# The Transactions Theory of the Demand for Money: A Reconsideration

Robert W. Clower
*University of California, Los Angeles*

Peter W. Howitt
*University of Western Ontario*

This paper deals with a class of models of the demand for money that includes the Baumol-Tobin and other inventory-theoretic models as special cases. Among other things, the analysis shows that many supposedly robust comparative-statics propositions derived by earlier writers do not survive even modest generalization. More generally, the results of the paper strongly reinforce other recent research in indicating the need for a wholesale reconstruction of the microfoundations of contemporary monetary theory.

Most of what passes for contemporary wisdom about the transactions demand for money derives directly from the classic but transparently rudimentary inventory-theoretic model of Baumol (1952) and Tobin (1956). Our purpose in this paper is to outline and explore the implications of a more general class of models in which individual economic agents freely choose not only the frequency of both sales and purchases but also the time phasing of transactions.[1] Our first and most significant findings

We are indebted to Joel S. Fried, R. A. Jones, J. M. Ostroy, Mack Ott, and John Riley for suggestions and critical comments on earlier versions of the paper. We have also benefited from reading the results of investigations on a closely related problem as reported in a privately circulated memorandum from D. W. Bushaw (see n. 2 below).

[1] The Baumol-Tobin model has been extended in various directions; see especially Clower (1970), Johnson (1970), Feige and Parkin (1971), Perlman (1971), Fried (1973), Grossman and Policano (1975), Barro and Santomero (1976), and Policano (1977). In every instance, as also in the original Baumol-Tobin model, formal analysis has been confined to special cases in which individual traders choose a trading frequency for just one commodity, and in which the time-phasing problem is ignored. Some of these papers (e.g., Feige and Parkin 1971) deal nominally with trading frequencies for more than one commodity, but in each such case trading frequencies of all goods are assumed to satisfy integer constraints so that each frequency is an exact multiple or divisor of every other.

[*Journal of Political Economy*, 1978, vol. 86, no. 3]

are that demands for trade inventories, that is, trade-related holdings of both goods and money, are in general discontinuous functions of the relative frequency of sales and purchases, and that commonly accepted comparative-statics propositions about the demand for money are of dubious validity except in cases that are too specialized to be of serious practical interest. Our analysis also sheds fresh and occasionally unexpected light on a variety of other issues: for example, the conditions in which individual traders will choose to hold money in positive amounts, the nature of potential gains from paying competitive interest on money, the economics of hyperinflation, the effects upon individual demands for money of the availability of trade credit and bonds, and the inadequacy of received price theory for analyzing even the most elementary aspects of interpersonal exchange in economies characterized by individual diversities in the frequency and time phasing of sale and purchase transactions.

## I. The Basic Model

Our first task is to characterize the stationary-equilibrium behavior of a representative trader in an idealized exchange economy where goods can be traded in organized markets only in exchange for units of a pure stock commodity called "money."[2] In keeping with familiar doctrine, we suppose that on each purchase and sale the trader incurs a set-up cost that is independent of the quantity traded. As a consequence, the trader will choose to execute trades only at discrete points rather than continuously in time and so may be expected to hold positive stocks of goods or money (or both) between successive sale and purchase dates. The holding of trade inventories is presumed to impose other costs on the trader, reflecting foregone consumption opportunities and the use of resources to maintain storage facilities. Thus to maximize long-run real income the trader must strike a balance between frequent transactions with low holding but high trading costs and infrequent transactions with low trading but high holding costs. We now proceed to formalize these general ideas.

### A. Fundamental Assumptions

We begin by considering an individual trader who produces and sells units of just one stock-flow good ($S$) and who purchases and consumes units of just one other stock-flow good ($D$). To avoid superfluous notation, we assume that the money prices of both goods are equal to unity, and we use generic symbols $S$, $D$, and $M$, respectively, to denote measurable quantities

[2] A nonstationary version of the problem analyzed here was studied by D. W. Bushaw and five other mathematicians during a summer institute in applied mathematics at Washington State University in 1972. The difficulty of the problem is reflected in the paucity of unambiguous results obtained by this group (which included some leading specialists in dynamical polysystems). One of us has made some progress on a simpler problem involving only two goods (see Howitt 1977).

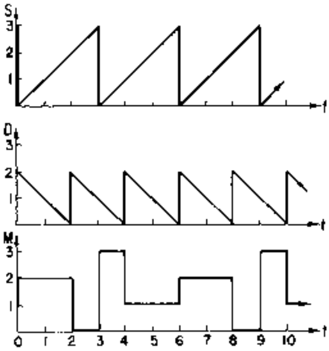


FIG. 1.—Time paths of inventories in the special case: $S = 3$, $D = 2$, $y = 1$, $m = 0$, $M_0 = 1$.

of $(S)$, $(D)$, and the money commodity $(M)$. For similar reasons (also to avoid dealing with feedbacks of trading and holding costs upon the trader's stationary-state level of consumption) we treat all costs as subjective (foregone leisure or consumption) or as charges that are incurred "outside" the model. We further suppose that the trader's rate of production of $(S)$ is predetermined at a constant level of $y$ units per unit of time, and that the trader holds no precautionary stocks of goods or money—that is to say, a purchase is made only when the trader's stock of $(D)$ has just been exhausted, and a sale is made only when the trader wishes to dispose of the whole of his accumulated stock of $(S)$. Finally, as conditions for stationary equilibrium we require that the trader consume $(D)$ at the constant rate $y$, sell $(S)$ in lots of constant size $S$ at uniform time intervals of length $S/y$, and purchase $(D)$ in lots of constant size $D$ at uniform time intervals of length $D/y$.

We suppose initially that at date zero a purchase and sale occur simultaneously. On this and earlier assumptions, the time paths of the trader's holdings of $(S)$ and $(D)$ are given by the familiar sawtooth patterns of figure 1. Average holdings of $(S)$ and $(D)$ are $\bar{S} \equiv S/2$ and $\bar{D} \equiv D/2$. Since the trader receives $S$ units of money on each sale date and gives up $D$ units of money on each purchase date, the time path of money holdings is represented by the irregular step function of figure 1. This path begins just prior to date zero, with an amount of money $M_0$ that is just sufficient to prevent the path ever from becoming negative.

For our purposes, the most important feature of the path of money holdings is its average value, $\bar{M}$. This value obviously is defined unambiguously by the choice variables $S$ and $D$, a fact that we may express formally by writing $\bar{M} = F(S, D)$. Indeed, it can be shown that the function $F$—hereafter referred to as the finance function—may be written explicitly as

$$\bar{M} = F(S, D) = \bar{S} + \bar{D} - G(S, D), \tag{1}$$

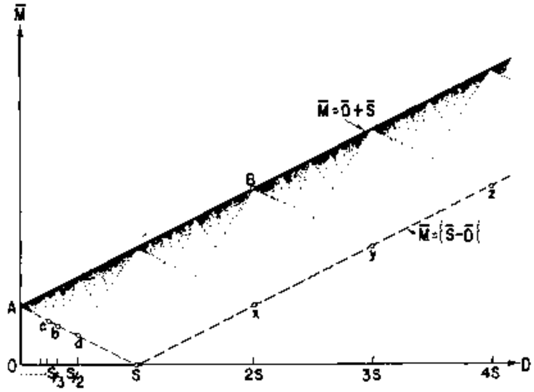


FIG. 2.—The finance function

where $G(S, D)$—hereafter referred to as the divisor function—is equal to the greatest common divisor of $S$ and $D$ if $S/D$ is rational and otherwise is equal to zero.[3]

The graph of the finance function (1) corresponding to any given (constant) value of $S$ is illustrated in figure 2.[4] Its most notable characteristic is that for any given value of $S$ the finance function contains a jump discontinuity at every value of $D$ for which $S/D$ is rational. These discontinuities have a straightforward economic interpretation. The trader can economize on money balances most effectively by so coordinating purchases and sales that sales often or always occur simultaneously with purchases, in which case $S$ and $D$ will have a common divisor that differs significantly from zero. If, instead, most or all sales are poorly coordinated with purchases, most purchases will have to be financed with money carried over from previous sales, and $S$ and $D$ will have a common divisor that is close to or equal to zero. For example, suppose that $S = D$; then $G(S, D) = S$, and the trader can avoid entirely the holding of money. But if $D$ is reduced by a very small amount, then $G(S, D)$ will fall very close to zero and the trader's money balances will rise discontinuously because the first several purchases after date zero will occur just before the corresponding sales.

The lower boundary of the finance function (points $x, y, \ldots$, and $a, b, \ldots$)

[3] For a discussion of some elementary properties of the divisor function, see Niven and Zuckerman (1972, pp. 4–6). The interested reader will find it easy to verify with specific numeric examples that the required value of $M_0$ equals $D - G(S,D)$. Thus the total value of all inventories just before date zero is $S + M_0 = S + D - G(S,D)$. But, this total value is constant over time because of our stationarity assumption; so it must equal $\bar{M} + \bar{S} + \bar{D}$. Formula (1) follows immediately. General proofs of these formulas can be found in Clower and Howitt (1976). These proofs necessarily involve the use of number theory—a branch of mathematics unfamiliar to most economists.

[4] Since $G(S,D)$ is homogeneous of degree one in $S$ and $D$, the absolute magnitude of $S$ is of no significance.

is attained when $D$ is an exact multiple or divisor of $S$, in which case $\bar{M} = |\bar{S} - \bar{D}|$. The upper boundary (the extended line $AB$) is attained when $S/D$ is irrational, in which case $\bar{M} = \bar{S} + \bar{D}$. All other cases (i.e., all cases in which $S/D$ is rational but neither $S/D$ nor $D/S$ is an integer) yield points that lie between these two extremes.

So far we have assumed that the trader has no reason to avoid making simultaneous purchase and sale transactions. In some situations this assumption may be acceptable, for example, in cases where sales and purchases take place at a single location. But in many circumstances it is plausible to suppose that traders purposely separate sale and purchase transactions in order to avoid bunching costs—that is, costs of rapid communication or travel between different locations, costs associated with rapid clearance of payments, and so forth. Formally we can allow for such costs by considering the time phasing as well as the frequencies of purchases and sales. Specifically, suppose that a sale occurs at date zero but that the first purchase does not occur until date $m > 0$. Since the choice of an origin to our time scale is arbitrary, we may also suppose that $m$ equals the minimum distance between any contiguous sale and purchase. Obviously $m$ will be constrained by the choice of $S$ and $D$. More precisely, it can be shown that:

$$m \leqq G(S, D)/2y. \tag{2}$$

As before, the time paths of holdings of $(S)$ and $(D)$ are represented by the sawtooth patterns shown in figure 1, except that the time path for $(D)$ is everywhere displaced to the right by the amount $m$. The value of $\bar{M}$ in this case is given by

$$\bar{M} = F(S, D) + ym, \tag{3}$$

which of course reduces to (1) if at some date a sale and purchase occur simultaneously so that $m = 0$.

### *B. Solution of the Model*

Stated formally, the trader's decision problem is to choose $S$, $D$, and $M$ so as to minimize the sum of trade-related costs:

$$C = \rho(\bar{D} + \bar{S} + \bar{M}) + \beta\bar{S} + \alpha\bar{D} + \gamma\bar{M} + ay/D + by/S + f(S, D, m). \tag{4}$$

The first term represents waiting costs, where $\rho$ is the trader's rate of time discount and, by virtue of our stationarity assumption, $\bar{D} + \bar{S} + \bar{M}$ is the total value of inventory holdings at each point in time. The next three terms represent storage costs, where storage-cost coefficients $\alpha$, $\beta$, and $\gamma$ are expressed as rates per unit time per unit commodity. The next two terms represent set-up costs per unit time, where $a$ and $b$ are the set-up costs per transaction, and where $y/S$ and $y/D$ are the frequencies of sale and purchase. The last term represents bunching costs per unit time. The exact form of the bunching-cost function $f(S, D, m)$ is left unspecified, except for the obvious requirement that it be decreasing with respect to $m$. Analysis of specific examples makes it clear that the bunching-cost function will generally exhibit the same kind of jump discontinuities as the finance function (see Clower and Howitt 1976). We refer to the sum of waiting and storage costs as holding costs and to the sum of set-up and bunching costs as trading costs.

The specification of the cost function (4) is extremely simple and conventional in form, but it generalizes earlier work in four significant respects. First, we allow the trader to choose the frequency of sales as well as purchases. The usual procedure in the past has been to regard the frequency of sales (the income period) as predetermined.[5] Second, we impose no a priori restrictions on the relative trading frequency, $S/D$. The standard approach in earlier work with similar models has been to suppose that $S/D$ can take on only integer values (see Grossman and Policano 1975). Third, we allow the trader to choose the time phasing of purchases and sales, an issue that has been ignored by previous writers, all of whom assumed, implicitly or explicitly, that $m = 0$.[6] Fourth and finally, we permit the previously ignored phenomenon of bunching costs to influence the trader's decision.

None of these generalizations is radical; indeed, each is suggested naturally by previous work or by the internal logic of our model. Yet in combination these generalizations yield a model of the timing of transactions with implications that differ fundamentally from those of any model considered in the existing literature.

To simplify the analysis of the trader's decision problem, suppose initially that there are no bunching costs, that is, that $f(S, D, m) \equiv 0$. In this case it is clear from (3) and (4) that the trader will choose $m = 0$, because that choice will permit holding costs associated directly with money to be minimized. The trader's decision problem then reduces to

$$\underset{(S,D)}{\text{minimize}}(2\rho + \alpha + \gamma)D/2 + (2\rho + \beta + \gamma)S/2 - (\rho + \gamma)G(S, D) + ay/D + by/S. \tag{5}$$

Because the divisor function $G(S, D)$ is discontinuous, standard calculus techniques cannot be used directly to establish the existence or properties of solutions to (5). This problem can be effectively circumvented here, however, by working with graphical methods. Let the curves $ab$, $cd$, etc. in figure 3 represent iso-trading-cost ($ITC$) loci corresponding to alterna-

[5] See Baumol (1952), Feige and Parkin (1971), and Perlman (1971). Barro and Santomero (1976) deal with the determination of the income period.

[6] See any of the references cited in n. 1 above.

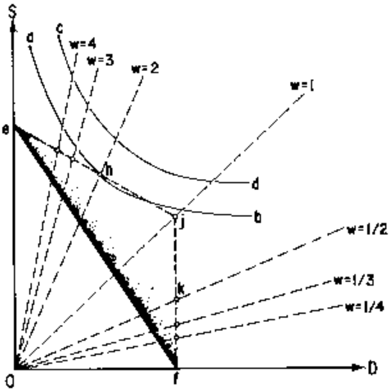


FIG. 3.—The first-stage decision problem

tive constant values of trading costs. These loci are continuous, downward sloping, and convex to the origin; and "higher" loci (larger values of $S$ and $D$) correspond to lower values of total trading cost. Let the points on or within the triangle *efj* represent an arbitrarily chosen iso-holding-cost (*IHC*) set corresponding to some given value of holding costs. The points at which $D$ is an exact multiple or divisor of $S$ lie on the upper boundary of the *IHC* set, indicated by the lines *ej* and *jf*, the slopes of which are, respectively, $(\gamma - \alpha)/(2\rho + \beta + \gamma)$ and $(2\rho + \alpha + \gamma)/(\gamma - \beta)$. The points at which $S/D$ is irrational lie on the lower boundary of the *IHC* set, indicated by the line *ef*, the slope of which is $-(2\rho + \alpha + \gamma)/(2\rho + \beta + \gamma)$. All other points lie strictly inside the triangle *efj*. Because the finance function is discontinuous, the *IHC* set consists entirely of isolated points except along its lower boundary (for a proof, see Clower and Howitt [1976]).

Using this diagram, we can illustrate the solution to (5) in two stages. A necessary condition for any solution is that trading costs be minimal for any given level of holding costs. Thus in searching for a solution we may restrict attention to points such as $h$ in figure 3 that lie on the highest possible *ITC* curve that intersects the given *IHC* set. If bunching costs are identically zero (as we are presently assuming), then trading costs and holding costs will both be homogeneous functions of the variables $S$ and $D$; hence the set of all points that minimize trading costs for given levels of holding costs will lie on a common ray through the origin. The slope of this ray, $w = S/D$, is therefore equal to the relative trading frequency of the solution to (5). Thus the first stage in solving (5) is to determine the optimal relative trading frequency, $\hat{w}$. Having determined $\hat{w}$ we may treat the equation $\hat{w} = S/D$ as a constraint and use it to rewrite (5) as:

$$\underset{\{D\}}{\text{minimize}}\ [(2\rho + \gamma)(1 + \hat{w}) + \alpha + \beta\hat{w} - 2(\rho + \gamma)G(\hat{w}, 1)]D/2 + (a + b/\hat{w})y/D. \tag{6}$$

The solution to (5) can then be derived as:

$$\hat{D} = \sqrt{\frac{2y(a + b/\hat{w})}{\alpha + \beta\hat{w} + (2\rho + \gamma)(1 + \hat{w}) - 2(\rho + \gamma)G(\hat{w}, 1)}}, \quad (7)$$

$$\hat{S} = \hat{w}\hat{D}, \quad (8)$$

and

$$\hat{M} = \hat{S}/2 + \hat{D}/2 - G(\hat{S}, \hat{D}). \quad (9)$$

Regrettably, these formulas do not permit us to calculate numerical solution values for $\hat{D}$, $\hat{S}$, and $\hat{M}$ corresponding to given values of the parameters $y$, $a$, $b$, $\alpha$, $\beta$, $\gamma$, and $\rho$. The difficulty lies in the first-stage problem—the determination of $\hat{w}$. Of course, in any particular problem we could attempt to find a numerical value for $\hat{w}$ by trial and error, but except in contrived special cases such a procedure is unlikely to yield anything but frustration.

A more promising approach is to establish limits on admissible solution values of $\hat{w}$ by imposing artful and (strictly speaking) invalid restrictions on the finance function (1). Suppose, for example, that we follow earlier literature and approximate the finance function by fitting straight lines through the lower boundary (i.e., the points satisfying integer constraints). This is equivalent to approximating the upper boundary of the *IHC* set in figure 3 by the straight lines *ej* and *jf*. Following the same two-stage procedure as before yields the solutions:

$$(\hat{S}, \hat{D}) = \begin{cases} [\sqrt{2by/(2\rho + \beta + \gamma)}, \sqrt{2ay/(\alpha - \gamma)}], \text{ if this makes } \hat{S} > \hat{D} > 0; \\ [\sqrt{2by/(\beta - \gamma)}, \sqrt{2ay/(2\rho + \alpha + \gamma)}], \text{ if this makes } \hat{D} > \hat{S} > 0; \\ [\sqrt{2(a + b)y/(2\rho + \alpha + \beta)}, \sqrt{2(a + b)y/(2\rho + \alpha + \beta)}], \end{cases} \quad (10)$$

otherwise;

$$\hat{M} = |(\hat{S}/2) - (\hat{D}/2)|. \quad (11)$$

As may be easily verified, the value of $\hat{w}$ (or $\hat{w}^{-1}$, if $\hat{w} < 1$) determined by these equations approximates its "true" value in the general solution to (5) to the nearest integer. In virtually every case this approximation will lead quickly (two trials) to an exact numerical solution, for because of the discontinuities it is clear from figure 3 that the "true" solution to (5) usually will correspond to a point on the upper boundary of some *IHC* set, and all such points define integer values of $w$ or $1/w$. But the rule suggested by this line of reasoning is not universal, because examples can be constructed in which neither $\hat{w}$ nor $1/\hat{w}$ is an integer.[7]

If we now relax the assumption that bunching costs are negligible, the decision problem becomes more complicated, but the same general principles apply. Generally speaking the *ITC* curves will consist of sets of isolated points, similar to the *IHC* sets, because of the discontinuities in the bunching-cost function. This produces an even greater presumption that the solution to the trader's problem will satisfy an integer constraint. However, the two-stage procedure described above no longer applies, for there is no reason to think that the bunching-cost function will possess the required homogeneity properties.[8] Another obvious implication of the assumption of positive bunching costs is that the trader will be more likely to choose positive average money holdings, for to avoid holding any money at all, the trader must not only choose $S = D$ but must also choose a corner solution, $m = 0$ (see Section II*B*).

## II. Implications and Extensions

In subsequent pages we discuss some implications and extensions of our basic model. Our analysis is in no sense exhaustive; its purpose is not so much to elucidate logical properties of our basic model (though we do this) as to illustrate the wide range of economic insights—critical as well as constructive—that may be gained by adding even modest elements of generality to earlier and more specialized inventory-theoretic models of the transactions demand for money.

### *A. Comparative Statics: General Observations*

It cannot be emphasized too strongly that the discontinuities in our basic model arise not from strained assumptions about the discreteness of time or the atomistic character of commodity and money units, but rather from the fact that trades involve stocks rather than flows so that small changes in the relative timing of transactions can produce large jumps in average finance requirements and in average bunching costs. These jumps would be less obvious if our model dealt with nonstationary processes so that between-trade time intervals were not necessarily uniform. As a matter of logic, however, jumps analogous to those implied by our model must occur

[7] Suppose that $y = 1$, $a = 5/11$, $b = 1$, $\rho = .01$, $\alpha = \beta = .2$, $\gamma = 0$. Pick the *IHC* set with holding cost equal to 1.0. Then the *ITC* curve with trading cost equal to (3.36)/11 touches two adjacent points on the upper boundary of the given *IHC* set for which $w = 1$ and $w = 2$; but the same *ITC* curve passes to the left of the point with $w = 3/2$ in the given *IHC* set, so $\hat{w}$ must lie strictly between 1 and 2.

[8] The example of Clower and Howitt (1976) produces a nonhomogeneous trading-cost function.

in any ongoing economy where trades take place at discrete points rather than continuously in time. Appearances to the contrary notwithstanding, therefore, the comparative-statics implications of our model are of more than purely academic interest.

The most important comparative-statics conclusion to be drawn from our model is negative, namely, the consequences of parameter changes upon equilibrium values of $S$, $D$, and $\bar{M}$ are generally ambiguous. The source of these ambiguities lies mainly in the different effects of parameter changes upon relative and absolute transactions frequencies. As is clear from earlier graphical analysis (fig. 3), small changes in parameters may leave the relative frequency $\hat{w} = \hat{S}/\hat{D}$ unchanged because one or both of the touching frontiers of the $IHC$ and $ITC$ sets consist of isolated points. In such cases the only effect will be to change the absolute frequencies $y/\hat{D}$ and $y/\hat{S}$, which are determined in the second-stage maximization problem. It is worth remarking that the standard assumption of regarding the absolute frequency of sales (the income period) as predetermined hides this sometimes crucial distinction by making a change in the absolute frequency of purchases equivalent to a change in the frequency of purchases relative to sales.

To illustrate the preceding remarks, let us consider a change in $\gamma$, the storage-cost coefficient on money holdings. If we employed the usual approximation giving rise to the square-root formulas (10) above, we would infer that when $S > D$ an increase in $\gamma$ would lead to an increase in $\hat{D}$. If the change in $\gamma$ did not affect the relative frequency $\hat{w}$, however, we would infer from the "correct" formula (7) that $\hat{D}$ would decrease. To get the result implied by the usual approximation, we should require the rise in $\gamma$ to produce (i) a large enough increase in $D$ (decline in $w$) in the first-stage decision problem to offset (ii) the decrease in $D$ at the second stage. The standard formulas (10) presuppose that the relative effect (i) always dominates the absolute effect (ii), but this cannot always be so because sometimes a change in $\gamma$ will produce no change in $\hat{w}$.

In simpler models involving integer constraints on $S/D$ or its reciprocal, individual demand functions will exhibit discontinuous steps, but the smoothing properties of aggregation may be invoked to argue that the aggregate demand function is smooth and behaves qualitatively, as indicated by the usual square-root formula (see Barro 1976). The present model shows that this is generally not valid. For example, when $\gamma$ increases, some traders will increase $\hat{D}$ and others will decrease $\hat{D}$; the aggregate effect will depend crucially upon the form of the distribution of traders between the two categories. The form of the distribution is not so crucial in simpler models because they do not allow for the possibility of traders moving in different directions, just for some not moving at all.

This is not to say that comparative-static results cannot be derived from the present approach. On the contrary, the approach allows us to isolate those results that are robust enough to survive some degree of generality. For example, Samuelson's well-known technique (1947, pp. 46–52) for manipulating inequalities associated with an extremum allows us to conclude that an increase in either of the trading-cost coefficients $a$ or $b$ will lead to an increase in average holding of the associated commodity $\bar{D}$ or $\bar{S}$; that an increase in any of the storage-cost coefficients, $\alpha$, $\beta$, or $\gamma$, will lead to a decrease in average holdings of $\bar{D}$, $\bar{S}$, or $\bar{M}$; and that an increase in the rate of time preference (interest rate), $\rho$, will lead to a decrease in the total money value of inventories, $\bar{D} + \bar{S} + \bar{M}$. Note, however, that when $\rho$ increases, any single inventory holding may increase rather than decrease.

One other familiar result that holds in the general case (provided that $m = 0$) is that the income-elasticity of demand for each of the inventories has a value of one-half. This follows from the modified square-root formulas (7)–(9) and from the fact that, because of the homogeneity of trading costs and holding costs in $y$, $S$, and $D$, a change in $y$ will not affect the value of $\hat{w}$ determined in the first-stage decision problem.

### *B. Positive Money Holdings*

Though a trader might be willing to hold positive money balances simply to avoid bunching costs, it appears that such balances would not be held for any other reason unless money were less costly to hold than the most frequently traded good. Otherwise, the trader would choose $\hat{S} = \hat{D}$ and would hold no money at all, for there would be no advantage to holding money rather than goods as a store of purchasing power or consumption. This may be seen most easily from our analysis by supposing that the solution point illustrated in figure 3 yields a value of $\hat{w}$ greater than unity. Then this point must lie to the left of the ray with $w = 1$. But if $\gamma \geq \alpha$, the line $ej$ will be horizontal or even upward sloping, which implies that total marketing costs could be reduced by moving the solution to point $j$, where $w = 1$.

These considerations have some bearing on the familiar question, Why do people choose to hold money when all other assets have a higher net return?[9] Our answer is that they will not—at least not in a stationary state without bunching costs. More generally, it appears that the usefulness of money as a means of payment is limited by its costliness to store, which may help to explain why representative monies have tended to displace commodity monies and commodity-backed monies in modern times.

The same considerations also shed light on the holding of money balances during hyperinflation. The coefficient $\gamma$ in our model may be interpreted as the sum of physical storage costs plus the expected rate of inflation. The model then implies that when inflation reaches some critical

[9] For historical accounts of this question, see Gilbert (1953) and Patinkin (1965).

point people will hold no money at all, except for brief intervals between transactions which become shorter as the expected rate of inflation increases. This prediction of our model accords well with behavior observed during actual hyperinflations. During even the most severe hyperinflations, however, people appear to be extremely reluctant to forgo trade in organized markets that require them to use conventional media of exchange (see Cagan 1956). That is to say, money continues to circulate with finite velocity even when it is the most costly of all goods to store. This observation, combined with our theoretical analysis, casts serious doubt on the conventional assumption that bunching costs may be ignored. The evidence suggests, on the contrary, that bunching costs are substantial at least for "nearly simultaneous" sale and purchase transactions.[10]

### C. Competitive Interest on Money

It has been shown by many authors that social optimality requires the payment of competitive interest on money (see Samuelson 1968; Friedman 1969; Feige and Parkin 1971). Optimality also requires that the money commodity be as inexpensive as possible to produce (see Gramm 1974). Generally speaking, it is taken for granted that the money commodity is almost costless to store relative to the cost of storing a typical nonmoney commodity of equal money value.

We can analyze the effect of paying interest on money by interpreting the coefficient $\gamma$ as the money storage cost coefficient minus the rate of return paid on money. Suppose that it is possible to find a money commodity that is literally costless to store. Then the optimal situation is to have $\gamma = -\rho$, in which case the trader's decision problem may be written as:

$$\underset{(D,S,m)}{\text{minimize}}\ (\alpha + \rho)(D/2) + (\beta + \rho)(S/2) + a(y/D) + b(y/S) + f(S, D, m) \tag{12}$$

subject to (2).

If bunching costs are nonzero, then (2) becomes binding, and

$$\hat{m} = G(\hat{S}, \hat{D})/2y. \tag{13}$$

In this case the formulas for $\hat{D}$ and $\hat{S}$ depend upon the exact nature of the bunching-cost function.

If bunching costs are identically zero, standard calculus techniques can be used to derive the optimal solution:

$$\hat{D} = \sqrt{2ay/(\alpha + \rho)},\ \hat{S} = \sqrt{2by/(\beta + \rho)}. \tag{14}$$

[10] The failure of previous models to take into account bunching costs may also explain their failure to account for the absolute magnitude of the typical household's money holdings, which these models all tend to underestimate. See Barro and Fischer (1976).

This solution is the one that would be obtained if the trader ignored the interaction of trading frequencies and minimized total holding and trading costs of each nonmoney good separately. Thus the gain from the payment of competitive interest on money (if $f[S, D, m] = 0$) derives from the freedom such a policy gives the trader to choose trading frequencies independently, without regard to cash constraints.

For reasons indicated earlier, it is hard to say what the qualitative effects of such interest payments might be on specific variables. In particular, our example in Section II*A* shows that a reduction in $\gamma$ will have an ambiguous effect on $\hat{D}$ even if we know a priori that $\hat{S} > \hat{D}$. However, two conclusions hold under quite general circumstances. The first is that a reduction in $\gamma$ will increase $\bar{M}$ (see Section II*A*). The second result is that if $f(S, D, m) = 0$, then total trading costs must fall. This follows from the fact that, as may be verified from (4) and (7)–(9), the optimal solution $(\hat{S}, \hat{D})$ always has the property that total trading costs equal holding costs. Since the sum of these costs must fall when $\gamma$ falls, both component elements of cost must also fall. Of course, it follows that either $\hat{S}$ or $\hat{D}$ must rise (otherwise trading costs could not fall); but in specific examples it is possible also for either $\hat{S}$ or $\hat{D}$ (but not both) to fall.

### *D. Bond Holdings*

The transactions theory traditionally has treated the demand for money as arising from the problem of finding the least costly combination of money and bond holdings with which to bridge time gaps between purchases and sales. The present paper shows how the same approach can be used to address the more fundamental problem of determining the size of the gaps to be bridged. The observation that few people actually engage in temporary bond transactions between monthly pay checks, rather than refuting the transactions approach, as has been frequently asserted, demonstrates the need to focus attention on this more fundamental problem. Nevertheless, since some agents (particularly large firms) are known to engage in frequent bond transactions, it is worth investigating the consequences of extending the present model to allow the trader to hold bonds as well as money as a temporary abode of purchasing power.

To this end, let us suppose that the trader is able to buy or sell bonds at a fixed price, plus or minus a brokerage fee on each transaction. To ensure the existence of a stationary solution, suppose also that the bond rate of interest, $i$, is no greater than the trader's rate of time discount, $\rho$. Then it can be shown that the trader will choose to buy bonds at every sale date of $(S)$ following which enough money would otherwise be held for so long that its opportunity cost in the form of bond interest would more than offset the double brokerage fee associated with buying and selling the bonds, and will sell bonds at every purchase date of $(D)$ at which money balances would otherwise run out. If bunching costs are incurred on bond transactions (and to suppose otherwise would be incongruous in a model where it is assumed that such costs are incurred for simultaneous goods transactions), there is an incentive to delay scheduled bond purchases and to advance scheduled bond sales up to the point where the marginal reduction in bunching costs on bonds is just offset by the marginal loss of interest income.

Let $\bar{B}$ denote the average holding of bonds. As in (3) above the average holding of financial assets is given by:

$$\bar{M} + \bar{B} = F(S, D) + ym. \tag{15}$$

In many circumstances, especially if bond-related bunching costs are significant, the trader will choose $\bar{B} = 0$, which is consistent with empirical evidence concerning the behavior of most households. Even if $\bar{B} > 0$, none of the findings reported earlier will be significantly affected by this fact. In particular, the trader's decision problem will be relatively insensitive to small changes in parameters because of discontinuities in the finance function and in the bunching-cost function. As in earlier discussions, therefore, we cannot expect to obtain unambiguous comparative-statics results on the basis of a priori considerations. A case in point is the effect of a change in the bond interest rate, $i$, on average holdings of money balances, $\bar{M}$. One would expect the sign of this partial derivative to be negative (see Grossman and Policano 1975, p. 1110), but that need not be so. If $i$ increases, this might have no effect on relative transactions frequencies. But by using Samuelson's technique of manipulating inequalities, it is easily seen that $\bar{B}$ must increase. If we assume that $m = 0$ before and after the change, then $\bar{D}$, $\bar{S}$, and $\bar{M}$ must all increase in the same proportion as $\bar{B}$.

If the introduction of bond holdings does not make earlier comparative-statics results less ambiguous, neither does it force us to revise earlier conclusions that were unambiguous. For example, it can be shown that the income elasticity of all inventory demands is equal to one-half (if bunching costs may be ignored). Moreover, a necessary condition for positive money holdings in the absence of bunching costs on goods is, again, that the storage-cost coefficient on money balances be less than the corresponding coefficient on the most frequently traded of the two nonfinancial commodities.[11] Finally, the payment of competitive interest on money that is costless to store would still eliminate the discontinuity in the *IHC* sets. In this situation no bonds would ever be held because to do so would be to incur needless trading costs; so the behavior of the trader would be precisely as described in the preceding section.

[11] One minor change in our results is that the sufficient conditions for positive money holdings are stronger than earlier. The trader may now choose $\bar{M} = 0$ even if $S \neq D$, provided that (i) bunching costs are all zero and (ii) a bond transaction accompanies every commodity transaction.

*E. Trade Credit*

The introduction of temporary bond holdings into our model does not significantly alter any of our earlier findings. More interesting results follow, however, if we extend our analysis to include a different kind of money substitute. Specifically, suppose that the trader has access to trade credit or bank overdraft facilities that permit the deferral of cash payment for goods through the nonnegotiable option at each purchase date of incurring a debt (to the seller of goods in one case, to a banker in the other) up to a predetermined limit, $L$. Suppose also that interest is charged on used credit at the rate $s$. Let $\bar{C}$ denote the average amount of unused credit and let $L - \bar{C}$ represent the average amount of used credit. As in (3) and (15) above, the sum of the trader's holdings of net financial assets is given by:

$$\bar{M} + L - \bar{C} = F(S, D) + ym. \tag{16}$$

As long as $s \leqq \rho$, it will be optimal for the trader to arrange for as much financing as possible to be done by credit. If $L$ satisfies

$$L \geqq D + S - 2G(S, D) + ym, \tag{17}$$

then holdings of money may be avoided altogether, for trade credit can be used to finance all of the trader's purchases without exceeding the credit limit (17), provided that all sales of ($S$) are accompanied by a running down of debt rather than an accumulation of money balances.

On these assumptions the trader's total holding cost will be given by

$$\rho(\bar{D} + \bar{S} + \bar{M} + \bar{C}) + \alpha\bar{D} + \beta\bar{S} + \gamma\bar{M} + s(L - \bar{C}). \tag{18}$$

The term $\rho\bar{C}$ is included in waiting costs because not to use possible credit lines involves the trader in the same abstention cost as for other commodities. In the case where (17) holds, of course, we have $\bar{M} = 0$. If, in addition, we have $s = \rho$, the holding-cost function (18) reduces to

$$(\rho + \alpha)\bar{D} + (\rho + \beta)\bar{S} + sL, \tag{19}$$

which is, except for the irrelevant constant term $sL$, identical to the function implied by our earlier discussion of the payment of competitive interest on money. In other words, competitive trade credit and bank overdraft arrangements provide two alternative routes to monetary optimality equivalent in effect to paying competitive interest on money. This result helps to rationalize—at least on an individual level—the presence in all advanced economies of a wide and (apparently) still expanding array of specialized credit facilities.

*F. The Coordination of Individual Trading Activities*

Though much of our analysis of individual trading behavior appears to have significant import also for market behavior, we shall limit our discussion of such matters here to a few general observations. To go beyond this would be ill advised, for the existing literature does not contain a satisfactory theoretical account of the overall working of an economy in which the resource costs of trading activity depend in an essential way on the frequency of exchange transactions. (For further elaboration, see Heller [1972].) In the absence of such a theory there is reason to believe that premature generalization of results derived from individual experiments will deal either superficially or not at all with what appears to be a fundamental externality (see Hahn 1973, pp. 229–34, 241; Perlman 1973; Russell 1974).

This externality arises from the fact that, in an economy with set-up costs of trading, individuals will trade only at isolated points in time; hence the set of transaction dates that are feasible for one trader cannot in general be specified independently of choices made by other traders. As Perlman (1971, p. 235) has put it, there exists in such an economy not only a problem of double coincidence of wants but also a problem of double coincidence of timing.

The timing-externality problem cannot be avoided by supposing (à la Debreu-Arrow) that all trades are the result of prearranged contractual obligations that specify exactly the dates at which trades are to be executed, for this procedure begs the question of how traders coordinate the timing of contract negotiations. The neo-Walrasian "auctioneer" is not an answer to but rather an evasion of this question. The only plausible and logically satisfactory solution is to posit the existence of specialist traders who, unlike the primary traders of the present paper, agree to do business continuously at dates chosen by the primary traders with whom they deal.[12]

Such specialist traders—shopkeepers, wholesalers, brokers, agents, marketing managers of manufacturing concerns—play a central role in every developed economy. Their usefulness arises from their willingness in normal circumstances to quote buying or selling prices (or both) at which they are ready to trade in large quantities at dates that suit their customers. By so doing they not only solve the double-coincidence-of-timing problem but also allow primary traders to plan and execute trades in accordance with budget constraints that do not require contingent allowance for possible nonprice rationing.[13] In practice, therefore, specialist traders act as visible fingers for Smith's invisible hand. Our argument suggests that a special class of market organizers and operators should be assumed to play a similar role in formal economic theory.

[12] This type of market arrangement is described in more detail by Howitt (1974), Clower (1975), and Clower and Leijonhufvud (1975).

[13] This logical difficulty in specifying the standard budget constraint is discussed by Clower (1965); the existence of specialist traders permits primary traders to make "notional" plans "effective" except in abnormal situations, where stocks of inventories held by specialists are temporarily exhausted or grossly in surplus.

**References**

Barro, Robert J. "Integral Constraints and Aggregation in an Inventory Model of Money Demand." *J. Finance* 31 (March 1976): 77–88.

Barro, Robert J., and Fischer, S. "Recent Developments in Monetary Theory." *J. Monetary Econ.* 2 (April 1976): 133–67.

Barro, Robert J., and Santomero, A. M. "Output and Employment in a Macro Model with Discrete Transaction Costs." *J. Monetary Econ.* 2 (July 1976): 297–310.

Baumol, William J. "The Transactions Demand for Cash: An Inventory Theoretic Approach." *Q.J.E.* 66 (November 1952): 545–56.

Cagan, P. "The Monetary Dynamics of Hyperinflation." In *Studies in the Quantity Theory of Money*, edited by M. Friedman. Chicago: Univ. Chicago Press, 1956.

Clower, Robert W. "The Keynesian Counterrevolution: A Theoretical Appraisal." In *The Theory of Interest Rates*, edited by Frank H. Hahn and Frank P. R. Brechling. New York: St. Martin's, 1965.

———. "Is There an Optimal Money Supply?" *J. Finance* 25, no. 2 (May 1970): 425–33.

———. "Reflections on the Keynesian Perplex." *Zeitschrift Nationalokonomie* 35 (July 1975): 1–24.

Clower, Robert W., and Howitt, Peter W. "Money, Credit, and the Timing of Transactions." Discussion Paper no. 72, Dept. Econ., Univ. California, Los Angeles, 1976.

Clower, Robert W., and Leijonhufvud, Axel. "The Coordination of Economic Activities: A Keynesian Perspective." *American Econ. Rev. Papers and Proc.* 65 (May 1975): 182–88.

Feige, Edgar L., and Parkin, Michael. "The Optimal Quantity of Money, Bonds, Commodity Inventories, and Capital." *A.E.R.* 61, no. 3 (June 1971): 335–49.

Fried, Joel S. "Money, Exchange and Growth." *Western Econ. J.* 11 (September 1973): 285–301.

Friedman, Milton. "The Optimum Quantity of Money." In *The Optimum Quantity of Money and Other Essays*, by M. Friedman. Chicago: Aldine, 1969.

Gilbert, John C. "The Demand for Money: The Development of an Economic Concept." *J.P.E.* 61, no. 2 (April 1953): 144–59.

Gramm, William P. "Laissez-Faire and the Optimum Quantity of Money." *Econ. Inquiry* 12 (March 1974): 125–32.

Grossman, Herschel I., and Policano, Andrew J. "Money Balances, Commodity Inventories, and Inflationary Expectations." *J.P.E.* 83, no. 6 (December 1975): 1093–1112.

Hahn, Frank H. "Foundations of Monetary Theory." In *Essays in Modern Economics*, edited by Michael Parkin. London: Longmans, 1973.

Heller, Walter P. "Transactions with Set-up Costs." *J. Econ. Theory* 4, no. 3 (1972): 465–78.

Howitt, Peter W. "Stability and the Quantity Theory." *J.P.E.* 82, no. 1 (January/February 1974): 133–51.

———. "Intertemporal Utility Maximization and the Timing of Transactions." *A.E.R.* 67, no. 2 (March 1977): 156–65.

Johnson, Harry G. "Is There an Optimal Money Supply?" *J. Finance* 25, no. 2 (May 1970): 435–42.

Niven, Ivan M., and Zuckerman, Herbert S. *An Introduction to the Theory of Numbers.* 3d ed. New York: Wiley, 1972.

Patinkin, Don. "An Indirect Utility Approach to the Theory of Money, Assets, and Savings." In *The Theory of Interest Rates*, edited by Frank H. Hahn and Frank P. R. Brechling. New York: St. Martin's, 1965.

Perlman, Morris. "The Roles of Money in an Economy and the Optimum Quantity of Money." *Economica*, n.s. 38 (August 1971): 233–52.

———. "The Roles of Money in an Economy and the Optimum Quantity of Money: Reply." *Economica*, n.s. 40 (November 1973): 432–41.

Policano, Andrew. "An Inventory-theoretic Model of Trade Credit Transactions." *Econ. Inquiry* 15 (April 1977): 166–82.

Russell, Thomas. "Feige and Parkin on the Optimal Quantity of Money." *A.E.R.* 64 (December 1974): 1074–76.

Samuelson, Paul A. *The Foundations of Economic Analysis*. Cambridge, Mass.: Harvard Univ. Press, 1947.

———. "What Classical and Neoclassical Monetary Theory Really Was." *Canadian J. Econ.* 1 (February 1968): 1–15.

Tobin, James. "The Interest Elasticity of Transactions Demand for Cash." *Rev. Econ. and Statis.* 38 (August 1956): 241–47.