ELSEVIER Journal of Banking & Finance 21 (1998) 1573–1624

Journal of BANKING & FINANCE

Payment transactions, instruments, and systems: A survey

Diana Hancock [a,*], David B. Humphrey [b]

[a] *Federal Reserve System, Board of Governors, Mail Stop 188, Washington, DC 20551, USA*
[b] *Florida State University, Tallahassee, FL 32306-1042, USA*

Abstract

The payments literature ranges from theoretical general equilibrium models to practical payment issues related to the day-to-day operation of various national networks for the transfer of money. It is an area where economic theory and institutional structure are often closely intertwined and it is currently undergoing significant change, shifting from costly paper-based systems to technologically advanced electronic payments. The extant literature is surveyed here with the aim of integrating the various strands of payment research which have been largely pursued separately. In addition, we present newly available data to illustrate and investigate a number of underdeveloped areas in this literature.

JEL classification: G200; E590

Keywords: Cash; Checks; Giro; Settlement; Risk

1. Introduction

Market economies rely on the payment system to facilitate trade and exchange among firms and consumers in product markets. These systems also assist in transforming domestic and international savings flows into productive investments through financial markets. Payment services are supplied primarily

* Corresponding author. Tel.: 202 452 3019; fax: 202 452 6474; e-mail: mldxh@fed.frb.gov.

0378-4266/97/$17.00
PII S0378-4266(97)00046-0

by the banking (or postal) system and typically are regulated by the central bank as an important adjunct to implementing monetary policy.

Payment services are not cheap: 5% or more of the value of an average consumer's purchase is eaten up in payment costs while the total cost of a country's payment system may account for about 3% of the value of its GDP. Some of this expense goes to the government in the form of seignorage revenues from issuing currency. Since electronic payments usually cost only one-third to one-half that of paper-based transactions, substantial savings in social costs can be realized in shifting from paper to electronic payments. Indeed, a few countries are currently actively promoting such a shift.

The payments literature is extremely diverse, ranging from general equilibrium models of payment instrument use to practical issues on how best to control risk on payment networks. Overall, the academic literature has focused on three basic payment issues. The first has dealt with theoretical general equilibrium models investigating the trade-off between cash and other assets as alternative means of payment. One problem has been to explain why cash – with a positive opportunity cost – continues to be used for transactions when other assets earn a positive return. The second issue concerns the demand for money and the implications that substitution among different current payment media (cash, checks, and credit and debit cards) have on asset holdings (demand deposits, savings deposits, time accounts, and bonds), on money supply measurement, and on the implementation of monetary policy. The third issue deals with the risk of settlement failure on large-value funds transfer networks. This is the so-called systemic risk problem which generates a moral hazard for the banking system's government-provided safety net (deposit insurance and discount window borrowing). Academic research on the second and third issues have a direct and important bearing on government policy.

Other payments research issues also exist but investigation has been greatly hindered by a lack of available data. We have theoretical models of the transactions demand for money but very little data with which to test and refine the models that have been developed. We also have good information regarding the scale, scope, and efficiency aspects of providing payment services on the part of the US central bank which processes and settles for a number of different payment instruments, but have almost no information regarding these relationships for commercial banks which in all countries provide the majority or all of the payment services used by consumers and businesses. Similarly, there is little public information concerning the determinants of payment instrument use by either consumers or businesses. We can easily make lists of likely influences but almost no estimates of their relative magnitudes have been made. The closest that research has come to this issue has been through a careful empirical analysis of consumer/business asset substitution in response to interest rates and other influences (e.g., the rise of cash management) to address questions regarding money supply measurement and the stability of velocity. Finally, although the popular press is full of references to the potential use

of new types of payment arrangements – from the effect of ATMs and smart cards on cash use to the potential for electronic payments over the Internet to replace checks or credit cards for bill payments – little theoretical (and even less empirical) work has been done in this area.

In this survey we will summarize the current state of the payments literature regarding these and other payment issues. Specifically, we describe and contrast the structure of payments in the US, Europe, and Japan in Section 2. In Sections 3 and 4, we survey issues relating to the demand for cash and non-cash payments respectively, while Section 5 focuses on such supply related issues as production efficiency and technological change in payment processing operations. Systemic risk, and what to do about it on payment networks, is covered in Section 6. The monetary policy and other implications of new and emerging payment methods, such as stored-value cards, are discussed in Section 7. Finally, Section 8 presents information on the social and private costs of paper and electronic payment methods and notes the barriers which exist in moving to lower cost electronic payments. Areas where further research is needed are outlined in Section 9, along with some conclusions.

2. The payment system: Structure and policy issues

The payments system, which consists of a legal framework, rules, institutions, and technical mechanisms for the transfer of money, is an integral part of the monetary and financial systems in a smoothly operating market economy. [1] The ability to make retail payments at low cost and with moderate risks facilitates the exchange of goods and services. Likewise, a more efficient and less risky means for making large-value wholesale payments (e.g., wire transfers and securities transfers) can make financial markets deeper and more liquid. To a large extent, the efficiency of transactions in a market economy is determined by the efficiency of the payments system.

2.1. Payment structure: Cash and non-cash payments

A country's payment system relies on both cash and non-cash payments. Cash is provided by the government, from which it obtains seignorage revenues, while non-cash payments are typically provided by the banking system using either deposit balances in transaction accounts or credit balances.

The limited data that exist indicate that cash transactions represent by far the largest volume of payment transactions but only a very small percent of the value. Cash transactions are estimated to represent 78% of the volume of all transactions in The Netherlands, 83% in Finland, 86% in Germany, and

[1] See Summers (1994) for a discussion of the nature of transactions in a market economy and the usage of money and credit through the payments system to facilitate transactions.

Table 1
The US payment system: Volume, value, and average transaction amount [a]

	Volume (billions)	Percent volume (%)	Value (billions)	Percent value (%)	Average value
Debit card	1	1	$45	0	$41
Credit card	14	18	$731	0.1	$53
Check	62	78	$71,500	9.3	$1159
ACH	2	3	$8370	1.1	$3639
Securities transfer	0.1	0	$178,700	23.3	$1,506,745
Wire transfer	0.1	0	$506,645	66.1	$3,992,467

[a] Numbers have been rounded off; data are for 1994. Source: Bank for International Settlements (1995b).

90% in the UK (Boeschoten, 1992; Virén, 1994). Cash use in the US varies from 50% of all transactions (Evans and Schmalensee, 1993) to 75% (US Department of the Treasury, 1995), to 83% (Humphrey, 1994) depending on the method used to generate the estimate. In value terms, however, cash accounts for less than 5%, and in the US less than 1%, of payment expenditures. This is because the average value of a cash transaction is relatively small, perhaps only $5 in the US although higher in other countries. Unfortunately, data on cash transactions represent estimates from one-shot surveys or expert guesses and are woefully inadequate compared to data for non-cash transactions.

2.2. Non-cash payments in the US

It is estimated that 79.2 billion non-cash payments totaling more than 750 trillion dollars were made in the US during 1994. Table 1 provides aggregate data on the number of and value of selected non-cash payments of various types. These are ranked by the aggregate value of transactions over each transfer system. The left panel has the estimated number of transactions (in billions) for payments by card, for check clearings, for automated clearing house (ACH) transfer systems, for securities transfer systems, and for wire transfers. The right panel has estimates of the dollar value of transactions for these respective payment mechanisms. While check clearings represent about three-quarters of the number of transactions on these payment systems, the value of transactions cleared through the check mechanism is estimated to be less than 10% of the value transferred on such systems. [2] In contrast, the number of transactions

[2] If securities and wire transfers are excluded from non-cash payments, then debit card, credit card, check, and ACH transfers represent 1%, 18%, 78% and 3% of the volume of transactions and 0.05%, 0.09%, 80%, and 10% of the value of transactions, respectively. As noted above, estimates of the volume of transactions effected using cash range from 50% to 83% of transactions. Cash accounts for less than 1% of the value of payment expenditures. If, for example, an estimate of 73% is used for the volume of cash transactions, then cards, checks, and ACH transfers represent about 5%, 21%, and 1% of transactions, respectively.

Table 2
Annual number and composition of non-cash transactions per person, 1994

	Total [a]	Check	Credit card	Debit card	ACH and giro	Percent electronic [b] (%)
US	302	237	52	4	9	22
Europe [b]	115	32	4	11	68	63
Japan	41	3	3	0.1	35	78

[a] Data have been rounded off so that they will add to the total. Source: Bank for International Settlements (1995b) and correspondence with central banks. Europe is defined in footnote 3.
[b] Paper-based payments include checks and paper giro transactions. Electronic payments include credit cards, debit cards, and electronic giro and ACH transactions.

cleared on large-value wire transfer systems and securities transfer systems is estimated to be less than one percent of the number of non-cash payments, but their value represents more than 80% of non-cash payments value. As noted in more detail below, the average value of a payment and its purpose – to initiate a point-of-sale transaction, to pay a bill received in the mail, to pay employees and business suppliers, or to complete a large-value financial transaction – largely determines the payment instrument that is typically used.

2.3. Differences in payment structure across countries

As illustrated in Table 2, there are important differences in the level of and composition of non-cash payments across developed countries. The US payment system relies heavily on checks while Europe and Japan focus on bank giro or postal giro payments. [3] Credit cards are used more frequently in the US, whereas Europeans more frequently use debit cards. Japan's use of either card is quite small, primarily due to its preference for cash transactions at the point-of-sale.

While, in the US, 237 (consumer and business) checks are written per person per year, 32 checks per person are written on average in Europe and only three in Japan. Overall, in the US, more than 2.8 times the number of checks per person are written than in Canada, France, or the UK, the three other developed countries that have the next highest use of checks. The US also writes 20 or more times more checks per person than does Germany, Japan, Italy, Belgium, The Netherlands, Sweden, or Switzerland. In contrast, Europe and Japan have come to rely on credit transfer giro payments for the majority of their non-cash transactions. Because of giros, other developed countries have a considerably higher share of their non-cash payments in electronic form than does the US. However, since the number of non-cash payments per person in the US is so great – over 2.5 times that of Europe and 7 times that of Japan, the actual

[3] Europe is composed of an average of 11 countries: Belgium, Denmark, Finland, France, Germany, Italy, The Netherlands, Norway, Sweden, Switzerland, and the UK.

number of US electronic payments per person is close to those of Europe and exceeds those of Japan.

3. Demand for cash payments

3.1. Cash use, seignorage, and general equilibrium payment models

As noted above, cash transactions are usually estimated to comprise from 75% to 90% of all transactions in developed countries. Cash transactions account for an even greater portion of all transactions in developing countries since the non-cash side of their payment systems is considerably smaller. While the average value of a cash transaction is the lowest of any payment instrument, seignorage benefits for governments can be large. The cumulative benefit from seignorage in the US has been around $400 billion, which is equal to the value of US currency and coin in the hands of the public in 1996. The annual benefit from seignorage is approximately equal to the value of newly issued currency, which averages almost $30 billion a year. [4] In addition to this direct benefit to the government from seignorage, there is an interest benefit as well since the $400 billion represents an interest-free, long-term loan from the public to the government. At an opportunity cost of 5.5% (the current average federal funds rate), this generates $22 billion a year in (implicit) revenues for the US Treasury. Of course, from the perspective of cash users, these benefits to the Treasury represent a cost of using currency, along with the loss in purchasing power due to inflation. [5]

General equilibrium analyses of payment arrangements have typically focused on the use and trade-offs associated with currency and other asset holdings, noting that earning assets can be held and accessed to make payments using existing non-cash payment instruments. Early general equilibrium studies, typified by the overlapping generations models of Samuelson (1958) and

[4] The Federal Reserve places currency into circulation at face value but purchases it at production cost from the Bureau of Printing and Engraving. Coin is purchased at face value from the US Mint. The difference between production cost and face value represents seignorage, but is overstated somewhat since a certain amount of currency needs to be physically replaced each year as it wears out. Hence the desire to replace the $1 note with a coin, a recent attempt at which the US – in contrast to other countries – has failed (Caskey and St. Laurent, 1994).

[5] Interestingly, some 60% of the value of US currency is estimated to be held outside of the country (Porter and Judson, 1996), so the majority of seignorage cost is borne by non-US citizens (largely in Mexico and South America). Germany is similarly advantaged since around 35% of it's currency is also held by foreign nationals (Boeschoten, 1992), most likely concentrated in Eastern Europe. In these instances, Grisham's Law leads to a significant transfer of income from developing countries to a developed country but creates the possibility that some time in the future – when currency outside a country "comes home" – this foreign-financed debt will be retired, although no sinking fund is being created for this eventuality (Sprenkle, 1993).

the spatial separation model of Townsend (1980), concentrated on the role of money as a medium of exchange. Such studies, noted that agents sometimes trade goods for fiat money and then trade fiat money for the goods they desire. These models were concerned with why currency is held, despite the fact that currency is dominated by Treasury bills that earn positive nominal yields. Indeed, Kocherlakota (1996) dubbed this fundamental question the "rate of return dominance puzzle". [6]

True, cash use has an opportunity cost which rises with the interest rate but this cost needs to be compared with the inconvenience of making more trips to an ATM and the expenses incurred – directly and indirectly – for use of non-cash instruments. [7] In addition, the opportunity cost of holding "excess" cash for personal use is surprisingly low. As Sprenkle (1993) has noted, a person who withdraws $200 every other week from a NOW account paying 3%, instead of $100 each week, will raise his average cash holdings from $50 to $100 and incur an opportunity cost of only $1.50 a year. As it happens, $100 is the average value of US household cash holdings in 1995. This is a $48 reduction from 1984 when household cash holdings averaged $148, a reduction which saved only around $1.50 a year per household. With such a small benefit for consumers from conserving on cash use, it is clear that factors other than simple opportunity cost must play the primary role in determining US and other country holdings and use of cash. However, opportunity cost and inflation have been the prime considerations in the extant general equilibrium literature, most likely because of a lack of analysis or data to identify or determine the empirical importance of various "non-price" determinants of consumer use of cash.

More recent general equilibrium models also reflect this concentration as they have focused on analyzing the welfare effects of inflation, as well as incorporating the payment transaction and seignorage costs associated with cash holdings. Cooley and Hansen (1989), for example, incorporated money into a real business cycle model that used a cash-in-advance constraint. Their results suggested that the business cycle would be the same regardless of whether there was a high inflation regime or low inflation regime, but that an "inflation tax" resulting from seignorage costs would cause agents to consume more leisure than otherwise and would imply higher welfare costs in the high inflation regime. Using a different framework, Imrohoroglu and Prescott (1991) examined economies where agents hold currency and deposits at financial institutions to smooth out their consumption over time. Their model suggests that from a

[6] Kocherlakota also argues that the so-called "rate of return dominance" puzzle motivates much of modern monetary theory and that transactions cost models and money-in-the-utility-function models fail to capture "the essential technological forces (for example, lack of communication) that disrupt the process of fully centralized exchange and thereby generate a demand for money".

[7] See Santomero and Seater (1996).

welfare standpoint, taxes on labor are preferable to "seignorage taxes" that are borne by agents that hold currency for transactions purposes. Also, Lacker and Schreft (1996) have used a model with a stochastic environment in which the mix of credit and cash were determined endogenously to demonstrate that inflation would influence *real* interest rates and would have sizeable welfare costs as a result.

The effects of technological innovations on the economy, which influence the costs of alternative transactions technologies, have been studied in a general equilibrium context by Prescott (1987), Cole and Stockman (1992), and Marquis and Reffett (1994). For example, Prescott (1987) demonstrated that checks will be used more frequently and for larger purchases in countries that have a relatively higher cost of capital. The model of Cole and Stockman (1992) suggests that the probability that transactions will not be paid for with fiat money rises with a lower cost of an alternative transactions technology or with a higher nominal interest rate. Marquis and Reffett (1994) focused explicitly on the welfare costs of inflation and on how cash management techniques that rely on costly accounting systems are used by market participants. They argue that when payment system participants alter their means of payment in order to avoid the tax that inflation imposes on cash transactions, then inflation leads to more of an economy's real resources being channeled into the payment system and into the financial sector more generally. The growth in payment and financial activities comes at the expense of output and at the expense of activities that would enhance economic growth. As a consequence, consumption is lower than it otherwise would be and a slower growth rate of the economy leads to cumulative losses over time. In sum, inflation causes a misallocation of resources towards payment system activities and a reduction in welfare.

3.2. Benefits of cash use

Among developed countries, Japan, in contrast to Europe and especially the US, continues to rely heavily on cash for transactions. This is inferred from Table 2 where Japan only had 41 non-cash payments per person per year while Europe had 115 and the US had 302. Low use of non-cash payment instruments implies a high use of cash. This is supported below when cash holdings across countries are compared.

Among the advantages of cash use is the fact that it represents final payment. Thus cash is often more readily accepted than non-cash instruments which, except for giro payments or a real time direct debit, are provisional or reversible under certain conditions. [8] Consequently, the creditworthiness

[8] Checks may not be paid when presented (due to insufficient funds in a payor's account), firms may refuse to pay a bill of exchange, customers can reverse credit card transactions, and makers of promissory notes may default.

of the payor often needs to be determined with a non-cash instrument while credit risk for cash (from counterfeiting) is lower. Using a general equilibrium model, Bernhardt (1989), has demonstrated that in larger economies, with limited communication, information concerning the repayment of credit may diffuse too slowly to deter payment system participants from reneging unless credit is restricted in its magnitude. In these larger economies, cash, which does not have a redeemability problem, and credit both play essential roles. However, in close-knit economies, where enduring trade relationships are valued, credit is optimal.

Another important feature of cash is that it can be reused immediately while collection of non-cash instruments can be delayed when drawn on non-local payor institutions. Models developed by Sargent and Wallace (1982), Prescott (1987), Townsend (1989), Mitsui and Watanabe (1989), Hodrick et al. (1991), Cole and Stockman (1992), Schreft (1992) and Lacker and Schreft (1996), among others, have examined the coexistence of private debt (e.g., resource-costly bank drafts and credit transactions) with a demand for fiat money. A critical feature of such multiple payment method models is that currency holdings depreciate with inflation and that checks drawn on the deposit liabilities of banks, or alternatively credit arrangements, have transaction costs associated with information gathering, clearance and settlement. [9] The chief role for fiat money is to economize on costly information collection and processing costs.

Lastly, cash is divisible and can be used to complete other transactions of differing amounts while checks or bills of exchange (when signed over to another payee) are only poorly adapted for transactions of a differing amount. Credit and debit card payments, of course, do not have this attribute at all.

Offsetting these benefits, cash is difficult (and expensive) to use when the amount needed for a payment is very large. Cash is also inconvenient when payment is required at a substantial distance from the payor, as occurs for most bill payments. More importantly, in countries with a relatively high crime rate, the risk of loss and/or threat to personal safety is a primary reason why users demand non-cash payment instruments. Indeed, if there was little risk of theft or loss, there would be little need to use non-cash instruments at the point-of-sale because the disadvantages of cash use – large values, distance, and loss – would be minimal while the advantages – payment finality, timeliness, and divisibility – would dominate.

Japan continues to rely on cash for point-of-sale transactions primarily because it is a much safer place than the US or even Europe (Federation of Bankers Associations of Japan, 1994a, b). But when the disadvantages to cash use loom large, non-cash instruments will be favored. Today, even in Japan, this occurs for bill payments (inconvenient distance) as well as for business

[9] Currency depreciation resulting from inflation is sometimes called an "inflation tax" or a "seignorage tax" in this literature.

Table 3
Average value per transaction for six payment instruments (1993, US dollars)

	Cash [a]	Debit card	Credit	Check	Giro	Wire [b] Transfer (million)
US	5	44	45	1147	4602	4.2
Europe [c]	6–14	52	91	3405	14,423	4.9
Japan	25	165	163	79,754	3820	93.1

[a] Boeschoten (1992), Table 1-1, p. 200.
[b] Europe averages CHAPS (UK), SAGITTAIRE and TBF (France), EAF (Germany), and ME (Italy); Japan is BOJ-NET; the US averages CHIPS and Fedwire.

disbursements and financial payments (large values and inconvenient distance). In other countries, particularly the US, a higher rate of violent crime is a statistically significant incentive to replace all but the lowest value cash transactions with non-cash instruments. [10]

3.3. Average value per transaction

The average value per transaction for six payment instruments for the US, Europe, and Japan are shown in Table 3. As seen, cash transactions have the lowest average value, debit and credit cards are next highest, followed by check and giro payments which are markedly higher, with wire transfer payments in the millions of dollars per transaction. [11] Consequently, cash and cards are typically tied to smaller value point-of-sale transactions while wire transfers are almost exclusively used for large-value financial (foreign exchange, stock, or commodity market) payments as well as interbank funding and business transactions. Checks, of course, are also heavily used at the point-of-sale but their higher average value indicates that they are also used for larger value business payments. ACH transactions include an even larger business payment component, such as payroll payments and cash management transfers, along with government benefit payments (retirement, Social Security) and a sprinkling of consumer pre-authorized bill payments.

3.4. Cash holdings as an indicator of cash use

As data on the volume or value of cash transactions (a flow) are very sparse and uncertain, the use of cash for payment transactions has instead been in-

[10] A higher incidence of violent crime is positively and significantly associated with the increased use of non-cash payment instruments across 14 developed countries (Humphrey et al., 1996).

[11] Japan's average value per wire transfer is high because many of its transfers represent settlement payments for sums of transactions rather than individual transfers. Japan's average check value is high because it is almost exclusively used for a small number of large value business/financial transactions.

Table 4
Indicators of cash use, 1993

	Cash/M1	Cash/GDP
US	0.11	0.02
Europe	0.14	0.04
Japan	0.27	0.08

ferred from the value of cash holdings (a stock), either relative to the money supply (M1) or relative to GDP. These ratios are shown in Table 4 for the US, Europe, and Japan for 1993. [12] Cash holdings in Japan are effectively twice those in Europe or the US, suggesting a much greater use of cash for payments. But Europe is far from homogeneous. Within Europe, Switzerland, The Netherlands, and Belgium all have a higher currency/M1 ratio than Japan while Sweden, Norway, Denmark, the UK, and Finland have a lower ratio than the US. This result is evident from Table 5 which shows the cash/M1 and cash/GDP ratios for 14 countries during 1993 as well as a decade earlier. [13]

3.5. Changes in cash holdings over time

Comparing the level of these ratios between 1983 and 1993, the ratio of cash to M1 fell – sometimes dramatically so – for all but one European country (Italy), rose slightly in Japan, and was constant in the US. Similar results were obtained for the ratio of cash to GDP over the same period: this ratio fell for all but one European country (Germany) while both Japan and the US experienced a slight rise. [14] The clear implication from changes in these two ratios is that cash use is falling in Europe. [15]

[12] European data reflects the total US dollar value of coin and currency (1993 exchange rate) for 11 countries divided by the total dollar value of M1 or GDP. It is equivalent to a weighted (not a simple) average of individual country ratios.

[13] Since cash balances can be held for prudential or speculative purposes (hoarding), in addition to pure transaction purposes, observed levels of cash holdings only approximate the level of cash use. Cross-country evidence suggests that hoarding is especially important in Switzerland, The Netherlands, and Germany but not important in Denmark, Finland, France, Norway, or the UK (Boeschoten, 1991, 1992). As seen in Table 5, countries where hoarding is thought to be important also have relatively large cash/M1 and cash/GDP ratios distorting, somewhat, the implied differences in cash use across countries.

[14] Interestingly, the opportunity cost of cash, measured by the difference between a three month interest rate and the inflation rate, rose in Europe – on a population weighted basis – and fell in the US and Japan over this period.

[15] The results for Germany and the US must be taken with caution since around 35% of the value of German currency and 60% of US currency is held outside the country (Boeschoten, 1992; Porter and Judson, 1996). Only if these percentages have remained stable over 1983–1993 is it possible to accurately infer domestic use of cash in either the cash/M1 or cash/GDP ratios.

Table 5
Cash holdings and annual non-cash transactions per person

	Cash [a] to M1 [b] ratio		Cash to GDP ratio (%)		Value of cash holdings per person (US 1993 dollars)		Non-cash transactions per person	
	1983 (1)	1993 (2)	1983 (3)	1993 (4)	1983 [c] (5)	1993 (6)	1983 (7)	1993 (8)
Switzerland	0.35	0.34	9.90	7.86	3524	2676	na	67
The Netherlands	0.33	0.25	6.52	6.49	1155	1300	85	146
Belgium	0.41	0.30	9.12	5.95	1411	1187	35	101
Italy	0.15	0.16	6.37	5.75	610	1191	13	29
Sweden	0.10	0.09	6.21	4.63	940	1210	na	93
Norway	0.33	0.10	5.41	4.61	967	1196	na	97
Germany [d]	0.23	0.20	3.53	3.89	742	995	86	139
France	0.17	0.15	5.08	3.58	810	759	88	157
Denmark	0.12	0.09	3.01	2.88	604	775	na	124
UK	0.26	0.08	3.72	2.84	410	520	74	115
Finland	0.26	0.07	2.07	1.97	277	438	83	120
Japan	0.25	0.27	7.59	8.43	1795	2387	19	39
Canada	0.29	0.20	3.16	3.43	529	716	88	128
US [d]	0.11	0.11	1.79	2.06	328	509	184	292

Sources – IMF: International Financial Statistics; BIS: Payment Systems in the G-10 countries (various issues).

[a] Cash held by the public (equals currency and coin in circulation less that held by the banking system and the government).

[b] M1 equals demand deposits (transferable deposits) plus cash (defined above).

[c] The 1983 figures are deflated using changes in each country's CPI.

[d] It is estimated that 35% of German currency and 60% of US currency is held outside these countries, and the figures are adjusted accordingly (Boeschoten, 1992; Porter and Judson, 1995).

An alternative way to illustrate the general reduction in cash use is to compare the value of cash holdings per person (in constant US dollars) with the number of non-cash transactions per person. Across countries, a high number of non-cash transactions per person is associated with a low stock of cash per person (comparing Column 8 with Column 6 in Table 5). The estimated relationship across 14 developed countries annually over 1987–1993 suggests that a 6.8% rise in non-cash transactions is associated with a 10% reduction in cash holdings (Humphrey et al., 1996).

3.6. Cash holdings and availability of POS terminals

The decreased reliance on cash for transactions is associated with the increased availability of credit and debit card terminals at the point-of-sale. This is seen in Fig. 1(a) which illustrates the negative and statistically significant relationship between the availability of electronic funds transfer (EFT) POS ter-

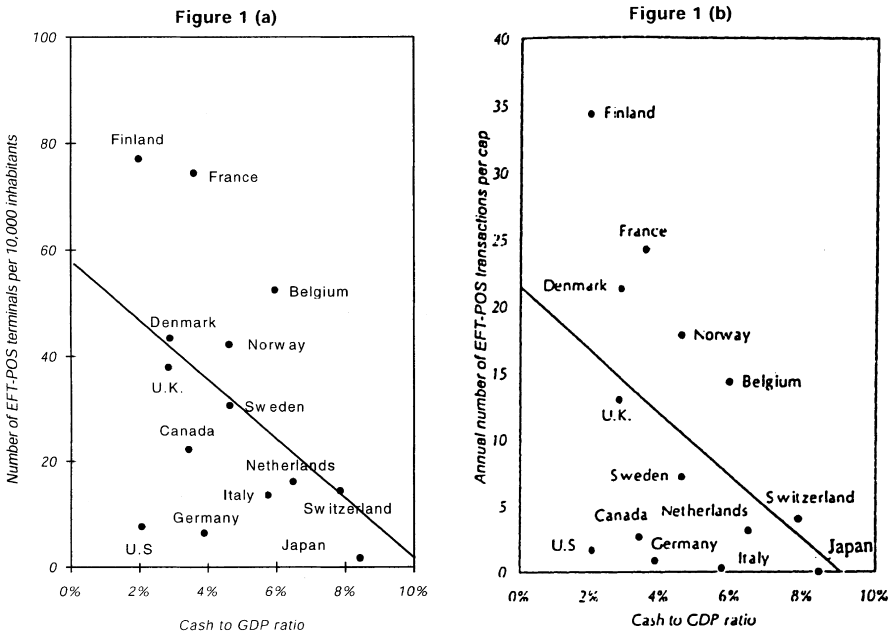


Fig. 1. EFT-POS terminals and transactions per capita with cash to GDP ratios, 1993. Source: Bank for International Settlements (1994) and Humphrey et al. (1996).

minals (terminals that accept credit and debit cards) across 14 developed countries and the ratio of cash holdings to GDP for 1993 ($R^2 = 0.21$). The increased availability of EFT-POS terminals leads directly to increased use, as evidenced from a very similar significantly negative relationship in Fig. 1(b) between the annual number of EFT-POS transactions per person and the same cash/GDP ratio ($R^2 = 0.25$). [16]

[16] The substitutability of credit cards for cash is dependent upon pricing and a country's cultural attitude toward credit. Per transaction credit card fees are largely borne by retailers that accept them for payment, so credit cards are perceived by consumers as a low cost, delayed payment substitute for an immediate cash payment. Countries that have the highest incidence of credit card use per person (the US and Canada) also have a culture in which personal credit is quite acceptable. Credit card use is lower in Europe in part because of a history of relying on savings (rather than credit) for many consumer expenditures, a tradition reinforced by a reliance on giros – a method of immediate and final payment – for consumer bill payments. Giro transactions rarely involve the extension of credit while payments by check – a deferred and provisional payment – provide for a short (1–2 day) extension of credit before a customer's account is actually debited.

3.7. ATMs and cash use

While debit and credit card transactions on POS terminals can directly substitute for cash payments, ATMs make it easier to obtain cash: this can have two opposite effects on cash holdings. [17] By making cash more accessible, ATMs would permit consumers to increase their use of cash for smaller value transactions and so may raise cash holdings. Alternatively, the greater convenience of ATMs as a way of obtaining cash may mean that consumers withdraw smaller amounts each time and make up the difference by visiting an ATM more frequently, and so may on balance reduce average cash holdings.

The relationships shown between cash holdings and the number of ATM terminals in Fig. 2(a) and the number of ATM transactions in Fig. 2(b) is in both cases insignificantly positive or negative (neither R^2 is greater than 0.02). Thus, in this cross-country comparison, ATMs have no significant effect on cash holdings (in contrast to the EFT-POS results where cash holdings are reduced). However, two studies of ATMs and cash use for individual countries suggest that cash holdings fall as ATMs have expanded, both in the US (Daniels and Murphy, 1994) and in Israel (Paroush and Ruthenberg, 1986). In addition, direct survey evidence indicates that average cash holdings per US household have been reduced from $148 to $100 between 1984 and 1995, with the percent of household expenditures made in cash falling from 30% to 20% (Porter and Judson, 1996, Table 1). This reduction occurred at the same time as the availability of ATMs more than doubled, although other factors – such as expanded use of credit cards and the introduction of debit cards – represent contributing influences to the decline in cash holdings and use. [18]

3.8. ATMs and the delivery of banking services

In terms of transaction frequency, the primary use of a bank branch is to withdraw cash and make deposits. ATMs represent a way to do this at a substantially lower cost per transaction. Indeed, an ATM cash withdrawal costs

[17] Most ATM transactions are cash withdrawals. In the US, for example, 86% of ATM transactions are cash withdrawals, 10% are deposits, 3% are account transfers, and 1% are bill payments – typically bank mortgage payments (Board of Governors of the Federal Reserve System, 1991).

[18] An additional influence on the demand for cash concerns hoarding and illegal activities. In many countries currency and coin denominations follow a 1-2-5 sequence, as in $1, $2, $5, $10, $20, $50, $100, $200, $500. Surveys of currency use by denomination are rare but when done they find that the vast majority of currency transactions involve the use of low value denominations (Boeschoten and Fase, 1989; Virén, 1994). High value denominations are invariably used for the underground economy to avoid taxes and for illegal activities and so are "hoarded". For 14 OECD countries, Boeschoten (1992) has estimated that such hoarding may account for one-third of the total value of currency in the hands of the public.

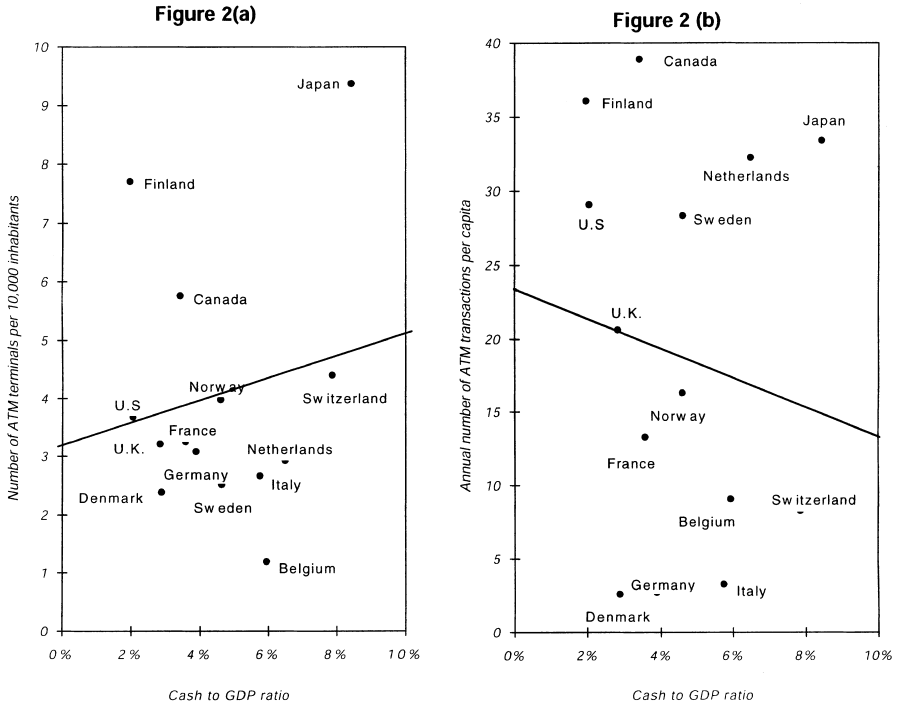


Fig. 2. ATMs and ATM transactions per capita with cash to GDP ratios, 1993. Source: Bank for International Settlements (1994) and Humphrey et al. (1996).

from one-third to one-half the cost of a withdrawal at a full service branch. The cost savings expected by banks from shifting consumers to ATMs have not been generally realized, however. This is because depositors have found ATMs so convenient that they use them 2–3 times more often than they did a bank branch, essentially offsetting the lower cost per ATM transaction. Econometric analysis suggests that, on balance, substituting ATMs for branches has not reduced costs (Humphrey, 1994). Even so, due to the fees now being charged for ATM use, profits from substituting ATMs for branches are estimated to be positive and significant, although it is not possible to determine if this level of profit represents a normal return on the capital invested in ATMs since their introduction.

4. Demand for non-cash payments

The demand for non-cash payment instruments is the second part of essentially a two-step nested decision process. The first step involves consumer and

business demand for transaction balances while the second step involves the choice of the non-cash payment instrument which, when used, debits this balance. Because of its implications for money supply measurement, and thus the effective application of short-term monetary policy, the focus of research to date has been on the first step, not the second.

4.1. Demand for money and asset substitution models

Accurate measurement of a monetary aggregate involves the determination of how assets – cash, demand deposits, other transaction balances, savings and time deposits, etc. – substitute with one another in response to changes in interest rates (reflecting one aspect of user costs) and income level. These substitution relationships are in turn affected by institutional developments such as: (a) the expansion of money market mutual funds and corporate cash management in the US during the 1970s; (b) the establishment of new interest earning transaction accounts and the removal of interest rate ceilings in the early 1980s; and (c) the development of sweep accounts in the 1990s. [19]

The effects of transaction costs borne by payments system participants on the optimal growth of the money supply has been studied by Prescott (1987) and Mitsui and Watanabe (1989). Their basic argument is that the growth in the money supply affects inflation which, in turn, influences the nominal rate on deposits. The larger the deposit rate, the greater is the use of checks (with a corresponding reduction in the use of fiat money), and the smaller is steady state consumption. Hodrick et al. (1991), Schreft (1992), and Lacker and Schreft (1996) examine the effects of payment system transactions costs on the velocity of money. Hodrick et al. (1991) used a standard cash-in-advance model to argue that the velocity of money depends on the substitutability between "cash goods" and "credit goods". Schreft (1992) and Lacker and Schreft (1996) examine a stochastic economic environment in which the mix of credit and fiat money are determined endogenously. Variability in the velocity of money depends on the structure of credit costs and the variability of nominal interest rates. If the marginal cost of credit rises slowly with its increased usage, then the velocity of money is quite interest-elastic (i.e. sensitive) and variable relative to interest rates. In the stochastic model, inflation affects real interest

[19] Money market mutual funds and corporate cash management expanded rapidly in response to the historically high inflation rates initiated primarily by oil price increases during the 1970s. As bank deposits shifted to money market mutual funds and other higher interest earning assets, legislation was passed permitting banks to establish interest earning (consumer) transaction accounts and depression-era interest rate (Regulation Q) ceilings on savings and small denomination (less than $100,000) time deposits were removed in the early 1980s. Sweep accounts automatically shift the transaction balances to savings accounts and vice versa, dramatically lowering the average reservable balances (reducing non-interest earning reserve balances by two-thirds by early 1997).

rates, and welfare costs are greater than would be derived using a non-stochastic model.

Asset substitution effects are also important because their size and stability determines how various asset holdings could be aggregated into a proper measure of the money supply (which assists in formulating an appropriate interest rate target to achieve through open market operations). Overall, estimated asset substitution elasticities are typically low, vary over time, and even suggest that some assets are complements rather than substitutes (see the survey by Barnett et al., 1992). [20] As a result, a Divisia index of monetary assets is supported theoretically and empirically over the simple-sum aggregation procedure currently in use (but justified only when asset substitution elasticities are quite large, in fact infinite). [21] While this literature has investigated the substitution between cash and transaction balances (along with other assets), more detail is needed to determine the demand for and substitution among the various non-cash payment instruments which rely on transaction balances to make payments. [22]

4.2. Surveys of payment instrument use

Only recently have data been available on payment instrument use within and across countries (Bank for International Settlements, 1993b; European Monetary Institute, 1996). Other surveys have (in the US) looked at family – but not business – use of cash and different payment instruments (Avery et al., 1986, 1987), examined changes in bank fees and service levels (Hannan, 1994; Elliehausen and Wolken, 1992), and discussed credit card use and pricing (Canner and Luckett, 1992). [23] Information on payment instrument use in some other countries is noted in Lempinen and Lilja (1989) and especially Vesala (1993).

4.3. Demand for non-cash 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

[20] Hancock (1985) demonstrates these results at the bank level as well.

[21] A related literature, surveyed in the following section, has used bank data to determine scale, scope, and frontier efficiency results associated with bank production of deposit, loan, and other services. However, these studies are supply or cost-orientated and, due to a lack of data, are unable to distinguish among the various payment instruments supplied to users.

[22] Although ATM terminals are used to debit transaction (or savings) balances, and have been called an "electronic funds transfer", their main purpose is to substitute for bank branches in providing cash. Consequently, ATMs are not included here as a non-cash payment method.

[23] Additional surveys cover a broad range of small business and consumer finance issues, but only touch on the use of checking accounts and other bank services (Elliehausen and Wolken, 1990; Kennickell and Starr-McCluer, 1994). Credit card data has long been available in Nilson (1995), *The Nilson Report* (an industry newsletter).

Table 6
Correlation matrix for per person use of payment instruments (14 developed countries; 1987–1993)

	Paper giro (1)	Electronic giro (2)	Credit card (3)	Debit card (4)	Cash (5)
Check	−0.37	−0.39	0.81	−0.11	−0.42
Paper giro		−0.09	−0.23	0.30	−0.14
Electronic giro			−0.50	0.12	0.19
Credit card				−0.17	−0.38
Debit card					−0.39

payment instruments from pooled time-series cross-country analysis using data developed by the BIS (Bank for International Settlements, 1993b) or by using case studies. The relationships among the per person use of five non-cash payment instruments plus cash is illustrated in Table 6. [24] Of the 15 pairwise correlation coefficients, 11 are negative suggesting substitution. However, these substitution effects are weak since the largest negative value is −0.50. [25] The simple elasticity between real cash holdings per person and the annual number of non-cash payments per person is −0.68, 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 data within a standard demand function framework suggested that while cash and non-cash payment instruments use are negatively related, the implied substitution between them is due more to the differences in use across countries than it is to changes in use over the seven-year time period, a result that also applies to the substitution among the five non-cash instruments as well.

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 −0.09 to −0.26 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 et al., 1996).

[24] The correlations relate the annual number of transactions per person across 14 developed countries in each year over 1987–1993. Cash use reflects the real value of cash holdings per person, also across countries and for each year.

[25] The strongest relationship is between checks and credit cards which, with an r of 0.81, suggests strong complementarity. When the US (which uses the most checks and credit cards) is excluded, the complementarity between checks and credit cards is reduced but not eliminated ($r = 0.52$). 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.

Payment instrument price information is more suitable in Scandinavia, since those countries are among the few that actually charge a price per transaction and the price is higher for paper-based non-cash instruments to better reflect the lower supply costs of most electronic payment alternatives. Based on bank price and quantity data for ATM cash withdrawals, checks, and point-of-sale electronic payments in Norway over 1989–1995, the own price elasticities were, respectively, −0.96, −0.75, and −0.87 and all were statistically significant (Humphrey et al., 1997). With this degree of price sensitivity, Norway has moved from having 90% of its non-cash payments in paper form in 1987 to having 60% in electronic form in 1996. Unfortunately, we know of no other payment price elasticity estimates. [26]

4.4. Historical changes in payment use

Taking a historical and more descriptive view of non-cash payment use, developed country payment systems shifted long ago from barter to widespread cash use and, about 150 years ago, replaced many cash transactions with paper-based instruments (primarily for business transactions over large distances – Spahr, 1926). The further shift from paper to electronic payments is quite recent, starting only about 20 years ago. Importantly, the shift from cash to non-cash payments has progressed at a different rate and taken a different path in Japan, Europe, and the US Japan has high cash holdings per person, low non-cash use, and a high percent of electronic payments. As noted above, Japan's low rate of violent crime suggests that the need to develop and use alternative non-cash instruments for retail payments has been weak: cash use is high for point-of-sale transactions while non-cash payments are largely electronic and almost entirely for bill payment and employee payroll disbursements. Europe, in contrast, has essentially an intermediate level of cash holdings per person, an intermediate level of non-cash use, but a similarly high percent of electronic payments. Like Japan, electronic payments in Europe are used mostly for bill payments and payroll deposits through a bank or postal giro, but there is a heavier reliance on electronic point-of-sale payments using debit cards. Finally, the US has low cash holdings per person, high non-cash use, and a low percent of electronic payments.

4.5. The shift away from cash and paper toward electronic payments

The trade-offs between cash holdings, the use of paper-based instruments, and electronic payments are illustrated in Fig. 3. This figure is divided into four

[26] Research in this area is severely hampered by a lack of appropriate data. Information relating to how banks price their payment services is outlined in Weinberg (1994), Llewellyn and Drake (1993), Murphy (1988), Carraro and Thornton (1986) and Humphrey (1984).

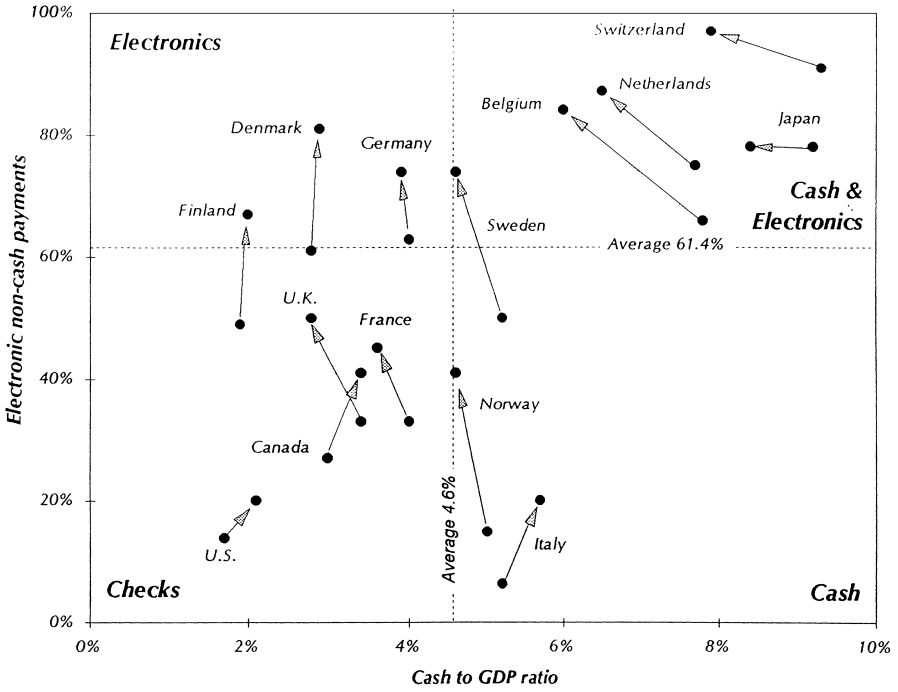


Fig. 3. Percent of electronic payments and cash to GDP ratios, 1987 and 1993. Note: The averages for the percent of electronic payments (61.4%) and for the cash to GDP ratios (4.6%) indicated in this figure are for 1993. In 1987, these averages were 47.3% and 5% respectively. Source: Humphrey et al. (1996).

quadrants using the average percentage electronic payments with respect to total non-cash payments (61.4%) and the average cash to GDP ratio (4.6%) for 14 developed countries in 1993. An arrow is drawn from the point each country was in 1987 to the point it moved to in 1993.

At either point in time, most countries fall into the lower left and upper right quadrants, indicating that most countries either use mostly paper-based non-cash payments and little cash (US, Canada, France, and the UK) or use mostly electronic non-cash payments and have high cash holdings (Sweden, Belgium, The Netherlands, Switzerland, and Japan). In contrast, three countries use mostly electronic methods and have low cash holdings (Finland, Denmark, and Germany) while two countries appear to rely strongly on paper-based non-cash payments and also have relatively high cash holdings (Italy and Norway). Except for Japan, all countries have increased their relative use of electronic non-cash payments and all but the US, Canada, and Italy have at the same time reduced or essentially maintained their cash holdings as a percent of GDP. Overall, it is clear that the general trend in Fig. 3 over the 1987–

1993 period has been to shift from holding cash to increasing the use of electronic payments. One factor influencing this shift is the demand-side externality which requires that both parties to a transaction have access to the same network for the network to have value. This affects the US more than other countries because few businesses have adopted pre-authorized electronic debit programs for their bill paying customers and, in addition, even when they have such a program offer little or no incentive for these customers to use it compared to mailing a check.

5. Supply of non-cash payment services

Modern market economies typically have a two-tiered banking structure through which payment services are provided. The first tier is composed of depository institutions (banks, savings institutions, credit co-operatives) who supply payment services to end users (consumers, financial and non-financial firms). [27] The second tier includes the central bank and other payment intermediaries who often supply clearing and settlement services to banks.

Freeman (1996a, b) not only focused on money's role as a medium of exchange, but also was interested in the role of banks in the clearing of private debt. His general equilibrium model captures the following basic features of the payments system: (1) some purchases are made with debt; (2) debts are repaid with fiat money; and (3) when the arrivals of creditors and debtors at a central clearing area are not perfectly synchronized, then there emerges an active market in second-hand debt (i.e., some proportion of debt is cleared through third parties, which buy the debt of those in immediate need of currency). In Freeman's model, fiat money plays a special role with regard to the payment of debts.

Freeman demonstrated that in an economy where fiat money serves both as a medium of exchange and as a means by which debts are repaid, non-optimal equilibria of constrained liquidity (where risk-free debts are not cleared at par value), may arise. His model suggests that such liquidity constrained equilibria, with their distortionary impact on credit markets, could be eliminated by policies that permit an elastic currency whereby the monetary authority temporarily provides enough currency to clear all debts at par. [28] Importantly, Freeman (1996a, b) also demonstrates that payment arrangements based on intermediated debt (and settled using fiat money) can achieve higher welfare than

[27] Some consumers rely on non-bank financial institutions (e.g., Western Union and money order issuers), but we do not focus on such non-cash payments here.

[28] The provision of an elastic supply of daylight credit to a large-value funds transfer system (such as Fedwire), that must be repaid with an inflow of reserves by the close of business, typifies the type of credit arrangement discussed in Freeman (1996b).

when payments are made exclusively with fiat money. Therefore, the monetary authority can sometimes improve welfare and achieve greater efficiency by participating in the secondary market for debt, but such positive welfare effects come with risks that are discussed in Section 6.

5.1. Clearing and settlement services

The supply cost of payment services for the first tier described above has been difficult to determine due to a lack of available data, a problem not encountered for the second tier. Payment clearing is an information-processing activity that is performed by depository or non-depository institutions, bank associations (such as local clearing houses or national organizations such as VISA), non-financial institutions (third-party processors), or by the central bank. Payment settlement involves the transfer of monetary assets, usually deposit balances, between the payor's bank and the payee's bank. [29]

5.2. Check scale economies

Banks typically subcontract parts of the payments process to take advantage of economies of scale, network effects, and technological change. Paper-based payment processing relies on manual labor, mechanical processing equipment, buildings, and a land/air transportation network. Electronic-based payment processing, in contrast, relies on computing equipment and communications networks and is more capital intensive. Paper-based payment processing systems initially experience significant economies of scale. However, because of their reliance on mechanical equipment and on land/air physical transportation networks, which are sensitive to the spatial distribution of delivery endpoints, these economies are eventually exhausted and additional volume is processed at essentially constant unit costs. Overall, and on average, each 10% increase in check processing volume is associated with a 9.7% increase in total production cost (Bauer and Ferrier, 1996).

Studies of the cost structure for check processing operations have of necessity focused on Federal Reserve facilities since data on commercial bank check operations are not publicly available (Humphrey, 1981a, ; Zimmerman, 1981; Bauer and Hancock, 1993; and Bauer and Ferrier, 1996). While these studies

[29] Settlement can be through book transfers using accounting entries on the books of one bank (if the payor and payee maintain deposit accounts at the same bank – an "on-us" transaction) or through book transfers on accounts at different banks (if the payor and payee maintain deposit accounts at different banks – an "on-others" transaction). In the US, the latter method represents the settlement procedure for around 75% of all non-cash payments, since around 25% of payments are on-us transfers. Settlement of an on-others interbank transfer can occur using correspondent balances held bilaterally at each of the settling banks, held at a third-party correspondent, or held with the central bank.

differ in terms of time period covered (from the 1970s to the 1990s), data used (single cross-section or panel), cost functional form (translog, Fourier), and parametric or non-parametric representations of a cost frontier, they all find similar results. Specifically, the average cost curve for interbank check processing operations is (after the pricing of Federal Reserve payment services) essentially "L-shaped" so that initial scale economies at low volumes become constant unit costs and later weak diseconomies at very high volumes. While bank and third-party check processors may reap some scope economies by providing additional "back-office" services (such as demand deposit account processing, proof of deposit, corporate check, lockbox, and disbursement services) in addition to standard check processing, it is unlikely that their broader product mix generates something other than a similar L-shaped average cost curve. This is because check processing operations – no matter who does it – involves the same technology, transportation arrangements, and labor skills.

5.3. Electronic payment scale economies

Electronic-based processing systems typically have high start-up costs because of their relative capital intensity. [30] Therefore, at low processing volumes unit costs for electronic payments are generally higher than those for checks, a ranking which is reversed when electronic payment volume becomes large. Importantly, scale economies in processing electronic payments depend upon the relative cost of centralized versus distributed processing which, in turn, is determined by the relative costs and scale economies of communication links versus computer processing. As communication costs have fallen over time, centralization of electronic payment processing has become increasingly cost effective, increasing the scale economies for processing electronic payments.

Several studies have investigated the cost structure for ACH electronic transfers (Humphrey, 1981b, , 1985; Zimmerman, 1981; Bauer and Hancock, 1995a; Bauer and Ferrier, 1996). The Federal Reserve's ACH system is a value-dated electronic funds transfer system that uses "batch" processing methods. These studies all report strong ACH scale economies, with the most recent study indicating that a 10% increase in processing volume is associated with only a 4.8% rise in production costs. The estimated average cost curve for ACH transfers is thus strongly downward sloping, rather than L-shaped like that for checks.

Scale economies of the Federal Reserve's electronic book-entry government securities transfer system has also been investigated (Belton, 1984). Book-entry securities transfer operations experienced significant scale economies whether

[30] Indeed, the cost share of communication, transit, and building expenses is 23% for checks, 38% for ACH, and 54% for wire transfers. The cost share for labor is, respectively, 49%, 21% and 14% (Bauer and Ferrier, 1996).

or not the largest office (New York, which processed about 70% of the total volume) was included. It was estimated that a 10% increase in processing volume increased costs by only 6.5%. [31]

Finally, two studies have investigated the cost structure of the Federal Reserve's Fedwire funds transfer network (Humphrey, 1982; Bauer and Ferrier, 1996). Fedwire is an electronic, real-time gross settlement (RTGS) interbank wire transfer network which directly posts transactions to central bank accounts as they occur, rather than by using a batch processing method with settlement of only the net position at a given point during the day. [32] Overall, it was found that average Fedwire costs were not significantly different from constant unit costs regardless of processing volumes. However, this result is suspect because of the manner in which the cost accounting data are reported: many Fedwire costs are distributed to processing sites on the basis of the average cost incurred for all Fedwire transfers. This arrangement biases a cross-section estimation of scale economies toward finding constant unit costs.

Although scale economies exist for most types of non-cash payments, at least up to some point, empirically significant scope economies do not appear to exist among broad classes of these payments such as checks, ACH, and wire transfers (Ferrier and Bauer, 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.

5.4. Network effects

Network size, measured by the number of delivery points, affects payment costs. There are both cost scale economies and diseconomies associated with network size. For example, the relationship between average check transportation costs and the log of the number of endpoints is U-shaped so that (typically) the lowest costs occur when checks are presented to only the highest volume endpoints. Analogously, the number of electronic endpoints affects the cost of Fedwire funds transfers (Bauer and Hancock, 1995b). Adding banks to a funds transfer network at first requires little additional hardware, but eventually,

[31] Scale economies also exist for ATM networks, which provide an electronic method for withdrawing cash and making certain other depositor transactions. Three studies have used survey data and estimate that ATM costs typically rise by only 3–5% for every 10% increase in transaction volume (Walker, 1980; Van de Velde, 1985; and Humphrey, 1994). However, these economies are limited: after reaching a certain level of monthly transaction volume, a new ATM is typically installed in order to avoid queuing problems.

[32] Such a system requires that computing and communications facilities accommodate spikes in transaction demand associated with daily peak volumes. Time-of-day pricing could reduce these spikes. Alternatively, like the bank-owned Clearing House Interbank Payment System (CHIPS), Fedwire could require that progressively higher fractions of each day's wire transfer value be sent prior to specified times each day (smoothing the demand spikes).

congestion on the network requires substantial additional amounts of hardware and software. Even so, it was found that scale economies were greater for electronic payment communication networks than for check transportation networks which require physical delivery.

5.5. Technological change

Over time, technological change can lower the cost of payment services so that with the same quantities of inputs, more payments may be processed. Technological change has typically been estimated by determining how the cost function for payment services shifts down over time. This result implies an outward shift in the supply function for non-cash payments.

Currently, the technologies used for check processing systems are moving these items through mechanical/electronic reader–sorters as quickly as is feasible without ripping the check and jamming the machine. [33] Improvements in the technologies used to transport checks – such as the electronic presentment of payment information on a check – would lower costs (Stavins, 1997) and reduce fraud but banks (and their corporate customers) have generally been unwilling to give up the (perceived) benefit of check float and only about 2% of checks are collected electronically in the US. [34] This issue has been addressed successfully in Germany where some 70% of checks below a certain value ($3700) are collected electronically. Over the last decade studies using Federal Reserve data have *not* found evidence that there has been statistically significant technological change for check processing operations (Bauer and Hancock, 1993; Bauer and Ferrier, 1996).

In contrast, computers today are able to process information faster and communication networks have faster speeds and lower costs than they did a decade ago. Thus, it is not surprising that econometric studies have generally found that technological change has been an important contributing factor in reducing the cost of processing electronic payments. Technological change is estimated to have reduced the cost of ACH transfers by about 10% per year since 1989 and Fedwire costs are estimated to have declined by 8% annually during the first half of the 1990s (Bauer and Hancock, 1995a; Bauer and Ferrier, 1996).

Overall, the effects of scale economies, network effects, and technological change have altered the level and relative costs of paper and electronic payment processing. In the mid-1980s, the real unit cost of processing an ACH transfer

[33] A thicker check and higher quality paper would be required to significantly improve reader–sorter productivity.

[34] Check float is largely a zero sum game since the float benefit from making business payments by check is, for most firms, offset by receiving checks from other businesses.

was about 6 cents, but the real cost of processing a check was about 2.5 cents. [35] Today, using the same measures for unit costs – on a per item basis, ACH transfer processing is cheaper than check processing. The supply function for ACH transfers has been downward sloping due to strong scale economies and has shifted down over time due to technological change. The supply function for Fedwire funds transfers has also been shifting downward over time but may or may not be downward sloping. In contrast, check processing costs lack scale economies (after a given volume is reached) and technological change has been minimal over the last decade.

5.6. Effects of banking structure and antitrust considerations on coordination and technological adoption

For historical and institutional reasons, a giro system never developed in the US, most likely because laws and regulations have long blocked (until recently) the emergence of nationwide banks which, in turn, led to a very unconcentrated banking system. With over 10,000 local and regional banks, strong antitrust laws, and a very large geographic area (which no doubt affected the optimal network size), the cooperation necessary to form a nationwide, centralized entity – such as a bank giro – never developed. [36] Europe and Japan, in contrast, had far fewer banks, permitted nationwide banking, had weaker antitrust laws, and a smaller geographic area. An important additional component for many of these countries was that commercial banks faced competition from postal banks which, being operated by a single government entity, had already set up (first on a local basis and then nationwide) an effective giro system for their depositors. Commercial banks, seeking a way to increase their market share of these low cost deposits, subsequently set up bank giro systems as well.

Giro payments have two benefits over checks. A (bank or postal) giro is a centralized entity, so it is relatively simple to determine if there are sufficient funds in a payor's account *before* a payment is made. Since payments are (typically) not made when payor balances are insufficient, there are no return items – or associated risk of loss – as there is with a check. In addition, being a centralized entity, it has been less costly to shift from paper-initiated giro payments to electronics. While funds are transferred among accounts electronically (as an internal accounting entry) between branches of the same bank in the US, this is more expensive to do (and more difficult to arrange) be-

[35] See Bauer and Hancock (1995a). These real unit costs are in 1995 dollars and include direct and support processing costs but leave out some overhead expenses. Importantly, the costs to payors and payees and their commercial banks of processing payments are not included here and, as illustrated in Section 9, Federal Reserve costs are only a small portion of the total social costs of providing payment services.

[36] Berger and Humphrey (1988) have used logit analysis to study the effects of interstate banking on the payment system.

tween independent banking organizations. This is a primary reason for a continued heavy reliance of checks in the US. [37]

6. Risk in the payment system

6.1. Types of payment risk

Fraud, operational, and systemic risks exist in all payment systems. Fraud risk is addressed through proper internal payment processing procedures and insurance. Operational risk is addressed through limiting access to sensitive facilities and, especially for large-value payment networks, the establishment of back-up or duplicate computer facilities. [38] Systemic risk involves the potential for a domino-like series of failures among participants on a payment network resulting from the *unexpected* failure of one or more participants to settle their intra-day net debit positions. That is, systemic risk results from making payments in a way that involves the creation of debt at an interim stage. [39] Particularly for large-value transfer systems, the creation of debt carries at least a small risk of the inability to settle, and since the real cost of making a transaction is typically very small, the mere fact that netting economizes on the number of transactions is unlikely to be sufficient to argue that uncollateral netting arrangements would dominate over real time gross settlement (RTGS) arrangements. [40]

On the non-giro large-value payment networks that have permitted net debits, and account for the vast majority of the trillions of dollars of large-value payments made in the world each day, intra-day net debit exposures often exceed bank capital positions and there has been uncertainty regarding how losses may be apportioned in the event of a settlement failure. On many such networks, legal agreements had called for the unwinding of payments both to and from a participant that failed to settle at the end of the business

[37] When accounting information for branches of the same bank is not centralized (as occurs in Russia, China, and other emerging market economies), the transfer of funds between accounts within the same bank resembles the transfer of funds between accounts between different banks. In both cases the check becomes the most cost-efficient method of payment (Sato and Humphrey, 1995).

[38] In some cases, the duplicate facilities are "hot" meaning that in the event of a breakdown there is an automatic transition to the other facility.

[39] Early analysis of systemic risk is discussed in various Federal Reserve and Bank for International Settlements discussion papers and publications (cf. Belton et al., 1987; Lindsey et al., 1988; Borio and Van den Bergh, 1993; Bank for International Settlements, 1993a).

[40] See Group of Experts on Payment Systems of the Central Banks of the Group of Ten Countries (1989) for a discussion of how the failure of a clearinghouse to settle its obligations created by novation and substitution causes concern for systemic risk.

day. [41] On one such network (CHIPS), simulations of a payment unwind resulting from the unexpected failure of a participant often raised (lowered) the net debit (net credit) positions of other participants by more than their capital positions and eliminated about one-third of day's payment value (Humphrey, 1986). [42] However, similar simulations for an Italian wire transfer network led to a considerably smaller number of systemic failures and a correspondingly smaller elimination of payment values (Angelini et al., 1996). [43] These two simulations support a conclusion that systemic risk is clearly smaller and more manageable when a large-value network is located in a country that does not service sizable domestic or international money or financial markets (so interbank net debits are not too big) and when the banking system is concentrated (raising bank capital positions to levels more likely to cover the net debits created).

The four main areas of research related to systemic risk have concerned developing theoretical and/or empirical models which address: (1) the cost-risk trade-offs associated with the different methods developed to contain systemic risk; (2) assessing the effects from pricing payments risk; (3) the potential money market and monetary policy effects if an intra-day funds market should arise where payment participants purchase funds to cover their intra-day net debits; and (4) the moral hazard implications of systemic risk for the banking system's government-provided safety net (discount window borrowing and deposit insurance). Each of these areas, and their main results, are now outlined and areas where additional research may prove useful are implied in the discussion.

6.2. Modeling the cost-systemic risk trade-off

Theoretical models of the trade-off between reducing systemic risk and the costs of the numerous ways this can be done have been developed by Schoenmaker (1996) and Angelini (1996). Large-value payment arrangements are less costly if they are structured to settle net (bilateral or multilateral) payment positions at the end of the business day – a net settlement (NS) network. It is

[41] Insurance is not a realistic option here because the exposures are so large and the incidence of actual settlement failures are so small that the situation is not actuarially sound (see Bernanke, 1990 for a similar view regarding systemic risk in the futures market).

[42] A payment unwind would not cause any problems *if* banks receiving funds released them to customers only after settlement occurred at the end of the day. However, since there are really no overnight investment opportunities after the wire networks close down and it is illegal to pay interest on bank demand deposits, customers would lose one day's interest if they had to wait until settlement occurred before using their funds. While some banks do offer some of their customers the opportunity to earn overnight interest on idle demand balances by "sweeping" these balances into interest earning accounts/assets at the end of the day, this arrangement has not been available to all and many customers feel that they can earn higher returns by investing funds received in other assets earlier in the day.

[43] On average, less than 1% of participants failed and this affected less than 3% of a day's payment value.

more costly to settle each payment as it occurs during the day – an RTGS network. [44] RTGS is more expensive because participants have to hold enough idle balances to cover all payments made during the day (as occurs on the Swiss Interbank Clearing (SIC) system or, equivalently, post enough collateral to cover all net debit positions generated when a balance is insufficient (as occurs on CHAPS). Holding an adequate idle balance is more expensive than posting sufficient collateral but either procedure eliminates systemic risk. Systemic risk is reduced on NS networks, but not eliminated, when enough collateral is posted to cover only the largest two net debits allowed (as occurs on CHIPS, which previously only covered the single largest net debit). [45]

Although an RTGS system is free of systemic risk, it will generate credit risk when a central bank operator permits payments to be made without sufficient funds – creating a daylight overdraft – but guarantees the finality of the payment (as occurs on Fedwire). Alternatively, this credit risk can be absorbed by those network participants in a net credit position who lend these funds to banks with net debits, enabling settlement to occur (as occurs on BOJ-NET). In either case, a price is being charged for the provision of daylight liquidity necessary to make payments. [46]

Both Schoenmaker (1996) and Angelini (1996) model the cost structure of NS and/or RTGS systems as well as the cost of delayed payments, raising

[44] Well-known net settlement networks are CHIPS in New York, CHAPS in London, and most of the BOJ-NET system in Tokyo. RTGS systems are Fedwire in the US, SIC in Switzerland, and a small part of BOJ-NET in Japan. Large value networks in other countries are described in Bank for International Settlements (1990a, 1993b), and European Monetary Institute (1996). Technical and legal details are outlined in Summers (1994).

[45] Additional risk reduction procedures (adopted in the mid-1980s) are also in place on CHIPS. These include the real-time control of limits placed on participants' net debit positions (net debit caps) and the ability of payment receivers to limit the value of payments they are willing to receive from other participants (bilateral net credit limits). These arrangements, plus the posting of collateral (in 1990) sufficient to cover the largest net debit, were subsequently adopted by European central banks as a recommended minimum standard for their own NS systems (Bank for International Settlements, 1990b – the so-called Lamfalussy report).

[46] Green (1996) provides a general equilibrium welfare analysis of central bank intervention and of clearinghouse intervention in payment arrangements. Extending the model of Freeman (1996a, b) to situations where private intermediaries issue debt payable by themselves (novation) in exchange for debt payable by original issuers (substitution) – an arrangement that corresponds to the clearinghouse practice of novation and substitution – Green demonstrates that the laissez-faire market equilibrium is efficient without the intervention by a monetary authority, even in rare cases where intermediaries are unable to settle some of the debt that they issue. Green concludes that neither an RTGS arrangement with central bank intervention nor a clearinghouse arrangement with novation and substitution has an absolute advantage over the other. Whether or not efficiency might require a central bank to participate in the payments system depends on whether the central bank can credibly promise and reliably reabsorb money (possibly daylight credit) that it issues to facilitate payments and also on whether the commercial-law framework governing the operations of a privately owned clearinghouse using novation and substitution is sufficient to warrant agents' use of debt issued by the clearinghouse as a money-like medium of exchange.

the possibility of a payments gridlock. [47] Schoenmaker argues that from an empirical cost–benefit standpoint, the probability of a single failure to settle and a single net debit loss needs to be contrasted with the lower cost of an NS system. As well, the (unknown) probability of multiple (and independent) settlement failures and a multiple net debit loss needs to be contrasted with the higher cost of an RTGS system. While we know which system is more costly, we do not know which system is more cost-effective given the risk and loss reduction achieved by the different settlement arrangements. [48]

Angelini (1996) focuses on reserve management and argues that payment senders face an incentive to delay payments until later in the day when more incoming funds have been received and the need to purchase additional cover is reduced. However, the cost-minimization behavior by payment senders creates uncertainty for payment receivers in managing their end-of-day central bank reserve positions. This can lead to greater variance in overnight reserve positions and higher average holdings of excess reserves which, in turn, creates information problems for short-term monetary policy and lowers bank profits. To eliminate these incentives for payment senders, Angelini proposes to charge payment receivers rather than senders (as is now done) for overdrafts created when a payment is sent. [49]

[47] A payment gridlock is where no payments are made because participants, conserving on their holdings of costly idle balances, collateral, or use of priced daylight credit, are unwilling to fund further payments. The threat of gridlock, however, has been addressed in practice (a) as sending banks realize that payment and other business is lost to competitors when customer-requested payments are not made, or (b) by splitting up an especially large payment into smaller pieces and sending the pieces at different times of the day as sufficient incoming funds are received, or (c) by having the central bank provide emergency daylight liquidity at a price.

[48] Settlement failure probabilities (Pr(loss)) are difficult to estimate. An ad hoc approach is to equate the collateral cost of an NS arrangement with the expected benefit from preventing systemic risk (which is not losing collateral by having to cover a failing participant's net debit). The implied cost–benefit relationship is Pr(loss) × net debit loss = cost of collateral, which can be solved for Pr(loss) since the collateral cost can be determined and the net debit loss approximated by the largest net debit permitted on a network. Multiplying Pr(loss) by the number of network participants and inverting the product gives the expected annual frequency of a failure to settle when the costs just equal the benefits (Humphrey, 1995). If central bankers or network participants (subjectively) expect settlement failures to be more frequent, then the benefits of the systemic risk reduction arrangement exceed the costs and should be implemented (otherwise not).

[49] Unfortunately, receivers would then have an incentive to request that senders delay their payments to them. While senders may ignore requests from banks they are sending customer-requested payments to, they will likely abide by such requests from banks they have borrowed inter-bank funds from (such borrowings account for the majority of payments over Fedwire – see Federal Reserve Bank of New York, 1987). Otherwise, a receiving bank will include its overdraft charge in its quoted overnight borrowing rate when returned borrowings are not delayed. This shifts the incidence of the overdraft charge back to the sending bank, providing an incentive for the sender to delay payments. Since the delayed send problem is essentially the same as the payment gridlock problem noted in an earlier footnote, the same solutions are available. In addition, delayed sends have been addressed more directly on CHIPS (which focuses on foreign exchange payments) by enforcing requirements that large percentages of each day's payment value must be sent prior to specified times during the day.

6.3. Effects from pricing payment risk

Currently, only two large-value payment networks assess a direct fee for daylight overdrafts. According to Kamata (1990), net creditors who sell funds to net debtors on BOJ-NET have charged intra-day fees of between 4 and 15 basis points (annual rate). The current price on Fedwire for credit supplied by the central bank is 15 basis points (annual rate) and applies to both funds and (US government book-entry) security transfers. [50] Almost no published information exists regarding the market or institutional response to pricing on BOJ-NET. And while the possible effects from pricing daylight overdrafts have been outlined in general terms (Mengle et al., 1987; Simmons, 1987; Evanoff, 1988; Gilbert, 1989; VanHoose and Sellon, 1989; and others), only recently has any analysis been available regarding the actual effects from pricing.

Data on the level of daylight overdrafts, the timing of various Federal Reserve risk reduction efforts, and the general effects of pricing these overdrafts on Fedwire are described in Richards (1995). Hancock and Wilcox (1996) go further and provide a more detailed statistical analysis of the effects of pricing and net debit caps.

Net debit caps on Fedwire were initially set high and at multiples of banks' equity capital positions so as not to disrupt the operation of financial markets. Consequently, their imposition in 1986, and a subsequent 25% reduction in 1988, had some effect on the growth of daylight overdrafts but none on its level. Overdrafts were priced by the central bank in 1994, a policy shift that allowed overdrafting banks to choose the lowest cost alternative between reducing their overdrafts or paying a fee. The initial price of 10 basis points (annual rate) per dollar of average overdraft reduced the average overdraft by 40% (or by $43 billion). The price was raised to 15 basis points in 1995 but the additional reduction was proportionally less ($8 billion).

Hancock and Wilcox estimate that the price elasticity of demand for daylight overdrafts is inelastic at −0.52 (for the average overdraft). A closer examination revealed that securities overdrafts fell by 45% while funds transfer overdrafts fell by 25%, giving a price elasticity of −1.01 for securities but only −0.11 for funds. This differential response is likely associated with institutional differences in the securities and interbank funds markets.

Securities dealers and their clearing banks have found it easier to divide up large customer orders for securities and deliver them in pieces earlier in the day as they are accumulated (reducing securities overdrafts), rather than wait until

[50] The cost of holding additional idle reserve balances to cover payments made on SIC is, of course, the overnight rate (opportunity cost). The cost of posting liquid collateral on CHIPS is estimated to be 25 basis points (annual rate). This represents the average spread between the overnight federal funds and repo rates. The logic is that banks could borrow in the federal funds market, purchase liquid government securities, and instead of entering into a repurchase agreement with these securities, post them as collateral.

the entire order is filled before delivering it to a customer late in the day. [51] The cost of partial deliveries in the interbank funds market, along with increased use of overdraft-reducing rollovers and/or continuing contracts for overnight federal funds or Eurodollar borrowings (Humphrey, 1989), are apparently still not strong cost-effective substitutes for priced overdrafts. Anecdotal information suggests that the primary response to overdraft pricing in the funds market has been a rise in delayed sends of non-time critical payments and an institutional adjustment in the overnight federal funds market whereby funds are now returned 24 hours after being received, rather than the former 18 hour arrangement of afternoon receipt and next morning return. If the overdraft price is increased markedly in the future, then an intra-day interbank funds market may develop after these lower-cost responses to pricing have been exhausted. [52]

6.4. An intra-day funds market

An intra-day funds market exists in Japan to fund bank morning withdrawals of cash from the central bank and to fund afternoon net debit settlement positions associated with clearing checks and payments over electronic networks. [53] This is in addition to other interbank payments over BOJ-NET which also may require funding. There appears to be no published source (in English) regarding the effects that a working intra-day funds market has for the implementation of monetary policy. However, prior to the actual use of overdraft pricing in the US, VanHoose (1991) and Lindsey et al. (1988) investigated how an intra-day market for interbank funds may affect short-run monetary policy and control.

In VanHoose (1991), a risk-neutral bank maximizes each day's expected profits subject to a need for and costs of intra-day and (two types of) overnight funds. Intra-day funds cover payments that would otherwise generate daylight

[51] Book-entry US government securities are transferred in a delivery-versus-payment environment. Thus the accumulation of securities prior to delivery to customers means that money has been paid out, usually creating an overdraft, and will only be eliminated when the securities are delivered. Earlier, partial, delivery means that less overdrafts are incurred since a smaller value of securities are accumulated prior to delivery (and payment). A $50 million limit on securities transfers facilitated this result (so a $200 million securities order would involve four separate transfers, regardless of the time of day each one was initiated).

[52] In this respect, Chakravorti (1996) argues that an interbank funds market would be less costly than having each participant hold idle balances adequate to clear its expected payments. This follows from the fact that funds lent intra-day would still be available for overnight use, while idle balances would incur the overnight rate as an opportunity cost.

[53] The Japanese Zengin system is equivalent to the US ACH network while the Gaitame–Yen system handles cross-border payments using foreign exchange. Kamata (1990) is a good source on this and other aspects of the Japanese payment system.

overdrafts while overnight funds cover loan takedowns and changes in deposits, and all are uncertain at the time funding decisions are made. [54] The model reproduces the result, based on intuitive and institutional considerations in Lindsey et al. (1988), that variations in the intra-day rate will likely affect the variance of the overnight rate.

The link between intra-day and overnight rates is essentially broken if the central bank stands ready to supply changes in daylight credit at a constant or administered price below a market-determined rate (as occurs on Fedwire, since no private intra-day market has yet developed). Here shifts in the demand for payment-related intra-day funds is matched with a perfectly elastic supply (up to a net debit cap). Since the intra-day rate is constant, only changes in funding requirements determine changes in the overnight rates and short-term monetary policy is not degraded. [55]

A different aspect of an intra-day funds market is developed in Rochet and Tirole (1996b) and Calomiris and Kahn (1996). These authors, relying on theory and historical experience respectively, suggest that important private risk monitoring benefits are discarded when daylight credit is supplied by central banks at an administered price. They support the private monitoring that takes place on a large-value payment network when one set of participants provides intra-day funds to others. This occurs today on BOJ-NET and, less directly, on CHIPS when receiving banks set their exposure to sending banks via a bilateral net credit limit. Rochet and Tirole (1996a) argue that these credit limits would be more useful if they operated as bilateral net credit lines. In practice, when a sending bank expects that it may be limited in sending payments to other CHIPS participants, it contacts the receiving bank prior to the determination each morning of the net credit limit and requests that the limit set be sufficient to receive the funds to be sent. As far as we know, no fee is paid for this expansion of net credit (as would be the case in a true net credit line). While private monitoring does occur, the net credit extended by receiving banks acts as a quota rather than a price sensitive supply curve.

Unfortunately, there are trade-offs. The private monitoring benefits achievable from having intra-day credit supplied in a market setting conflicts with the desire to not contaminate the information available for short-term monetary policy. In principle, this is an empirical question since the extra variance injected into the overnight rate by having a variable, market-determined, intra-day funds rate may not be so large as to offset the possible benefits associated with private monitoring and risk-absorption. By contrasting the monetary policy and private monitoring experience of Japan (where an intra-day rate is deter-

[54] Both net borrowers and net lenders in each market face this uncertainty, so the modeling of these two sides of the market is similar.

[55] The 24 hour rate would now include a constant premium, equal to the administered overdraft price, and exceed the 18 hour funds rate by at least this amount.

mined by the market) with that of the US (which has an administered price), it may be possible to determine more accurately the benefits and costs of these two alternative policies. [56]

6.5. Moral hazard for the bank safety net

To the extent that procedures adopted to reduce and/or contain payment system systemic risk significantly underprice this risk, moral hazard is created for the bank safety net of discount window loans and deposit insurance. [57] This was certainly the case before the problem of systemic risk was addressed on NS systems or credit risk was dealt with on those RTGS systems which allowed free daylight overdrafts but guaranteed the finality of the payments made. In the event of a possible settlement failure, banks expected that this would in all likelihood trigger a discount window loan to the failing participant. On an NS network, such a loan transforms systemic risk into credit risk for the central bank. However, since discount window loans are collateralized, the ultimate risk of loss instead falls first on the bank's equity holders and secondarily (if losses exceed equity value) on the bank's uninsured creditors and (possibly) on the provider of deposit insurance. [58]

It is not well recognized, but moral hazard for the safety net is due to a central bank policy of "conscious ambiguity" regarding bank access to the discount window and a historically demonstrated predilection of direct intervention should serious banking problems arise. In a less protected and more certain environment, banks would have had stronger incentives to monitor one another and to institute their own set of payment risk-reduction procedures. This, in fact, has been the case for clearing and settlement arrangements in stock, commodity, futures, options, and other financial markets. With less protection, but greater certainty that they were on their own, participants in these markets have devised their own collateral, loss-sharing, and cross-guarantee procedures. [59]

Independent of the moral hazard issue, Flannery (1996) has argued that discount window loans can play an important role in stabilizing the payment system during crisis periods, a view seconded by Bernanke (1990) in the context of the 1987 stock market crash. When markets are in disarray and potential exposures are large, private lenders may rationally withdraw or severely limit credit

[56] To our knowledge, such a contrast has not yet been attempted.

[57] Szegö (1997) argues that the bank safety net of discount window loans and deposit insurance has led to a bank monopoly of the payment system, particularly for large-value transfer systems.

[58] Prior to the passage of FDICIA in 1991, the deposit insurance provider in the US (the FDIC) shared losses proportionally with other general creditors to a failed bank. Today, the FDIC will be paid before the other general creditors, reducing (but not eliminating) its risk of loss.

[59] See Mengle (1995) and Eisenbeis (1995) for a similar view and a detailed example using the market for derivatives.

lines, forcing solvent but illiquid payment participants into settlement failure unless discount window credit is available. The provision of such credit is viewed by Flannery as a lower cost method of resolving serious payment system uncertainty and disruption problems. Losses incurred on discount window loans and/or by a government deposit insurance agency are considered as transfer payments from society as a whole enabling the payment system to continue to function.

7. Emerging payment methods: Electronic money

Recently, various innovations for making retail payments, known collectively as electronic money (e-money) payment products, have been proposed and piloted. [60] There are two broad categories of such products: (1) pre-paid stored-value cards, and (2) pre-paid software products that use computer networks such as the Internet. [61] There are also two ways that stored value is represented. One way uses a single account balance that is stored and updated with each transaction. The other way uses electronic notes or tokens, each of which has a fixed value and serial number, that are transferred from one device to another. Cryptography (including digital signatures) is commonly used to authenticate messages and devices and protects the integrity and confidentiality of data for stored-value systems. While e-money is still at a relatively early stage of development, it has the potential to challenge the predominant role of cash for making small-value payments, to permit consumers to purchase goods and services over the Internet, to lessen risks by speeding up payment settlement, and to reduce the cost of making payments both domestically and internationally.

E-money research to date has focused on how e-money (1) would substitute with traditional forms of money, (2) would circulate at par with government-issued currency, (3) would affect monetary policy, and (4) would impact government seignorage revenues. These four areas are now briefly reviewed.

[60] By early 1997, general-purpose stored-value cards had been introduced in regional pilot tests in all of the G-10 countries (Group of Ten, 1997; Congressional Budget Office, 1996).

[61] Traditional electronic payment transfers, such as those with debit or credit cards, typically require on-line authorization and involves debiting a consumer's bank account after a transaction. In contrast, electronic money products are stored-value products in which a record of the funds available to a consumer is stored on an electronic device in the consumer's possession. In some cases, these products are able to increase their stored value from a dispenser connected to an on-line authorization network. The funds so stored typically represent a general or "pooled" liability of an issuer (Bank for International Settlements, 1996a; Congressional Budget Office, 1996; and Group of Ten, 1997).

7.1. The substitution of E-money for other monies

The willingness of a (representative) consumer to substitute e-money for cash has been explored in a Baumol–Tobin type model of money demand by Santomero and Seater (1996). They argued that the amount of pre-paid value stored on e-money products by households will likely be functions of the types of consumer goods that can be purchased using them, the availability of terminals that accept them, and the compatibility of competing e-money products with each other. [62] Exploring plausible determinants of the demand for e-money further, Kane (1996) reasoned that time-of-day flexibility and the protection from violent crime provided by electronic banking and video shopping may be deemed to be desirable services that paper-money transactions simply cannot offer.

Survey data on households has also been used to focus on how demographic characteristics might affect the demand for electronic banking products. For example, Kwast and Kennickle (1997) have illustrated that income, financial assets, age, and education all play important roles in determining household use of e-money products. They concluded that the potential US market for such products is still highly specialized, with the demand coming almost entirely from higher income, younger, and more educated households that have accumulated significant financial assets.

In order to significantly replace other money instruments, e-money systems will not only need to offer enough features to induce consumers to use these products, they will also need to induce merchants and issuers to bear some costs (Wenninger and Laster, 1995; Working Group on EU Payment Systems, 1994). E-money products could save merchants time and expense in handling cash, and there is a potential for lower transaction fees than are now charged for on-line debit card use. Issuers of e-money products can potentially reduce on-line network charges, gain new sources of fee income from merchants, and may collect float from investing their pooled e-money liability. [63]

[62] Empirical evidence also suggests that the availability of terminals provides a demand-side network externality that spurs the volume of electronic-based transactions (Humphrey et al., 1996, contains a discussion of ATM terminal availability and card usage).

[63] Seignorage issues are noted below. The distribution of benefits among e-money issuers and the dynamics of the competitive process will be important in determining the speed of technological adoption. Innovators must contend with reactions from existing payment service providers that may have incumbent advantages, such as scope economies with other services (e.g., debit cards with credit cards), name recognition, and that *existing suppliers* may have already borne some fixed costs of providing competing products (Berger et al., 1996).

7.2. Circulation of e-money at par

The transaction costs and risks associated with substituting privately issued e-money for government-backed money will depend in part on whether e-money circulates at par. Factors that determine par circulation include the transparency of the e-money system being used, the financial integrity of the issuer, the technical security measures undertaken, and the design of the product. Potential customers and merchants can make informed choices about the relative merits of an e-money product only if its features, costs, and risks are known. This concerns disclosures about the issuer's obligations and liability to users, the intended use of personal transactions data that can be collected and monitored, and the applicability of FDIC deposit insurance (Congressional Budget Office, 1996).

In this regard, a survey of current consumer protections within the G-10 countries found that most countries were relying on existing laws and regulations in addressing risks – such as loss, fraud, and issuer insolvency – rather than enacting comprehensive new measures specifically designed for e-money (The Group of Ten, 1997). In the US, for example, the primary issue has been the applicability of the Electronic Funds Transfer Act to electronic stored-value cards (Board of Governors of the Federal Reserve System, 1997). Overall, financial institution authorities in the US have taken a "wait and see" approach to regulating these emerging electronic payment instruments and the FDIC has determined that deposit insurance does not apply to the value stored on cards.

Europe, in contrast, has expressed the view that public confidence in the payment system may be compromised if non-regulated institutions are allowed to issue stored-value cards and that consumers should be protected against the consequences of issuer failure (Working Group on EU Payments, 1994). [64] Technical security measures – such as physical barriers for tampering with the e-money devices, cryptographic algorithms, and security verification between devices or on-line verification – are important for protecting against fraud or counterfeiting attacks that can threaten the overall integrity of an e-money product. Even so, differences in the risk of default, fraud, and product malfunction across issuers may result in some discount from par for the use of certain e-money instruments, thereby affecting transaction costs. [65]

Taking a historical view, Rolnick et al. (1996, 1997) have questioned whether market mechanisms will always lead to market equilibriums with par exchange between privately issued monies and government-backed money. Using

[64] That is, e-money issuers should be restricted to banks that would be protected by the government provided safety net.

[65] A deeper discussion of security risks, fraud risks, and product malfunction risks associated with e-money products is contained in Bank for International Settlements (1996b).

information on private bank notes in the US prior to 1864, they found that privately issued bank notes frequently did not circulate at par outside of the local areas where they had been issued. Their study suggests that par circulation with multiple issuers of e-money may require a coalition of market participants and may not necessarily result from market forces. Although such cooperation might reduce transactions costs and risks associated with non-par banking, other observers (e.g. Salop, 1990) caution that it is difficult to maintain an adequate level of competition and still permit the cooperation that enhances efficiency and consumer welfare. More optimistically, it is noted that traveler's checks, which are generally offered by non-banks, typically circulate at par with traditional payment media without a coalition of market participants.

7.3. Monetary policy issues

Several studies have discussed whether the introduction and adoption of e-money products will have an effect on the ability of central banks to implement monetary policy (Working Group on EU Payment Systems, 1994; Congressional Budget Office, 1996). It is expected that, initially, stored-value products will be mainly substitutes for small-value cash transactions. With such limited use the impact on monetary policy would also be limited. Indeed, if issuers of e-money hold 100% cash reserves for balances on stored-value devices, then the money supply will not change at all. As well, if non-depository e-money issuers hold their e-money balances in their own checking account, the money supply will not be altered. Only when issuers invest their e-money balances will the money supply be increased. Even so, it is likely that the initial effects on monetary policy will be manageable and have less of an overall impact than the earlier shift of funds by consumers from traditional deposit accounts to money market funds during the 1970s and 1980s (Congressional Budget Office, 1996).

If, in the longer term, e-money products substantially replace cash and other payment media, then it could be more difficult to measure and control the money supply. [66] This would be the case if non-depository issuers of e-money continue to remain under no legal obligation to report outstanding stored-value balances to the Federal Reserve (Congressional Budget Office, 1996). In this (probably extreme) case, the widespread use of e-money could cause some instability in velocity and the behavior of monetary aggregates so that other intermediate targets may have to be employed by the central bank. Indeed, The Working Group on EU Payment Systems (1994) has argued that if e-money strongly replaced cash, the size of some country's central bank balance sheets may be so reduced as to hamper their ability to influence interest rates through open market operations. In this event, central banks would have to find alter-

[66] Some observers argue that there exists a lower bound for currency holdings so that e-money will not fully replace currency.

native ways to control the monetary base, one of which could be making e-money products reservable. [67]

7.4. Seignorage revenues

The substitution of privately issued e-money for government-issued currency reduces seignorage. Inflation provides an incentive to substitute private payment instruments, which typically pay interest, for currency which does not. Lacker (1996a, b) has applied this result in a general equilibrium model to show that inflation provides an incentive to create new payment instruments like e-money that capture the seignorage tax – the foregone nominal interest earnings from holding cash. The argument is that the seignorage tax associated with inflation is an opportunity cost to private participants in the payments system, but it does not reflect "true social costs" of using money, which are the costs to the government for providing the monetary base. If the cost of the monetary base to the government is very low, then it is possible that – from an efficiency perspective – too much e-money stored value is substituted for government money. In such instances, Lacker (1996b) proposes that taxes or reserve requirements on privately issued money be used to constrain its "excessive" use and limit further innovation along this line.

E-money products are at an early stage of development. While the technology and software are new, the research and policy issues are largely familiar ones that were applicable to earlier payment system innovations, such as privately issued banknotes, the heavy use of checks, the expansion of ATMs, and the development of alternative account-based payment instruments such as credit and debit cards. In this respect, there should be few surprises in the analysis of e-money products, even though all the characteristics of this emerging payment method are not yet known.

8. Social versus private costs and the efficiency of payment use

Social costs reflect the real resources involved in initiating, accepting, processing, and settling a payor payment at the point-of-sale, through the mail, or a (payee-initiated) pre-authorized direct debit. Private costs represent the costs charged to – and thus seen by – the payor, payee, and bank parties to a transaction. In a perfectly competitive world, with no government interven-

[67] In Congressional testimony, for example, Alan Blinder (then Vice Chairman of the Federal Reserve) indicated that "under current regulations, stored-value balances issued by depository institutions would be treated as transaction account balances and hence subject to reserve requirements". Presumably, the same principles would apply to pre-paid software products that use computer networks such as the Internet.

tion, the private and social costs to each party would be equal so price reflects marginal cost and resource allocation would be efficient. Banks, as providers of non-cash payment instruments, would pay market-determined rates on deposits and charge resource-based prices for payment services (cf. Johnson, 1968). Due to institutional barriers and user preferences outlined below, only a few countries come close to such an idealized world. As a result, although electronic debit card, ACH, or giro transactions generally cost from one-third to one-half that of check or paper-based giro payments, resource savings in shifting from paper to electronic payments have often been difficult to realize. In what follows, we illustrate how costs differ among various payment instruments, explain the reasons why (for some instruments) social and private costs differ, and outline the main barriers to shifting from paper to electronic payments.

8.1. Estimates of payment costs

Data on payment costs are fragmented and often proprietary, so no cross-section or time-series data exists on the payment costs incurred by payors, payees, or banks. However, some limited survey information does exist regarding payee and bank costs. Table 7 presents the (average) cost of accepting different payment instruments at supermarkets in the US (Food Marketing Institute, 1994). On either a per transaction basis or by $100 of sales value, cash is always cheapest, followed by debit cards, then checks, with credit cards being by far the most expensive. [68] While cash is used for 57% of transactions, its small average value per transaction reduces its importance in value terms to 36%. Checks account for a third of transactions but almost half of the value purchased. On a weighted average basis, retail receivers expend around $0.20 per transaction, or about 1% of sales value ($0.97 per $100 in sales). [69]

In developed countries, production functions for payment services are quite similar but, because of scale effects and other differences, realized production costs among countries can differ. [70] The bank cost of payment services provided to payees are to a large extent already included in the payee payment costs

[68] These costs include the time it takes to complete a transaction, the wage and fringe benefits of the various accepting/verifying/accounting employees, armored courier costs (for cash), check and other fraud expenses, bank charges, electronic network transaction fees, and credit card fees.

[69] Credit cards, the most expensive payment instrument, are used less intensively at supermarkets than they are in other retail settings (oil and gas, restaurants, hotels/motels, durable goods, and specialty retailers). Applying payment instrument weights representative of a broad array of retail establishments to the cost figures in Table 7 raises the weighted average receiver cost to around $0.30 per transaction, accounting for 1.05% of sales value (see Caskey and Sellon, 1994, and the source of their Chart 3 for the weights).

[70] In addition to scale effects, there are differences in the average wage, telecommunication costs, and the expense of transferring funds between accounts at different institutions. The US has an advantage in the first two areas while Europe has an advantage in the third (since Europe has a centralized giro-based, credit transfer, payment system with more "on-us" payments and no expensive return items).

Table 7
Payee costs of receiving different payment instruments by transaction and sales value (for US supermarkets)

	By transaction volume		By $100 of sales value	
	Unit cost	Percent (%)	Cost	Percent (%)
Cash	$0.07	57	$0.52	36
Debit card [a]	$0.30	2	$0.94	3
Check [b]	$0.43	33	$1.20	49
Credit card	$0.81	3	$2.27	5
Other [c]	–	5	–	7

Source: Food Marketing Institute (1994).
[a] Refers to an on-line debit, with settlement through the ACH.
[b] A verified check reduces fraud losses and it's cost per transaction (per $100 of sales) is somewhat lower at $0.37 ($1.05).
[c] Primarily food stamps, a government-sponsored food welfare program.

shown in Table 7. [71] In a recent survey of Norwegian postal and commercial banks (Robinson and Flatraaker, 1995), the total of payor bank and payee bank costs of an electronic bill payment through a giro was $0.49 while its paper-based substitute averaged $1.34. As well, a transaction at the point-of-sale using a debit card was $0.63 while check expenses were $1.97. Overall, electronic giro payment costs are 37% of their paper equivalent while an electronic debit card transaction costs 32% of that for a check. [72] Similarly, in the US, estimates of the total (payor, payee, and bank) cost of an electronic ACH payment in the US average $1.34 per transaction while check costs average $2.97 (Wells, 1996). This gives a ratio of electronic to paper payment cost of 0.45. [73]

Based on these and similar cost estimates, it is now a matter of (bank and government) policy to provide direct price incentives to shift users from paper to electronic payments in Norway (Flatraaker and Robinson, 1995). In addition, recent legislation in the US has mandated that federal government check payments (accounting for less than 1% of the total) be shifted to electronic payments by 1999 (and similar legislation is pending at the state government level).

[71] Banks provide payor services to their depositors by providing branch offices, ATMs, and cash, along with paying check, ACH, giro, and wire transfer debits. These payor services have costs and prices that are separate from services the same bank provides to (typically business) payees, where excess cash is deposited, checks are collected, and ACH, giro, and wire transfer credits are received.

[72] Norway uses far fewer checks per person than the US (six checks per person per year versus 237) and has a centralized (and therefore lower cost) giro system for electronic payments. Thus, due to scale effects alone, bank check (electronic payment) costs in Norway would be expected to be higher (lower) than those in the US.

[73] An earlier study of US debit card and check costs gave a ratio of 0.59 (Humphrey and Berger, 1990).

8.2. Social versus private costs

Although the social (real resource) cost of electronic giro or ACH payments is less than that for paper-based giro or check payments, this cost difference may not be *seen* by the payor who, in many transactions, can choose which payment instrument to use. This occurs for check and electronic credit card payments due to the existence of check float and the fact that payees – not payors – pay the credit card transaction fee. [74] When interest rates are high, the benefit to payors from check float have in the past reduced the payor cost of using a check to a negative amount, a level at which no amount of scale economies for electronic payments can compensate (Humphrey and Berger, 1990). [75] Check float and credit card pricing arrangements both reduce, and thus distort, the relative private costs to payors of using these payment instruments. The general result is that checks and credit cards will be overused while other, lower-cost instruments (debit card, ACH, electronic giro) will be underutilized compared to their relative social cost. Because the cost of accepting checks or credit cards by payees exceeds that of other payment instruments (Table 7), and this cost difference does not lead to price differences at the retail level when different instruments are tendered, users of cash and debit cards cross-subsidize users of checks and credit cards (Humphrey, 1984). [76]

8.3. Barriers to use of electronic payments

As seen, one barrier to shifting from checks to lower cost electronic payments is the float wedge between the social and private costs faced by payors. [77] Another, even broader and stronger barrier, lies in how payment services are priced by banks. While payees usually pay a price per transac-

[74] Check float reflects the time lag (often 1–2 days) from when a check is exchanged for a product or service and when the payee actually receives funds that can be used. Float is a benefit for the payor and its size depends on the value of a check, the interest rate, and the length of the time lag involved to deposit, process, and present checks for payment at a paying bank. In the US, businesses write 40% of all checks but receive 90% of the float.

[75] More recently, the combination of lower interest rates (Wells, 1996) plus the ability of bill payment payees to largely eliminate 1–2 days of mail float has dramatically reduced the benefit of check float to consumers. Mail float was unilaterally reduced when consumer bill payment payees altered their payment due dates from the implicit "past-due if *mailed after*" a given date, a position supported by much case law, to "past-due if *received after*" a given date.

[76] Cross-subsidization is reduced when discounts are given for payments by cash, although this is not common.

[77] Canada has eliminated the check float wedge for consumers (which was small in any event) as well as for businesses. Canadian banks give payees same-day funds availability for deposited checks and charge firms for the float they create. By shifting the cost of float to payors, Canada has served to promote electronic payments. Indeed, per person check use in Canada reached a peak in 1988 and is falling while the US (as of this writing) has not yet reached this point although is forecasted to do so around 1999–2000.

tion, payor pricing – because of user preferences – typically involves a minimum balance and monthly account fees. [78] Except for Scandinavia and a few other countries, where banks charge (and payors pay) a per transaction fee that differs across payment instruments, consumer payors effectively face the same explicit marginal cost – zero – regardless of which instrument they choose to use.

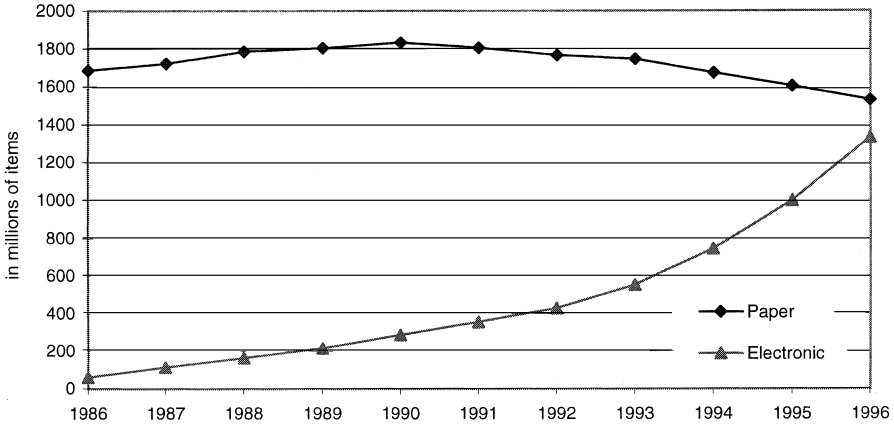
A zero explicit marginal cost for payment instrument use, paired with float benefits for checks and frequent flyer miles (or other rebates) for credit card use, serve to hinder the adoption of other types of electronic payments with lower social costs. [79] While the market has failed to allocate payment resources adequately in many countries, this has been less a strict case of market failure than the result of earlier government intervention restricting the concentration of the banking system, limiting its geographical dispersion, and regulating how it can attract deposits. Europe in general, and Scandinavia in particular, have fewer of these barriers and so have a higher incidence of electronic payments. And, by generally providing a safer environment in which to make transactions, these countries – along with Japan – have a higher incidence of low-cost cash use as well. The US, with the world's highest per person use of non-cash payments (see Table 2) has the most to gain from a more cost-effective use of payment resources. Currently, the weighted average cost of a non-cash transaction is around $2.60, generating over $200 billion a year in payment system expenses (Humphrey and Pulley, 1997). This represents 3% of GDP and could be reduced substantially if price and other incentives were developed to shift a larger percent of transactions to more cost-effective electronic payments.

8.4. Expanding electronic payments: The experience of Canada and Norway

In Canada, paper transactions (almost entirely checks) no longer dominate clearings as they once did. Indeed, as shown in Fig. 4, the peak in paper transactions was reached in 1990 and by 1996 the number of paper and electronic transactions were almost equal. Anecdotal information suggests that Canadian banks encouraged the use of debit cards, ATM services (e.g., utility payments) and computer banking services and discouraged check writing by simulta-

[78] Although banks often offer payors a choice of (a) paying a price per transaction or (b) holding a minimum balance and being able to incur an "unlimited" number of transactions at zero marginal cost, the vast majority of consumer depositors choose the minimum balance option. Business depositors have much higher transaction volumes and are charged on a per transaction basis, but may pay for the service through a debit to their account or by holding a compensating balance (where the balance held is set to generate the same revenue for the bank when evaluated by a market interest rate).

[79] In the US, rebates for card usage include gifts of goods and services, as well as discounts off of the purchase price of automobiles, restaurant meals, and telephone charges.



Source: Canadian Payments Association, FORUM (March 1997).

Fig. 4. Canadian clearing and settlement items: Paper versus electronic.

neously lowering fees for electronic banking products and raising fees for checks. In addition, the banks increased dramatically the availability of EFT-POS terminals over 1991–1995. The number of EFT-POS terminals per million inhabitants rose from 491 in 1991 to 6394 in 1995 – a 13-fold increase – while the number of transactions per inhabitant per year rose from 0.4 to 13.3. A second reason for the shift to electronics, although a minor one, has been in the movement away from exchanging checks for large-value transfers (and interbank settlement) to relying on a new large-value wire transfer network.

Norway has gone even further in promoting the shift to electronic payments by relying heavily on differential pricing to consumers and businesses for each payment they make, with a substantially lower price being charged for electronic payments. Fig. 5 indicates that paper transactions in Norway (primarily paper-based giro payments) have fallen since 1986 and that by 1995 were absolutely less than the number of electronic transactions. As noted earlier, the banks in Norway were encouraged to provide direct price incentives – based on actual variable costs – to shift users toward lower cost electronic payments. The experience of Canada and Norway support the assertion that users of payment instruments are price sensitive and, if prices are directly faced by users for each payment they make, they will shift to the lowest cost alternative.

9. Some conclusions and areas for further research

Our survey has covered many, but not all, of the issues associated with payment transactions, instruments, and systems. In what follows, we select four

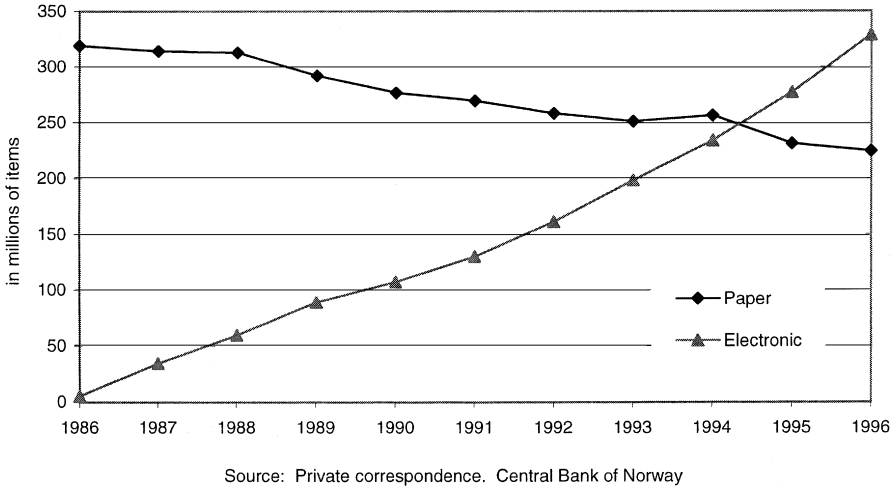


Fig. 5. Norwegian clearing and settlement items: Paper versus electronic.

general conclusions that deserve emphasis and outline areas related to them needing additional research.

One conclusion is that while all developed countries have essentially the same set of payment instruments available to them, the intensity of use is often quite different. In the past, evidence of different payment use was primarily anecdotal but now a growing body of detailed data on payment instrument use across countries now exists. Paired with appropriate institutional, historical, and legal information, this data can be used to identify the main determinants of a country's choice of payment instruments. Such an analysis would clarify whether there is a single general pattern to the evolution of a country's payment system (suggesting that current differences in payment use across countries indicates that they are at different stages along a common path) or whether there are truly unique evolutionary paths and why this is so. This question is of some importance to both developing and emerging market economies as they attempt to establish and modernize their payment systems to provide the infrastructure necessary for sustained growth within a market-based economy. Or, as is frequently asked in the heavily cash-based, formally centrally planned economies: "Is it necessary to go through a stage – even a short stage – where consumers and businesses learn to rely on check or paper giro payments, as preparation for more complex electronic payments, or can we skip paper payments entirely?".

A second conclusion deals with the negative effects of inflation on social welfare by shifting users away from fiat money toward more costly payment instruments associated with accounts that earn interest. While this demand-side general equilibrium result has been extensively analyzed under different condi-

tions and frameworks, its supply side implications have been undeveloped. Specifically, general equilibrium theory could be used to analyze the effects of scope economies between payments and other banking activities (such as deposit-taking or lending activities) to determine (a) if restricting access to the payments system to only banks is warranted on efficiency grounds and (b) whether synergies with other central banking activities provides a comparative advantage to the monetary authority in the provision of some payment activities. A related issue in many countries concerns the mispricing of payment services to users, particularly paper-based services. General equilibrium analysis can be used to determine better the distribution of costs and benefits among payment system participants of using different payment instruments. Finally, general equilibrium theory could be applied to improve our understanding of the relationship between payment system infrastructure investment and its impact on economic growth, productivity, and the depth and liquidity of money and capital markets.

A third conclusion, although one not solidly based, is that users of payment services seem to be quite responsive to price signals when they have been used to shift users toward lower cost electronic payments. Unfortunately, there is little public information concerning the determinants of payment instrument use by either consumers or businesses and this has hindered quantification of the effect of pricing on payment choice. Lists of the various likely influences on payment choice are easy to make but almost no estimates of their relative magnitudes exist. Given limited cross-country data in this area, a case study approach of the experience of payment pricing in one or a few countries would provide a reasonable basis from which to draw generalizations. Such an analysis, to be most effective, would also seek to estimate the social costs of the various payment instruments – especially the contrast between paper-based and electronic payments – to quantify the possible social benefits of shifting users away from more costly instruments. A related issue here would be in assessing the potential social benefit (if any) of adopting new and emerging payment arrangements, such as stored-value cards and Internet payments.

Finally, a fourth conclusion concerns how best to address systemic risk on large-value payment networks. It is generally argued that net settlement – when augmented with real-time overdraft controls and backed up with some sort of collateralization arrangement – can be made to be just as successful in reducing systemic risk, but at a lower cost, as RTGS (which is the arrangement favored by most central banks). What is not known is how far the collateralization process needs to go. While it seems clear that the social benefits of collateralizing the single largest net debit on a large-value network exceeds the costs, it is not known whether the additional benefit from collateralizing the largest two net debits, or all of the net debits, will also exceed the costs. The answer, which is so far unavailable, depends on the likelihood of multiple rather than single bank failures and the probability that such failures will lead to a failure to settle net debits on a large-value payment network.

Acknowledgements

For their comments and suggestions we thank Allen Berger, James Bohn, Jeffrey Marquardt, and David VanHoose. Superb research assistance was provided by Kara Meythaler and Stephanie Russek. All opinions and errors contained here are solely the responsibility of the authors and do not indicate concurrence by the Federal Reserve Board, The Reserve Banks, or their staffs.

References

Angelini, P., 1996. A simple model of the interbank payment system, Working Paper, Bank of Italy, Rome.

Angelini, P., Maresca, G., Russo, D., 1996. Systemic risk in the netting system. Journal of Banking and Finance 20, 853–868.

Avery, R., Elliehausen, G., Kennickell, A., Spindt, P., 1986. The use of cash and transaction accounts by American families. Federal Reserve Bulletin 72, Board of Governors of the Federal Reserve System, Washington, DC, pp. 87–108.

Avery, R., Elliehausen, G., Kennickell, A., Spindt, P., 1987. Changes in the use of transaction accounts and cash from 1984 to 1986. Federal Reserve Bulletin 73, Board of Governors of the Federal Reserve System, Washington, DC, pp. 179–196.

Bank for International Settlements, 1990a. Large-Value Funds Transfer Systems in the Group of Ten Countries. Basle, Switzerland.

Bank for International Settlements, 1990b. Report of the Committee on Interbank Netting Schemes of the Central Banks of the Group of Ten Countries. Basle, Switzerland.

Bank for International Settlements, 1993a. Minimum common features for domestic payment systems, Report of the Committee of Governors of the Central Banks of the Member States of the EEC, Working Group on EC Payment Systems, Basle, Switzerland.

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

Bank for International Settlements, 1994. Payment Systems in the Group of Ten Countries. Basle, Switzerland.

Bank for International Settlements, 1995a. Payment Systems in Norway. Basle, Switzerland.

Bank for International Settlements, 1995b. Statistics on Payment Systems in the Group of Ten Countries. Basle, Switzerland.

Bank for International Settlements, 1996a. Security of Electronic Money. Basle, Switzerland.

Bank for International Settlements, 1996b. Settlement Risk in Foreign Exchange Transactions. Basle, Switzerland.

Barnett, W., Fisher, S., Serletis, A., 1992. Consumer theory and the demand for money. Journal of Economic Literature 30, 2086–2119.

Bauer, P.W., Hancock, D., 1995b. Estimation of the Fedwire Funds Transfer Cost Function, Mimeo. Payment System Studies Section, Board of Governors of the Federal Reserve System.

Bauer, P.W., Hancock, D., 1993. The efficiency of the Federal Reserve in providing check processing services. Journal of Banking and Finance 17, 287–311.

Bauer, P.W., Hancock, D., 1995a. Scale economies and technical change in the Federal Reserve automated clearinghouse payment processing. Federal Reserve Bank of Cleveland Economic Review 31, 14–29.

Bauer, P.W., Ferrier, G.D., 1996. Scale economies, cost efficiencies, and technological change in Federal Reserve payments processing. Journal of Money, Credit and Banking 28, 1004–1039.
Belton, T.M., 1984. Economies of Scale in Book-Entry Securities, Mimeo. Financial Studies Section, Board of Governors of the Federal Reserve System .
Belton, T., Gelfand, M., Humphrey, D., Marquardt, J., 1987. Daylight overdrafts and payment system risk. Federal Reserve Bulletin 73, 839–852.
Berger, A., Humphrey, D., 1988. Interstate banking and the payments system. Journal of Financial Services Research 1, 131–145.
Berger, A.N., Hancock, D., Marquardt, J.C., 1996. A framework for analyzing efficiency, risks, costs, and innovations in the payments system. Journal of Money, Credit and Banking 28, Part II, 696–732.
Bernhardt, D., 1989. Money and loans. Review of Economic Studies 56, 89–100.
Bernanke, B., 1990. Clearing and settlement during the crash. Review of Financial Studies 3 (1), 133–151.
Board of Governors of the Federal Reserve System, 1991. Functional Cost Analysis. Washington, DC.
Board of Governors of the Federal Reserve System, 1997. Report to the Congress on the Application of the Electronic Fund Transfer Act to Electronic Stored-value Cards. Board of Governors, Washington.
Boeschoten, W., 1991. National trends in payment systems and the demand for currency and banknotes. De Nederlandsche Bank Research Memorandum no. 9108.
Boeschoten, W., 1992. Currency use and payment patterns. Financial and Monetary Policy Studies, vol. 23. Kluwer Academic Publishers, Norwell, MA.
Boeschoten, W., Fase, M., 1989. The way we pay with money. Journal of Business and Economic Statistics 7, 319–326.
Borio, C.E.V., Van den Bergh, P., 1993. The nature and management of payment system risks: an international perspective, BIS Economic Papers no. 36, Bank for International Settlements, Basle, Switzerland, p. 86.
Calomiris, C., Kahn, C., 1996. The efficiency of self-regulated payment systems: Learning from the Suffolk system. Journal of Money, Credit, and Banking, Part II, 766–797.
Canner, G., Luckett, C., 1992. Developments in the pricing of credit card services. Federal Reserve Bulletin 78, Board of Governors of the Federal Reserve System, Washington, DC, pp. 652–666.
Carraro, K., Thornton, D., 1986. The cost of checkable deposits in the United States. Economic Review, Federal Reserve Bank of St. Louis, 19–27.
Caskey, J., Sellon, Jr., G., 1994. Is the debit card revolution finally here?. Economic Review, Federal Reserve Bank of Kansas City 79, 79–95.
Caskey, J., St. Laurent, S., 1994. The Susan B. Anthony dollar and the theory of coin/note substitutions. Journal of Money, Credit, and Banking 26, 495–510.
Chakravorti, S., 1996. Analysis of systemic risk in the payments system, Working Paper, Federal Reserve Bank of Dallas.
Cole, H.L., Stockman, A.C., 1992. Specialization, transaction technologies, and money growth. International Economic Review 33, 283–298.
Congressional Budget Office, 1996. Emerging Electronic Methods for Making Payments. US Government Printing Office, Washington.
Cooley, T.F., Hansen, G.D., 1989. The inflation tax in a real business cycle model. American Economic Review 79, 733–748.
Daniels, K., Murphy, N., 1994. The impact of technological change on the currency behavior of households: an empirical cross-section study. Journal of Money, Credit, and Banking 26, 867–874.
Eisenbeis, R., 1995. Private sector solutions to payments system fragility. Journal of Financial Services Research 9, 327–349.

Elliehausen, G., Wolken, J., 1990. Banking markets and the use of financial services by small and medium-sized businesses. Federal Reserve Bulletin 78, Board of Governors of the Federal Reserve System, Washington, DC, pp. 801–817.

Elliehausen, G., Wolken, J., 1992. Banking markets and the use of financial services by households. Federal Reserve Bulletin 76, Board of Governors of the Federal Reserve System, Washington, DC, pp. 169–181.

European Monetary Institute, 1996. Payment Systems in the European Union. Frankfurt.

Evanoff, D., 1988. Daylight overdrafts: rationale and risks. Federal Reserve Bank of Chicago Economic Perspectives 12, 18–29.

Evans, D., Schmalensee, R., 1993. The Economics of the Payment Card Industry. National Economic Research Associates, Cambridge, MA.

Federation of Bankers Associations of Japan, 1994a. The Banking System in Tokyo. Tokyo, Japan, p. 95.

Federation of Bankers Associations of Japan, 1994b. Payment Systems in Japan. Tokyo, Japan, p. 1.

Federal Reserve Bank of New York, 1987. A Study of large-dollar payment flows through CHIPS and Fedwire. New York.

Flannery, M., 1996. Financial crises, payment system problems, and discount window lending. Journal of Money, Credit, and Banking, Part II, 804–824.

Flatraaker, D., Robinson, P., 1995. Income, costs and pricing in the payment system. Norges Bank Economic Bulletin 66, 321–332.

Food Marketing Institute, 1994. Benchmarking Comparative Payment Methods: Costs and Case Studies. Research Department, Washington, DC.

Freeman, S., 1996a. Clearinghouse banks and banknote over-issue, Journal of Monetary Economics 38, 101–115.

Freeman, S., 1996b. The payments system, liquidity, and rediscounting. American Economic Review 86, 1126–1138.

Gilbert, A., 1989. Payments system risk: what is it and what will happen if we try to reduce it? . Federal Reserve Bank of St. Louis Economic Review 71, 3–17.

Green, E.J., 1996. Money and debt in the structure of payments. Federal Reserve Bank of Minneapolis Working Paper.

Group of Experts on Payment Systems of the Central Banks of the Group of Ten Countries, 1989. Report on Netting Schemes, Bank for International Settlements, Basle, Switzerland.

Group of Ten, 1997. Electronic money: consumer protection, law enforcement, supervisory and cross border issues, Report of the Working Party on Electronic Money.

Hancock, D., 1985. The financial firm: production with monetary and non-monetary goods. Journal of Political Economy 93, 859–880.

Hancock, D., Wilcox, J., 1996. Intraday bank reserve management: the effects of caps and fees on daylight overdrafts. Journal of Money, Credit, and Banking, Part II, 870–908.

Hannan, T., 1994. Recent trends in retail fees and services of depository institutions. Federal Reserve Bulletin 80, Board of Governors of the Federal Reserve System, Washington, DC, pp. 771–781.

Hodrick, R.J., Kocherlakota, N.R., Lucas, D., 1991. The variability of velocity in cash-in-advance models. Journal of Political Economy 99, 358–384.

Humphrey, D., 1981a. Economies to scale in Federal Reserve check processing operations. Journal of Econometrics 15, 155–173.

Humphrey, D., 1981b. Scale economies at automated clearinghouses. Journal of Bank Research 12, 71–81.

Humphrey, D., 1982. Costs, scale economies, competition, and product mix in the US payments mechanism. Federal Reserve Bulletin 68, 215.

Humphrey, D., 1984. The US payments system: costs, pricing, competition and risk. Monograph Series in Finance and Economics, vol. 1/2. New York University, New York.

Humphrey, D., 1985. Resource use in Federal Reserve check and ACH operations after pricing. Journal of Bank Research 16, 45–53.

Humphrey, D., 1986. Payments finality and risk of settlement failure. In: Saunders, A., White, L. (Eds.), Technology and the Regulation of Financial Markets: Securities, Futures, and Banking. Lexington Books, Lexington, MA, pp. 97–120.

Humphrey, D., 1989. Market responses to pricing Fedwire daylight overdrafts. Economic Review, Federal Reserve Bank of Richmond, 23–34.

Humphrey, D., 1994. Delivering deposit services: ATMs versus branches. Economic Quarterly, Federal Reserve Bank of Richmond 80, 59–81.

Humphrey, D., 1995. Payment systems: Principles, practices, and improvements. The World Bank, Washington, DC.

Humphrey, D., Berger, A., 1990. Market failure and resource use: economic incentives to use different payment instruments. In: Humphrey, D. (Ed.), The US Payments System: Efficiency, risk, and the Role of the Federal Reserve. Kluwer Academic Publishing, Norwell, MA, pp. 45–86.

Humphrey, D., Pulley, L., 1997. The check's in the mail: why the US lags in the adoption of cost-saving electronic payments, Working Paper, Florida State University, Tallahassee, FL.

Humphrey, D., Pulley, L., Vesala, J., 1996. Cash, paper, and electronic payments: a cross-country analysis. Journal of Money, Credit, and Banking Part II 28, 914–939.

Humphrey, D., Kim, M., Vale, B., 1997. Realizing the gains from electronic payments: Costs, pricing and payment choices, Working Paper, Norges Bank, Oslo, Norway.

Imrohoroglu, A., Prescott, E., 1991. Seignorage as a tax: a quantitative evaluation. Journal of Money, Credit, and Banking 23, 462–475.

Johnson, H., 1968. Problems of efficiency in monetary management. Journal of Political Economy 76, 971–990.

Kamata, S., 1990. Managing risk in Japanese interbank payment systems. Federal Reserve Bank of San Francisco Economic Review, 18–32.

Kane, E.J., 1996. Comment on alternative monies and the demand for media of exchange. Journal of Money, Credit and Banking Part II 28, 961–964.

Kennickell, A., Starr-McCluer, M., 1994. Changes in family finances from 1989 to 1992: evidence from the survey of consumer finances. Federal Reserve Bulletin, Board of Governors of the Federal Reserve System, Washington, DC, pp. 861–882.

Kocherlakota, N.R., 1996. The equity premium: It's still a puzzle. Journal of Economic Literature 34, 42–71.

Kwast, Myron, Kennickle, A., 1997. Who uses electronic banking? Results from the 1995 survey of consumer finances, Mimeo. Division of Research and Statistics, Board of Governors of the Federal Reserve System, Washington, DC, (also forthcoming in Technology, Proceedings of the 33rd Annual Conference on Bank Structure and Competition).

Lacker, J., 1996a. Electronic Money and Monetary Policy, Mimeo. Federal Reserve Bank of Richmond.

Lacker, J., 1996b. Stored value cards: costly private substitutes for government currency. Federal Reserve Bank of Richmond Economic Quarterly 82, 1–25.

Lacker, J., Schreft, S., 1996. Money and credit as means of payment. Journal of Monetary Economics 38, 3–23.

Lempinen, U., Lilja, R., 1989. Payment Systems and the Central Bank. Bank of Finland, Helsinki, Finland.

Lindsey, D., Gelfand, M., Barrett, C., Gilbert, A., Humphrey, D., Ireland, O., Lyons, J., Marquardt, J., 1988. Controlling Risk in the Payments System, Board of Governors of the Federal Reserve System, Washington, DC.

Llewellyn, D., Drake, L., 1993. The Economics of Bank Charges for Personal Customers. Banking Center Monograph no. 9. Loughborough University, Loughborough, UK.

Marquis, M., Reffett, K., 1994. New technology spillovers in the payments system. Economic Journal 104, 1123–1138.
Mengle, D., 1995. Regulatory solutions to payment system risks: lessons from privately negotiated derivatives. Journal of Financial Services Research 9, 381–391.
Mengle, D., Humphrey, D., Summers, B., 1987. Intraday credit: risk, value, and pricing. Economic Review, Federal Reserve Bank of Richmond 73, 3–14.
Mitsui, T., Watanabe, S., 1989. Monetary growth in a turnpike environment. Journal of Monetary Economics 24, 123–137.
Murphy, N., 1988. Determinants of household check writing: the impacts of the use of electronic banking services and alternative pricing of checking services, Board of Governors of the Federal Reserve System Finance and Economics Discussion Series 38.
Nilson, S., 1995. The Nilson Report. HSN Consultants, Inc., Oxnard, CA.
Paroush, J., Ruthenberg, D., 1986. Automated teller machines and the share of demand deposits in the money supply: the Israeli experience. European Economic Review, 1207–1215.
Porter, R., Judson, R., 1996. The location of US currency: how much is abroad? Federal Reserve Bulletin 82, 883–903.
Prescott, E.C., 1987. A Multiple means-of-payment model. In: Barnett, W.A., Singleton, K.J. (Eds.), New Approaches to Monetary Economics. Cambridge University Press, New York, pp. 42–51.
Richards, H., 1995. Daylight overdraft fees and the Federal Reserve's payment system risk policy. Federal Reserve Bulletin 81, 1065–1077.
Robinson, P., Flatraaker, D., 1995. Costs in the payments system. Norges Bank Economic Bulletin 66, 207–216.
Rochet, J., Tirole, J., 1996a. Controlling risk in payment systems. Journal of Money, Credit, and Banking, Part II, 832–861.
Rochet, J., Tirole, J., 1996b. Interbank lending and systemic risk. Journal of Money, Credit, and Banking, Part II, 733–762.
Rolnick, A.J., Smith, B.D., Weber, W.E., 1996. The United States' experience with state bank notes: lessons for regulating E-cash (A progress report), Mimeo.
Rolnick, A.J., Smith, B.D., Weber, W.E., 1997. Lessons from a laissez-faire payments system: the Suffolk Banking System (1825–1858), Working Paper 584, Federal Reserve Bank of Minneapolis, Minneapolis, MN.
Salop, S.C., 1990. Deregulating self-regulated shared ATM networks. Economics of Innovation and New Technology 1, 85–96.
Samuelson, P.A., 1958. An exact consumption-loan model of interest with or without the social contrivance of money. Journal of Political Economy 56, 467–482.
Santomero, A.M., Seater, J.J., 1996. Alternative monies and the demand for media of exchange. Journal of Money, Credit and Banking Part II 28, 942–960.
Sargent, T.J., Wallace, N., 1982. The real-bills doctrine versus the quantity theory: a reconsideration. Journal of Political Economy 90, 1212–1236.
Sato, S., Humphrey, D., 1995. Transforming Payment Systems: Meeting the Needs of Emerging Market Economies. The World Bank, Washington, DC.
Schoenmaker, D., 1996. A Comparison of alternative interbank settlement systems, Working Paper, Financial Markets Group, London School of Economics, London.
Schreft, S.L., 1992. Transactions costs and the use of cash and credit. Economic Theory 2, 283–296.
Simmons, R., 1987. Would banks buy daytime Fed funds? Economic Perspectives, Federal Reserve Bank of Chicago 11, 34–36.
Spahr, W., 1926. The Clearing and Collection of Checks. The Bankers Publishing Co., New York.
Sprenkle, C.M., 1993. The case of the missing currency. Journal of Economic Perspectives 7, 175–184.
Stavins, J., 1997. A comparison of social costs and benefits of paper check presentment and ECP truncation. Economic Review, Federal Reserve Bank of Boston, July/August, 27–44.

Summers, B. (Ed.), 1994. The Payment system: Design, Management, and Supervision. International Monetary Fund. Washington, DC.

Szegö, G., 1997. A critique of the Basel regulation, or how to introduce additional (im) moral hazards. Proceedings of the Conference on Risk Management and Regulation in Banking, Bank of Israel, Jeruselem (forthcoming).

Townsend, R.M., 1980. Models of money with spatially separated agents. In: Kareken J., Wallace, N. (Eds.), Models of Monetary Economics. Federal Reserve Bank of Minneapolis, Minneapolis, MN, pp. 265–303.

Townsend, R.M., 1989. Currency and credit in a private information economy. Journal of Political Economy 97, 1323–1344.

US Department of the Treasury, 1995. Financial Crimes Enforcement Network, Exploring the World of Cyberpayments. Washington, DC.

Van der Velde, M., 1985. ATM Cost Model: A Survey of Fully Weighted and Incremental ATM Transaction Costs – 1984. Bank Administration Institute, Rolling Meadows, IL.

VanHoose, D., 1991. Bank behavior, interest rate determination, and monetary policy in a financial system with an intraday federal funds market. Journal of Banking and Finance 15, 343–365.

VanHoose, D., Sellon, Jr., G., 1989. Daylight overdrafts, payments system risk, and public policy. Economic Review, Federal Reserve Bank of Kansas City 74, 9–29.

Vesala, J., 1993. Retail Banking in European Financial Integration. Bank of Finland, Helsinki, Finland.

Virén, M., 1994. Demand for Different Payment Media in Finland. Bank of Finland Bulletin, 12–16.

Walker, D., 1980. Electronic funds transfer cost models and pricing strategies. Journal of Economics and Business 33, 61–65.

Wells, K., 1996. Are checks overused? Federal Reserve Bank of Minneapolis Quarterly Review 20, 2–12.

Weinberg, J., 1994. Selling Federal Reserve payment services: one price fits all? Federal Reserve Bank of Richmond Economic Quarterly 80, 1–23.

Wenninger, J., Laster, D., 1995. The electronic purse. Current Issues in Economics and Finance vol. 1. Federal Reserve Bank of New York, New York.

Working Group on EU Payment Systems, 1994. Report to the Council of the European Monetary Institute on Prepaid Cards. European Monetary Institute, Frankfurt.

Zimmerman, G.C., 1981. The pricing of Federal Reserve services under MCA. Federal Reserve Bank of San Francisco Economic Review, 22–40.