Econometric Research and Special Studies Department

# Cash Management and the choice of payment media: a critical survey of the theory

C. K. Folkertsma and G.E. Hebbink

*[Translation of Research Memorandum WO&E nr 532]*
*January 1998*

# Cash management and choice of payment media: a critical survey of the theory

C.K. Folkertsma and G.E. Hebbink

Translation of
Research Memorandum WO&E nr 532/9802

De Nederlandsche Bank NV
Afdeling Wetenschappelijk
onderzoek en econometrie
Postbus 98
1000 AB  AMSTERDAM

Cash management and choice of payment media: a critical survey of the theory

C.K. Folkertsma and G.E. Hebbink

## ABSTRACT

This report addresses the question whether economic theory can explain the choice of payment media for over-the-counter payments. A survey of the literature reveals that most of the existing microeconomic models are limited to the explanation of the size of cash balance. Only two models describe the optimal choice between different modes of payment such as cash payment, payment by electronic money and debit card payment. One of these approaches has been extended in this report, in order to study the effects on transactions balances and optimal mode of payment, for example, caused by interest rate changes or the introduction of new payment media.

Keywords: cash management, demand for money, payments behaviour
JEL codes: D11, E41

Betaalgedrag: een kritisch overzicht van de theorie met het oog op empirische toepassing

C.K. Folkertsma en G.E. Hebbink

## SAMENVATTING

Dit rapport betreft de bijdrage van de economische theorie aan de analyse van het betaalgedrag van consumenten bij toonbankbetalingen. Uit een literatuuroverzicht van micro-economische modellen blijkt dat de theorie zich voornamelijk beperkt tot verklaring van de omvang van chartale transactiekassen. De keuze van consumenten tussen verschillende betaaltechnieken zoals contante, electronische en debetkaart-betaling, wordt echter nauwelijks gemodelleerd. Een van de twee theoretische modellen die deze keuze wel verklaren wordt in het rapport uitgebreid, zodat kan worden aangegeven hoe transactiekassen en gebruik van betaaltechnieken reageren op de rente of de introductie van betaaltechnieken. Verder is dit uitgebreide model geschikt voor empirische toepassing om deze effecten te kwantificeren.

Trefwoorden: betaalgedrag, geldvraag, betalingsverkeer
JEL codes: D11, E41

# 1 INTRODUCTION

Until some decades ago, consumers wishing to make over-the-counter (OTC) payments had but one option, viz. to use cash, i.e. banknotes and coins. The economic analysis of consumers' payment behaviour was therefore concerned mainly with the demand for the optimal stock of banknotes and coins in transactions balances. The best-known model was introduced by Baumol (1952) and Tobin (1956). In the Netherlands and in particular at the Nederlandsche Bank, research on this subject has so far concentrated mostly on demand for banknotes and coins (Fase and Van Nieuwkerk, 1977; Fase, 1985; Boeschoten and Fase, 1992) and the cash management and payment behaviour of households on the basis of survey data (Fase and Boeschoten, 1985; Boeschoten and Fase, 1989; Boeschoten, 1998). Since then, much has changed where payments are concerned and the near future will only be seeing further developments in that respect. Nowadays, consumers can choose between cash and non-cash payment modes, between different monies, such as currency and money in a debitable account (Fase and Vleminckx, 1995) and between different payment techniques, such as currency, credit card or debit card payments. The number of options has recently increased further as a result of the introduction of chip cards or electronic purses (Boeschoten, 1996; Fase, 1996). Another new money is network money, which so far is used on a limited scale only. In this report, chip cards and network money together form electronic money.

The increasing number of instruments for cash management available to consumers gives rise to the question to what extent economic theory can help to provide more insight into the choices which consumers make when engaging in OTC payments. This report attempts to answer that question. The type of this question requires a micro-economically-oriented, theoretical investigation. We opted, moreover, to attune the analysis from the outset to the institutional aspects of payments.

The next section delineates the problem in greater detail, and presents a list of the monies currently in use in the Netherlands and the payment techniques employed. Section 3 gives a brief overview of the economic literature on the choices made by consumers when undertaking payments. Section 4 goes more extensively into two models with an approach underlain by inventory theory, based predominantly on the Baumol-Tobin model. It also looks at the extent to which these models can be empirically applied. Finally, the conclusions are presented in section 5.

## 2 TERMINOLOGY OF OTC PAYMENTS

For a theoretical analysis of the choices made with respect to over-the-counter payments, the institutional characteristics of payments will form the basic starting point, where possible. It is stressed once more that the scope of our analysis is restricted to *consumer* choices with regard to *over-the-counter* payments. It is estimated that slightly less than half of total consumer spending in the Netherlands (in guilders) is accounted for by OTC payments, viz. to businesses, shops, restaurants etc. The remainder consists of non-cash payments in the form of standing orders, in payment transfers and direct debits. The bulk of total payments, i.e. interbank transfers, is not discussed here. Payments between businesses, likewise mainly non-cash, are not considered here, either. Consumers' OTC payments may be broken down into cash payments (banknotes and coins) and non-cash payments made by PIN or debit card, credit card, cheques and chip cards.[1]

Apart from the above distinction between cash and non-cash payment modes, it is also possible to differentiate between various monies and payment techniques; a distinction also made by Fase (1996). Monies are the various liquid assets that constitute a share of consumer wealth. Monies are stock variables. The first type of money is currency which is circulated by the monetary authorities. The second type of money, demand deposits, relates to bank accounts. As far as electronic money is also related to a bank account, there is much to be said for regarding electronic money as a special form of demand deposits (Fase and Vleminckx, 1995). But electronic money can in principle also be issued by non-banks. Furthermore, chip cards may be directly charged using currency at specialised teller machines, a transaction which does not involve any bank deposit. For these reasons we distinguish between demand deposits and electronic money.

There are various techniques for using the three different types of money to make payments. Payments made with the aid of banknotes and coins are obviously cash payments. Here, the type of money and the payment technique coincide. That cannot be said of demand deposits and electronic money. Table 1 gives an overview of the payment modes mentioned above, the monies and the related payment techniques. The intensity with which consumers use these techniques can be shown by flow data, as in Table 2, which also presents total consumption.

[1] This breakdown is based in part on European Monetary Institute (1996), pp. 399-402.

TABLE 1 Payment methods, geldvormen and payment techniques used in OTC transactions

| Payment mode | Money | Payment technique |
|---|---|---|
| Cash | Banknotes and coins | Banknotes and coins |
| Non-cash | Transfers | Cheques |
| | | Credit cards |
| | | PIN cards |
| Non-cash | Electronic money | Chip cards |
| | | Network payments |

The breakdown of monies and payment techniques in Table 1 forms the framework for this study, both to characterise existing models from the literature and to provide an incentive for further modelling. As we can see in the next paragraphs, the existing models address notably the choice between different monies. In particular the choice of the optimal money stocks. As the overview in Table 1 shows, there is also a choice to be made between payment techniques. An expansion of the existing inventory models so that the choice between payment techniques is explained would seem the most appropriate strategy towards more complete modelling of the choices made in the case of OTC payments.

TABLE 2 Turnover of OTC payments and consumption in the Netherlands Billions of guilders

| Year | Cash | Cheque | Credit card | debit card | Chip card | Total | Consumption |
|---|---|---|---|---|---|---|---|
| 1991 | 112 | 33 | 3 | 2 | - | 150 | 322 |
| 1992 | 115 | 32 | 4 | 4 | - | 155 | 341 |
| 1993 | 120 | 26 | 6 | 7 | - | 159 | 352 |
| 1994 | 122 | 21 | 7 | 14 | - | 164 | 369 |
| 1995 | 123 | 16 | 7 | 24 | 0.00 | 170 | 382 |
| 1996 | 123 | 13 | 8 | 35 | 0.01 | 179 | 398 |

Explanatory note: The turnover of cheques and debit cards has been observed directly (see for example Bank for International Settlements, 1996). The other data are based on Nederlandsche Bank estimations. The last column shows the total consumption of households.

# 3 MODELS OF PAYMENT BEHAVIOUR: AN OVERVIEW OF THE LITERATURE

## 3.1 Demand for money for transactions purposes: theory

A broad interpretation of payment behaviour also encompasses demand for money. In fact, payment behaviour and the demand for money for transactions purposes coincide when there is no choice between different types of money. Theoretical studies on this subject largely restrict themselves to the demand for money for transactions purposes. Here a transactions motive for holding money is considered exogenous. At the micro-economic level, this leads, in its most general form, to a relationship between the real money stock, the volume of transactions and the interest rate (McCallum and Goodfriend, 1994).

On the basis of Hicks (1939), Baumol (1952) and Tobin (1956) introduced a formal model of the demand for money, specifically for transactions purposes, in the Anglo-Saxon literature.[2] They use a mathematical approach from inventory theory, which explains why these models are called inventory models. They assume that money can be converted into an interest-bearing investment at fixed costs per conversion. In addition, there is no synchronisation of (exogenous) income and expenditure flows. The outcome of the model is optimum average cash holdings, which depend on income, on the interest rate on savings deposits, and on the fixed costs of visiting a bank (with elasticities of 0.5, -0.5 and 0.5, respectively).[3] This means that the model predicts economies of scale; higher income leads to a less than proportional increase in cash holdings.

The Baumol-Tobin inventory model forms the basis for most of the later models of the demand for money for transactions purposes. Likewise, the next paragraph and section 4 of this report address variations which correspond almost entirely in their set-up with the original Baumol-Tobin inventory model. Another, smaller group of models explaining the demand for money for transactions purposes shows more discernible differences vis-à-vis the original inventory model. The differences lie notably in the underlying assumptions about the role of money. For instance, the Baumol-Tobin inventory model assumes that all payments are made with currency. A similar assumption can be found in the cash-in-advance models designed by, for instance, Lucas (1980) and Svensson (1985). By contrast with the inventory models, the demand for money in these models is determined simultaneously with the demand for consumer goods, yet it is assumed that

---

[2] However, in Baumol and Tobin (1989), they acknowledge that the essential formulas from what has become known as the Baumol-Tobin model (see footnote 4 below) were derived from a French publication by Allais (Allais, 1947; pp. 238-241).

[3] The optimum number of visits to the bank during the period in which income $Y$ is received is $n = \sqrt{r\,Y/(2a)}$, with $a$ the fixed costs of visiting the bank and $r$ the interest rate on savings. The optimum cash holdings are then $1/2\,Y/n$.

(part of) the consumer goods are paid for exclusively in cash. Other models do not assume that money is indispensable to certain transactions, but that money reduces the costs of transactions. Sidrauski (1967), for instance, incorporates cash holdings in the utility function of consumers. This is also the case, albeit indirectly, in McCallum and Goodfriend (1994), where the use of money reduces the time needed for transactions and hence augments the quantity of leisure in the utility function.

### 3.2 Inventory models of payment behaviour

This paragraph presents an overview of recent extensions to the Baumol-Tobin inventory model. Baumol (1952) and Tobin (1956) were the first to explain the amount of cash held for transactions purposes from a micro-economic angle. They argue that a cost-minimising consumer will choose his cash holdings so that the marginal costs of larger holdings will, as a consequence of the interest forgone on savings, equal the marginal earnings, because fewer visits to the bank are needed to convert savings into money. Over the past few years, the Baumol-Tobin model has been developed further in various directions. Notably the theoretical contributions by Whitesell (1989 and 1992) and Santomero (1974, 1979 and 1996 in conjunction with Seater) are relevant to the explanation of payment behaviour. Other extensions to the model relate to the uncertainty about income and expenditure flows (Miller and Orr, 1966 and 1968; Whalen, 1966), a continuous time dimension (Romer, 1986) and the addition of the interest rate on demand deposits as explanatory variable (Barro and Santomero, 1972).

Apart from optimal cash holdings, Barro and Santomero's (1972) inventory model includes optimum holdings of deposits. Both turn out to depend on income and the difference between the attending interest rates (with elasticities of 0.5 and -0.5, respectively). In this model, the part of consumption that is paid for cash and that paid for non-cash are, however, determined exogenously. Whitesell (1989) goes further in his explanation of payment behaviour. Given the size distribution of payments, Whitesell's model explains which transactions are paid for in cash and which by cheque. It is assumed that, owing to the lag in debiting, payments by cheque are attended by interest earnings, but also by costs. It follows that small amounts are paid for in cash and larger by cheque. Consumers optimise their payment behaviour in terms of both the number of visits to the bank and the amount above which payments are made through transfers. The results are more complex than in the simple inventory model. There is, for instance, generally no symmetry in the elasticity of cash holdings with regard to income and the interest rate. The higher the costs of using cheques, the higher the maximum amount paid for in cash. This amount shows a negative correlation with the interest rate on demand deposits and the costs of visiting a bank.

A more recent study by Whitesell (1992) not only addresses transfers and banknotes and coins, but also breaks down payment techniques into cheques and credit cards. It explains for what transaction amounts consumers opt for one out of three payments techniques. Given the assumed cost differences between the various payment techniques, the lowest amounts are paid for in cash; up to a certain amount payments are subsequently made by cheque, while in excess of that amount use is made of credit cards. Like Whitesell (1989), this model assumes that the size and frequency of the transactions are exogenous for payment behaviour. In addition, the given distribution of transactions must meet stringent conditions. The choice of payment technique for a transaction is determined by the amount of the transaction, because every payment is attended not just by fixed costs, but also by costs which vary in accordance with the amount. Whitesell's (1992) model does not make any pronouncements on the optimal number of visits to banks; it is assumed that the initial holdings of demand deposits and of cash are sufficient for all transactions in the period observed. With the aid of the given distribution of transactions over the period, the average holdings of banknotes and coins and of demand deposits can then, in principle, be calculated.

In a series of articles, Santomero develops the most extensive model to explain payment behaviour. Santomero and Seater (1996) describe a model which explains simultaneously the size of cash holdings, the balance on the demand deposit and that on the savings account, as well as households' stocks of goods. The model also endogenously determines the number and size of transactions, as well as the payment techniques used for each transaction. In their payment behaviour, consumers balance, on the one hand, the interest earnings of the various monies and of the stock of goods, and, on the other, the costs arising from the conversion of savings deposits into monies and from the purchase of goods, depending on the various payment techniques.

Therefore, we may conclude that in the recent literature on payment behaviour there is a tendency towards models which include a larger number of monies and payment techniques. At the same time, the analytical emphasis is shifting in the models, from determining the optimal stock of money - in the original models - to the determination of the optimal cash management. The last two models mentioned (Whitesell, 1992; Santomero and Seater, 1997) will be analysed in greater detail in Section 4, notably with a view to their possible empirical application. Prior to that, the next paragraph gives an overview of empirical research into payment behaviour.

### 3.3 Empirical models explaining payment behaviour

At the Nederlandsche Bank, the empirical research of payment behaviour was originally aimed notably at explaining and predicting the circulation of banknotes and coins. Fase and Van

Nieuwkerk (1977) explain the circulation of banknotes per denomination for four countries. The interest rate elasticity turns out to be negligible, and the income elasticity shows a correlation with the denomination. Fase (1981a and 1981b) made use of various techniques to predict the circulation per denomination of Dutch coins and banknotes.

Another aspect of payment behaviour, i.e. the relationship between the amount paid and the degree to which the payment is made in cash, by cheque or by transfer, is addressed in a descriptive study by Fase and Boeschoten (1985), based on a survey of the total payments of consumers. They also study to what extent the existing range of banknote denominations is efficient. On the basis of the same survey, Boeschoten and Fase (1989) make a first attempt at an empirical study of the explanation of payment behaviour. According to them, the share of cash payments shows a negative correlation with the amount of the transaction. Other relevant factors identified by them are age, income and the number of giro accounts, which all have a negative effect on the share of cash payments.

The empirical studies by Mot and Cramer (1992), Boeschoten (1995 and 1998) and Koning (1997) all seek to explain payment behaviour. In these studies, the choice of techniques for OTC payments is described with the aid of a multinomial logit model. In the models designed by Mot and Cramer and by Koning, consumers can choose between cash payments and payments with various types of cheques. They determine the probability that a consumer uses a payment technique on the basis of the characteristics of that technique (risk and inconvenience) and of the consumer (e.g. age, sex, and income). Of the above studies using logit models, only Boeschoten's also look at POS terminals as a payment technique. They explain the shares in the bulk of OTC payments of cash, cheques, credit cards and POS terminals. The explanatory variables for the choice of payment technique studied by Boeschoten also include, next to the amount involved in the transaction and personal characteristics, the number of payment accounts and an indicator for the use of cash dispensers. The transaction amount turns out to be the main explanatory factor in all the studies mentioned which contain a logit model.

Like the theoretical models described in the preceding paragraphs, the multinomial logit model also assumes utility-maximising behaviour on the part of the consumer, although it differs substantially from the inventory models. First of all, the underlying decision process as such is not modelled in these empirical studies; instead a linear indirect utility function is postulated. Secondly, this model does not explain the size of the cash holdings for the various monies, but only the choice of the payment technique. This approach therefore has no direct implications for the demand for money. Finally, the model assumes that the choice between payment techniques is a choice between clearly differing alternatives. The model assumes that a high probability that

the consumer pays with a certain type of cheques has no implications for the probability that the consumer uses one out of the other two types. However, if the consumer considers these techniques close substitutes, this is an erroneous assumption. This property of the model implies that the probability of a new payment technique being used, let us say the chip card, cannot be predicted as consumers regard that technique as virtually identical to existing alternatives.[4]

We see from the above that by itself the logit analysis is incapable of making a theoretical contribution to explaining payment behaviour. The analysis is, however, interesting because of the empirical results obtained. These are, however, limited in nature because it is assumed that the personal characteristics of the consumer apart, the choice of payment technique is influenced only by the transaction amount. Mot and Cramer (1992) admittedly introduce the degree of risk and inconvenience of a payment technique as an explanatory variable, but they postulate that risk and inconvenience are determined entirely by the level of the amount. Thus they present a hypothetical explanation for the empirical relationship between the transaction amount and the choice of payment technique found by them and others.

Finally, it is worth looking at international comparative empirical research. A study by Boeschoten (1992) shows that in the 1970s and 1980s the use of credit cards and debit cards had a negative, but minor influence on the use of cash. Of the fifteen industrialised countries studied, it turned out that the greatest influence exercised by debit cards on the circulation of banknotes and coin was recorded in the Netherlands. The impact of credit cards, on the other hand, was relatively small in the Netherlands. The analysis supports the Baumol-Tobin inventory model, with economies of scale and a negative impact of the interest rate. Another international study (Humphrey, Pulley and Vesala, 1996) explains the use of five payment techniques in fourteen countries between 1987 and 1993. The explanatory variables include the costs of using the various payment techniques. These turn out to have a significant negative effect on the use of transfers and credit cards. No price effect is found in the case of cheques and debit cards. The considerable differences among countries found are explained mainly by cultural and institutional factors.

[4] The property of the multinomial logit model referred to is known as the independence of irrelevant alternatives.

# 4 MORE DETAILED ANALYSIS OF INVENTORY MODELS

For the benefit of a possible empirical application, this section deals extensively with two theoretical models which were discussed in the previous section. The most recent model, Santomero and Seater (1996) is interesting because of its completeness, though its complexity could thwart empirical application. Alternatively, Whitesell's (1992) model will be looked at, which is able to explain the choice between payment techniques in spite of a simpler set-up.

## 4.1 The model of Santomero and Seater (1996)

Santomero and Seater (1996) have introduced the - so far most general - model to explain payment behaviour. The first part of this paragraph describes this model and the results of the comparative static analysis. This part employs a more technical approach than the literature survey in the previous section. The second part of this paragraph contains a critical discussion and deals with the problems which might arise in the empirical application of this model and possible solutions to these problems.

### Description of the model

The model covers a single period. At the beginning of this period, the consumer receives an income which he spends on goods that he uses during the same period. The spending decision can be separated from the payment behaviour, so that expenditures per good $g$, $X_g$ are exogenous for the determination of payment behaviour. For each payment of a good (a transaction) the consumer chooses one of the total number $L$ of different monies. If $X_{gi}$ is the expenditure per good $g$ paid for with money $i$, then:

$$X_g = \sum_{i=1}^{L} X_{gi}, \quad g = 1,\ldots,G \tag{1}$$

The consumer can hold his income not just as money, but also in the form of savings. Holding money $i$ and having the disposal of a savings balance are attended by *fixed costs* $F_i$ and $F_S$. These costs are, for instance, the costs of a credit card or the administrative expenses charged by a bank for a demand deposit. As transactions can be made only with money, part of the savings must be converted into money at regular intervals. The *costs of conversion* are independent of the amount and vary per type of money. These costs ($\alpha_i$) consist of administrative costs and the money value of the time needed to make the conversion. By contrast with other authors, Santomero and Seater assume that every payment is attended by fixed costs as well, which vary

for every good and depend on the type of money used ( $\beta_{gi}$ ). Here, too, administrative costs are involved and the opportunity costs of the time used to effect the transaction. One implication of this cost structure is that expenditure and actual consumption need not coincide over time and that the consumer will hold a certain stock of every commodity. After all, the consumer will attempt to minimise the number of transactions.

The costs are offset by real interest earnings. Interest is paid not just on savings deposits ( $r_S$ ), but also on the holding of different monies, and there is a return on every good, $r_{Mi}$ and $r_{Xg}$, respectively. The interest rates on monies are non-negative and lower than the interest rate on savings. Such interest payments arise from, for instance, the fact that the transaction and its settlement do not coincide, as in the case of payments by cheque or credit card. The interest rates on commodity stocks may be positive if inflation arises, but will be mostly negative owing to decay, storage costs and depreciation.

If $T_i$ equals the number of conversions of savings into money $i$, $Z_{gi}$ equals the number of transactions made with money $i$ for good $g$; if average savings, money holdings and goods stocks equal $\overline{S}$, $\overline{M}_i$ and $\overline{X}_g$, respectively, then the gains made through the payment behaviour equal

$$\pi = r_S \overline{S} + \sum_{i=1}^{L} r_{Mi} \overline{M}_i + \sum_{g=1}^{G} r_{Xg} \overline{X}_g - \sum_{i=1}^{L} T_i \alpha_i - \sum_{i=1}^{L} \sum_{g=1}^{G} Z_{gi} \beta_{gi} - F_S I(\overline{S}) - \sum_{i=1}^{L} F_i I(\overline{M}_i) \tag{2}$$

where *I(x)* is an indicator function with the value 1 if the argument is positive and nil otherwise. The consumer's decision problem is to maximise the profit from his payment behaviour via the choice of the number of conversions per type of money, the expenditures and the number of transactions per good and per type of money $T_i$ , $X_{gi}$ and $Z_{gi}$ , respectively. In addition, the consumer decides whether he uses a savings account for his cash management. The solution to this decision problem is not trivial. The analysis shows that the optimum is a corner solution, at which the consumer pays for every good with one type of money only, but using different monies to pay for different goods.

This property of the model has three major consequences. To begin with, it is impossible to determine an optimum using standard methods (first-order conditions); the optimal payment behaviour must be found by comparison of each of the $2 \times L^G$ corner solutions. The model's

complexity consequently increases rapidly when the number of goods or monies goes up; with six goods and five monies, for instance, more than 30,000 corner solutions must be evaluated and compared. Secondly, a small change in one of the exogenous variables, an interest rate, say, may shift the optimum to another corner, causing discontinuous changes in payment behaviour. This makes it difficult to perform the usual analysis of the optimum by calculating the effect of exogenous variables on optimal payment behaviour (comparative statics); only several general qualitative observations can be made, as we will see below. This also means that the elasticities of the demand for the various monies cannot be calculated. Thirdly, the model cannot be seen as a model for a representative consumer, explaining, for instance, the aggregated demand for various monies, because a discontinuous money demand function does not occur in practice.

The effect of exogenous variables on optimal payment behaviour

In spite of the above-mentioned problems involved in analysing the optimum, the model yields several interesting insights. If, for instance, the costs of converting savings into money 1 ($\alpha_1$) go up, then the average holdings of money 1 will increase and savings will shrink. This reaction is plausible because the consumer tries to save on conversion costs by increasing his cash holdings. However, such a reaction occurs only if the increase in $\alpha_1$ does not lead to a change in the payment pattern (i.e. the decision on which goods are paid for with which money). After all, the rise in conversion costs makes money 1 relatively unattractive, so that it is in principle possible that in the first optimum, money 1 is used for all transactions, whereas in the new optimum no transaction at all is settled with this money. If the transactions costs which arise when good $g$ is paid for with money $i$ ($\beta_{gi}$) increase, then the optimal holdings of that money will decline and the probability that the consumer pays for that good with another money will increase. Higher fixed costs for holding savings or cash do not affect average holdings if the payment pattern does not change. Obviously it becomes more plausible that the consumer chooses a payment pattern which allows him to avoid the higher costs. A rise in the interest paid on savings leads to an increase in savings deposits, and enhances the probability that the consumer will make use of that option. A rise in the interest paid on cash will induce the consumer to reduce his savings and to increase his holdings of the money in question. However, this finding again holds only if the payment pattern does not change.

Criticism and possible empirical applications

From a theoretical viewpoint, some criticism is in order where Santomero and Seater's model is concerned. To begin with, the model distinguishes only between various monies and not between monies, on the one hand, and payment techniques, on the other. This difference is most apparent with regard to demand deposits as a type of money, and debit card, cheque and credit card payments as three demand-deposit-related payment techniques. If these payment techniques

were included in the model, then every payment technique would have to be considered a distinct type of money, linked to its own stock, interest rates and conversion costs. It is possible to add an extension to the model which makes explicit allowance for the difference between monies and payment techniques. However, this modified model also has as a characteristic that optimal payment behaviour is a corner solution, i.e. that for the payment of every good only one payment technique is used. As a result, the three problems mentioned earlier also apply in full here.

Secondly, Santomero and Seater assume that the conversion and transactions costs and the interest earnings are independent of the amounts converted and paid. Generally speaking, there are, however, also variable costs and earnings. The greater the amount of cash obtained from a bank by the customer, the greater the expected loss due to theft etc. The higher the amount paid by the consumer by way of a credit card or a cheque, the greater his interest earnings because his account is debited with a lag. The introduction of variable expenses and earnings, however, would make the model analytically intractable.

The last remark concerns the concept of a good used by Santomero and Seater. They assume that every transaction causes fixed costs $\beta_{gi}$ which may vary with the type of money used. This is not plausible if good is taken in the Arrow-Debreu sense. On the one hand, not every good is paid for separately while, on the other, the administrative costs attending payments are not the same at very point of sale. It is more useful to assume that, in this model, a good indicates a package of goods which the consumer purchases at a certain type of shop.

When the model is applied empirically, for instance, in order to make a quantitative analysis of the effects of the introduction of a new type of money, one encounters two problems. The first is that, as noted earlier, the model describes the behaviour of a single consumer or a single household. For a quantitative analysis of the model's implications for macro-economic demand for monies, the demand of various homogeneous household groups would first have to be derived in isolation and then added up. When determining homogeneous household groups, allowance must be made for the fact that payment behaviour depends on spending patterns, i.e. on preferences and income. Allowance must also be made for the fact that payment behaviour depends, via conversion and transactions costs, $\alpha_i$ and $\beta_{gi}$, on the opportunity costs of time (i.e. the marginal net wage) and on the household's location vis-à-vis banks and shops.

The second problem is that, before the payment behaviour of an individual household can be determined, the various costs $F_S$, $F_i$, $\alpha_i$, $\beta_{gi}$, interest rates $r_S$, $r_{Mi}$, $r_{Xg}$ and expenditures per good $X_g$ must be known. However, usually conversion and transactions costs $\alpha_i$ and $\beta_{gi}$ cannot be observed directly. It is therefore required to estimate these costs with the aid of cross-section or panel data. Here the costs of the conversion of savings into monies $\alpha_i$

and the costs of the transactions $\beta_{gi}$ are seen as a function of the opportunity costs of time. Demographic household characteristics may also be included as explanatory variables. The parameters of these functions must then be estimated indirectly on the basis of the observable variation of payment behaviours between individual households in the random sample.

Apart from the problems outlined, it seems impossible to actually estimate the Santomero and Seater model. The reason is that the model is highly non-linear, owing to the discontinuous change in the payment pattern and the large number of possible corner solutions. One could, however, attempt to circumvent the estimation problem by calibrating the model, so that the simulated payment behaviour corresponds with the main characteristics of the payment behaviour which has been empirically observed. It goes without saying that in that case the simulation results can only give a rough indication of the consequences of, for instance, the introduction of a new type of money.

### 4.2 Whitesell's approach

Like Santomero and Seater, Whitesell (1992) assumes that a consumer minimises his transactions costs when making payments. In both models, transactions costs are costs in terms of the time and money used for the transaction itself, and the interest earnings forgone on money holdings. However, Whitesell assumes that the size distribution of the transactions is given, whereas Santomero and Seater regard spending per good as a given. They determine the frequency distribution endogenously, because consumers minimise the costs of their portfolio, which encompasses not only several types of money, but consumer goods as well. Because Whitesell considers the size distribution of transactions exogenous, his model is analytically simpler. Consequently, it is possible to include other aspects of payment behaviour in greater detail, such as costs varying with the amount of the transaction. This paragraph first describes Whitesell's (1992) model, which has been extended on two essential points. Then the model outcomes and possible applications will be discussed.

#### Description of the extended model

As the model of Santomero and Seater, this model covers a single period. During this period, the consumer makes payments whose frequency distribution is a given from his point of view. The frequency distribution of the transactions $F(b)$ indicates how often the consumer will effect transactions with a value of $b$ during the period. It is furthermore assumed that the transactions are spread evenly over the period. The consumer can make the payment in one of three ways: he can pay in cash, by cheque or debit card or by credit card. Each of these techniques is attended by costs. The fixed costs consist of time to process and record the transaction, but exclude the time needed to fill out a cheque. For a credit card payment, these costs have been normalised at

one. The fixed costs ( $f$ ) for a transfer are lower than those for a credit card transaction because the former usually takes less time. For cash payments, the fixed costs ( $-a$ ) are negative, because they take very little time to process and record. Apart from fixed costs, the use of these payment techniques also entails costs for holding cash and demand deposits for cheque and debit card payments. After all, no interest is paid on average cash holdings, while only a low rate $r_D$ is paid on demand deposits. Only credit card payments do not entail a loss of interest earnings, because debiting takes place at the end of the period. Finally, the consumer takes into account the risk inherent in cash payments. The higher the amount paid in cash, the higher the expected losses due to theft etc. Table 3 summarises the costs per transaction.

TABLE 3 Transactions costs of payment techniques in Whitesell's extended model

| Payment technique | Costs of $n$ transactions with value $b$ |
|---|---|
| Cash payment | $-a\,n+\left(\frac{1}{2m}r_S+c\right)b\,n$ |
| Payment by cheque/debit card | $f\,n+\left(\frac{1}{2m}(r_S-r_D)\right)b\,n$ |
| Payment by credit card | $n$ |

Explanatory note: The first term shows the fixed transactions costs, the second the costs of holding cash and the variable transactions costs. $m$ is the number of conversions of savings deposits into cash or demand deposits.

Since cash payments are attended by the lowest fixed costs and credit card payments by the lowest variable costs consumers will pay small amounts in cash and large amounts by credit card. If $m$ is the number of times that the consumer goes to the bank to take out savings or to transfer them to his demand deposit account, and such a conversion is attended by costs $q$, the total transactions costs are

$$C = qm + \int_0^{cash}\left((c+\frac{r_s}{2m})b - a\right)F(b)\,db + \int_{cash}^{credit}(\frac{r_S - r_D}{2m}b + f)\,F(b)\,db + \int_{credit}^{\infty}F(b)\,db \qquad (3)$$

where *cash* and *credit* are the highest and lowest amount, respectively, paid for by the consumer in cash or by credit card, respectively.[5]

[5] A sufficient condition for the requirement that $credit \geq cash$, is $f \leq 1-(1+a)\frac{r_S - r_D}{2c+r_S}$. It is assumed from here on that this condition is met.

The consumer's decision problem consists of determining the cost-minimising limits above which he pays by transfer or by credit card, and the number of times that he converts savings. Given these decisions, the average cash holdings and the average demand deposit balance are also determined. They are, respectively:

$$\frac{1}{2m}\int_0^{cash} b\,F(b)\,db \tag{4}$$

$$\frac{1}{2m}\int_{cash}^{credit} b\,F(b)\,db \tag{5}$$

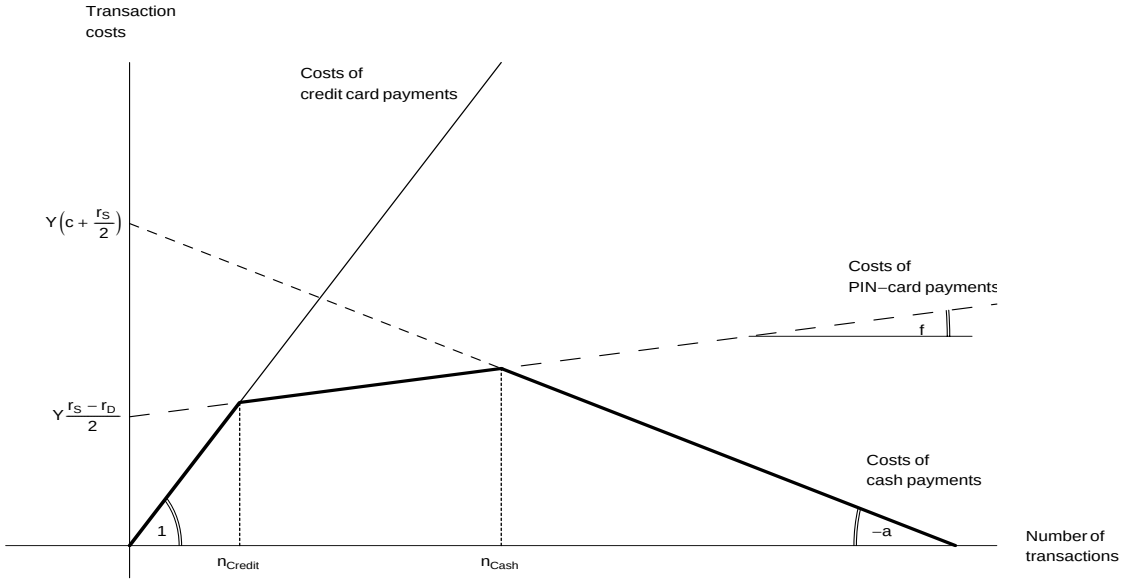
The model described above has been extended materially on Whitesell (1992) in two ways. Whitesell assumes first of all that consumers can take out savings or transfer them to their demand deposit account only at the beginning of the period, whereas all payments are effected at the end of the period. Secondly, Whitesell assumes that the frequency of payments is inversely proportional to the amount of the payment. Both assumptions are highly restrictive and are empirically rejected.[6] They were therefore abandoned in the extended model.

Optimal payment behaviour and reactions to changes in exogenous variables

In the original Whitesell model, the number of visits to the bank has been fixed to 1, and the frequency distribution of the transactions has the property that the total value of all transactions with size $b$ is constant, i.e. that $nb = Y$ for all $b$. The solution to the optimisation problem and the reactions to changes in exogenous variables can then be easily shown graphically. Figure 1 shows the transactions costs for the various payment techniques as a function of the frequency of the transactions. As $n = Y/b$ in Whitesell, there is a one-to-one relationship between the frequency and the value of the transactions while the transaction amount increases as the frequency declines. The figure shows the optimal payment techniques: high amounts (low frequencies, up to $n_{credit}$) are paid by credit card, average amounts (frequencies between $n_{credit}$ and $n_{cash}$) are paid by debit card and the lowest amounts (frequency in excess of $n_{cash}$) are paid in cash. The figure also shows that an increase in the interest rate on savings deposits shifts the cost functions of cash and debit card payments up, so that $n_{credit}$ rises and $n_{cash}$ remains unchanged. It follows that a rise in the interest rate on savings deposits makes the amount paid by credit card instead of debit card go down, while the highest amount paid in cash does not change. As Whitesell assumes the number of visits to the bank to be a given, it follows that the average demand deposit decreases and the average cash holding remains on the old level.

[6] Boeschoten and Fase (1989) consider the lognormal distribution an adequate description of the transaction distribution (see also Virén, (1994)).

FIGURE 1 Choice of payment technique in the original Whitesell model




When the interest rate paid on demand deposits goes up, the cost function for debit card payments shifts down, so that $n_{credit}$ declines and $n_{cash}$ rises. In other words, if the interest rate on demand deposits rises, consumers tend to switch sooner from cash to debit card payments, but later from debit card payments to credit card payments. At the same time, the average demand deposit balance goes up, decreasing the average cash holdings. If parameter $a$ is reduced, i.e. if the fixed costs of cash payments increase, then the straight line representing the costs of cash payments rotate upwards around its intersection with vertical axis. This means that higher fixed costs on cash payments imply a decline of the highest amount paid in cash and a rising average demand deposit balance. By analogous reasoning one can deduce the consequences of changes in the other parameters.

In the model extended by us, most of the properties of the original Whitesell model remain intact. In spite of the fact that we do not specify the size distribution of transactions $F(b)$, the amounts at which a switch to a different payment technique is optimal can be derived from the first-order conditions for a cost minimum of (3), depending on the optimal number of visits to the bank $m$:

$$cash(m) = 2m\frac{f + a}{2cm + r_D} \tag{6}$$

$$credit(m) = 2m\frac{1 - f}{r_S - r_D} \tag{7}$$

TABLE 4 Effects of changes in model parameters

| | Model parameter | | | | | | |
|---|---|---|---|---|---|---|---|
| | a | c | f | | q | $r_S$ | $r_D$ |
| Endogenous variable | | | (10) | $r_D=0$ | | (8) | (8,9) |
| Number of bank visits (m) | + | - | - | - | - | - | + |
| Upper limit for cash payments | + | - | ? | + | - | - | ? |
| Lower limit for credit card payments | + | - | - | - | - | - | + |
| Average cash holdings | + (9) | - (9) | ? | + | + | + | - |
| Average demand deposit balance | ? | ? | - (8,9) | - (8) | - (8,9) | - (8,9) | + |

Explanatory note: + (-) indicates an increase (decrease) in the endogenous variable when the model parameter in question goes up. The cases for which no unequivocal statement can be made have been marked ?. (8), (9), (8,9) and (10) indicate that the effect has been derived under the conditions concerned.

It follows that no payments are made by credit card if the interest rate on demand deposits equals that on savings deposits, and that the maximum amount paid in cash is independent of the number of visits to the bank if no interest is paid on demand deposits. It furthermore turns out that the more often that consumers convert savings into monies, the higher the amounts at which consumers switch from cash to debit card payments and from debit card to credit card payments.In order to solve the model entirely, the number of visits to the bank still needs to be specified. This is obviously not possible so long as the distribution of the transactions has not been specified in greater detail. It is, however, possible to deduce a number of comparative static results without knowing the distribution.[7] A decrease in the fixed costs of cash payments (increase in $a$), for example, leads independently of the distribution of transactions to an increase in the lowest amounts paid by debit or credit card. The same effect arises if the variable costs of cash payments decline. To derive most other results, only a number of qualitative assumptions relating to the distribution of transactions is required.

[7] Substitution of (6) and (7) in (3) yields the 'concentrated' cost function $C^*(m)$, which depends on $m$ only. If the implicit function theorem is applied to the first-order condition for the optimum number of visits to the bank, then it is possible to find what consequences variations of the parameters have for the optimum number of bank visits. For instance, $\frac{\partial m}{\partial a} = -\frac{\partial^2 C^* / \partial m \partial a}{\partial^2 C^* / \partial m^2}$. In this formula, it is known that $\partial^2 C^* / \partial m^2$ is positive because $m$ minimizes the costs. The reactions of the other endogenous variables, such as the optimum average cash holdings and the peripheral amounts can then easily be determined. For the peripheral amount of cash payments, it holds that

$$\frac{\partial cash}{\partial a} = \frac{\partial cash}{\partial a}\bigg|_{m=constant} + \frac{\partial cash}{\partial m}\frac{\partial m}{\partial a}.$$

For instance, the effect of changes in the interest paid on savings is not independent of the distribution of the transactions. It is possible to show that, when the interest rate $r_S$ is higher, the consumer pays lower amounts by debit or credit card, if it holds that

$$\left.\frac{\partial \ln \int_0^x b\, F(b)\, db}{\partial \ln x}\right|_{x=credit} \geq 1. \tag{8}$$

This means that if the lowest amount paid for by credit card goes up by 1%, the total value of all payments for which a credit card is not used should expand by over 1%. A comparable condition is sufficient to show that the average cash holdings increase when the fixed costs of cash payments decrease, viz.

$$\left.\frac{\partial \ln \int_0^x b\, F(b)\, db}{\partial \ln x}\right|_{x=cash} \leq \frac{2cm + r_D}{r_D}. \tag{9}$$

In other words, a 1% increase in the highest amount paid in cash makes the total value of cash payments go up by at most $1 + \frac{2cm}{r_D}$%. In order to be able to determine the effect of changes in the fixed costs of debit card payments, the condition

$$\frac{\partial \ln F(b)}{\partial \ln b} \leq -1 \tag{10}$$

suffices. This condition is the most restrictive and entails that the number of transactions falls by at least 1% when the transaction amount rises by 1%. Every distribution function which does not decline monotonously, such as the log-normal distribution, does not meet (10).[8] It may be noted that all of the conditions (8) to (10) are met in the original Whitesell model. Table 4 summarises the effects which arise as a result of variations in the model's parameters.

Further extensions and possible empirical application

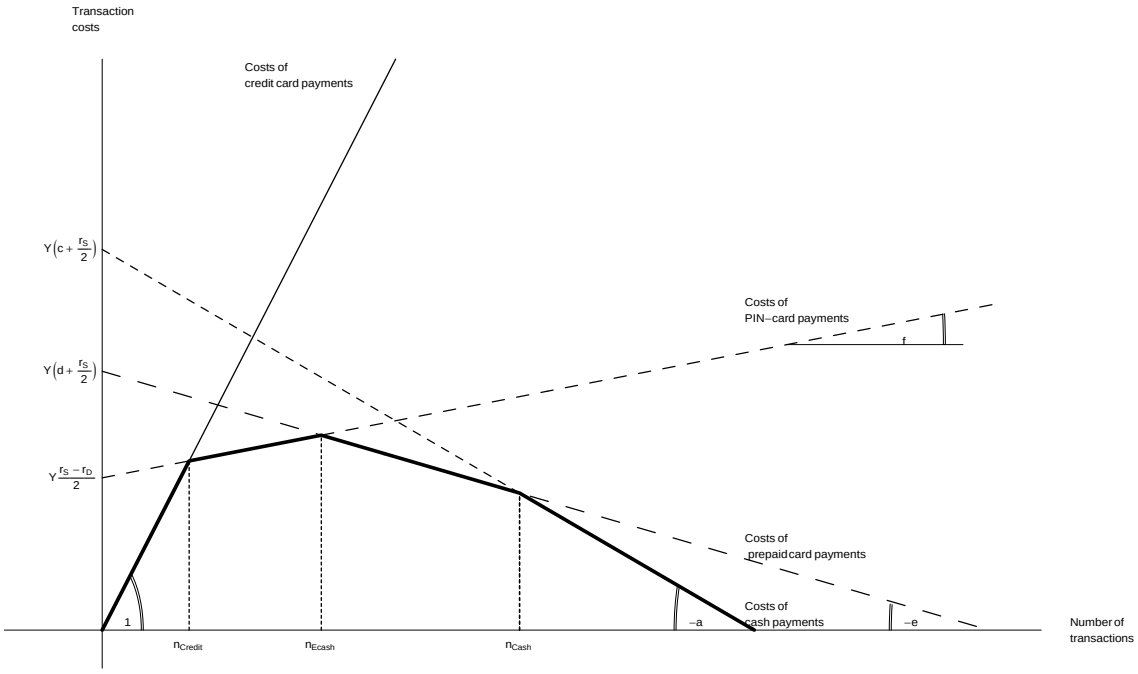
As can be seen from the above discussion, Whitesell's original model is the basic starting point for interesting theoretical models regarding payment behaviour. It may be noted that the model is not restricted to three payment techniques. Let us assume that electronic payments are attended by the same transactions costs as cash payments, but that the fixed costs are higher (for instance because the chip card has been secured by way of a PIN code) and the variable costs lower. In that case the payment behaviour in the original Whitesell model is shown in figure 2.

[8] An example of a distribution which meets (10), is the Pareto distribution. Condition (10) was chosen on account of its interpretation possibility, but is more restrictive than necessary. A necessary and sufficient condition is $r_D/(2\,c\,m + r_D)\, cash\, F(cash) \leq credit\, F(credit)$.

In the situation portrayed, introduction of the chip card would crowd out the highest cash payments and the lowest debit card payments. It goes without saying that more than three payment techniques can be analysed by the extended Whitesell model.

As in Santomero and Seater's model, empirical application of Whitesell's model means that the behaviour of a single consumer or household is explained. That is why the model parameters should be estimated with the aid of micro-economic data, where the parameters can also be regarded as functions of social and economic characteristics of households. With the aid of a model version which has thus been estimated, it is possible to simulate the effects of, let us say, interest rate changes or changing cost parameters on money holdings.

FIGURE 2 Payment behaviour after the introduction of chip cards



# 5 CONCLUSIONS

The objective of this report was to determine to what extent economic theory can contribute to the analysis of consumers' payment behaviour when making OTC payments. Although a large number of studies on this subject have appeared in the literature since the inventory-theoretical explanations by Baumol and Tobin, only two models prove capable of describing the choice between the various payment techniques as well. The most recent, that by Santomero and Seater, is also the most comprehensive. A major disadvantage of this model is that one can derive only a few properties of payment behaviour theoretically. The model's complexity is due to the fact that payment habits tend to react abruptly if the model parameters vary. Such discontinuity also makes it impossible to make an empirical estimation of the model, so that it cannot be easily applied.

In its original form, Whitesell's model, i.e. the second model studied, is based on very restrictive assumptions. That is why his model was considerably generalised in this report, so that it can be empirically applied. It was furthermore shown that the extended model version is also capable of explaining various aspects of payment habits. For instance, it is possible to determine analytically how the size of transactions balances and the use of payment techniques change when there are changes in, for instance, the interest rate on savings or when new payment techniques are introduced. It would also be possible to quantify these effects with an empirically estimated version of the model. The version of Whitesell's model presented here is therefore better able than the model designed by Santomero and Seater to make a useful contribution to the theoretical and empirical analysis of OTC payment behaviour.

REFERENCES

**Allais, M.**, 1947, *Economie et Intérêt*, Imprimerie Nationale, Paris.

**Bank for International Settlements**, 1996, Statistics on payment systems in the Group of Ten countries, Basel, December.

**Barro, R.J. and A.M. Santomero**, 1972, Household money holdings and the demand deposit rate, *Journal of Money, Credit, and Banking*, 4, 397-413.

**Baumol, W.J.,** 1952, The transactions demand for cash - An inventory theoretic approach, *The Quarterly Journal of Economics*, 66, 545-556.

**Baumol, W.J. and J. Tobin**, 1989, The optimal cash balance proposition: Maurice Allais priority, *Journal of Economic Literature*, 27, 1160-1162.

**Boeschoten, W.C.**, 1992, *Currency Use and Payment Patterns*, Kluwer, Dordrecht.

**Boeschoten, W.C.**, 1995, Kasmanagement en betaalgedrag van gezinnen in 1994, Onderzoeksrapport WO&E, 419/9501, De Nederlandsche Bank N.V.

**Boeschoten, W.C.**, 1996, Modern betalingsverkeer en het chartale geld, *Economisch Statistische Berichten*, 81, nr. 4046, 162-166.

**Boeschoten, W.C.**, 1998, Cash management, payment patterns and the demand for money, *De Economist*, 146, 117 - 142 .

**Boeschoten, W.C. and M.M.G. Fase**, 1989, The way we pay with money, *Journal of Business & Economic Statistics*, 7, 319-326.

**Boeschoten, W.C. and M.M.G. Fase**, 1992, The demand for large banknotes, *Journal of Money, Credit and Banking*, 24, 319-337.

**European Monetary Institute**, 1996, Payment systems in the European Union, Frankfurt am Main.

**Fase, M.M.G.**, 1981a, Forecasting the demand for banknotes: some empirical results for the Netherlands, *European Journal of Operational Research*, 6, 269-278.

**Fase, M.M.G.**, 1981b, Forecasting the need for coins: a case study for the Netherlands, *Technological Forecasting and Social Change*, 19, 147-159.

**Fase, M.M.G.**, 1985, The Fl.1000 bank note: means of payment or means of hoarding? *Quarterly Bulletin*, March, De Nederlandsche Bank N.V., 31-38.

**Fase, M.M.G.**, 1996, Vervagend geld en de dynamiek van het betalingsverkeer, in: *100 Jaar Markus Verbeek, Jubileumbundel*, Wolters-Noordhoff, Groningen, 40-55.

**Fase, M.M.G. and W.C. Boeschoten**, 1985, The distribution of payments and the use of bank notes, *Quarterly Bulletin*, December, De Nederlandsche Bank N.V., 43-53.

**Fase, M.M.G. and M. van Nieuwkerk**, 1977, The demand for bank notes in four countries, *Quarterly Statistics*, June, De Nederlandsche Bank N.V., 84-98.

**Fase, M.M.G. and A. Vleminckx**, 1995, *Geld in Veelvoud*, Lemma, Utrecht.

**Hicks, J.R.**, 1939, *Value and Capital*, Oxford, Oxford University Press.

**Humphrey, D.B., B. Pulley and J.M. Vesala**, 1996, Cash, paper, and electronic payments: a cross-country analysis, *Journal of Money, Credit, and Banking*, 28, 914-939.

**Koning, R.H.**, 1997, Utility maximization and mode of payment, Tinbergen Institute Discussion Paper TI 97-092/4, Tinbergen Institute, Amsterdam/Rotterdam.

Lucas, R.E., Jr. 1980, Equilibrium in a pure currency economy, in: *Models of Monetary Economics*, J.H. Kareken and N. Wallace (eds.), Federal Reserve Bank of Minneapolis, Minneapolis.

**McCallum, B.T. and M.S. Goodfriend**, 1994, Demand for money: Theoretical studies, in: P. Newman, M. Milgate, J. Eatwell (eds.), *The New Palgrave Dictionary of Money and Finance*, III, Macmillan, Londen, 611-617.

**Miller, M.H. and D. Orr**, 1966, A model of the demand for money by firms, *Quarterly Journal of Economics*, 80, August, 413-435.

**Miller, M.H. and D. Orr**, 1968, The demand for money by firms: extensions of analytic results, *Journal of Finance*, 23, 735-759.

**Mot, E.S. and J.S. Cramer**, 1992, Mode of payment in household expenditures, *De Economist*, 140, 488-500.

**Romer, D.**, 1986, A simple general equilibrium version of the Baumol-Tobin model, *Quarterly Journal of Economics*, 101, 663-686.

**Santomero, A.M.**, 1974, A model of the demand for money by households, *The Journal of Finance*, 29, 89-102.

**Santomero, A.M.**, 1979, The role of transaction costs and rates of return on the demand deposit decision, *Journal of Monetary Economics*, 5, 343-364.

**Santomero, A.M. and J.J. Seater**, 1996, Alternative monies and the demand for media of exchange, *Journal of Money, Credit, and Banking*, 28, 942-960.

**Sidrauski, M.**, 1967, Rational choice and patterns of growth in a monetary economy, *American Economic Review*, 57, Papers and Proceedings, 534-544.

**Svensson, L.E.O.**, 1985, Money and asset prices in a cash-in-advance economy, *Journal of Political Economy*, 93, 919-944.

**Tobin, J.**, 1956, The interest elasticity of transactions demand for cash, *Review of Economics and Statistics*, 38, 241-247.

**Virén, M.**, 1994, Demand for different payment media in Finland, *Bank of Finland Bulletin*, 68(2), 12-16.

**Whalen, E.L.**, 1966, A rationalization of the precautionary demand for cash, *Quarterly Journal of Economics*, 80, 314-324.

**Whitesell, W.C.**, 1989, The demand for currency versus debitable accounts, *Journal of Money, Credit, and Banking*, 21, 246-251.

**Whitesell, W.C.**, 1992, Deposit banks and the market for payment media, *Journal of Money, Credit, and Banking*, 24, 483-498.