Chapter 1

Consumer Theory and the Demand for Money

*William A. Barnett, Douglas Fisher, and Apostolos Serletis**

1.1 Introduction

The demand for money has been at the center of the macro-policy debate ever since Keynes's *General Theory* set down the initial version of what has become the standard macroeconomic version of the theory. Over the years it has become almost a dictum that a necessary condition for money to exert a predictable influence on the economy is a stable demand function for money, as often emphasized by Milton Friedman. While 'stability' hardly means 'simplicity,' it has also become believed, rather optimistically in our view, that this self-same demand function should be linear (or linear in the logs) and should have as arguments a *small* number of variables, themselves representing significant links to spending and economic activity in the other sectors of the economy. This complete argument appears in numerous places in the literature, from basic textbooks to research monographs (see, for example, the statement in John Judd and John Scadding (1982, p. 993)).

The theoretical literature on money demand does not contain the result that a linear function of a few key variables would be expected to serve as the demand for money. In particular, there exist a large number of potential alternatives to money, the prices of which might reasonably be expected to

*Originally published in the *Journal of Economic Literature* 30 (1992), 2086–2119. Reprinted with permission.

influence the decision to hold money. Furthermore, microeconomic theory rarely produces linear demand functions for rational economic agents. Even so, linear single-equation estimates of money demand with only a few variables continue to be produced, in spite of serious doubts in the literature about their predictive performance. Stephen Goldfeld (1976) brought wide attention to the poor predictive performance of the standard function. The result was a large literature that introduced new variables and/or transformations of the old; this largely inconclusive literature is surveyed in Judd and Scadding.[1]

There is another problem with this literature, and this is that the studies of the demand for money — and the many studies of the influence of money on the economy — are based on official monetary aggregates (currently M1, M2, M3, and L) constructed by a method (simple-sum aggregation over arbitrary financial components) that does not take advantage of the results either of existing aggregation theory or of recent developments in the application of demand theory to the study of financial institutions. Part of the problem is certainly perceived in the literature: the possible influences of financial innovation and regulatory changes. How this is usually handled is that money is frequently redefined — in the sense of composing new arrangements of the component assets — in order to capture the appearance of changing characteristics of the representative monetary product. This largely unstructured approach appears not to have produced an agreed-upon monetary measure. Instead, what we have seen over time is a considerable array of what actually turn out to be temporary monetary aggregates whose existence creates both unnecessary baggage in empirical studies as well as obvious problems for monetary-policy decision makers.

More central to the objectives of this survey are problems arising from the simple-sum method of aggregation itself. There are conditions under which this approach is appropriate, as we will explain, but if the relative prices of the monetary components fluctuate over time, then neither this method nor the Hicksian approach to aggregation will produce theoretically satisfactory definitions of money. The problem is the incorrect accounting for substitution effects that these methods entail, and the result is a set of monetary aggregates that do not accurately measure the actual quantities of the monetary services that optimizing economic agents select (in the aggregate). To underscore this issue we note that the empirical work discussed and illustrated below suggests that actual fluctuations in the relative prices of the monetary products of the U.S. financial system definitely

[1] Separately, Thomas Cooley and Stephen LeRoy (1981) question the econometric methodology common in the literature. A key concern is whether classical statistical properties can be attributed to estimators obtained by the "grid search" approach common to many of these studies.

are sufficient to generate concern about the method of aggregation.

Until recently the existing attempts to structure the search for a stable money demand — and a satisfactory measure of moneyness — using traditional macroeconomic paradigms (e.g., Keynesian, monetarist) do not seem to have provided any very firm assistance for empirical purposes. In contrast, as this survey will make clear, there is in place a steadily growing literature that does offer a solution; this is the integrated literature on monetary aggregation and the demand-systems approach to money demand. What we propose to do here is to lay out the theory and main empirical results of this important 'microfoundations' approach to money demand. Microfoundations approaches to macro topics usually imply either disaggregated general equilibrium modelling or the use of simple forms of aggregation over goods and over economic agents, but the literature we have in mind uses aggregation theory in a way that enables the researcher to test for the existence of both the postulated aggregate good and the aggregate economic agent while estimating the demand for (aggregate) financial services. Success here could well provide significant gains in both the study of monetary phenomena and in the application of monetary policy.

This new literature is actually an ongoing one that has only just begun to produce empirical results worthy of the effort required to understand it. The main research lies in two areas: the construction of *monetary aggregates* that conform to the specifications of demand theory and the estimation of systems of financial asset-demand equations in which the restrictions of demand theory are incorporated in such a manner as to assure consistency with the optimizing behavior of economic agents. Of course, there are useful paradigms in existence that employ a simultaneous-equations structure — notably the asset and transactions approaches to money demand — as typified by the mean-variance model of James Tobin (1958) or the transactions model as explained by Jürg Niehans (1978). But these approaches do not integrate the choice of monetary aggregate with the consumer choice problem and, depending on the version, often do not take full advantage of the simultaneous-equations structure inherent in the choice of a portfolio of monetary assets.[2]

We have four tasks before us. First, we discuss the problem of the definition (aggregation) of money; after a consideration of the theoretical problems, we will propose the use of the Divisia method of aggregation for the construction of monetary aggregates. Second, we show how a simultaneous-equations financial assets structure both fits neatly into the

[2]The two general approaches mentioned can be shown to be special cases of the approach we are surveying once risk is introduced into decisions; see James Poterba and Julio Rotemberg (1987); Barnett, Melvin Hinich, and Piyu Yue (1991a); Barnett and Apostolos Serletis (1990); and Barnett, Hinich, and Yue (2000).

definitional approach we are recommending and also provides a structure that can be used to measure income and interest rate elasticities as well as the important elasticities of substitution among financial entities. An econometric digression here will emphasize the contribution that can be made by employing one of the flexible functional forms at the estimation stage. Third, in our discussion of the empirical literature, we will emphasize how the theory might be implemented (and briefly show some of the results); the purpose of this discussion will be to illustrate the theory rather than to survey what is a rapidly changing empirical record. Finally, we briefly discuss ongoing research and extensions of the literature. The work discussed here includes the use of formal methods of aggregating over economic agents as well as the incorporation into the general framework of risk aversion and rational expectations (both for consumers and firms).

1.2 The Definition of Money

The natural place to begin is with the definition of money. Currently, the common practice among central banks is to construct monetary aggregates from a list of entities by adding together those that are considered to be the likely sources of monetary services. That is, commercial banks and other financial intermediaries provide demand deposits, certificates of deposit and the like, and it is from this list that the (usually simple-sum) monetary aggregates are composed. At the Federal Reserve, for example, there are currently 27 components in the entire collection of liquid financial assets — as shown in Table 1.1 — and, as also shown, the four popular aggregates of M1, M2, M3, and L are constructed directly from this list by means of a recursive form of accounting that starts with M1 (the inside block) and adds blocks of items to M1 until all 27 entities are included (in L, for 'liquid assets').

What is important about these components for what follows in this chapter is that the quantities of each vary at different rates over time (and so do their 'prices').[3] To see the behavior of the quantities, consider the collection in Figure 1.1 of monthly Federal Reserve data.[4]

Here Figure 1.1a shows the behavior of the components of M1 in recent years, while Figures 1.1b and 1.1c show that behavior for the items added

[3] As we will describe below, from the point of view of this survey, the appropriate price for each entity in Table 1.1 is the *user cost* of the asset. These are based partly on the nominal interest rate just mentioned. Some of the more liquid items (such as currency) do not possess an "own" interest rate and so a zero rate is usually assumed. We illustrate the behavior of user costs in Figure 1.3, below.

[4] These numbers were supplied by the Federal Reserve and are available from the authors.

to M1 to construct M2. Not only are the fluctuations of the quantities different for different assets, especially since 1979, but also new assets appear in the list from time to time. These sorts of changes potentially complicate the calculation of a unique measure of moneyness, although broader measures might well perform better than narrower ones simply because the new assets, e.g., are likely to draw funds from other entities within the broader collection of potential substitutes.

As noted, the monetary aggregates currently in use by the Federal Reserve are simple-sum indices in which all monetary components are assigned a constant and equal (unitary) weight. This index is M in

$$M = \sum_{i=1}^{n} x_i, \tag{1.1}$$

where x_i is the i^{th} monetary component of, say, the subaggregate M1; it clearly implies that all monetary components are weighted linearly and equally in the final total. This sort of index has some use as an accounting measure of the stock of nominal monetary wealth, of course, and this is an important advantage. More tellingly, this form of aggregation implies that all components are dollar-for-dollar perfect substitutes, since all indifference curves and isoquants over those components must be linear with slopes of minus 1.0 if this aggregate is to represent the monetary service flow selected by the economic agents. This is the source of its potential weakness.

The main problem with the simple-sum index arises from the fact that in aggregation theory a quantity index should measure the income effects (i.e., welfare or service flow changes) of a relative price change but should be unresponsive to pure substitution effects (at constant utility), which the index should internalize. The simple-sum index cannot untangle income from substitution effects if its components are not perfect substitutes.[5] In the face of what appear to be significant changes in the relative prices of financial assets and with increasing numbers of apparently imperfect substitutes among the relevant short-term financial assets, it is not surprising that attempts are now common to arrive at a definitional procedure that will accommodate this volatility. Milton Friedman and Anna Schwartz, in their monumental survey of the literature on the definition of money, when discussing the simple-sum approach, discuss the basic issue in the following terms:

[5] What is required for a consistent aggregation is an *aggregator function*, which the simple sum is. The problem is that the natural choice for an aggregator function is a utility or production function; only if the components are perfect substitutes is the simple-sum the appropriate utility or production function.

TABLE 1.1

Official Monetary Aggregates/Components
U.S. Federal Reserve

L

M 3

M 2

M 1

- Currency and travelers' checks
- Demand deposits held by consumers
- Demand deposits held by businesses
- Other checkable deposits
- Super NOW accounts held at commercial banks
- Super NOW accounts held at thrifts

(M 2, in addition to M 1)

- Overnight RPs
- Overnight Eurodollars
- Money market mutual fund shares
- Money market deposit accounts at commercial banks
- Money market deposit accounts at thrifts
- Savings deposits at commercial banks
- Savings deposits at savings and loans (S&Ls)
- Savings deposits at mutual savings banks (MSBs)
- Savings deposits at credit unions
- Small time deposits and retail RPs at commercial banks
- Small time deposits at S&Ls and MSBs and retail RPs at thrifts
- Small time deposits at credit unions

(M 3, in addition to M 2)

- Large time deposits at commercial banks
- Large time deposits at thrifts
- Institutional money market funds
- Term RPs at commercial banks and thrifts
- Term Eurodollars

(L, in addition to M 3)

- Savings bonds
- Short-term Treasury securities
- Bankers' acceptances
- Commercial paper

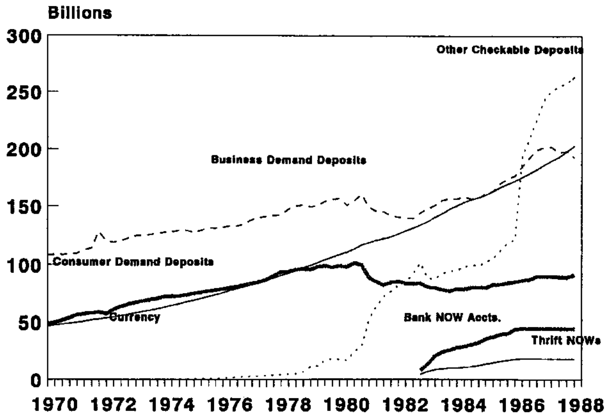

Figure 1.1a: Components of M1 in the United States.

> "This (summation) procedure is a very special case of the more general approach. In brief, the general approach consists of regarding each asset as a joint product having different degrees of 'moneyness,' and defining the quantity of money as the weighted sum of the aggregate value of all assets, the weights for individual assets varying from zero to unity with a weight of unity assigned to that asset or assets regarded as having the largest quantity of 'moneyness' per dollar of aggregate value. The procedure we have followed implies that all weights are either zero or unity. The more general approach has been suggested frequently but experimented with only occasionally. We conjecture that this approach deserves and will get much more attention than it has so far received." (Friedman and Schwartz 1970, pp. 151-152)

Their observation is deficient only in failing to point out that even weighted aggregation implies perfect (but not dollar-for-dollar) substitutability unless the aggregation procedure is nonlinear.[6]

[6] For example, introduction of estimated multiplicative coefficients into Equation (1.1) would retain the linearity of the aggregation and hence the implication of perfect substitutability.

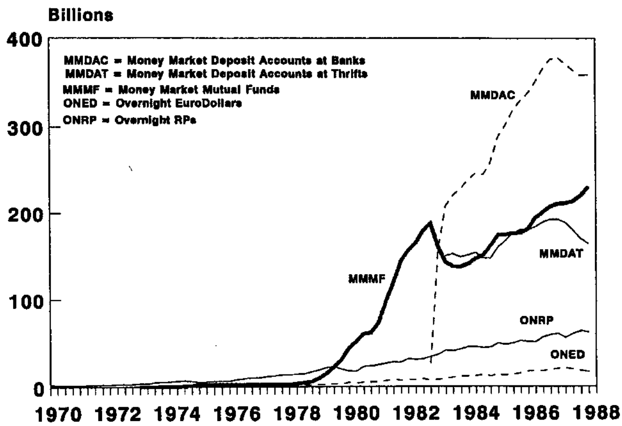

Figure 1.1b: Liquid instruments in the United States.

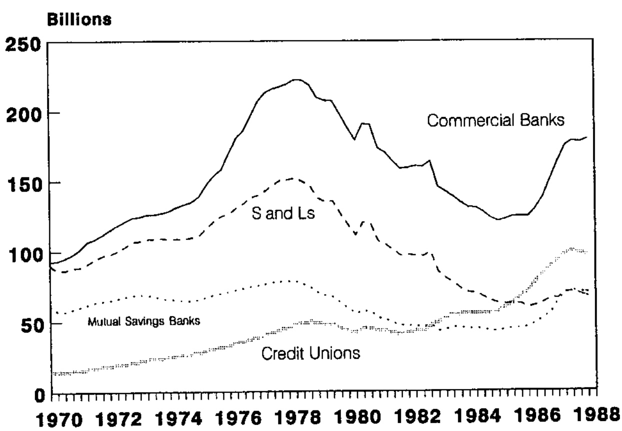

Figure 1.1c: Savings deposits in the United States.

Over the years, there have been a series of attempts to achieve a rule for aggregating monetary components without abandoning the simple-sum structure. There is, indeed, a theory that might seem to back this up: that of Hicksian aggregation (John Hicks 1946; Don Patinkin 1965). The main difficulty is that for Hicksian aggregation to be possible it is required that the relative prices (user costs) of the financial commodities not change over the sample period. Even if that assumption were true, Hicksian aggregation alone is not sufficient for simple-sum aggregation; it also is necessary that the constant user cost between any two assets be equal to 1.0. Once again, this condition can be expected to hold only if the component assets are indistinguishable perfect substitutes; this is unlikely if only because all financial assets one can think of provide different services and hence have different 'own rates' of return — and these change over time. The user costs depend upon those yields.

An attractive alternative to the simple-sum approach is to use microeconomic aggregation theory to define money. This theory has two branches, one leading to the construction of index numbers using methods derived from economic theory and one leading to the construction of money-demand functions in the context of a system of equations modelling the wealth-holder's allocation of funds among money and non-money assets. The two branches are supported by the same structure, in that the supporting theory in both cases is that of the constrained maximization of the aggregate consumer's intertemporal utility function. The following section, in spelling out the theory, emphasizes this theoretical coherence.

1.3 The Microeconomic Theory of a Monetary Economy

Consider an economy with identical individuals, having three types of goods: consumption goods, leisure, and the services of monetary assets. Assuming that the services of these three entities enter as arguments in the individual's utility function, the utility function can be written as[7]

$$u = U(\boldsymbol{c}, L, \boldsymbol{x}) \tag{1.2}$$

where $\boldsymbol{c}$ is a vector of the services of consumption goods, L is leisure time, and $\boldsymbol{x}$ is a vector of the services of monetary assets. For money, the services

[7]Putting money into the utility function, for some utility functions, is observationally equivalent to putting money (solely) into the constraints. This result is established by Robert Feenstra whose demonstration applies for a broad class of utility functions and a broad class of transactions cost models (including the inventory-theoretic model and the Robert Clower (1967) cash-in-advance constraint formulation). Much of this is also discussed in Patinkin (1965). Feenstra notes...

could be convenience, liquidity, and information (as in Karl Brunner and Allan Meltzer 1971).

The utility function in Equation (1.2) can be assumed to be maximized subject to the full income constraint

$$q'c + \pi'x + wL = y \tag{1.3}$$

where y is full income (i.e., income reflecting expenditures on time as well as on goods and services); q is a vector of the prices of c; π is a vector of monetary asset *user costs* (or rental prices); and w is the shadow price of leisure (see Barnett 1981b). The i^{th} component of π is given by (see Barnett 1978 and Donal Donovan 1978),

$$\pi_i = p^* \left(\frac{R - r_i}{1 + R} \right) \tag{1.4}$$

This formula measures the opportunity cost — at the margin — of the monetary services provided by asset i. It is calculated as the discounted value of the interest foregone by holding a dollar's worth of that asset. Here r_i is the expected nominal holding-period yield on the i^{th} asset, R is the maximum expected holding-period yield available on an alternative asset (the 'benchmark' asset) and p^* is the true cost-of-living index.[8] Note, especially, that this formula is not arbitrary but can be derived from an in-

> "We demonstrate a functional equivalent between using real balances as an argument of the utility function and entering money into liquidity costs which appear in the budget constraints." (p. 271)

This should dispose of the issue for purposes of this survey, since the utility-maximizing models discussed here are of the very general sort that Feenstra discusses. In a general equilibrium context, the same result also has been proved by Kenneth Arrow and Frank Hahn (1971); a parallel proof in a production context is due to Stanley Fischer (1974). See also Louis Philips and Frank Spinnewyn (1982) and Poterba and Rotemberg (1987). Once money has been put into the utility function, the inverse mapping to the motivating transactions constraint is not unique, so that the reason for holding money is lost. But we do not seek to explain the reason for holding money. We simply observe that money does have a positive value in equilibrium so that we can appeal to the Arrow and Hahn proof.

[8]The benchmark asset is specifically assumed to provide no liquidity or other monetary services and is held solely to transfer wealth intertemporally. In theory, R is the maximum expected holding period yield in the economy. It is usually defined in practice in such a way that the user costs for the monetary assets are positive. The true cost of living index, p^*, is defined for the consumer goods, c; we use the term "true" merely to warn the reader that this is an *exact* price index, as this concept is described below. Note that if p^* is deleted from the user cost formula, the formula produces real rather than nominal user cost. The interest rates are nominal so that inflationary expectations appear here (in the denominator, since the effects in the two rates in the numerator of the formula may well cancel out).

tertemporal optimization problem of a standard sort (see Barnett 1981b).[9]

In order to focus on the details of the demand for monetary services, a good starting point is the theory of two-stage optimization investigated initially in the context of consumer theory by Robert Strotz (1957, 1959) and William Gorman (1959). The theory describes a sequential expenditure allocation in which in the first stage (that of budgeting or 'price aggregation') the consumer allocates his expenditure *among* broad categories (consumption goods, leisure, and monetary services in the context of the model used here) and then in the second stage (that of 'decentralization') allocates expenditures *within* each category. In the first stage his decision is guided by price indices among the three categories, while in the monetary part of the decentralized decision, he responds to changes in the relative prices of the monetary assets (π_i/π_j as defined above).

Decomposition of the consumer choice problem along these lines is possible only if the individual's utility function (1.2) is *weakly separable* in the services of monetary assets. That is, it must be possible to write the utility function as

$$u = U[\boldsymbol{c}, L, f(\boldsymbol{x})] \tag{1.5}$$

in which f defines the monetary subutility function. As laid out originally by Masaza Sono (1961) and Wassily Leontief (1947a), the condition of weak separability described in (1.5) is equivalent to

$$\frac{\partial}{\partial \phi}\left(\frac{\partial U/\partial x_i}{\partial U/\partial x_j}\right) = 0 \tag{1.6}$$

[9]For example, suppose that the representative consumer (under continuous replanning) maximizes utility over three periods $(1, 2, 3)$, subject to a budget constraint. For simplicity, we can fix the time path of leisure and consider only one monetary asset and one consumption good. The one-period budget constraint for period 2 is

$$p_2^* c_2 = (1 + r_1)p_1^* x_1 - p_2^* x_2 + [(1 + R_1)A_1 - A_2]$$

where A denotes per capita holdings of an alternative asset (say, a bond), while R is the expected nominal yield on A and r is that on x. The other variables are defined the same way as in the text, above.

Solve the period 2 budget constraint for A_2 and write the resulting equation for each of the three periods. Then by back substituting for A, starting from A_3 and working down to A_1, we obtain (after some manipulation) the consumer's budget constraint in present value form:

$$(1+R_0)A_0 + (1+r_0)p_0^* m_0 = \frac{A_3}{\rho_3} + \sum_{s=1}^{3} \frac{p_s^*}{\rho_s} c_s + \sum_{s=1}^{3} \left[\frac{p_s^*}{\rho_s} - p_s^* \frac{(1+r_s)}{\rho_{s+1}}\right] x_s + \frac{p_3^*(1+r_3)}{\rho_4} x_3$$

where the discount factor, ρ_s, equals 1 for $s = 1$, $(1 + R_1)$ for $s = 2$, etc. .... In the equation just given, the term in the bracket is the user cost of x. Writing the user cost for $s = 1$ (the current period), we obtain equation (1.4).

for $i \neq j$, where ϕ is any component of $\{c, L\}$. This condition asserts that under weak separability the marginal rate of substitution between any two monetary assets is independent of the values of c and L.[10]

If we have established a separable subset of assets, whether by separability test or by assumption, we then can continue in the framework of the following neoclassical consumer problem,

$$\max f(\boldsymbol{x}) \qquad \text{subject to } \pi'\boldsymbol{x} = m \tag{1.7}$$

where m is the total expenditures on monetary services, a total that is determined in the first stage of the two-level optimizing problem. It is this simple structure that we will need to recall in the work in the following sections.

Whether or not the utility function (1.2) is weakly separable in monetary services is, ultimately, an empirical question. Ideally, instead of treating equation (1.5) as a maintained (and therefore untested) hypothesis as so much of the money-demand literature implicitly does, one could test whether the utility function (1.2) is appropriately separable in monetary services — the assumption implicit in the traditional 'money-nonmoney' dichotomization. The existing methods of conducting such tests are not, however, very effective tools of analysis, as discussed below.

1.3.1 The Aggregation-Theoretic Approach to Money Demand

In the preceding discussion we have shown the steps that are normally taken to reduce a very general consumer choice problem to an asset choice problem. At this point, we are prepared to proceed to results in the 'aggregation-theoretic' literature on money demand. This literature ties up the theory just sketched with that on the macroeconomic demand for money in an obvious and logical way. What we are after here are monetary aggregates that are consistent with the optimizing behavior of rational economic agents.

We begin with the *aggregator function.* In the aggregation-theoretic literature for the consumer choice problem, the quantity aggregator function has been shown to be the subutility function defined solely over the individual monetary components (as listed in Table 1.1). This function, for the

[10] Note that the separability structure is asymmetric. That is, c is not separable from $\boldsymbol{x}$ and L in U unless there exists a function $g(c)$ such that $u = U(c, L, \boldsymbol{x}) = U[g(c), L, f(\boldsymbol{x})]$. For an extensive discussion of separability, see Charles Blackorby, Daniel Primont, and Robert Russell (1978).

monetary services problem defined in equation (1.7), is $f(x)$.[11] Using a specific and differentiable form for the monetary services aggregator function $f(x)$, and solving decision (1.7), we can derive the inverse and/or the direct demand-function system. Using these derived solution functions and specific monetary data, we then could estimate the parameters and replace the unknown parameters of $f(x)$ by their estimates. The resulting estimated function is called an *economic* (or functional) monetary *index*, and its calculated value at any point is an economic monetary-quantity index number.

We have noted that the researcher would choose a specific and differentiable functional form if he were to follow this approach. The problem is that the use of a specific function necessarily implies a set of implicit assumptions about the underlying preference structure of the economic agent. For example, as we have already emphasized, the use of a weighted-linear function for $f(x)$ implies perfect substitutability among the monetary assets. If the weights are all unity so that we get the widely used simple-summation functions, the consumer will specialize in the consumption of the highest yielding asset. The use of the more general Cobb-Douglas function imposes an elasticity of substitution equal to unity between every pair of assets. To continue, the constant elasticity of substitution function, although relaxing the unitary elasticity of substitution restriction imposed by the Cobb-Douglas, nevertheless imposes the restriction that the elasticity of substitution is the same between any pair of assets. The list of specific functional forms is, of course, boundless, but the defining property of the more popular of these entities is that they imply strong limitations on the behavior of the consumer. While the issue of their usefulness is ultimately an empirical question — and we shall treat the issue that way below — we feel that most members of this class of functions should be rejected for estimation of money demand, partly in view of the restrictive nature of their implicit assumptions, and partly because of the existence of attractive alternatives.

Among the alternatives is a member of the class of quadratic utility functions. With a member of the quadratic class, we would be using a *flexible functional form* to approximate the unknown monetary-services aggregator function. Flexible functional forms — such as the translog — can locally approximate to the second order any unknown functional form for the monetary services aggregator function, and even higher quality approximations

[11] The argument just given requires that $f(\cdot)$ be homothetic. In the nonhomothetic utility case, the aggregator function is the distance function (see Barnett 1987). However, since the resulting index is the same in either case, the conclusions of this section are unaffected by this assumption. This topic is considered further, below.

are available.[12] We will consider the details of this method below.

If one is to do away with the simple-sum method of aggregating money and replace it with a nonlinear aggregator function as suggested, one will be able to deal with less than perfect substitutability and, for that matter, with variations over time in the elasticities of substitution among the components of the monetary aggregates. There is a problem, however, and this is that the functions must be estimated over specific data sets (and re-estimated periodically) with the attendant result that the index becomes dependent upon the specification. This dependence is particularly troublesome to government agencies that have to justify their procedures to persons untrained in econometrics. This is a reasonable concern — and it is exacerbated by the fact that there are many possible nonlinear models from which to choose. Under these circumstances, government agencies around the world have taken a more direct approach and use index number formulas from statistical index number theory for most of their calculations. We will explain how this approach can be implemented in a way that simultaneously deals with the theoretical and the practical issues. We will not, however, be able to explain why there is not more use of the approach by the monetary authorities of these same governments.

1.3.2 Index Number Theory

Statistical index-number theory provides a variety of quantity and price indices that treat prices and quantities as jointly independent variables. Indeed, whether they are price or quantity indices, they are widely used, since they can be computed from price and quantity data alone, thus eliminating the need to estimate an underlying structure. In fact, since the appearance of Irving Fisher's (1922) early and now classic study on statistical index number theory, nearly all national government data series have been based upon aggregation formulas from that literature. Well-known examples are the Consumer Price Index (a Laspeyres price index), the Implicit GNP Deflator (a Paasche price index), and real GNP (a Laspeyres quantity index). The simple-sum index often used for monetary quantities is a member of the broad class. But the simple sum is a degenerative measure, since it contains no prices.

Statistical indices are distinguished by their known statistical properties. These properties are described in detail by Irving Fisher (1922), and in that work he provides a set of tests (known as Fisher's System of Tests)

[12] Such as the Minflex Laurent (see Barnett 1983a, 1985; Barnett and Yul Lee 1985), the Fourier (see Ronald Gallant 1981), and the Asymptotically Ideal Models (see Barnett and Andrew Jonas 1983; Barnett and Yue 1988; Yue 1991; and Barnett, John Geweke , and Michael Wolfe 1991a).

useful for assessing the quality of a particular statistical index.[13] The index that he believes often to be the best in the sense of possessing the largest number of satisfactory statistical properties, has now become known as the Fisher Ideal Index. Another index found to possess a very large number of these properties is the Törnqvist discrete-time approximation to the Divisia Index.[14] We note that Fisher found the simple-sum index and to be the worst of the literally hundreds of possible indices that he studied.

Let x_{it} be the quantity of the i^{th} asset during period t, and let π_{it} be the rental price (that is, user cost) for that good during period t. Then, the *Fisher ideal index* (Q_t^F) during period t is the geometric average of the Laspeyres and Paasche indices:

$$\frac{Q_t^F}{Q_{t-1}^F} = \left[\frac{\sum_{i=1}^{n} s_{i,t-1} \left(\frac{x_{it}}{x_{i,t-1}} \right)}{\sum_{i=1}^{n} s_{it} \left(\frac{x_{i,t-1}}{x_{it}} \right)} \right]^{1/2} \tag{1.8}$$

where

$$s_{it} = \frac{\pi_{it} x_{it}}{\sum_{k=1}^{n} \pi_{kt} x_{kt}}.$$

On the other hand, the discrete time (Törnqvist) *Divisia index* during period t is Q_t^D, where

$$\frac{Q_t^D}{Q_{t-1}^D} = \prod_{i=1}^{n} \left(\frac{x_{it}}{x_{i,t-1}} \right)^{(1/2)(s_{it}+s_{i,t-1})} \tag{1.9}$$

It is informative to take the logarithms of each side of (1.9), so that

$$\log Q_t^D - \log Q_{t-1}^D = \sum_{i=1}^{n} s_{it}^* (\log x_{it} - \log x_{i,t-1}) \tag{1.10}$$

where $s_{it}^* = (1/2)(s_{it} + s_{i,t-1})$. In this form, it is easy to see that for the Divisia index the growth rate (log change) of the aggregate is the share-weighted average of the growth rates of the component quantities.

A characteristic of the Fisher Ideal Index is that the Fisher Index is 'self dual.' In such a case, if the quantity index is the Fisher *quantity* index, then the implied price index — defined by dividing total expenditure on

[13]Fisher's tests for statistical indices are proportionality, circularity, determinateness, commensurability, and factor reversal. For recent discussions, see Eichhorn (1976, 1978), Diewert (1992), and Balk (1995).

[14]The Divisia index was originated by the French economist Francois Divisia (1925).

the components by the quantity index — is the Fisher *price* index. Hence, the Fisher price and quantity indices comprise a dual pair in the sense that their product equals total expenditure on their components; this is known as Fisher's *factor reversal test.* In contrast, the price index that is dual to the Divisia quantity index is actually not a Divisia price index. Nevertheless, even if the Divisia price index were used to measure the price of the Divisia quantity index, the size of the error produced as a result of the violation of the factor reversal test would be very small (third order in the changes). Indeed, the failure to be self-dual is common among the most popular index numbers.[15] In view of the fact that the Divisia quantity index has the very considerable advantage of possessing an easily interpreted functional form, as in equation (1.10), it is now often employed in the emerging literature on monetary aggregation. It has another desirable property, as we shall see in a moment.

1.3.3 The Links Between Aggregation Theory, Index Number Theory, and Monetary Theory

Until relatively recently, the fields of aggregation theory and statistical index number theory developed independently. Erwin Diewert (1976, 1978), however, provided the link between aggregation theory and statistical index number theory by attaching economic properties to statistical indices. These properties are defined in terms of the statistical indices' effectiveness in tracking a particular functional form for the unknown aggregator function. Recall in thinking about this that the utility function is itself the appropriate aggregator function. What Diewert shows is that using a number of well-known statistical indices is equivalent to using a particular functional form to describe the unknown economic aggregator function. Such statistical indices are termed *exact* in this literature. Exactness, briefly, occurs when the specific aggregator function (e.g., the linear-homogeneous translog) is exactly tracked by a particular statistical index (e.g., the discrete-time Divisia); the parentheses illustrate one such case.[16]

Having the property of exactness for *some* aggregator function, however, is not sufficient for acceptability of a particular statistical index when the true functional form for the aggregator function is not known, *a priori.* What can be done in these circumstances is to choose a statistical index that is exact for a *flexible* functional form — a functional form that can

[15] For example, neither the Paasche nor the Laspeyres index is self-dual, although the Paasche and the Laspeyres are a dual pair. Hence it is common to use Laspeyres (not Paasche) quantity indexes with Paasche (not Laspeyres) price indexes.

[16] Diewert also shows that the Fisher-Ideal Index is exact for the square root of a homogeneous quadratic function.

provide a second-order approximation to any arbitrary unknown aggregator function. Taking this approach cuts through the problem of not knowing the underlying structure. Diewert terms such a statistical index *superlative.* As it turns out, the Divisia Index is exact for the linearly homogeneous (*and flexible*) translog and is, therefore, superlative; that is, it approximates an arbitrary unknown exact aggregator function up to a third-order remainder term.[17] What one gains from this is the ability to do precise work even when the form of the underlying function is not known.

With Diewert's successful merging of index number theory and economic aggregation theory and the rigorous derivation of the appropriate price of monetary services in the form of the user cost of these services (Donovan 1978; Barnett 1978), the loose link between monetary theory and economic aggregation theory has been turned into a firm one and the scene has been set for the construction of theoretically inspired monetary aggregates. In our discussion, we have pointed out that either the Divisia Index or the Fisher-Ideal Index of monetary services would be superlative. Actually, it has been demonstrated (Barnett 1980a) that the difference between the two is typically less than the roundoff error in the monetary data. The Federal Reserve, indeed, has employed both procedures as alternatives to the much more widely-known simple-sum aggregates [18]

There is one other aggregate in at least limited use in the monetary literature, and that is MQ, the monetary quantities index. In the particular form computed by Paul Spindt (1985a), MQ is measured as in equation (1.8), but with the user costs replaced by monetary-asset turnover rates. The problem with this procedure is that the MQ index, unlike the Divisia, is inconsistent both with existing aggregation and index number theories. The relevant foundations (both index-number theoretic and aggregation theoretic) for the Fisher-Ideal Index *require* the use of prices and quantities and not turnover rates and quantities.[19] An attempt to define money in yet another way is due to Don Roper and Stephen Turnovsky (1980). In their paper the object is to determine the optimal monetary aggregate for stabilization policy assuming a one-goal objective function — which is the minimization of the variance of income. This, too, is an atheoretical approach from the point of view of this survey. It has the added difficulty that there are significant theoretical and empirical problems associated with expanding the objective function to incorporate multiple goals (and with

[17] In fact, even if the aggregator is not homogenous, the Divisia Index remains exact — but for the distance function, which is the economic aggregator function in that case.

[18] However, the Federal Reserve does not publish these numbers.

[19] For a proof, see Barnett (1987, 1990). If nothing else, MQ can be said to be no less arbitrary than the official simple-sum aggregates. But the number of such atheoretical aggregates is infinite, as is the number of possible nonlinear functions of quantities and turnover rates.

dealing with a large number of monetary assets and other policy tools).

There has recently been some interest in measuring wealth effects produced from changes in expected monetary service flows. The formula for the monetary capital stock was derived by Barnett (1991, eq. 2.2), who proved that it equals the discounted present value of the expected future Divisia monetary service flow. Under the assumption of stationary expectations, this expected discounted present value simplifies to an index called the CE index (see Barnett 1991, theorem 1).[20] Rotemberg (1991) and Rotemberg, Driscoll, and Poterba (1995) present some interesting empirical results with the CE index. While the implicit stationary expectations assumption is troubling, the CE index, unlike the MQ and the Roper-Turnovsky indexes, is not atheoretical. We anticipate that future applications of Barnett's capital stock formula, with less restrictive assumptions on expectations than used in generating the CE special case, will produce further advances in measuring monetary wealth effects.

At this point we might pause to consider the composition of these new indices and what their recent behavior has been. The indices are constructed from the same components as the traditional measures (e.g., Divisia M1 is constructed from the components list used for M1, as described in Table 1.1); they employ user costs, as defined above, in their calculations. As the graphs in Figure 1.2 indicate, these numbers differ, sometimes considerably, from the summation indices. This difference is especially large for the broader measures and for the years since 1978.

The simple correlations that go with these figures are

	Level	Differenced
Sum/Divisia M1	.998	.964
Sum/Divisia M2	.986	.542
Sum/Divisia M3	.982	.407
Sum/Divisia L	.985	.487

While these correlations are quite high in the trend-dominated level figures, when differenced, the broader series show considerably smaller correlations. Because it is in differenced form that the monetary data are usually studied, there is sufficient reason in these figures to dig further into their relative performances. Note also that the user costs (of the components of these aggregates) are often not highly correlated (see Figure 1.3 below).

[20]There is also a considerably less attractive interpretation of the CE index as a flow index. See Barnett (1995, section 5) regarding the flow interpretation, which requires stronger assumptions than those needed to derive the Divisia flow index.

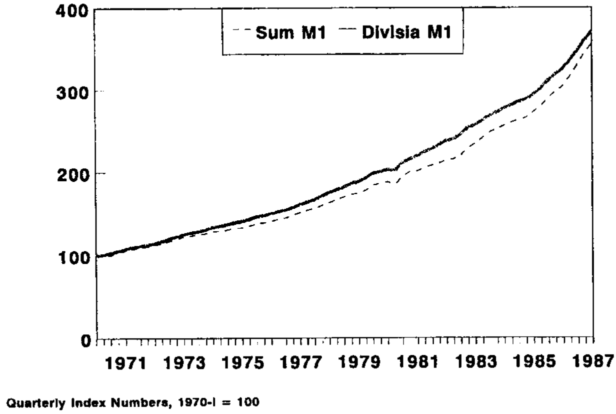

Figure 1.2a: Sum M1 versus Divisia M1 in the United States.

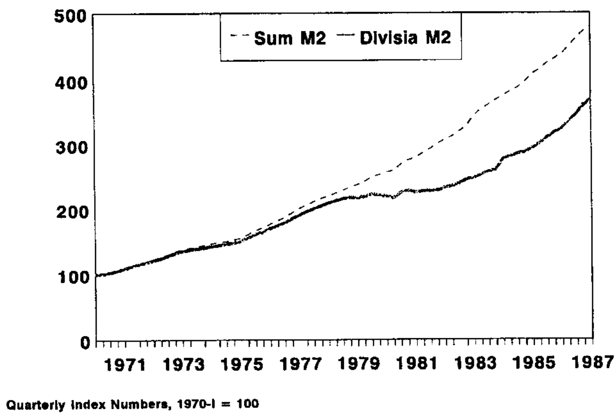

Figure 1.2b: Sum M2 versus Divisia M2 in the United States.

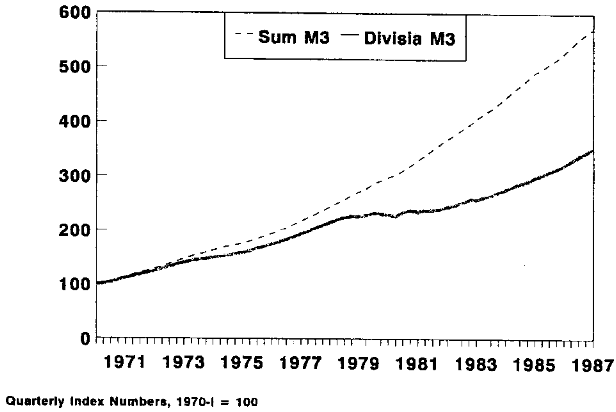

Figure 1.2c: Sum M3 versus Divisia M3 in the United States.

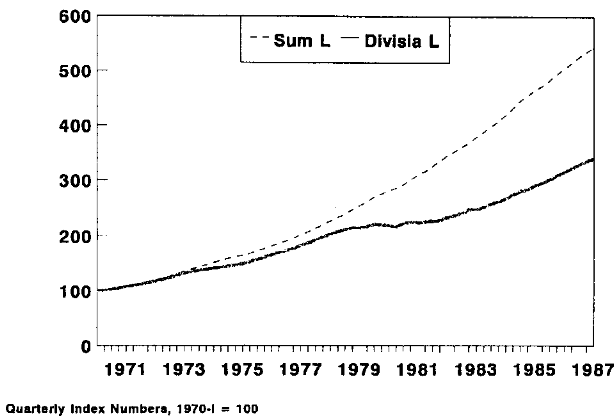

Figure 1.2d: Sum L versus Divisia L in the United States.

1.3.4 Understanding the New Divisia Aggregates

To understand the new Divisia aggregates, we must consider the underlying microeconomic theory behind the Divisia Index. Let us return to the consumer's utility function over monetary assets as defined above (this was $f(\boldsymbol{x})$). Writing out the total differential of $f(\boldsymbol{x})$, we obtain

$$df(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right) dx_i \tag{1.11}$$

where the partial derivatives are marginal utilities (which are functions themselves containing the unknown parameters of the function $f(\boldsymbol{x})$). From the first-order conditions for the expenditure-constrained maximization of $f(\boldsymbol{x})$, we can write the marginal utilities as

$$\lambda \pi_i = \frac{\partial f}{\partial x_i} \quad i = 1, \ldots, n \tag{1.12}$$

Here λ is the Lagrange multiplier, and π_i is the user-cost (i.e., the rental price) of asset i. This expression can then be substituted into equation (1.11) to yield

$$df(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda \pi_i \, dx_i \tag{1.13}$$

which is written not in unknown marginal utilities but in the unknown Lagrange multiplier, user costs, and changes in quantities.

In equation (1.13) the Lagrange multiplier is itself a function of unknown tastes and thereby a function of the parameters of the unknown utility function. Hence, we have one more step to go; we must eliminate the Lagrange multiplier. This involves the assumption that the economic quantity aggregate, $f(\boldsymbol{x})$, is linearly homogeneous in its components. This is, indeed, a reasonable assertion, since it would be very curious indeed if linear homogeneity of $f(\boldsymbol{x})$ failed. It is reasonable because if it does not hold, the growth rate of the aggregate differs from the growth rates of its components, even if all components are growing at the same rate.[21]

Let us now define $P(\pi)$ to be the dual price index satisfying Fisher's factor-reversal test as this was described above.

$$P(\pi) f(\boldsymbol{x}) = \sum_{i=1}^{n} \pi_i x_i \quad [= m]. \tag{1.14}$$

[21] In fact, the linear homogeneity assumption not only is necessary, but also is harmless. In the general case of nonhomothetic tastes and technology, the aggregator function is the distance function — which *always* is linearly homogeneous.

It can then be shown that[22] $\lambda = 1/P(\pi)$, in which case equation (1.13) can be written as

$$df(\boldsymbol{x}) = \sum_{i=1}^{n} \frac{1}{P(\pi)} \pi_i \, dx_i. \tag{1.15}$$

Manipulating equation (1.15) algebraically to convert to growth rate (log change) form, we find that

$$d \log f(\boldsymbol{x}) = \sum_{i=1}^{n} s_i \, d \log x_i \tag{1.16}$$

where

$$s_i = \frac{\pi_i x_i}{\sum_{k=1}^{n} \pi_k x_k}$$

is the i^{th} asset's value share in the total expenditures on monetary services. The result is the Divisia index, as defined in equation (1.10), where the log change in the utility level (and therefore in the level of the aggregate) is the weighted average of the log changes of the component levels, with expenditure shares providing the weights. This exercise demonstrates the solid micro-foundations of the Divisia index. It is, indeed, the logical choice for an index from a theoretical point of view, being exactly the transformed first-order conditions for constrained optimization. In addition, the derivation just completed demonstrates that the prices appearing in index numbers cannot be replaced by any other variables, such as turnover rates, bid-ask spreads, brokerage fees, etc. To do so would be to violate the first-order conditions, (1.12). In particular, for the case of monetary assets, it is user costs that preserve (1.12) and therefore permit (1.13).

1.3.5 The Optimal Level of Monetary Subaggregation

Even if the utility function, $U(\cdot)$, is weakly separable in its monetary assets group, there remains the problem of selecting monetary asset subgroups for inclusion in monetary *subaggregates.* In particular, the use of any monetary sub-aggregate (such as M1, M2, M3, or L) implies that the components of

[22]Let $\boldsymbol{x} = D(m, \pi)$ to be the solution to the maximization of $f(\boldsymbol{x})$ subject to $\pi'\boldsymbol{x} = m$. The linear homogeneity of $f(\boldsymbol{x})$ implies that there must exist a vector of functions $\boldsymbol{h}(\pi)$ such that $\boldsymbol{x} = m\boldsymbol{h}(\pi)$. Substituting for $\boldsymbol{x}$into $f(\boldsymbol{x})$, we find that

$$f(\boldsymbol{x}) = f[m\boldsymbol{h}(\pi)] = mf[\boldsymbol{h}(\pi)].$$

As $\lambda = \partial f / \partial m$, we have from the last equation that $\lambda = f[\boldsymbol{h}(\pi)]$. In addition, from Equation (1.14), $P(\pi)f(\boldsymbol{x}) = m$. Hence from $f(\boldsymbol{x}) = mf[\boldsymbol{h}(\pi)]$ we have that $f[\boldsymbol{h}(\pi)] = 1/P(\pi)$. Hence, $\lambda = 1/P(\pi)$.

the subaggregate are themselves weakly separable within $\boldsymbol{x}$. This additional nested separability condition is required, regardless of the type of index used *within* the subaggregate. Weak separability over the component assets is both necessary and sufficient for the existence of stable preferences (or technology) over those components of the subaggregate. This implies that without separability such a subaggregate has no meaning in theory, since the subaggregate fails the existence condition.

Even so, weak separability establishes only a necessary condition for subaggregation in its simplest form. In particular, if we wish to measure the subaggregate using the most elementary method, we would require the additional assumption that the separable subfunction within f be homothetic. Then f is said to be homothetically weakly separable. Indeed, homothetic weak separability is necessary and sufficient for the simplified form of subaggregation.[23]

For illustration, let us describe the Federal Reserve Board's *a priori* assignment of assets to monetary subaggregates. As illustrated in Table 1.1, their method is based on the implicit assumption that the monetary services aggregator function, $f(\boldsymbol{x})$, has the recursive weakly separable form of

$$f(\boldsymbol{x}) = f_4(\boldsymbol{x}^4, f_3(\boldsymbol{x}^3, f_2(\boldsymbol{x}^2, f_1(\boldsymbol{x}^1)))). \tag{1.17}$$

This clearly implies that the marginal rate of substitution between, say, an asset in $\boldsymbol{x}^1$ and an asset in $\boldsymbol{x}^2$ is independent of the values of assets in $\boldsymbol{x}^3$ and $\boldsymbol{x}^4$.

In equation (1.17), the components of $\boldsymbol{x}^1$ are those included in the Federal Reserve Board's M1 monetary aggregate, the components of $\{\boldsymbol{x}^1,\boldsymbol{x}^2\}$ are those of the M2 aggregate, the components of $\{\boldsymbol{x}^1,\boldsymbol{x}^2,\boldsymbol{x}^3\}$ are those of the M3 aggregate, and the components of $\boldsymbol{x}$are those of the L aggregate. The aggregator functions f_i, for $i = 1, \ldots, 4$, would rationalize the Federal Reserve's component groupings, M1, M2, M3, and L. Of course, the actual numbers produced for the official monetary aggregates further require the assumptions that f_1, f_2, f_3, f_4 (and hence f itself) are all simple summations.

[23] In its most general form, however, aggregation is possible without homotheticity but with only the minimal existence condition of weak separability. That generalization uses the distance function for the aggregator function and produces the Malmquist index (see Barnett 1987); see below for more on the distance function. Hence we currently are presenting a special case, to which the Malmquist index reduces if and only if f is linearly homogeneous.

1.4 Econometric Considerations

In recent years there have been a number of related developments that have increased the usefulness of the 'demand systems' modeling approach for monetary studies. The following discussion attempts to clarify just what these developments are and how they are tending to reorganize a very traditional literature on an important topic.

The first problem is involved with the definition of money: what assets should be selected and how should they be grouped? As a first pass, one might aggregate (by a Divisia index) those assets that have highly collinear rates of return; examples would be savings deposits held by consumers at various types of financial institutions or negotiable large-scale CDs and commercial paper. Similarly, small time deposits at various financial institutions would probably qualify for this sort of preaggregation. One is still left with a considerable number of other disaggregated assets when this is done, however. If a demand system is the object of the exercise, the problem becomes that of an excessive number of equations and thereby also of parameters to estimate (for the sample size available).

Instead of the *a priori* assignment of assets to monetary groups, the structure of preferences over monetary assets could be discovered by actually testing for weakly separable sub-groupings. There are problems, however, with the available tests for separability. For example, consider the most common separability pretest in this literature: Hal Varian's (1982, 1983) nonparametric (NONPAR) revealed preference procedure. This test examines the results of (what are assumed to be) actual consumer choices to see if there are any violations of consistency.[24] However, the NONPAR procedure possesses a number of undesirable features, possibly the most serious being its inherently nonstatistical nature.[25] Even so, there do exist studies in the monetary literature that employ this methodology, with a frequent result being that the traditional monetary subgroupings are not those that appear to be separable within wealth holders' choices.[26] This is

[24] A violation would occur if consumers actually choose new market baskets that make them worse off than their original choice (evaluated at the original prices).

[25] The NONPAR procedure produces a total rejection if there is a single violation; in this case, the data having been rejected, it is also impossible to test for separable groupings, even though the rejection could have been produced from purely white noise in the data. See Barnett and Seungmook Choi (1989a) for some rather pessimistic results drawn from Monte Carlo experiments on various methods used (including NONPAR) in testing for weak separability.

[26] James Swofford and Gerald Whitney (1988) *on annual data*, conclude that there exist relatively liquid sets of assets — one group being M1, other checkable deposits, and savings deposits at each institution type (taken separately) — that are separable from consumption expenditures and leisure. *On quarterly data*, they find that no interesting collection of financial assets passes both of Varian's necessary and sufficient conditions for

clearly a very preliminary observation, however.

There is a further problem, already discussed in a theoretical context, that concerns the separability of the monetary asset decision from the consumption/leisure decision. Most money-demand studies simply ignore the possible difficulties, but recent empirical tests, either utilizing the NONPAR procedure or embedding the hypothesis parametrically in a set of simultaneous equations, generally do not support this separability.[27] We should note, somewhat parenthetically, that tests that employ the wage rate in the money demand function — usually as a proxy for the 'value of time' — also could be interpreted as providing evidence of the lack of separability of the money-holding decision from the leisure (and hence consumption) decision.[28]

Moving on to the main issues of this section, the next topic concerns the relationship between the direct monetary services aggregator function and the indirect user cost aggregator function. Since the structural properties of the two are not necessarily the same, and since one generally theorizes about the direct function but estimates the indirect function, a correspondence between the two must be established. In the case at hand, the indirect utility function corresponding to the direct utility function in equation (1.2) would be weakly separable in expenditure-normalized monetary-asset user costs if there exists an indirect aggregator function $H(\mathbf{v})$ such that we can write

$$g = G(\boldsymbol{q}, w, H(\boldsymbol{v})) \tag{1.18}$$

where $\boldsymbol{q}$ and $\boldsymbol{v}$ are the expenditure-normalized price vectors for $\boldsymbol{c}$ and $\boldsymbol{x}$, respectively, and w is the expenditure-normalized wage rate.[29] The weak separability condition in equation (1.18) holds if and only if the marginal rate of substitution between any two user costs in the monetary index is

---

weak separability. This is discouraging, although the NONPAR procedure is definitely biased toward rejection (since one violation of consistency produces a rejection). Michael Belongia and James Chalfant (1989) also test for what they call "admissible" monetary groupings on quarterly U.S. data; the groupings that pass the necessary conditions are M1A (currency plus demand deposits, M1 (as currently defined), and M1+ (M1 plus interest-bearing checkable deposits currently included in M2.)

[27] The Swofford and Whitney paper just mentioned provides the nonparametric results as does Salam Fayyad (1986). For a system study that rejects the separability of consumption goods from monetary assets, also see Fayyad, who employs the Rotterdam model in his calculations.

[28] The rationale (see Edi Karni 1974; Thomas Saving 1971; or Dean Dutton and William Gramm 1973) is often that of saving time in transactions by employing money; time is then valued at the wage rate. A recent paper by Kevin Dowd (1990) continues this tradition. Note that the lack of separability between money holding and consumption is explicit in the Dutton and Gramm paper just referred to.

[29] In particular, if $(\boldsymbol{q}, w, \pi)$ are the corresponding nonnormalized prices and if y is total expenditure on $(\boldsymbol{c}, L, \boldsymbol{x})$, then the expenditure-normalized prices are $(\boldsymbol{q}, w, \pi)/y$.

independent of changes in prices outside the monetary group.[30]

1.4.1 Approximating the Monetary Services Subutility Function

In recent years a number of empirical studies have made use of the *flexible functional form* method to approximate unknown utility functions. The advantage of this method is that the corresponding demand system can approximate systems of demand equations (for liquid assets in this case) that arise from a broad class of utility functions. Flexible functional forms have the defining property that they can attain arbitrary level and both first- and second-order derivatives at a predetermined single point (see Diewert 1974); they are, in the terminology of the literature, 'locally' flexible, and they provide second-order local approximations to the desired function.

The two most commonly used flexible functional forms are the Generalized Leontief, introduced by Diewert (1971), and the translog, introduced by Laurits Christensen, Dale Jorgenson, and Lau (1975). These have been especially appealing in econometric work because they have close links to economic theory and because their strengths and weaknesses are generally understood. Below we will use the translog as an example that reveals the characteristics of this approach in modeling the demand for money.

The decision problem with which we are working is the maximization of 'second-stage' utility subject to the second-stage monetary expenditures constraint. That is, in the second stage of a two-stage maximization problem with weak separability between monetary assets and consumer goods, the consumer maximizes a direct utility function of the general form

$$f(x_{1t}, x_{2t}, \ldots, x_{nt})$$

subject to

$$\sum_{i=1}^{n} \pi_{it}^{*} x_{it} - m_t^{*} = 0$$

with $m^* = m/p^*$ being real expenditure on the services of monetary assets (determined in the first stage and hence given at this point), and where

[30]Lawrence Lau (1969, Theorem VI) shows that a homothetic direct aggregator function is weakly separable if and only if the indirect aggregator function is weakly separable in the same partition. Hence, if one wishes to test for homothetic separability of the direct aggregator function, one equivalently can test for homothetic separability of the more easily approached indirect aggregator function. This survey deals with the homothetic case, which has self-duality of separability between the direct and indirect utility function.

$\pi^*_{it} = \pi_{it}/p^*$ is the real user cost of asset i, so that,

$$\pi^*_{it} = \left(\frac{R_t - r_{it}}{1 + R_t}\right).$$

The user cost here would be calculated from the own rate of return and the return (R_t) on the benchmark asset. The latter would normally be the highest available rate in the set of monetary assets.[31]

Let the indirect utility function be

$$H(v_1, v_2, \ldots, v_n)$$

with v_i defining the expenditure-normalized user costs, as in

$$v_i = \frac{\pi_i}{m} \qquad i = 1, \ldots, n.$$

Then, by application of Roy's Theorem, we will be able to move directly from the estimates of the parameters of the indirect utility function to calculations of income and price elasticities and the elasticities of substitution among the various assets.[32]

The demand-systems approach provides the ability to impose, and for that matter to test, the set of neoclassical restrictions on individual behavior; here we are referring specifically to monotonicity and curvature restrictions.[33] In addition, the approach provides an infinite range of possible parametric functional forms, thereby affording a rich supply of alternative models for actual estimation. Indeed, in the monetary literature a number of such models have been employed in the attempt to represent consumer preferences. These alternative models transform the behavioral postulates of the theory into restrictions on parameters; they differ in the specific parameterization and approximation properties of the model.

In many studies of money demand, the restrictions of theory are *implicit* at best — as in the standard Goldfeld (1973) money-demand specification

[31] The role of the benchmark asset is to establish a nonmonetary alternative. It is acceptable for this to be a different asset in each period, since the maximization is repeated each period. In theory, any measurement of R_t could be viewed as a proxy for the unknown rate of return on human capital.

[32] Once the form of the indirect utility function is specified (and under the assumption that this function is differentiable), Roy's Theorem allows one to derive the system of ordinary demand functions by straightforward differentiation, as follows:

$$-x_i(\pi_1, \pi_2, \ldots, \pi_n, m) = \frac{\partial H}{\partial \pi_i} / \frac{\partial H}{\partial m} \quad (i = 1, \ldots, n).$$

[33] The monotonicity restriction requires that, given the estimated parameter values and given prices v_i, the values of fitted demand be nonnegative. It can easily be checked by direct computation of the values of the fitted budget shares. The curvature condition requires quasi-convexity of the indirect utility function.

— but because the connection with optimization theory is either unclear or nonexistent in such cases, we are often not in a position to test or impose those restrictions. A simultaneous-equations demand system is an effective alternative, because in this case the restrictions of theory become *explicit* in the context of a particular functional form. In addition, flexible functional forms, because they permit a wide range of interactions among the commodities being tested, are especially useful in this respect.[34]

1.4.2 An Example

Consider the popular basic translog model. The logarithm of the indirect utility function $\log h = H(\log v_1, \log v_2, \ldots, \log v_n)$ can be approximated by a function that is quadratic in the logarithms as in equation (1.19)

$$\log h = \log \alpha_0 + \sum_i \alpha_i (\log v_i) + \frac{1}{2} \sum_i \sum_j \gamma_{ij} (\log v_i)(\log v_j). \qquad (1.19)$$

In fact, the function just exhibited can be derived as a second-order Taylor-series expansion to an arbitrary indirect utility function at the point $v_i^* = 1$ $(i = 1, \ldots, n)$. The translog actually is likely to be an adequate approximation at that point, although away from the point its effectiveness decreases rapidly (see Barnett and Lee 1985).

One normally does not estimate the translog model in the form of equation (1.19). Rather, starting with the indirect utility function of

$$\log h = \log H(v_1, v_2, \ldots, v_n) \qquad (1.20)$$

and applying Roy's Theorem (see Varian 1984), the budget share for the j^{th} asset for translog tastes becomes

$$S_j = \frac{\alpha_j + \sum_i \delta_{ij} \log v_i}{\alpha + \sum_i \delta_i \log v_i} \qquad (1.21)$$

with $j = 1, \ldots, n$ where, for simplicity, $S_j = \pi_j x_j / m$, $\alpha = \sum_j \alpha_j$ and $\delta_i = \sum_j \delta_{ij}$. The equation system given in (1.21) is what is typically estimated.

There is, however, a significant weakness to the translog model just described, in that as a *locally* flexible functional form it is capable of an effective approximation of an arbitrary function only at or near a single point

[34] But note that the application of a separability restriction in this context will generally alter the flexibility characteristics of a flexible functional form (toward less flexibility); see Blackorby, Primont, and Russell (1977).

(v^*). This has inspired research on approximating the unknown monetary services aggregator function based on the use of flexible functional forms possessing *global* properties (in the limit implying an approximation at *all* points). Three such forms in use with U.S. monetary data are the Fourier, the Minflex Laurent generalized Leontief, and the Minflex Laurent translog. The Fourier form uses the Fourier series expansion as the approximating mechanism (see Gallant 1981, and footnote 40), while the Minflex Laurent models make use of the Laurent series expansion — a generalization of the Taylor series expansion — as the approximating mechanism (see Barnett 1985; Barnett and Lee 1985; and Barnett, Lee and Wolfe 1985, 1987).[35] We shall discuss their brief empirical record in the next section.

1.5 Empirical Dimensions

There are two major observations in what has gone before in this survey. These are (1) that ideal index numbers represent a theoretically attractive alternative to fixed weight or simple sum aggregation and (2) that a systems approach to studying the demand for money is consistent with the same theory that generates the ideal index numbers, and clearly provides a promising alternative strategy for locating the apparently elusive demand for money. These topics will be the theme of the following discussion.

1.5.1 Empirical Comparisons of Index Numbers

Divisia indices will be more effective — and simple-sum (and fixed weight) indices less, ceteris paribus — if user costs of the different component monetary assets are unequal and fluctuate to any degree over the sample period, whether one needs the aggregate for direct policy purposes or as an input into a money demand study. In Figures 1.3a and 1.3b, we graph the behavior of the user costs of some of the components of the monetary aggregates M1, M2, M3, as defined in Table 1.1. There are two collections, one with user costs picked from three simple-sum categories, and one with three user costs taken from among the components of M1.[36]

In Figure 1.3a, three series are graphed, one from M1, one from the additional set of assets that makes up M2, and one, similarly, from M3.

[35] An especially promising global approximation not yet applied to monetary data is the AIM ("asymptotically ideal model"), generated from the Müntz-Szatz series expansion (Barnett and Jonas 1983; Barnett and Yue 1988; and Barnett, Geweke, and Wolfe 1991a).

[36] These user costs were obtained from Gerald Whitney of the University of New Orleans; their method of construction is described by Swofford and Whitney (1987, 1988) but is, in any case, carried out by the method of calculation recommended in this survey.

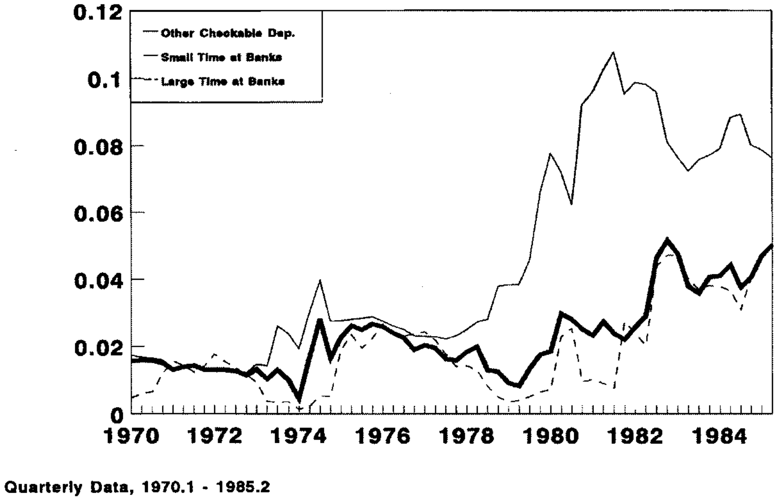

Figure 1.3a: Sample user costs, items from M1, M2, M3.

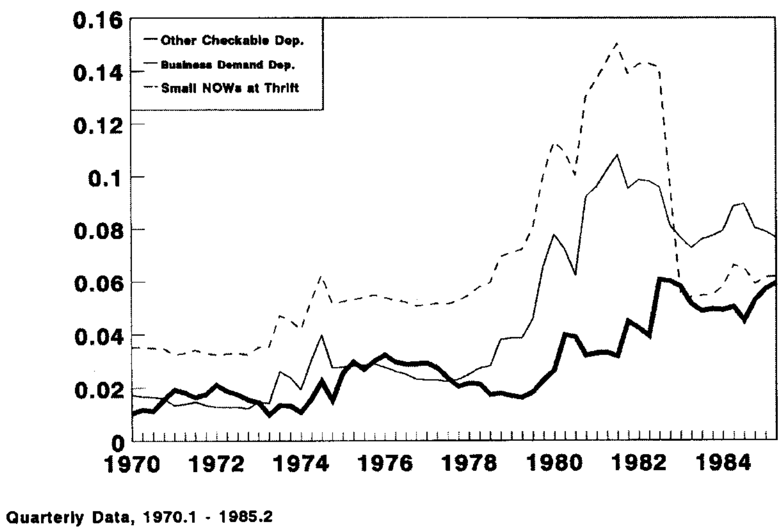

Figure 1.3b: Sample user costs, items from M1.

They are not very highly correlated (a significant fact for the construction of M3) and, what is particularly noticeable, they often go in different directions during periods of general economic disturbance (1970, 1973-5, 1977-85). Similar results hold for the three items selected from the six that make up M1 in the official aggregates.[37]

Because the user costs (as estimated) are neither equal to each other nor constant, an interesting question is opened up about the relative performance of the monetary aggregates over this period. The safest generalization at which we could arrive at is that the different aggregates behave differently (as we illustrated in Figure 1.2) and, when compared for particular money demand models, often perform differently there, too. For example, the Divisia numbers usually provide a better fit than the simple-sum measures, but the nature of the differences among the aggregates appears to depend on the particular specification at hand and upon the choice of components over which the aggregation occurs. Under these circumstances it would be premature to assert that one component grouping is best for all purposes and, in fact, that simply may not be the case.

For example, Barnett, Edward Offenbacher, and Spindt (1984) provide a comparison in the context of a Granger-causality test of money on income (and prices) — where simple sums sometimes show up well — while Serletis (1988a) employs the Akaike Information Criterion (as defined in H. Akaike 1969b) to effect statistical comparisons. In another study, Belongia and Chalfant test the St. Louis Equation and find evidence in favor of Divisia M1A; this measure is currency plus demand deposits. Cagan (1982) finds the velocity of money series more stable when Divisia M1 is employed, while Douglas Fisher and Serletis (1989) find that Divisia measures built from short-term assets work better than those constructed from longer-term assets in a framework that attempts to explain recent velocity behavior in terms of monetary variability (see Friedman 1983). Finally, Lawrence Christiano (1986), attempting to explain the alleged structural shift in the U.S. monetary data in 1979, finds that a differenced Divisia measure does just that.

---

[37]The correlation matrix for the level figures of the three entities in Figure 1.3a is:

	OCD	STDCB	LTDCB
Other Checkable Deposits	1.00		
Small Time Deposits at Banks	.71	1.00	
Large Time Deposits at Banks	.50	.90	1.00

The matrix for the items in Figure 3b is:

	BUSDD	OCD	SNOWT
Business Demand Deposits	1.00		
Other Checkable Deposits	.78	1.00	
Small NOW Accounts at Thrifts	.44	.82	100

1.5.2 Empirical Results for the Demand System Approach

The second major issue raised above concerns the advantages one might gain by employing the demand-systems approach to the study of money demand (and monetary interaction). The fluctuations in the user costs just referred to provide one possible reason for taking this route, since the same economic theory that recommends using the Divisia technique also supports the use of the demand systems approach coupled with a flexible functional form. Before beginning, however, the reader should be warned that the best of the models put a lot of strain on the data in terms of their need for large sample size (generally, the more flexible models have more parameters to estimate).[38]

The demand-system approach produces interest-rate and income elasticities — as well as the elasticities of substitution among monetary assets. The underlying question for the income elasticities is whether they are less than 1.0, particularly for the narrowest measures. On the whole, across this entire literature, the answer is 'yes.' With respect to the elasticities of substitution, one of the most curious — *and consistent* — results in all of monetary economics is the evidence in such studies of very low substitution or even (sometimes) complementarity among the liquid financial assets; this is a general result that occurs in all but the very earliest studies.[39] These results are robust across definitions of the money stock and across *flexible* functional forms. We should note that the reason for referring to this as a 'curious result' is that there is a traditional view in the monetary literature that most of these assets are very close substitutes, as is, in fact, necessary for simple-sum aggregation. The policy importance of this result, should it continue to stand up, can hardly be exaggerated in view of the dependence of existing policies on aggregation procedures that require very high (really infinite) elasticities of substitution.

A second equally important finding concerns the temporal behavior of

[38]Most seriously, when the simple-sum aggregates are included, tests for model failure (e.g., symmetry, monotonicity, and quasi-convexity) generally show such failures (sometimes even quite a few failures); see Donovan (1978), Nabil Ewis and Douglas Fisher (1984, 1985), and Douglas Fisher (1989, 1992).

[39]The early study by Karuppan Chetty (1969), which employed a constant elasticity of substitution format, found relatively high elasticities of substitution. Since then, Donovan (1978), Ewis and Douglas Fisher (1984, 1985), Fayyad (1986), Serletis and Leslie Robb (1986), Serletis (1988b), and Douglas Fisher (1992), have found the lower elasticities. These studies have in common their employment of one or another of the popular flexible functional forms.

Another controversial finding is that in several studies it is suggested that currency might be a closer substitute for time deposits than it is for demand deposits (Offenbacher 1979, Ewis and Douglas Fisher 1984).

these same elasticities. We noted in Figure 1.3 that user costs appear to have fluctuated considerably in recent years; the next question concerns the behavior of the various elasticities over the same period. One can certainly generate a time series of elasticities of substitution for models such as the translog, but a far more attractive approach is to use a model like Gallant's (1981) Fourier flexible form because it provides a global approximation (at each data point) rather than a local one.[40] A few of these results have been published; we reproduce one set of these (Douglas Fisher 1989) for the same U.S. data that figured in the calculations in Figure 1.3. At stake is the reputation of single equation money demand studies that (implicitly) rely on linearity of the demand equation in its parameters. The results, which appear in Figure 1.4, are drawn from a four-equation system (three financial assets and one set of consumption expenditures) for quarterly U.S. data.[41] S12 refers to the elasticity of substitution between cash assets and savings deposits (and money market accounts), S13 refers to the elasticity of substitution between cash assets and small time deposits, while Y_i is the income elasticity of the i^{th} category.[42]

---

[40]The Fourier model that is equivalent to Equation (1.19) for the translog is

$$h_k(\boldsymbol{v}\}) = u_0 + \boldsymbol{b}'\boldsymbol{v} + \frac{1}{2}\boldsymbol{v}'\boldsymbol{C}\boldsymbol{v} + \sum_{\alpha=1}^{A}\left(u_{0\alpha} + 2\sum_{j=1}^{J}\left[u_{j\alpha}\cos(j\boldsymbol{k}'_\alpha\boldsymbol{v}) - w_{j\alpha}\sin(j\boldsymbol{v}\boldsymbol{k}'_\alpha\boldsymbol{v})\right]\right)$$

in which

$$\boldsymbol{C} = -\sum_{\alpha=1}^{A} u'_{0\alpha}\boldsymbol{k}_\alpha\boldsymbol{k}'_a$$

This is a set of equations (which would be estimated in budget share form after application of Roy's identity) in which the parameters are the $b_i\}$, u_{ij}, and w_{ji}. The researcher picks the degree of the approximation (by picking j) and the particular form and number of the so-called multi-indices, (the $\boldsymbol{k}$ vectors). The latter are generally taken as 0, 1 vectors of length $n-1$ (n is the number of assets in the problem) whose purpose is to form simple indices of the normalized user costs. These decisions are made on goodness-of-fit criteria. See Gallant (1981) or Douglas Fisher (1989).

[41]The financial categores are:

Group 1: currency, consumer demand deposits, other checkable deposits
Group 2: small NOW accounts, money market deposit accounts, and savings deposits
Group 3: small time deposits

The fourth category, total consumption expenditures, was included in the system because of the failure to establish separability between financial asset holding and consumption for these data. See the discussion in Douglas Fisher (1989). There is another, similar, study by Douglas Fisher (1992).

[42]The model estimates a set of parameters employing the iterative seemingly unrelated equation approach. The elasticities and their standard errors are generated from these estimated parameters and the data at each point.

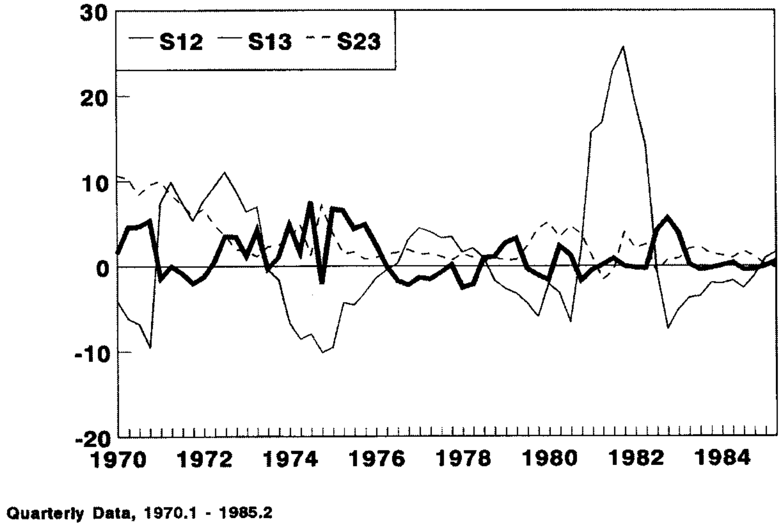

Figure 1.4a: Substitution elasticities.

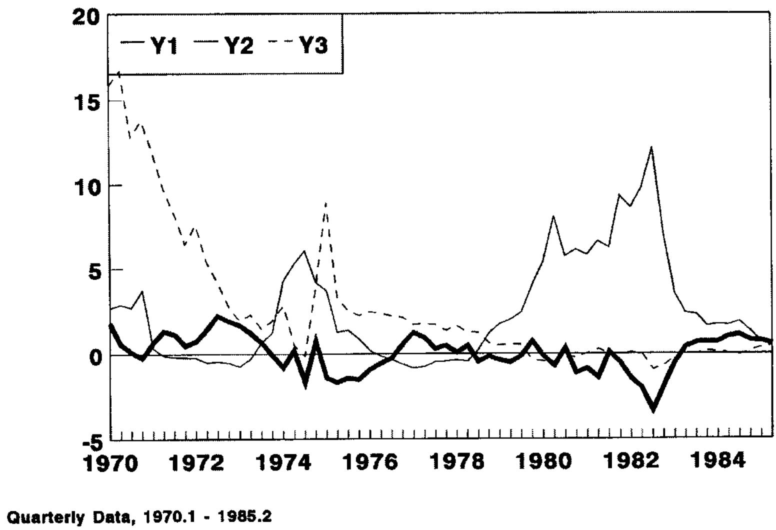

Figure 1.4b: Income elasticities.

It is readily apparent that none of the elasticities are constant for the comparisons made, and that, in particular, the elasticities of substitution change considerably over approximately the same subperiods of the data as did the user costs. This is no coincidence, of course, since the elasticities often change in response to changes in relative prices for plausible *fixed* tastes. What it does provide — and the differences were often statistically significant in the study from which they were taken — is a possible explanation of why approaches involving simple-sum aggregation and single-equation linear money demand functions might not perform particularly well, as indeed they often have not.[43]

We should note that most of the early studies of demand systems — such as that in equation (1.19) — have tended to be cast in static terms, although the issue of dynamic adjustment has been addressed frequently in the traditional money-demand literature.[44] Recently, though, attention has been focused on the development of dynamic generalizations of the traditional static systems models in such a way as to enable tests of the model and its theoretical restrictions and simplifications. Gordon Anderson and Richard Blundell (1982), for instance, develop an unrestricted dynamic formulation in order to accommodate short-run disequilibrium; this is effected through the use of lagged endogenous and exogenous regressors. In the same spirit, Serletis (1991a) applies the Anderson and Blundell approach to various demand-systems models and demonstrates that the dynamic generalizations of the traditional static models (e.g., the dynamic translog) provide more convincing overall fits as well as empirically credible estimates of the various elasticities.

1.6 Extensions

The purpose of this chapter is to survey and assess the mainstream of the recent literature on neoclassical system-wide modeling and testing in monetary economics. Because of the complexity of that topic — and the limitations of space — the discussion to this point has been limited to the most central and fundamental of the research in this area. Extensions including such important topics as uncertainty, rational expectations, nonhomotheticity, intertemporal optimization, and the possibility of extend-

[43]For similar results with other series expansion models, see Barnett (1983a) and Yue (1991). Briefly, the evidence of instability of money demand from the old linear models results from the fact that in general the slopes of local linear approximations to nonlinear functions vary over time as the tangency moves along the nonlinear function.

[44]For example, there is a "real adjustment" specification in Gregory Chow (1966), a "price adjustment" specification in Robert Gordon (1984), and the "nominal adjustment" specification in Goldfeld (1976). For a recent approach, see Choi and Sosin (1992).

ing the framework to study the supply of money function, have not been presented. Nevertheless, the reader should not thereby conclude that the literature has not in fact been extended in those directions.

In Section 1.3, we stated the consumer's decision problem to be the maximization of contemporaneous single-period utility, in equation (1.2), subject to the current period expenditures constraint, equation (1.3). In fact, it can be shown that this decision is consistent with rational *intertemporal optimization* by the consumer if the consumer's intertemporal utility function is intertemporally blockwise weakly separable, so that $U(\mathbf{c}, L, \mathbf{x})$ is a weakly separable block within the intertemporal utility function. The current period decision that we use in Section 1.3 then can be shown to be the second-stage decision in the two-stage decision of maximizing intertemporal utility subject to a sequence of single-period constraints (Barnett 1980a, 1987). Hence, all of the results surveyed above are consistent with rational intertemporal optimization by the consumer if the relevant weak separability assumption holds. As nearly all intertemporal decision models currently in use in economics are in fact intertemporally *strongly* separable, the assumption of intertemporal *weak* separability does not seem excessive.

The first stage of the two-stage decomposition of the intertemporal decision allocates total wealth over periods. In the second stage, the total expenditure allocated in the first stage to the current period is then allocated over current period consumption of individual goods, asset services, and leisure. The two-stage decomposition can be produced under intertemporal weak separability of tastes, if there is perfect certainty or risk aversion. However, the two-stage solution, producing the current period conditional second-stage decision, is not possible under risk aversion. With risk aversion, the intertemporal decision becomes a problem in stochastic optimal control, which can be solved by dynamic programming methods.

This implies that another productive direction for future research in this area is likely to be the search for nonparametric statistical index numbers that will track the exact *rational expectations aggregator function* in the risk-averse case as effectively as the Divisia can track the exact aggregator function in the risk-neutral case.[45] The existing literature on index number

[45] In the work referred to in earlier sections, there is the assumption of risk neutrality, so that the decision can be stated in certainty-equivalent form, with random variables replaced by their expectations. A means for extending this literature to include risk aversion and rational expectations has been proposed by Poterba and Rotemberg (1987). Subsequent work by Barnett, Hinich, and Yue (1991) and Barnett, Hinich, and Yue (2000) has produced a solution to the intertemporal stochastic optimal control problem characterizing a rational consumer, when current period consumption of monetary services is weakly separable within the consumer's intertemporal expected utility function. Those papers also contain an empirical implementation that produces the "exact" rational expectations monetary aggregates for M1. This does not appear to differ materially

theory, which has never before been extended to the risk-averse case, provides no guidance here. This is in striking contrast to the risk-neutral case, in which the existing literature on index numbers and aggregation theory has provided all of the tools that have been found necessary.

In the discussion so far, we have assumed that the aggregator function is linearly homogeneous. We also have assumed that the aggregator function is a utility function. There is a paradox in this. On the one hand, it is clear that the aggregator function does indeed have to be linearly homogeneous. Otherwise the aggregate will grow at a different rate than the components in the case of identical growth rates for all components. That, of course, would make no sense. On the other hand, linear homogeneity of a utility function is a strong assumption empirically. The solution to the paradox is that the aggregator function is a utility function only if the utility function is linearly homogeneous. If the utility function over component quantities is not linearly homogeneous, then it is known from aggregation theory that the aggregator function is the distance function.

The quantity aggregate defined by the distance function is called the Malmquist index.[46] The distance function is always linearly homogeneous in $\boldsymbol{x}$, regardless of whether or not the utility function, U, is itself linearly homogeneous. We concentrate above on the special case of linear homogeneous utility in this chapter for expositional reasons, but the generalization to the nonhomothetic utility case presents no problems at all. If we seek to estimate the aggregator function, then the literature described in the earlier sections is directly applicable to estimating the parameters of the utility function. Once the estimation is complete, the parameter estimates are substituted into the distance function rather than into the utility function to produce the estimated aggregator function. The situation is even simpler if we seek a nonparametric statistical index number. Diewert (1976, pp. 123-4) has proved that under an appropriate method for selecting the base level of utility, U_0, the Divisia index provides a superlative approximation to the distance function, just as it did to the utility function in the special case of linearly homogeneous utility. Hence, the second-order approximation property of the Divisia index to the exact aggregator holds true, regardless of whether or not utility is homothetic (Barnett 1987).

Since we see that homotheticity of utility is not needed at all — and was used in our earlier discussion only to simplify the presentation — we are left

---

from Divisia M1 (but does differ from the simple-sum version of M1.

[46]For the definition of the distance function, see Barnett (1987, Equations (7.1) and (7.2), pp. 146-47). For the Malmquist index, see Barnett (1987, Equation (7.7), p. 148). The distance function $d(u_0,\boldsymbol{x})$ relative to base utility level u_0 can be acquired by solving the equation $f(\boldsymbol{x}/d(u_0,\boldsymbol{x})) = u_0$ for $d(u_0,\boldsymbol{x})$ where f is the utility function. Hence, the distance function measures the amount by which the monetary asset vector $\boldsymbol{x}$must be deflated to reduce the utility vector to its base level u_0.

with the weak separability condition as the key indispensable assumption. Here it is important to observe that weak separability is not an additional assumption imposed to permit use of one particular approach to modeling or aggregation, but rather is the fundamental *existence* condition, without which aggregates and sectors do not exist. It should be observed that we do not require that weak separability hold for some particular prior clustering, but rather that there exists at least one clustering of goods or assets that satisfies weak separability. Empirically, when one is free in that way to consider all possible clusterings, weak separability is indeed a weak assumption.[47]

Another important extension that is just beginning to appear in this literature is the application of the available theorems on aggregation over economic agents. It should be observed that the theory of aggregation over economic agents is independent of the theory of aggregation over goods, and hence the theory discussed above on aggregation over goods for one economic agent remains valid for any means of aggregation over economic agents. However, the existing theory on aggregation over economic agents is much more complicated than that for aggregation over goods. While a unique solution exists to the problem of aggregation over goods, and the necessary and sufficient conditions are known, the same cannot be said for aggregation over economic agents. The solution to the latter problem is dependent upon the modeler's views regarding the importance of distribution effects.

In particular, an array of solutions exists to the problem of aggregation over economic agents, depending upon the model's dependence on distribution effects. At one extreme is Gorman's (1953) solution by means of the 'representative agent' who can be proved to exist if all Engel curves are linear and are parallel across economic agents. At the other extreme is Pareto's perfectly general stratification approach, which integrates utility functions over the distribution functions of all variables that can produce distribution effects; these would be such as the distribution of income or wealth and the distribution of demographic characteristics of the population. Between these two extremes are such approaches as John Muellbauer's, which introduces dependency upon the second moments as well as on the first moments of distributions but does so in a manner that preserves the existence of a representative consumer.[48] Somewhat closer to the general Pareto

[47] If the procedure is reversed, and a clustering is chosen by some other means prior to separability testing — and one clustering then is subjected to a test for weak separability — the assumption of weak separability becomes a strong one that is not likely to be satisfied empirically.

[48] In particular, Muellbauer preserves the representative consumer by retaining the dependence of the decision upon only one income index, although that income index, in turn, depends jointly upon the mean and variance of the income distribution.

approach is Barnett's (1981b, pp. 58-68) stochastic convergence method, which requires fewer assumptions than Muellbauer's method while, at the same time, not preserving the existence of the representative consumer.[49]

The importance of distribution effects is central to the choice between methods of aggregating over economic agents. The Gorman method assumes away all distribution effects and leaves only dependence upon means (i.e., per capita variables). The further away one moves from Gorman's assumptions, the further one must move along the route to the Pareto method, which requires estimation of all of the moments of the distribution functions.[50] At present, the empirical research on this subject seeks to test for the depth of dependence upon distribution effects. Barnett and Serletis (1990) have done so with monetary data by explicitly introducing Divisia second moments into models and testing for their statistical significance. The importance of the induced distribution effects was found to be low.[51]

In short, the empirical evidence does not yet suggest the need to move away from Gorman's representative consumer approach towards any of the generalizations. However, should the need arise, these generalizations do exist for any applications in which complex distribution effects might be suspected. There are surveys of many of these approaches in Barnett (1981b, pp. 306–07 and 1987, pp. 153–54) and Barnett and Serletis (1990).

Finally, we consider some results that come from the application of the neoclassical theory of the firm to the financial sector. For the supply of money problem there exists system-wide modeling that employs aggregation and index number theory that is analogous to that for the demand side. An especially interesting development from this literature is the proof that the exact supply-side monetary aggregate may not equal the exact demand-side monetary aggregate, even if all component markets are cleared. This situation is produced by the existence of a regulatory wedge created by the nonpayment of interest on the reserves required of financial intermediaries. The wedge is reflected in different user costs on each side of the market. The relevant theory is available in Diana Hancock (1985, 1991) and Barnett (1987), and the statistical significance of the wedge is investigated empirically by Barnett, Hinich, and Weber (1986). They find that the size of the

[49] In this case some of the properties of the Slutsky equation are retained after aggregation over consumers.

[50] Another highly general approach, requiring extensive data availability, is the Henri Theil (1967) and Barnett (1987, p. 154) approach to Divisia aggregation over economic agents. By this method, a single Divisia index can aggregate jointly over goods and economic agents, although detailed data on the distribution of asset holdings over economic agents is required.

[51] Similarly, Ernst Berndt, W. Erwin Diewert, and Masako Darrough (1977) use the Pareto appproach by integrating over the entire income distribution with Canadian consumption data and similarly find little significance to the distribution effects.

wedge is astonishingly large when measured in terms of the dollar value of the implicit tax on financial firms but nevertheless produces only insignificant divergence between the demand-side and supply-side exact monetary aggregates. In that study, the wedge was found to have potentially important effects in the dynamics of the monetary transmission mechanism only at very high frequency (i.e., in the very short run).

It is perhaps worth observing that analogous wedges are produced by differences in explicit marginal tax rates among demanders and suppliers of monetary services through differences in capital gains taxation, in local and federal income taxation, and in sales and corporate income taxation. The resulting divergence in user costs between the demand and supply side of the market can create the analogous paradox to the one produced by the implicit taxation of financial firms through the nonpayment of interest on required reserves. The empirical importance of these explicit wedges for the monetary transmission mechanism has not yet been investigated systematically.

1.7 Conclusions

In the history of economic thought, economic paradigms rise and fall based upon how well they actually work in the eyes of the public. The acid test usually is the connection between the paradigm and the performance of economies that adopt the paradigm for policy purposes. The approach that we survey in this chapter has been used in research in many countries, both in academia and in central banks [52] While these data along with some of the modeling principles described in this chapter are available and are being used internally within some central banks, the methods we have just surveyed have not yet, to our knowledge, been adopted publicly as formally announced targeting methods by any country's central bank or government. Hence, the ultimate acid test cannot yet be applied.

A way to visualize how this new work affects the traditional money-demand literature is to think — as we have occasionally done in this survey — in terms of the well-known 'missing money' puzzle (Goldfeld 1976). What is being suggested here is that a good part of the problem may be in the way money is measured — both in the choice of component groupings and in the method of aggregating over those groups — and in the way that the demand model's capabilities relate to the generally nonlinear optimizing behavior of economic agents. Unlike conventional linear money demand

[52] For example, Divisia monetary aggregates have been produced for Britain (Roy Batchelor 1989, Leigh Drake 1992, and Belongia and Alec Chrystal 1991), Japan (Kazuhiko Ishida 1984), Holland (Martin Fase 1985), Canada (Jon Cockerline and John Murray 1981b), Australia (Tran Van Hoa 1985), and Switzerland (Yue and Fluri 1991).

equations, a system of demand equations derived from a flexible functional form with data produced from the Divisia index can be expected to capture those movements in money holding that are due to changes in the relative prices among assets. In the increasingly unregulated and volatile financial markets that have followed the collapse of the Bretton-Woods system, this would seem to be useful. In addition, the approach offers a solution to the long-running money-demand puzzle: the observed variability of elasticities is actually consistent with the variability occurring naturally along stable, nonlinear, integrable demand systems. Linear approximations to nonlinear functions are useful only locally.

In sum, the successful use in recent years of the simple representative consumer paradigm in monetary economics has opened the door to the succeeding introduction into monetary economics of the entire microfoundations, aggregation theory, and microeconometrics literatures. The moral of the story is that the nonlinearity produced by economic theory is important.[53] We have surveyed a growing literature on the importance of the use of nonlinear economic theory in modeling the demand for money. We also have surveyed the recent literature on the use of aggregation theoretic nonlinear quantity aggregation in producing monetary data. Finally, we agree with Irving Fisher (1922) that the simple-sum index should be abandoned both as a source of research data and as an intermediate target or indicator for monetary policy.

[53] In fact there recently has been an explosion of interest in the dynamic implications of nonlinearity in economics, and the relationship with chaotic dynamics. It is possibly not surprising that the only successful detection of chaos with economic data has been with the Divisia monetary aggregates. See Barnett and Ping Chen (1988).