

**Retail Payments:  
Overview, Empirical Results, and Unanswered Questions\***

**Norges Bank Research Conference on Payment Systems  
Oslo, Norway  
November, 2008**

**David B. Humphrey  
Department of Finance  
Florida State University**

**\* The views expressed are my own but I benefited from comments by Wilko Bolt,  
Joanna Stavins, and Bent Vale.**

## **Retail Payments: Overview, Empirical Results, and Unanswered Questions.**

### **1. Introduction.**

Retail payments are small to medium value transactions using cash, check, debit and credit card payments at the point of sale. They also include medium value bill payments using checks, giro (in Europe) or automated clearing house (ACH in the U.S.) credit transfers and direct debits, with cards at times used on the Internet or over the phone. In contrast, wholesale payments cover wire transfers (in the U.S. and Europe) and giro payments (in Europe). These are large value transactions among businesses as well as between business and government and are used for almost all financial transactions in the foreign exchange, bond, equity, derivative, and other financial markets both for payment and settlement.

In developed countries, retail payment volume can comprise over 99% of total non-cash transactions but may account for less than 8% of the total value. Just the reverse occurs for wholesale transactions. They can account for less than 1% of non-cash transactions but over 90% of the value. Unfortunately, only guesstimates of retail cash use exist but, if included in retail payments, would push retail payments even closer to 100% of all payment transactions while having only a small effect on raising the value of all payments.

An additional difference between retail and wholesale payments concerns risk. While both generate credit and default risk for participants, only wholesale payment networks—due to the large values being transferred relative to the capital positions of the banks initiating the transfers—are considered to create systemic risk. Credit and default risk have long been recognized in retail payments and have been dealt with through contracts, case law, commercial codes, as well as regulation and legislation (e.g., bankruptcy law). In contrast, systemic risk has only been seriously addressed on wholesale payment networks since the Bankhaus Herstatt failure of 1974. While systemic risk has now largely been contained, liquidity issues and operational risk are the current focus. Although both networks are essential to an economy, this survey covers only retail payments and, in particular, empirical studies of retail payments.<sup>1</sup>

The goal of this survey is to provide an overview of the context in which retail payments occur in Section 2 and outline generally what we know about these payments in Section 3. The determinants of payment composition both across and within countries is noted in Section 4 while payment costs are covered in Section 5. Some areas where further research could prove useful are suggested in Section 6. As usual with empirical analysis, data availability is a serious constraint so empirical research naturally focuses on where data are publicly available.

---

<sup>1</sup> Numerous Bank for International Settlements, European Central Bank, and various country central bank publications detail the characteristics, operational procedures, and systemic risk, liquidity, and payment timing issues associated with wholesale payment networks.

## 2. Overview and the Setting in Which Payments Occur.

Retail payments occur almost entirely within a country's banking system rather than outside of it.<sup>2</sup> Fortunately for empirical researchers, the production functions for the various payment instruments, the way in which these instruments are used for different types of transactions, and the motivations of payment users themselves (mainly a desire for convenience and security) are very similar across countries. While some technical change does occur, it typically takes the form of speeding up the processing of paper-based instruments or shifting from a paper to an electronic version of the same instrument. Over long periods, the composition of instruments has basically evolved from a reliance on cash to checks to cards at the point of sale in the U.S. versus cash to cards (with some checks in between) for most of Europe. For bill payments the evolution has been from cash to checks to giro or ACH payments.<sup>3</sup> Over shorter periods, consolidation of processing centers—for both paper and electronic instruments—has significantly lowered unit costs while the main cost-reducing innovation has been the shift from paper to electronics and the reliance on ATMs (and “cash-back” at merchants) rather than bank branches to obtain cash. For Europe, these changes in payment technology and service delivery are estimated to have saved some \$32 billion in bank operating costs.<sup>4</sup>

Although payment costs have fallen over time, banks retain an effective monopoly in supplying payment services as only they provide the accounts used to transfer funds. In addition, due to most countries' central bank rules, banks have had monopoly access to central bank settlement facilities.<sup>5</sup> While cash does not require central bank settlement—since coin and currency already represent final payment, here too banks are effectively the monopoly retail supplier since to obtain cash depositors' accounts need to be debited. Thus competition within a country's payment system is limited to competition among banks. And competition among payment instruments (actually, the ability to substitute) is also constrained since certain instruments have characteristics that favor their use at the point-of sale but not for bill payments (cash, cards) rather than vice versa (giro and ACH).

Given the commonalities among payment instruments in terms of who supplies them, the types of transactions they are typically used for, and the similar set of characteristics desired by users, it is surprising

---

<sup>2</sup> Retail payments outside of the banking system concern the provision of travelers' checks, telegraph transfers or other remittances, and offering non-bank retail charge cards or bill aggregation. In addition, large credit card firms such as Visa and MasterCard currently are nonbanks although they started life as associations of banks. Person to person cash transactions also occur but cash is obtained first from the banking system.

<sup>3</sup> Many developing countries are focusing on trying to deemphasize checks and move directly from cash to cards or giro payments, a detour possible today but not before electronic payment methods were developed.

<sup>4</sup> The estimate is for changes over 1987-1999 (Humphrey, Willeson, Bergendahl, and Lindblom, 2006). No similar study has been done for the U.S. since time-series payment and ATM data are limited and these data do not exist by bank nor by states for use in a pooled analysis.

that payment composition differs as much as it does across countries. Some of these differences can be viewed as historical accidents associated with differences in bank concentration across countries (for check use in the U.S.) or the existence of a national system of postal banks (for giro use in Europe). Other differences are associated with a time preference for consumption now rather than later magnified by advertising and rewards, which largely explains the rise of credit cards. Finally, country differences in public safety appears to explain the heavy use of cash in Japan. Other than credit card rewards, these considerations are outside the usual set of economic determinants of product demand, such as own and substitute prices, income differences, and supplier production cost.

As payments are rarely directly priced for consumers, this useful single metric is not available to gauge the minimum value of unobserved benefits from using different instruments—the combined implied value of safety, convenience, time opportunity cost, habit persistence, etc., that comprise the user’s utility function. This makes it difficult to assess empirically the underlying economic determinants of payment instrument use. Although business use of different payment instruments is typically priced on a per transaction basis, cross-section or time-series information on price or business use is just not available. Although consumer payment costs are to a degree covered by bank minimum balance requirements, fixed monthly fees, float, etc., they are not clearly “seen” by consumers nor operate as a per transaction “price”. Consequently, the usual demand models economists apply to explain both past use and predict the future are not really available for consumer or business payments. Hence the reliance on consumer demographic survey data of payment instrument use and creative efforts to mine special one-time data sets for “price-equivalent” indicators of payment preference and use. Lacking explanatory information, future payment use is probably reasonably forecasted as a simple function of time using logistic or Gompertz S-curves, although some deeper explanatory models have been developed and are noted below.<sup>6</sup>

Understanding the reasons why different payment instruments are used the way they are is important in developing new instruments or improved methods of operation. This applies to security/fraud concerns in contrasting use of cards with a PIN entered at a terminal versus a signature or cards with a chip with an embedded PIN versus a magnetic stripe. It also applies to consumer convenience and privacy concerns regarding Internet bill payments, mobile phone transactions, or government issued cards replacing cash while retaining seigniorage benefits (Kok, 2002). Although developers of new instruments almost always satisfy the necessary conditions for replacing an existing instrument—that is, the new instrument could be a substitute since its characteristics are the same or somewhat better than those of an existing instrument—they often fail the sufficient condition of getting a large enough portion of users to actually

---

<sup>5</sup> Some recent exceptions exist (e.g., Canada, Australia) where non-bank financial firms can have limited access to central bank settlement services.

adopt them (a topic at the center of network economics). Since relative prices do not play an important role for consumer payments, only when a new instrument significantly improves convenience or security over the instrument it is designed to replace or offers a markedly lower bank cost has it been successful in having banks to supply it and users to adopt it.

### 3. Information on Payment Instrument Use.

**Cash Transactions.** The cost of delivering cash acquisition services to depositors has fallen significantly in the last 15 years as banks have progressively substituted ATM networks for networks of branch offices. Indeed, in many countries the number of expensive stand-alone offices has fallen absolutely or in relation to population growth. While the use of cash is often illustrated by the ratio of cash in circulation to GDP, this is only a first approximation.<sup>7</sup> Even so, Table 1 suggests that European countries in the euro area use more than twice the value of cash than does the U.S. while Japan uses more than twice that of Europe.<sup>8</sup> Using the currency/GDP ratio to rank the five countries/areas shown in the table according to their approximate cash use, this ranking is generally the reverse of each area's total non-cash use per person (in the last column).

Table 1: Payment Instrument Use Across Countries (2005)

	Cash/GDP Ratio	Annual Transactions Per Person:			
		Check	Card	Giro/ACH	Total Non-Cash
U.S.	2.6	112	145	43	300
U.K.	3.5	32	104	95	231
Canada	3.7	42	156	46	244
Euro Area	7.3	16	46	92	154
Japan	16.7	1	39	11	51

Source: BIS and ECB. Latest data available are for 2006, which is very similar to 2005 here.

Other approaches to estimating cash use in a country concern simply adding up the observed annual value of non-cash payments and subtracting it from the value of sales where cash, cards, and checks are known to be used at the point of sale or for bill payments (Humphrey, Kaloudis, Øwre, 2004).

<sup>6</sup> Knowledge of future payment use is useful because, when branch office and ATM networks that deliver payment services are included, payment-related services are by far the largest component of bank total operating expenditures.

<sup>7</sup> This is because the value of currency actually used in some countries can be considerably less than the value in circulation as currency issued in one country can be held and used in other countries.

<sup>8</sup> U.S. data has been corrected for the approximately 60% of cash held overseas rather than used domestically (Porter and Judson, 1996).

Alternatively, these data can be used in an econometric model along with the value of currency in circulation, the interest rate, and ATMs to estimate current cash use (Snellman, Vesala, and Humphrey, 2001).<sup>9</sup> Such econometric analysis suggests that the share of cash in point of sale payments is 20% for the U.S. but can be more than three times higher in Europe (Paunonen and Jyrkönen, 2002). This is consistent with a general aversion to consumer debt in Europe (hence their relatively low use of credit cards) and a history of relying on giro credit transfers for bill payments (where funds have to be in an account in order to make a transfer).

One area where cash is the clear instrument of choice is in tax evasion and other illegal transactions. Here large denomination notes are heavily used while smaller valued notes are common for legal activities and give an estimate of the legal demand for and use of cash in a country (c.f., Drehmann, Goodheart, and Krueger, 2002). The distinction between legal versus illegal cash use by currency note value is supported by the rough agreement between the value of large value notes in circulation times a cash turnover ratio with statistical and survey-based estimates of the value of illegal activity in some European countries (Isachsen and Strom, 1985). Consequently, the demand for cash for legal activities can be approximated by the value of lower value currency notes in circulation. This approach has recently been applied to 13 OECD countries with some support from the fact that the value of notes commonly dispensed by ATM's fall with the availability of debit card terminals while the value of large currency denominations do not (Amromin and Chakravorti, 2008). The value of cash used in illegal or "shadow" activities in a country has also been estimated by comparing changes in nationwide electricity use (reflecting reported plus unreported economic activity) with changes in reported incomes and tax revenues over time. Here the difference is presumed to be associated with the level of tax evasion and other shadow activities. Estimates of illegal/shadow market activity differ greatly but averages 19% of GDP for 12 European countries and 9% for the U.S. (Schneider and Enste, 2000, Table 7).<sup>10</sup> Other analyses have suggested that perhaps 70% of the value of currency in OECD countries is either held outside their country or used in illegal activities, two of which are the drug trade and tax evasion (Rogoff, 1998).

**Check, Card, Giro, ACH, and Other Non-Cash Transactions.** Differences in payment instrument use across countries is perhaps best seen when non-cash payments are in terms of annual transactions per person. As illustrated in Table 1, 112 checks are written per person in the U.S., 16 per person in the 15 country euro area, but only 1 per person in Japan. And while there are 39 card transactions per person in Japan and 46 in the euro area, card transactions are three times higher (at 145 and 156) in the

---

<sup>9</sup> Although theory suggests that the interest rate is an important opportunity cost for consumer demand for cash, if a person withdraws \$400 every two weeks from an account paying 3% rather than withdrawing \$200 once a week, the average cash balance would double from \$100 to \$200 but the opportunity cost is only \$0.50 per month (c.f., Sprenkle, 1993).

<sup>10</sup> See Schneider (2006) for similar results.

U.S. and Canada. In contrast, the U.K. and countries in the euro area make twice as many giro/ACH payments per person than does the U.S. or Canada and over eight times that for Japan. These comparisons illustrate the heavy reliance in Japan on cash for retail transactions, the focus of the U.S. on checks and cards, and the emphasis of the U.K. and the 15 countries in the euro area on giro/ACH payments. In contrast to data on cash transactions or value which have to be approximated by researchers before trends and influences on its use can be made, data on non-cash transactions and payment values are quite good and are published annually by the Bank for International Settlements (“Red Book”) and the European Central Bank (“Blue Book”). As a result, the literature has focused on non-cash payment instruments.

One newer way to make a payment at the point of sale is with e-money, such as a card with a chip. While this instrument currently accounts for less than 1% of non-cash transactions, it was viewed as a promising replacement for small value cash payments in terms of convenience as well as merchant and bank costs once (and if) large transaction volumes could be obtained. Indeed, estimates of the marginal production cost of a card with a chip (aka stored value or electronic purse) transaction in the Netherlands is only 30% of that for cash and 60% for Belgium (Brits and Winder, 2005, Table 4.3; Quaden, 2005, Table 3). After an initial flurry of activity, however, its rate of adoption slowed markedly and market penetration has not lived up to expectations. Indeed, some banks in Europe are now abandoning their earlier programs (Van Hove, 2006). While a stored value card can be an effective replacement for small value cash transactions, it needs a terminal to be accepted and often a different terminal to be refilled just as cash in a wallet needs replacing. Until more terminals/readers are in place and merchants can realize cost savings from larger volumes, cash seems to be preferred due to its broader acceptability and most consumers would rather use one instrument for small value transactions rather than two.

Another relatively recent payment development involves the use of the Internet in Europe to initiate giro bill payments (credit transfers). While Europe, Japan, and the U.S. have for some time had arrangements whereby billers could be pre-authorized to debit consumers’ deposit accounts for recurring bill payments (insurance, home loan, and utility payments), the Internet arrangement is usually preferred since this allows the consumer to initiate the debit to its account (generating a credit transfer) and thus returns control to the payor compared to a pre-authorized debit. Similar on-line banking arrangements have lagged in the U.S. since, unlike Europe with their pre-existing giro network, many U.S. billers are not set up to accept ACH transactions from individual consumers. While this issue is being addressed over time, problems arising from identity theft and fraud will have to be addressed better before Internet banking becomes the norm.<sup>11</sup>

---

<sup>11</sup> More recently, the possibility of conducting banking and payment activity using a mobile phone has stirred some interest. This is just an extension of hard-line telephone banking to a mobile phone paired with mobile Internet access similar to what one now has with a laptop computer.

#### 4. Determinants of Payment Composition.

**Cross-Country Differences in Payment Composition.** The major differences in payment composition between the U.S., Europe, and Japan are not due to price or cost differences but rather to what could be termed historical “accidents” of an institutional, geographical, or cultural nature rather than from following a clear plan to shape payment use. And once a particular payment arrangement becomes established, it is difficult to replace it with something different unless the newer arrangement is clearly and markedly superior. Even then, the substitution process is slow due to user habit persistence and, in some cases, the likelihood of technological “lock-in” effects.

One important institutional “accident” was the development of a postal giro system in Europe before 1900 but never in the U.S. Even though the U.S. postal service had the authority to establish postal banks early on, the farthest this got was selling war bonds at post offices during the 1940s as an adjunct to raising government funds for WWII. Before 1900, banks in Europe did not importantly serve the common man—only merchants and the wealthy. Post offices in Europe started offering savings accounts for the common man and, noticing that one person would withdraw funds to pay another who would then re-deposit the funds back with the post office, the giro concept was born. As the postal system is national in scope, the postal giro naturally evolved into a nationwide payment network, first using paper and now electronics. Banks, after a delay and eyeing this growing deposit base, later developed their own giro network.

In contrast, U.S. banks effectively served everyone but with low bank chartering requirements, thousands of small banks were established in a geographically very large country. With no nationwide banks due to restrictions on intra-state and interstate banking (associated with a fear of a concentration of economic power), the U.S. had no alternative but to rely on checks and correspondent banking for transactions among banks by individuals and businesses since no national payment supplier existed. Consequently, the ongoing shift to electronics for bill payments and disbursements has been slower in the U.S. than in Europe.

Another institutional example concerns the cost of obtaining a banking charter in Canada versus the U.S. over 100 years ago. The capital base required to establish a bank in Canada was dramatically greater than that for the U.S. with the result that few banks were established in Canada but many banks in the U.S. The end result is that Canada has a very concentrated banking system where the top 6 banks account for over 90% of banking assets while in the U.S. it would take 508 banks to achieve a 90% concentration. While Canada, like the U.S., also relied on checks for payments (replacing an earlier emphasis on cash in both countries), the efficiency of their check collection system turned out to be far superior. One important reason was the Canadian banking practice of charging firms for the check float they created, giving a strong incentive to save real resources and disburse checks close to where the payees were located. Also, the

central bank backdates interbank check settlements by one day, minimizing check float through an accounting procedure. Only in 2004 was the U.S. able to pass legislation to give collecting banks the legal authority under Check 21 to require paying banks to pay upon presentment of an electronically transmitted digital image of a check. Today about one-third of checks are paid with a one-day lag after being deposited and truncated at a bank of first deposit<sup>12</sup> while another one-fourth are “on-us” items and are paid on a same-day or next-day basis after being deposited in an account in the same bank they are drawn upon.

As mentioned earlier, differences in public safety—a cultural issue—are commonly held to be the primary reason for Japan’s strong reliance on cash for point of sale transactions with almost no use of checks or debit cards (the 39 annual card transactions per person in Japan is due almost entirely to credit cards in Table 1). Even so, most bill payments are made electronically at month end in Japan. Although some may view the heavy reliance on cash in Japan as an indication of “insufficient evolution” of their payment system, it is more instructive to see this reliance as evidence that Japan—unlike the U.S. and Europe—does not need to rely on non-cash payment instruments. Being a much safer country, there is little incentive to switch to more secure non-cash payment instruments at the point of sale (Federation of Bankers Associations of Japan, 1994).

A final “accident” example concerns the early U.S. practice of discounting the face value of a check to cover the cost of collection (non-par checking). This resulted in payment delays that were at times inordinately long and disrupted commerce. One reason the Federal Reserve was established was to transport and process checks at face (par) value and at no charge to eliminate circuitous routing of checks which hindered commerce. Non-par checking was eliminated but today, alone among central banks, the Federal Reserve processes about one-third of all checks and the majority of ACH payments (now for a fee). Europe, not having many (debit transfer) checks to begin with due to a reliance on cash and giro (credit transfer) payments, instead covered payment costs by debiting accounts prior to the value date to earn float revenues and central banks are not involved in processing retail transactions in competition with banks.<sup>13</sup>

**Within Country Changes in Payment Composition.** Attempts to determine the main drivers of changes in payment composition within a country have been hampered by a lack of historical data, which is basically limited to the past 15 years, and then only reported as annual national-level aggregates. This makes degrees of freedom a problem for time-series analysis. In addition, as noted above, data on cash use is highly approximate at best so that shifts away from cash toward cards, for example, are difficult to gauge. Finally, there is the problem that there is very little direct pricing of payment services (Norway excepted) so

---

<sup>12</sup> A similar truncation procedure has been in operation for some time in Germany.

<sup>13</sup> Information on the long-term evolution of payment instruments in different countries exists in the economic history literature but no paper focusing only on payments seems to exist. Consequently, the exposition in the text is from earlier work (Humphrey, Sato, Tsurumi, and Vesala, 1996).

that the standard economic approach to “explaining” changes in the demand for a good or service—own and substitute prices—is very limited. Since most pricing of consumer payments in the U.S. and Europe is indirect and relies on fixed monthly account fees, minimum balance requirements, no or low interest paid on deposits, and/or payment float, efforts have naturally focused on using these measures as price equivalents.

Although time-series or cross-section data on such indirect prices are rare, when they have been available they were earlier used to estimate the implicit cost to depositors of obtaining transaction, safekeeping, and other non-priced services from banks (Heffernan, 1992). Along this line, others have collected data on changes in various bank fees, deposit requirements, and service levels and associated these changes to variations in deposit composition for both consumers and businesses (Hannan, 2002; Anguelov, Hilgert, and Hogarth, 2004). While this can be useful for illustrating the apparent value of various non-priced banking services, the link to consumer use of cards versus checks versus cash is missing since this information is not available by bank.

Faced with these data problems researchers have turned to socio-demographic data from recurring surveys in Europe and the U.S. (Stavins, 2001; Hayashi and Klee, 2003; Mester 2006; Klee, 2006; and others too numerous to mention here). Analysis of demographic data indicate that younger individuals, those with more education, higher income, and work at white collar type jobs have higher adoption rates of the newer electronic forms of making a payment or obtaining cash from an ATM. However, as Schuh and Stavins (2008) have recently shown, demographic influences on payment instrument use are secondary compared to individuals’ assessment of the relative cost, convenience, safety, privacy, and other characteristics of different payment instruments (even when demographic influences are controlled for). While data on these characteristics come from a special survey of Federal Reserve System employees, other similar survey-based studies by Jonker (2007) and Borzekowski, Kiser, and Ahmed (2008) that have incorporated similar payment characteristics find that they importantly augment the socio-demographic determinants of payment instrument use.

What is useful about the payment characteristics approach (requiring survey information on why users favor one instrument over another) is that it is possible to estimate the “price equivalent” trade-off among payment instruments. This approach uses probit/tobit models to suggest price responsiveness relationships among payment instruments. For example, Borzekowski, Kiser, and Ahmed (2008) find a very elastic response to fees imposed on PIN debit transactions (an effort by banks to shift users to signature debit cards where merchant interchange fees are higher, raising bank revenue). In another application, Ching and Hayashi (2008) suggest that the likely effect of dramatically reducing or eliminating already existing credit card reward programs would have a relatively small effect on credit versus debit card or paper versus electronic payment composition. This result is consistent with the actual experience of Australia where merchant interchange fees and, consequently, card reward programs, have been markedly

reduced. But what if credit card reward programs did not exist in the first place? Addressing this question with Australian data, Simon, Smith, and West (2008) suggest that credit cards would never have overtaken debit cards to begin with and the latter would have been the dominant instrument. This illustrates the importance of payment habits: that is, reward programs were important in getting people to use credit cards but once commonly used, removing or constraining these programs seems to have little effect on credit card use.<sup>14</sup>

More indirect indicators of payment instrument substitution rely on surveys that indicate changes in consumer use over time or relative use of one instrument versus another. An early example was the almost 50% reduction in U.S. family cash holdings observed from surveys over 1984 to 1995 (Hancock and Humphrey, 1998), an analysis continued in Klee (2006) for more recent time periods. The survey approach also indicates that consumers who often use debit cards hold 20% less cash (Stix, 2003) while others verify again that the type of transaction and its value clearly favor one type of payment instrument over another (Bounie and Francois, 2006).

Generalizing from other markets, researchers have long believed that per transaction pricing of payment instruments would have had a marked impact on the composition of payment instruments within a country by speeding up the shift to cheaper electronic card payments away from cash and checks. Comparing Norway, which implemented per transaction pricing, and the Netherlands, which did not price, it appears that pricing speeded up this shift in Norway by about 20% (Bolt, Humphrey, and Uittenbogaard, 2008). This effect is smaller than one might expect but payment demand elasticities appear to be inelastic (Humphrey, Kim, and Vale, 2001) and non-price characteristics of electronic payments seem to be more important than the relative prices being charged in explaining volume growth. In this regard, terminal availability is a good summary indicator for the non-price attributes of debit card use since mere availability is quickly followed by use. Similarly, for ATMs convenience often trumps pricing which is why non-banks offer ATM services to any bank depositor for a fee and customers of one bank many times—especially when traveling—use another bank's ATMs even though a surcharge applies (Hannan, Prager, and McAndrews 2003). However, the importance of convenience is not always so strong. This is illustrated for the Netherlands using 2006 data where merchants are permitted to surcharge debit card use and consumer and retailer survey data indicate a substitution to cash at merchants that impose a surcharge. Since debit card use is cheaper than cash above a certain transaction size (and this breakeven point is now lower than the €8 to €12 threshold most merchants use), abolishing card surcharging could lower payment costs (Bolt, Jonker, and van Renselaar, 2008).<sup>15</sup>

---

<sup>14</sup> Such a non-symmetric response is common and was used long ago to explain the stability of consumption with reductions in income but a proportional increase in consumption as income rose.

<sup>15</sup> Earlier work on determining the merchant cash/debit card breakeven point using 1998 data is outlined in ten Raa and Shestalova (2004).

## 5. Payment Costs.

As a rough estimate, the full cost of a nation's payment system is around 1% to 2% of GDP annually depending on current payment composition (as electronic payments are only 1/3 to 1/2 the cost of paper-based instruments) and how much of bank branch and ATM network costs are included as essential for check deposit, cash withdrawal, and card issue and maintenance activity.<sup>16</sup> Another cost indicator is that the user plus merchant plus bank cost of making a \$50 or €50 payment is around 1% to 5% of the transaction value, depending on the instrument used. With improvements in accounting systems over the last 20 years, banks today have a better idea of what their payment production costs are but (as noted earlier) consumer per transaction fees are rarely assessed. Consequently, on the demand side consumers choose different payment instruments for different applications—cash, checks, cards for local point of sale transactions versus giro, ACH, and checks for more distant bill payments—depending on the characteristics of the different instruments. On the merchant side, there can be limits on cash use due to counterfeiting concerns or restrictions on card use for only higher value transactions to cover better merchant card-related fees paid to banks.

On the supply side, cost considerations first induced banks to shift cash acquisition by consumers away from branch offices to cheaper ATMs. Later, similar cost considerations led banks to try to replace cash and checks with cards using POS terminals even though that affects ATM use (Carbo and Rodriguez, 2008). Averaging two European cost accounting studies (for the Netherlands and Belgium) suggests that the bank plus merchant average cost of a cash transaction at the point of sale is around €0.42 while a debit card is €0.52 with incremental costs of €0.12 and €0.20, respectively (Brits and Winder, 2005; Quaden, 2005). Average credit card costs are €3.11 with incremental cost of €0.68. As the ratio of marginal to average cost reflects scale economies, these figures also suggest that cash and cards realize strong cost scale economies (from .29 to .38, respectively, for cash and debit cards and lower still at .22 for credit cards). Similar payment scale economy estimates for Europe are obtained using confidential bank payment processing cost data (Bolt and Humphrey, 2008) or publicly available cross country information (Bolt and Humphrey, 2007).<sup>17</sup> Thus consolidation of payment processing operations across borders in Europe offers the opportunity for substantial reductions in payment costs (the U.S. is already consolidated).

More recently, cost accounting analysis and other information has been applied to determine the unit cost and benefits of U.S. payment instruments at the bank and merchant level (Garcia-Swartz, Hahn,

---

<sup>16</sup> For example, payment cost estimates for some smaller European countries are much less than 1% (Takala and Viren, 2008).

<sup>17</sup> There are numerous papers which have estimated scale economies and technical change in Federal Reserve check, ACH, and wire transfer operations. One of the more recent studies is by Adams, Bauer, and Sickles (2002). These studies find strong payment scale effects but basically no scope economies.

and Layne-Farrar, 2006; Shampine, 2007). Consumer costs and benefits from using the various payment instruments, however, remain difficult to estimate accurately and different assumptions do lead to different conclusions, such as which is cheaper to use—cash or credit card?<sup>18</sup> In this regard the two papers just cited should be read together since their contrasting assumptions and their differential effects are well-explained. All assumptions, initially at least, appear to be reasonable but as more information is gathered and more work is done some will be seen to be more reasonable than others. In all likelihood which instrument turns out to be less costly (privately or socially) will turn on the transaction size being considered as well as the transaction volume being processed in a country (since scale economies are quite important).<sup>19</sup> At present, the most emotional issue is the relative cost—private or social—of credit cards compared to other payment instruments. Card firms have paid damages to merchants where debit card acceptance and its interchange fee was tied to merchant acceptance of a credit card and court cases are underway concerning how the interchange fee is set. In a related move, competition/regulatory authorities in some countries have argued that interchange fees considerably exceed their underlying cost (sans reward programs) and have pushed to have these fees reduced either directly by regulatory action (Australia) or indirectly by consultation (Spain).

## **6. Unanswered or Partially Answered Empirical Questions.**

While there has been an increasing amount of work done in the payments area over the last decade or so, some interesting issues remain. The eight topics listed below represent areas where additional research in retail payments may be productively focused. The topics differ in their complexity but all require some inventiveness in devising a workable empirical strategy both in terms of obtaining the requisite data and in applying proper econometric procedures and analysis.

1. Direct Versus Indirect Pricing of Payment Services. The main question here is: to what degree are resources already allocated efficiently at banks that currently use indirect pricing to influence payment user behavior? This efficiency issue can differ between consumer and business payment use. Posed differently, should bank regulators and antitrust authorities encourage banks to implement per transaction pricing as a substitute for no or low depositor interest returns, delayed availability of deposits, debiting prior to bill payment value dates, monthly fixed account maintenance fees? What are the revenue returns from these indirect fees and do they generally cover the underlying bank payment costs? How much more efficiently, if at all, could resources be allocated with direct fees? Would the direct fees have to cover most/all underlying costs to result in significantly more efficient payment use?

---

<sup>18</sup> A survey of this literature is contained in Koivuniemi and Kempainen (2007).

<sup>19</sup> Other payment cost studies exist for Norway (Robinson and Flatraaker, 1995; Gresvik and Øwre, 2003), Australia (Simes, Lancy, and Harper, 2006), Sweden (Bergman, Guibourg, and Segendorf, 2007), Spain (Carbo, Humphrey, Linares, and Rodriguez, 2007), and Finland (Takala and Viren, 2008).

2. Payment Privacy: What is it Worth and How has it Traded-off With Other Characteristics? In retail payments, debit and credit cards will continue to replace cash in Europe and cash and checks in the U.S. As card transactions are not anonymous, the continued replacement of cash will raise privacy issues, especially in countries with high sales taxes where cash is still heavily used (facilitating tax evasion). Similar rapid growth and privacy concerns will likely be seen in electronic retail payment systems used for transportation services such as toll roads and in-town bus/rail services. The question here is how important is the privacy issue? How much might (some) users pay to retain privacy relative to their likely gain in convenience, security, and other payment instrument characteristics? Namely, how much is privacy worth? Is this a cultural difference that differs across countries? Is there only a vocal minority who really care about privacy, either for personal or for tax evasion reasons?

3. Do Increased Merchant Sales Cover the Cost of Credit Card Reward Programs? This is the main argument card firms use to have merchants agree to accept their cards but there is little data or analysis available to support or reject this view. Indeed, often merchants seem to agree to accept cards as a defensive measure to maintain—rather than expand—their sales. So, do credit card reward programs expand individual merchant sales (or prevent their decline) and, if so, by how much? Are the profits (losses) from the expanded (decreased) sales and/or the ability of merchants to pass on card costs to customers enough to offset the bank fees for card use and terminal rental compared with merchant costs of accepting other payment instruments? How has this changed as credit card use has become more widespread? Are the effects of reward programs and no surcharge rules symmetrical so that removing them will lower merchant sales and card use to the same degree that sales and card use rose when rewards were first offered and no surcharge rules adopted?

4. How Close is the Correspondence Between Private/Social Payment Costs and Relative Payment Use? If the estimated private or social costs of different payment instruments have been correctly estimated, these relative costs—passed on via indirect bank fees to consumer users—should be reflected in the different relative consumer use of these instruments. Put differently, do the estimated net benefits of different payment instruments correspond to observed changes or levels of actual usage? If not, then why not—poor cost estimates, importance of non-priced payment characteristics, or poor pass-through of costs to indirect fees so users lack clear “price” incentives?

5. Bank Competition in Providing Payment Services. European banks are estimated to have saved some \$32 billion in operating expenses over 1987-1999 due to the shift from paper-based payments to

electronics and away from branches toward ATMs for cash acquisition. What portion of these estimated cost savings have been passed on to depositors and, if so, in what form? Lower monthly fees? Lower float revenues? Higher interest paid on deposits? Are any of these changes related to current indicators of banking market competition? Have the increased efficiencies been used to keep real bank fees roughly constant rather than rising over time?

6. Should the Government Continue to Supply Cash When Most of it is Used for Illegal Activities?

The share of cash used at the point of sale and other legal activities appears to be falling in most developed countries. This raises two public policy issues. First, using projections of the future legal use of cash, when might significant government revenues have to be used to redeem consumer use of cash? Second, as legal use of cash falls to low levels, should the government continue to supply it—especially large denomination notes--when these notes mainly facilitate illegal activities such as the drugs trade and tax evasion? The main problem raised with this topic concerns the ability to accurately estimate the shares of currency in circulation used in legal versus illegal activities and then to project the former forward in time. As well, what really are the costs and benefits of cash as a payment instrument, an estimate that would include seigniorage and privacy benefits versus printing, distribution, and replacement costs of cash?

7. Towards A Cardinal Measure of Payment Instrument Use. So far, surveys of what payment instruments are used by consumers and what characteristics of these instruments are important to users has basically been ordinal in nature. That is, these surveys attempt to obtain an ordinal ranking of payment instrument use by consumers and, effectively, an ordinal ranking of characteristics (where a characteristic is either considered to be important or not important). Hence the use of probit/tobit analysis to judge the influence of payment characteristics on payment choice. The next step would be to try to devise a survey that could gauge, in a rough manner, the degree of importance of a particular payment characteristic relative to a few others and in this way obtain an approximate cardinal indicator of payment instrument use.

8. Business Use of Payment Instruments. Consumer use of different payment instruments has been the focus of almost all empirical payment research, with business use—when mentioned at all—noted as an afterthought. But business use of different payment instruments should be more sensitive to relative costs/prices than consumer use for two reasons. First, business use of payment instruments is typically subject to per transaction pricing, so businesses “see” the prices and can respond better than consumers to minimize costs. Second, the volume of business payment transactions is large so that even small differences in costs/prices can have an important influence on business cash management behavior and firm profitability. Determining how business payment use has changed over time, and the reasons for these

changes, will not be simple since public information on business payment use will have to be gleaned from multiple sources as payment data are rarely separated by businesses versus consumers. Unfortunately, the choice that businesses make in using one instrument over another have until recently often been driven by an incentive to slow down payments rather than speed them up. Check 21 in the U.S. and Electronic Data Interchange in the U.S. and elsewhere, however, is changing these calculations and offers business the opportunity to achieve marked reductions in the cost of their cash management and account receivable/payable operations.

## 7. Bibliography

- Adams, R., Bauer, P., and Sickles R. (2002). Scope and scale economies in Federal Reserve payment processing, working paper 02-13, Research Department, Federal Reserve Bank of Cleveland, November.
- Amromin, G., and Chakravorti, S. (2008). Whither loose change? The diminishing demand for small denomination currency, working paper, Federal Reserve Bank of Chicago, February.
- Anguelov, C., Hilgert, M., and Hogarth, J. (2004). U.S. consumers and electronic banking, 1995-2003, *Federal Reserve Bulletin*, 90: 1-18.
- Bank for International Settlements, (various years). *Statistics on Payment Systems in the Group of ten Countries*. Basle (“Red Book”).
- Bergman, M., Guibourg, G., and Segendorf, B. (2007). The cost paying—Private and social cost of cash and card payments, working paper No. 212, Sveriges Riksbank, Stockholm, December.
- Bolt, W., and Humphrey, D. (2007). Payment network scale economies, SEPA, and cash replacement, *Review of Network Economics*, 6: 453-473.
- Bolt, W., and Humphrey, D. (2008). Payment scale economies from individual bank data, working paper, De Nederlandsche Bank, August.
- Bolt, W., Humphrey, D. and Uittenbogaard, R. (2008). Transaction pricing and the adoption of electronic payments: A cross-country analysis, *International Journal of Central Banking*, 4: 89-123.
- Bolt, W., Jonker, N., and van Renselaar, C. (2008). Incentives at the counter: An empirical analysis of surcharging card payments and payment behaviour in the Netherlands, working paper, De Nederlandsche Bank, March.
- Borzekowski, R., Kiser, E., and Ahmed, S. (2008). Consumers' use of debit cards: Patterns, preferences, and price response, *Journal of Money, Credit and Banking*, 40: 149-172.
- Bounie, D., and Francois, A. (2006). Cash, check or bank card? The effects of transaction characteristics on the use of payment instruments, working paper ESS-06-05, Telecom Paris Economics and Social Sciences.
- Brits, H. and Winder, C. (2005). *Payments Are No Free Lunch*. Occasional Studies, Vol. 3/Nr. 2, De Nederlandsche Bank.
- European Central Bank (various years). *Payment and Securities Settlement Systems in the European Union*. Frankfurt (“Blue Book”).
- Carbo, S., Humphrey, D., Linares, J., and Rodriguez, F. (2007). A cost-benefit analysis of a two-sided market, working paper, FUNCAS, Madrid, October.
- Carbo, S., and Rodriguez, F. (2008). ATMs vs. POS terminals: A horse race?, working paper, University of Granada.
- Ching, A., and Hayashi, F. (2008). Payment card rewards programs and consumer payment choice, working paper, Federal Reserve Bank of Kansas City, June.

- Drehmann, M., Goodhart, C., and Krueger, M. (2002). The challenges facing currency usage: Will the traditional transaction medium be able to resist competition from new technologies?, *Economic Policy*, Volume?: 195-227.
- Federation of Bankers Associations of Japan (1994). *Payment Systems in Japan*. Tokyo.
- Garcia-Swartz, D., Hahn, R., and Layne-Farrar, A. (2006). The move toward a cashless society: A closer look at payment instrument economics, *Review of Network Economics*, 5: 175-198.
- Gresvik, O., and Øwre, G. (2003). Costs and income in the Norwegian payment system 2001: An application of the activity based costing framework, working paper No 2003/8, Norges Bank, Oslo.
- Hancock, D., and Humphrey, D. (1998). Payment transactions, instruments, and systems: A survey, *Journal of Banking and Finance*, 21: 1573-1624.
- Hannan, T. (2002). Retail fees of depository institutions, 1997-2001, *Federal Reserve Bulletin*, 88: 406-413.
- Hannan, T., Prager, R., and McAndrews, J. (2003). To surcharge or not to surcharge: An empirical investigation of ATM pricing, *Review of Economics and Statistics*, 85: 990-1002
- Hayashi, F., and Klee, B. (2003). Technology adoption and consumer payments: Evidence from survey data, *Review of Network Economics*, 2: 175-190
- Heffernan, S. (1992). A computation of interest equivalences for non-price characteristics of bank products, *Journal of Money, Credit, and Banking*, 24: 162-172.
- Humphrey, D., Kaloudis, A., and Øwre, G. (2004). The future of cash: falling legal use and implications for government policy, *International financial Markets, Institutions and Money*, 14: 221-233.
- Humphrey, D., Kim, M., and Vale, B. (2001). Realizing the gains from electronic payments: Costs, pricing and payment choice, *Journal of Money, Credit, and Banking*, 33: 216-234.
- Humphrey, D., Sato, S., Tsurumi, M., Vesala, J. (1996). *The Evolution of Payments in Europe, Japan, and the United States*. The World Bank, Financial Sector Development Department, October.
- Humphrey, D., Willeson, M., Bergendahl, G., and Lindblom, T. (2006). Benefits of a changing payment technology in European banking, *Journal of Banking and Finance*, 30: 1631-1652.
- Isachsen, A., Strom, S. (1985). The size and growth of the hidden economy in Norway, *Review of Income and Wealth*, 31: 21-38.
- Jonker, N. (2007). Payment instruments as perceived by consumers: Results from a household survey, *De Economist*, 155: 271-303.
- Klee, E. (2006). Families use of payment instruments during a decade of change in the U.S. payment system, working paper, Federal Reserve Board Finance and Economics Discussion Series, 2006-1.

- Koivuniemi, E., and Kempainen, K. (2007). On costs of payment methods: A survey of recent studies, working paper, Bank of Finland.
- Kok, L.S. (2002). Singapore electronic legal tender (SELT)—A proposed concept, in: *The Future of Money*. Organization for Economic Co-Operation and Development, Paris: 147-161.
- Mester, L. (2006). Changes in the use of electronic means of payment: 1995-2004, Federal Reserve Bank of Philadelphia *Business review*, Q2: 26-30.
- Paunonen, H., and Jyrkönen, H. (2002). Cash usage in Finland—How much can be explained?, working paper, Bank of Finland Discussion Papers, 10/2002.
- Porter, R., and Judson, R. (1996). The location of U.S. currency: How much is abroad?, *Federal Reserve Bulletin*, 82: 883-903.
- Quaden, G. (2005). *Costs, Advantages and Drawbacks of the Various Means of Payment*. National bank of Belgium, English Summary, December: 41-47.
- Robinson, P., and Flatraaker D. (1995). Costs in the payment system, *Norges Bank Economic Bulletin*, 66: 207-216.
- Rogoff, K. (1998). Blessing or curse? Foreign and underground demand for Euro notes, *Economic Policy*, 26: 263-290.
- Schneider, F. (2006). Shadow economies and corruption all over the world: What do we really know?, working paper, IZA Discussion Paper No. 2315, September.
- Schneider, F., and Enste, D. (2000). Shadow economies: Size, causes, and consequences, *Journal of Economic Literature*, 38: 77-114.
- Schuh, S., and Stavins, J. (2008). Why are (some) consumers (finally) writing fewer checks? The role of payment characteristics, working paper, Federal Reserve Bank of Boston, March.
- Shampine, A. (2007). Another look at payment instrument economics, *Review of Network Economics*, 6: 495-508.
- Simes, R., Lancy, A., and Harper, I. (2006). Costs and benefits of alternative payments instruments in Australia, working paper, Melbourne Business School, April.
- Simon, J., Smith, K., and West, T. (2008). Price incentives and payment patterns: Money for nothing and your flights for free, working paper, Payments Policy Department, Reserve Bank of Australia, Sydney, June.
- Snellman, J., Vesala, J., and Humphrey, D. (2001). Substitution of noncash payment instruments for cash in Europe, *Journal of Financial Services Research*, 19: 131-145.
- Sprenkle, C. (1993). The case of the missing currency, *Journal of Economic Perspectives*, 7: 175-184.
- Stavins, J. (2001). Effect of consumer characteristics on the use of payment instruments, Federal Reserve Bank of Boston *New England Economic Review*, 3Q: 19-31.
- Stix, H. (2003). How do debit cards affect cash demand? Survey data evidence, *Empirica*, 31: 93-115.

Takala, K., and Viren, M. (2008). Efficiency and costs of payments: Some new evidence from Finland, working paper, Bank of Finland Research Discussion Papers 11, Helsinki.

ten Raa, T., and Shestalova, V. (2004). Empirical evidence on payment media costs and switch points, *Journal of Banking and Finance*, 28: 203-213.

Van Hove, L. (2006). Why electronic purses should be promoted, *Banking and Information Technology* (Universität Regensburg, Germany), Nr. 2, June: 20-31