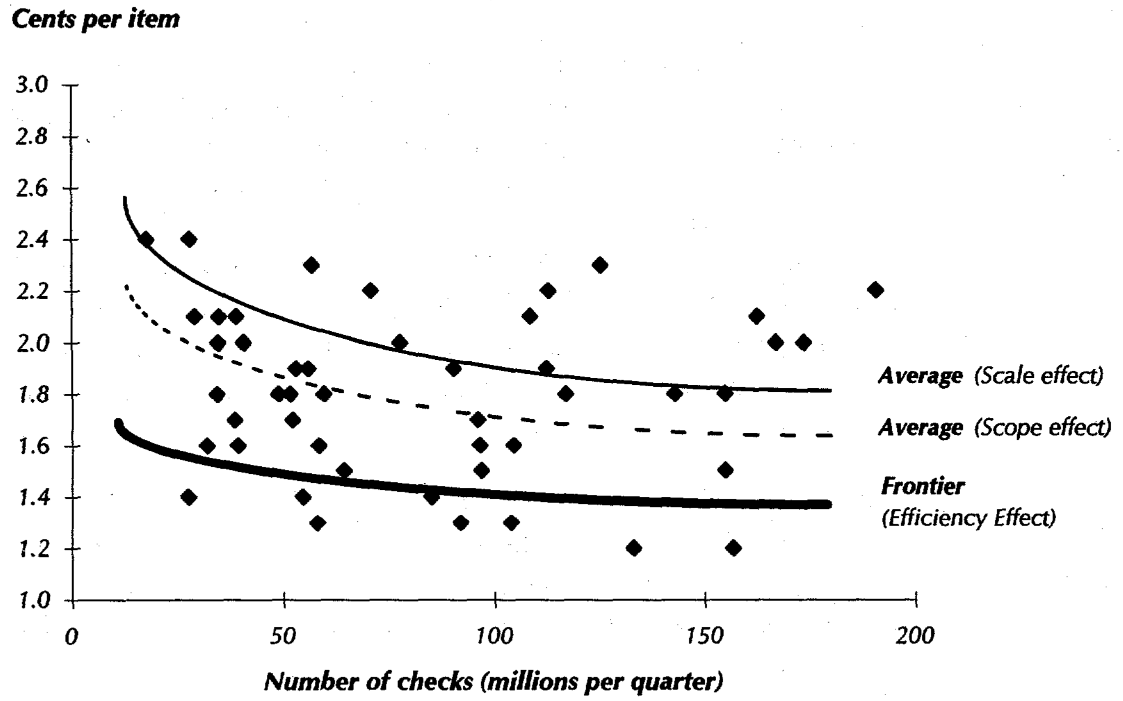
Scale economies exist up to a point for processing paper-based payments (Bauer, 1993). However, after a certain large volume is attained, unit costs no longer fall but remain relatively constant. In contrast, electronic payments appear to experience scale economies regardless of the volume of current transactions (Bauer and Hancock, 1995). This relationship may or may not continue if electronic payment volumes became extremely large.

Efficiency in producing payment services. Payment costs can be reduced when the productivity or efficiency of producing payment services is improved. This involves comparing the costs of producing a particular payment service among either different payment processors in the same country (e.g., Bauer and Hancock, 1993) or between processors in different countries. Holding constant those cost differences believed to be beyond the strong control of management, such as the local price of labor, the unit cost of computers and buildings, and the cost of local transportation and supplies, it is possible to identify statistically the set of production units which have the lowest costs due to "best practice" organizational structure or use of more efficient processing technology. When the identified best practices are applied to those production units where costs are currently higher than average, productivity should rise and costs should fall.[21]

Figure 3 illustrates the scale, scope, and best practice (or frontier payment cost) concepts. The scatter plot shows the average cost per check processed at 47 offices of the U.S. central bank over 1983-90. If a cost function is fitted to all of these observations, the average relationship between unit cost and payment volume is identified (holding input prices, product mix, and certain other cost influences constant). The downward slope of the fitted curve (thin solid line) indicates that scale economies exist but that at moderate volumes average cost flattens out and becomes constant. If volume rose beyond 80 million processed items (per office per quarter), average costs would essentially be constant as the volume of processed payments expands.

[21] *Such comparisons and productivity improvements have been performed successfully among branch offices of a single bank (surveyed in Berger and Humphrey, 1997). Since the primary function of a bank branch is to accept and make payments among depositor accounts, such restructuring has improved the efficiency of bank payment processing operations.*

Figure 3: Average check processing cost



Scope economies, if they exist, would lead to a downward shift in the fitted cost curve (dotted line). If the volume of one payment instrument was unchanged at 80 million items, its marginal and average cost could still be reduced (say, falling from the thin solid line to the dotted line) as the volume of a second instrument processed at the same facility was expanded. Here expanding the scope of operations, rather than their scale, can lower costs.

If attention is focused instead on those processing offices which experience the lowest unit costs, a "frontier" of best-practice offices can be identified (thick solid line). The downward slope of the frontier indicates that similar scale economies are experienced by the set of most efficient processing offices, although such a similarity need not hold in general. A similar scope effect may exist for frontier offices as well (not shown).

Bibliography

Bank for International Settlements, Payment Systems in Eleven Developed Countries. Third edition, Bank Administration Institute, Rolling Meadows, IL (May 1989).

Bank for International Settlements, Payment Systems in the Group of Ten Countries. Basle, Switzerland (December 1993).

Bank for International Settlements, Payment Systems in Finland. Basle, Switzerland (July 1994).

Bank for International Settlements, Statistics on Payment Systems in the Group of Ten Countries. Basle, Switzerland (December 1994).

Bank for International Settlements, Payment Systems in Norway. Basle, Switzerland (May 1995).

Bauer, P., "Efficiency and Technical Progress in Check Processing", Federal Reserve Bank of Cleveland Economic Review, 29 (Quarter 3, 1993): 24-38.

Bauer, P,, and G. Ferrier, "Multiproduct Frontier Cost Analysis of Federal Reserve Payments Processing", Working Paper, Federal Reserve Bank of Cleveland (December 1996).

Bauer, P., and D. Hancock, "The Efficiency of Federal Reserve Check Processing Facilities", Journal of Banking and Finance, 17 (April 1993): 287-330.

Bauer, P., and D. Hancock, "Scale Economies and Technological Change in Federal Reserve ACH Payment Processing", Federal Reserve Bank of Cleveland Economic Review, 31 (Quarter 3, 1995): 14-29.

Berger, A., and D. Humphrey, "Efficiency of Financial Institutions: International Survey and Directions for Future Research", European Journal of Operational Research, (April 1997), 175-212.

Board of Governors of the Federal Reserve System, "Proposed 1997 Fee Schedules for Priced Services", Washington, D.C. (October 1996).

Electricite de France, Tarification of electricity in France: Principles and construction of scales, Cahier de Tarification (June 1995)

Federal Reserve System, Functional Cost Analysis. Washington, D.C. (1994).

Flatraaker, D., and P. Robinson, "Income, Costs and Pricing in the Payment System" Norges Bank Economic Bulletin, 66 (September 1995): 321-32.

Food Marketing Institute, Benchmarking Comparative Payment Methods: Costs and Case Studies. Washington, D.C. (1994).

Humphrey, D., L. Pulley, and J. Vesala, "Cash, Paper, and Electronic Payments: A Cross-Country Analysis", Journal of Money, Credit, and Banking, 28 (November 1996, Part 2): 914-39.

Llewellyn, D., and L. Drake, "The Economics of Bank Charges for Personal Customers", Research Monograph No. 9, Loughborough University Banking Centre, Loughborough, U.K. (May 1993).

Robinson, P., and D. Flatraaker, "Costs in the Payments System", Norges Bank Economic Bulletin, 66 (June 1995): 207-16.

S.W.I.F.T., "Pricing & Invoicing", User Handbook (February 1997)

Weinberg, J., "Selling Federal Reserve Payment Services: One Price Fits All?", Federal Reserve Bank of Richmond Economic Quarterly, 80 (Fall 1994): 1-23.

Wells, K., "Are Checks Overused?", Federal Reserve Bank of Minneapolis Quarterly Review, 20 (Fall 1996): 2-12.

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1815	Unfair Trade? Empirical Evidence in World Commodity Markets Over the Past 25 Years	Jacques Morisset	August 1997	N. Busjeet 33997
WPS1816	Returns to Regionalism: An Evaluation of Nontraditional Gains from Regional Trade Agreements	Raquel Fernandez	August 1997	J. Ngaine 37947
WPS1817	Should Core Labor Standards Be Imposed through International Trade Policy?	Keith E. Maskus	August 1997	J. Ngaine 37947
WPS1818	What Affects the Russian Regional Governments' Propensity to Subsidize?	Lev Freinkman Michael Haney	August 1997	N. Campos 38541
WPS1819	The Argentine Pension Reform and Its Relevance for Eastern Europe	Dimitri Vittas	August 1997	P. Infante 37642
WPS1820	Private Pension Funds in Argentina's New Integrated Pension System	Dimitri Vittas	August 1997	P. Infante 37642
WPS1821	The "IPO-Plus": A New Approach to Privatization	Itzhak Goldberg Gregory Jedrzejczak Michael Fuchs	August 1997	I. Goldberg 36289
WPS1822	Intergovernmental Fiscal Transfers in Nine Countries: Lessons for Developing Countries	Jun Ma	September 1997	C. Ima 35856
WPS1823	Antidumping in Law and Practice	Raj Krishna	September 1997	A. Bobbio 81518
WPS1824	Winners and Losers from Utility Privatization in Argentina: Lessons from a General Equilibrium Model	Omar Chisari Antonio Estache Carlos Romero	September 1997	T. Malone 37198
WPS1825	Current Accounts in Debtor and Creditor Countries	Aart Kraay Jaume Ventura	September 1997	R. Martin 39026
WPS1826	Standards and Conformity Assessment as Nontariff Barriers to Trade	Sherry M. Stephenson	September 1997	M. Pateña 39515
WPS1827	The Determinants of Agricultural Production: A Cross-Country Analysis	Yair Mundlak Don Larson Ritz Butzer	September 1997	P. Kokila 33716

Policy Research Working Paper Series

	Title	Author	Date	Contact for paper
WPS1828	The Determinants of Banking Crises: Evidence from Developed and Developing Countries	Asli Demirgüç-Kunt Enrica Detragiache	September 1997	P. Sintim-Aboagye 38526
WPS1829	Economic Reform and progress in Latin America and the Caribbean	Norman Loayza Luisa Palacios	September 1997	E. Khine 37471
WPS1830	Private Ownership and Corporate Performance: Some Lessons from Transition Economies	Roman Frydman Cheryl W. Gray Marek Hessel Andrzej Rapaczynski	September 1997	B. Moore 38526
WPS1831	How Trade Patterns and Technology Flows Affect Productivity Growth	Wolfgang Keller	September 1997	J. Ngaine 37947
WPS1832	Pension Reform in Bolivia: Innovative Solutions to Common Problems	Hermann von Gersdorff	September 1997	C. Pavlak 82099