Milton Friedman's Monetary Framework

A Debate with His Critics

Edited by
Robert J. Gordon

Milton Friedman
Karl Brunner and Allan H. Meltzer
James Tobin
Paul Davidson
Don Patinkin

The University of Chicago Press
Chicago and London

To Miguel Sidrauski*

We are all of us teachers, and we all know that the greatest reward of our career is the rare student who has the touch of genius, who absorbs what we have to teach, and repays by teaching us—the intellectual child in whose reflected glory we bask. Miguel was such a one. Of Polish descent but reared and schooled in Argentina, Miguel came to the University of Chicago as a graduate student only a few years ago. He happened to take one of my courses in one of his first quarters at Chicago. It was a large class and he did not speak up much, so I did not single him out until the first examination. That opened my eyes, and thenceforth he was a marked man. The same experience was repeated in course after course, and all of us quickly came to know that here was a student who was destined not only to use economics but to be of that small band who make economics.

Miguel's intellectual drive, his urge to get things straight, to get to the heart of the matter, shone through in every paper he wrote, in every conversation with him. So did his character: straightforward, confident but not arrogant, warm, generous, and sympathetic. He was a truly fine human being, who quickly became a leader among his fellow students.

Miguel became interested in monetary theory. He then also came under the influence of my colleague, Hirofumi Uzawa, who has had such a far-reaching influence on so many young men. Miguel had not only a fine mathematical mind, dissatisfied with any analysis that lacked complete rigor, but like Hiro he had that clarity of mind that picks out the key elements of a problem, strips away all irrelevant complications, and produces the kind of simplicity that is the height of sophistication. Like Hiro also, he had the real instincts of an economist, so that he was dissatisfied with any purely mathematical analysis unless the results had a meaningful economic interpretation. His dissertation reveals these qualities. It is a sophisticated yet simple analysis of a complex economic problem, mathematically rigorous yet always economically motivated and interpreted. It has already had considerable influence and will have more.

Two years ago, just as he was finishing his degree, Miguel's lovely fiancée came to the United States. After their marriage, they moved to Cambridge, where Miguel started his career as an assistant professor of

*This statement was made in honor of Miguel Sidrauski at the American Bankers Association-University Professors' Conference, Ditchley House, England, which was held September 10-13, 1968. It is reprinted with permission of the *Journal of Money, Credit, and Banking* 1 (May 1969): 129-30.

economics at Massachusetts Institute of Technology. We wanted, in our own interests, to keep him at Chicago, and offered him the opportunity to stay. But we also advised him that our interests were not necessarily his and that it would very likely be better for him and more stimulating to move into a new and different intellectual environment. I was pleased when he visited us a year later to find that our advice to him had been good—that he had found stimulus and challenge, difference of opinion without intolerance, that he felt that both he and his colleagues had benefited from the intellectual diversity. And I was equally pleased to hear from his MIT colleagues how high an opinion they had come to form of him, both of his intellect and his character.

The death of any young man is a personal tragedy for his family and his friends. The death of this young man is a grievous loss to our profession and to the world. Here was a man who would have pushed out the frontiers of our subject, would have changed and added to economic analysis, would have enlightened and informed generations of students—struck down at the very beginning of his career, full of promise but as yet almost bereft of fulfillment.

MILTON FRIEDMAN
University of Chicago

Introduction

The publication in 1963 of *A Monetary History of the United States,* by Milton Friedman and Anna Schwartz (1963*b*), represents a landmark in the development of monetary economics. Through its demonstration of a consistent relationship between fluctuations in the money supply and national income over the period 1867–1960, and through its proposition that a one-third decline in the money supply between 1929 and 1933 was primarily responsible for the severity and duration of the Great Depression, it was probably the most important single contribution to the revival of interest in monetary economics which has occurred during the past fifteen years. Several reviewers of the book, however, criticized the absence of an explicitly stated theory of the role of money in income determination capable of generating the propositions supported by the extensive empirical investigation.[1] While Friedman had previously presented his general theoretical approach in "The Quantity Theory of Money—a Restatement," there was widespread interest in a more formal and complete statement in which Friedman would analyze the channels by which money influences income and spending, explore the dynamics of the money-income relationship, and compare his quantity theory approach with the neo-Keynesian doctrines which at that time dominated the teaching of macroeconomic theory in American universities.[2]

Friedman responded with his "Theoretical Framework for Monetary Analysis," which was first published in the *Journal of Political Economy* in early 1970 (Friedman 1970*a*). One year later he added another, shorter supplementary paper, "A Monetary Theory of Nominal Income," which provided an additional model of income determination (Friedman 1971). Both papers were then combined and republished in one volume by the National Bureau of Economic Research.[3] The NBER version of the two Friedman papers is the first item reprinted in this book.[4]

[1]See Culbertson (1964), Meltzer (1965), and Tobin (1965*a*). It should be noted that several reviewers also made substantive critical comments on the basic empirical propositions summarized in the previous sentence.

[2]See Friedman (1956).

[3]Milton Friedman, *A Theoretical Framework for Monetary Analysis,* Occasional Paper 112 (New York: National Bureau of Economic Research, 1971).

[4]Most of pp. 31–48 in the NBER version is taken from "A Monetary Theory of Nominal Income" (Friedman 1971), whereas the remainder of that version is taken from the 1970 version of "A Theoretical Framework" (Friedman 1970*a*), with the exception of four new paragraphs added to the NBER version (from the paragraph beginning "Another, more subtle, difference . . ." on p. 27 to the end of sec. 5 at the top of p. 29). The NBER version

The present version of "A Theoretical Framework" begins (pp. 1-15) with a comparison of four versions of the quantity theory, the Irving Fisher transactions version, the income version, the Cambridge cash-balances approach, and the Friedman restatement. Next (pp. 15-31) the basic elements in Keynes's *General Theory* are summarized, and the analytical approach of Keynes's followers (the "Keynesians") is contrasted with that of Friedman and his associates, with major emphasis placed on the tendency of the Keynesians to regard the price level as fixed and to concentrate on a narrower range of assets in their analysis of the channels by which money influences income.[5] The major elements in Keynesian analysis and the quantity theory are summarized in a "simple common model," involving seven unknown variables but only six equations, with the two theoretical approaches characterized by the different seventh or "missing" equation which must be added to make the "simple common model" determinate. The simple Keynesian theory fixes the price level, the simple quantity theory fixes the level of real output, and a third approach suggested by Friedman, "the simple monetary theory of nominal income" (pp. 34-48), fixes the difference between the anticipated real interest rate and the real secular rate of growth of output. A final section on dynamic adjustment analyzes the effects of a change in the rate of growth of the money supply on the rate of growth of nominal income, and on the division of the latter between changes in the rate of growth of real output and in the rate of growth of the price level.

Because Friedman's statement was important and controversial, both as a commentary on the history of economic thought and as a theoretical contribution to be considered in its own right, I began in late 1970 to solicit critical reviews from a number of noted monetary theorists. Fortunately, almost all of my invitations were accepted, and four papers, by Karl Brunner and Allan Meltzer, James Tobin, Paul Davidson, and Don Patinkin, were finished in summer 1971 and submitted to Friedman, whose reply was written in late 1971. One round of revisions followed during the winter of 1971-72 in which all authors revised their papers to minimize misunderstandings and maximize attention to substantive issues. The symposium appeared in print in the September/October 1972 issue of the *Journal of Political Economy*.[6]

---

will also appear in adapted form as chap. 2 of a National Bureau of Economic Research monograph by Milton Friedman and Anna J. Schwartz, *Monetary Trends in the U.S. and the U.K.*, which is near completion.

[5]It should be noted that the emphasis on the narrower range of assets was added in the NBER version and was not present in the initial *JPE* paper.

[6]The comments of the critics appeared in the *JPE* symposium in the order in which Friedman chose to respond to them. The listing of authors' names on the title page of this volume reflects this order.

My original invitation placed no restriction on the aspects of Friedman's papers to be considered by each reviewer, and so the critics' papers overlap to some degree and cover a variety of topics. Brunner and Meltzer concentrate on elements which they feel are important but which Friedman omits from his "simple common model," particularly the effects of fiscal policy, interest-bearing government debt, the distinction between money and bank credit, and the distinction between nominal and real interest rates.[7] In contrast to Friedman's charcterization of a Keynesian model with fixed prices and a quantity theory model with fixed output, Tobin argues that the crucial issue is not the fixity or flexibility of prices but, rather, the assumed insensitivity of monetary velocity to interest rates. Tobin also criticizes Friedman's third approach to the problem of the missing equation, the "monetary theory of nominal income." Davidson emphasizes several factors, particularly the role of uncertainty, that he believes are basic elements of Keynes's approach but that he claims are missing from Friedman's characterization of Keynesian economics. Patinkin devotes most of his attention to issues in doctrinal history and provides textual evidence to show that Friedman incorrectly interpreted the nature of both the quantity theory and Keynesian economics, thus continuing his theme in an earlier article that Friedman's theoretical framework is more in the tradition of Keynes than in the tradition of the quantity theory.

All of the major points raised by the critics are discussed in Friedman's reply, which clarifies and expands on his original themes. Some interesting new material is included, particularly his contrast (in his reply to Patinkin) of the different views held in the early 1930s by Chicago and "London-Austrian" economists regarding the causes of the Great Depression. Also of interest is his analysis of the elements in Keynes's *General Theory* that are not uniquely "Keynesian," that is, introduced first in the *General Theory*, but, rather, that were developed by Keynes and others in earlier "classical" writings.

The purpose of the present volume is simply to combine in one location Friedman's NBER monograph, the comments of the critics, and Friedman's reply, for the convenience of scholars and students. Although several of the critics requested the right to present rebuttals to Friedman's reply for this volume, these requests were denied on the twin grounds that

---

[7]They recently published a paper presenting an alternative model in which these items are incorporated (Brunner and Meltzer 1972).

The University of Chicago Press, Chicago 60637
The University of Chicago Press, Ltd., London
 Published 1974. Second Impression 1977. Third Impression 1980
Printed in the United States of America
Library of Congress Catalog Card Number: 73-92599
International Standard Book Number: 0-226-26407-6 clothbound
0-226-26408-4 paperbound

Contents

(1) the etiquette of academic journals reserves the final word to the original author and (2) Friedman had already devoted a great deal of time to revising the first draft of his reply to minimize misunderstandings and problems of communication with the critics. Thus, the critics have not been given the chance to evaluate the sections of Friedman's reply that in the previous paragraph I refer to as "interesting new material," and at least one critic has indicated to me his doubts as to whether the material is either new, valid, or relevant.

Since the critics were originally asked to respond to the two separate *JPE* articles before the NBER version was completed, in the present volume all of the critics have made minor changes in their comments to take account of the additional passages on pages 27-29 (see n. 4 above) that were added by Friedman in the process of combining the two separate articles into the NBER volume. The changes are in the form of additional footnotes in the comments by Davidson and Patinkin and short postscripts by Tobin and Davidson. Other than these minor changes, the debate remains as originally published, and readers are invited to reach their own conclusions on the many fascinating issues discussed between these covers.[8]

ROBERT J. GORDON, Coeditor
Journal of Political Economy, 1970-73

Evanston, Illinois
February 1974

[8]Other minor changes are the combined set of references at the end of the volume, the combined index, and page number references to reflect the NBER version instead of, as originally, the two *JPE* papers.

A Theoretical Framework for Monetary Analysis

Milton Friedman

Every empirical study rests on a theoretical framework, on a set of tentative hypotheses that the evidence is designed to test or to adumbrate. It may help the reader of the series of monographs on money that Anna J. Schwartz and I have been writing to set out explicitly the general theoretical framework that underlies them.[1]

That framework is the quantity theory of money—a theory that has taken many different forms and traces back to the very beginning of systematic thinking about economic matters. It has probably been "tested" with quantitative data more extensively than any other set of propositions in formal economics—unless it be the negatively sloping demand curve. Nonetheless, the quantity theory has been a continual bone of contention. Until the past three decades, it was generally supported by serious students of economics, those whom we would today term professional economists, and rejected by laymen. However, the success of the Keynesian revolution led to its rejection by perhaps most professional economists. Only recently has it experienced a revival so that it once again commands the adherence of many professional economists. Both its acceptance and its rejection have been grounded basically on judgments about empirical regularities.

1. The Quantity Theory: Nominal versus Real Quantity of Money

In all its versions, the quantity theory rests on a distinction between the *nominal* quantity of money and the *real* quantity of money. The nominal quantity of money is the quantity expressed in whatever units are used

[1]I am, as always, heavily indebted to Anna Schwartz. I have also benefited from discussion of some parts of this article in a number of classes in monetary theory at the University of Chicago and a number of meetings of the Workshop in Money and Banking of the University of Chicago. H. G. Johnson read the semifinal draft and made many useful suggestions for revision. I am grateful to the staff reading committee of the National Bureau: Irving B. Kravis, Chairman, Gary S. Becker, and Richard T. Selden, and to the reading committee of the National Bureau's Board of Directors: Otto Eckstein, Walter E. Hoadley, and James J. O'Leary.

to designate money—talents, shekels, pounds, francs, lire, drachmas, dollars, and so on. The real quantity of money is the quantity expressed in terms of the volume of goods and services that the money will purchase.

There is no unique way to express the real quantity of money. One way to express it is in terms of a specified standard basket of goods and services. That is what is implicitly done when the real quantity of money is calculated by dividing the nominal quantity of money by a price index. The standard basket is then the basket the components of which are used as weights in computing the price index—generally, the basket purchased by some representative group in a base year.

A different way to express the real quantity of money is in terms of the time durations of the flows of goods and services the money could purchase. For a household, for example, the quantity of money can be expressed in terms of the number of weeks of the household's average level of consumption that it could finance with its money balances, or, alternatively, in terms of the number of weeks of its average income to which its money balances are equal. For a business enterprise, the real quantity of money it holds can be expressed in terms of the number of weeks of its average purchases, or of its average sales, or of its average expenditures on final productive services (net value added) to which its money balances are equal. For the community as a whole, the real quantity of money can be expressed in terms of the number of weeks of aggregate transactions of the community, or aggregate net output of the community, to which it is equal.

The reciprocal of any of this latter class of measures of the real quantity of money is a velocity of circulation for the corresponding unit or group of units. In every case, the calculation of the real quantity of money or of velocity is made at the set of prices prevailing at the date to which the calculation refers. These prices are the bridge between the nominal and the real quantity of money.

The quantity theory of money takes for granted that what ultimately matters to holders of money is the real quantity rather than the nominal quantity they hold and that there is a fairly definite real quantity of money that people wish to hold under any given circumstances. Suppose that the nominal quantity that people hold at a particular moment of time happens to correspond at current prices to a real quantity larger than the quantity that they wish to hold. Individuals will then seek to dispose of what they regard as their excess money balances; they will try to pay out a larger sum for the purchase of securities, goods and services, for the repayment of debts, and as gifts than they are receiving

from the corresponding sources. However, they cannot as a group succeed. One man's expenditures are another's receipts. One man can reduce his nominal money balances only by persuading someone else to increase his. The community as a whole cannot in general spend more than it receives.

The attempt to do so will nonetheless have important effects. If prices and income are free to change, the attempt to spend more will raise the volume of expenditures and receipts, expressed in nominal units, which will lead to a bidding up of prices and perhaps also to an increase in output. If prices are fixed by custom or by government edict, the attempt to spend more will either be matched by an increase in goods and services or produce "shortages" and "queues." These, in turn, will raise the effective price and are likely sooner or later to force changes in official prices.

The initial excess of nominal balances will therefore tend to be eliminated, even though there is no change in the nominal quantity of money, by either a reduction in the real quantity available to hold through price rises or an increase in the real quantity desired through output increases. And conversely for an initial deficiency of nominal balances.

It is clear from this discussion that changes in prices and nominal income can be produced either by changes in the real balances that people wish to hold or by changes in the nominal balances available for them to hold. Indeed, it is a tautology, summarized in the famous quantity equation, that all changes in nominal income can be attributed to one or the other—just as a change in the price of any good can always be attributed to a change in either demand or supply. The quantity theory is not, however, this tautology. On an analytical level, it is an analysis of the factors determining the quantity of money the community wishes to hold; on an empirical level, it is the generalization that changes in desired real balances (in the demand for money) tend to proceed slowly and gradually or to be the result of events set in train by prior changes in supply, whereas, in contrast, substantial changes in the supply of nominal balances can and frequently do occur independently of any changes in demand. The conclusion is that substantial changes in prices or nominal income are almost invariably the result of changes in the nominal supply of money.

2. Quantity Equations

The tautology embodied in the quantity equation is a useful device for clarifying the variables stressed in the quantity theory. The quantity

equation has taken different forms, according as quantity theorists have stressed different variables.

a) Transactions Equation

The most famous version of the quantity equation is doubtless the transactions version popularized by Irving Fisher (Fisher 1911, pp. 24–54):

$$MV = PT, \tag{1}$$

or

$$MV + M'V' = PT. \tag{2}$$

In this version, the elementary event is a transaction: an exchange in which one economic actor transfers to another economic actor goods or services or securities and receives a transfer of money in return. The right-hand side of the equations corresponds to the transfer of goods, services, and securities; the left-hand side, to the matching transfer of money.

Each transfer of goods, services, or securities is regarded as the product of a price and a quantity: wage per week times number of weeks, price of a good times number of units of the good, dividend per share times number of shares, price per share times number of shares, and so on. The right-hand side of equations (1) and (2) is the aggregate of such payments during some interval, with P a suitably chosen *average* of the prices, and T a suitably chosen *aggregate* of the quantities during that interval, so that PT is the total nominal value of the payments during the interval in question. The units of P are dollars per unit of quantity; the units of T are number of unit quantities per period of time. We can convert the equation from an expression applying to an *interval* of time to one applying as of a *point* in time by the usual limiting process of letting the interval of time for which we aggregate payments approach zero, and expressing T not as an aggregate but as a rate of flow (that is, the limit of the ratio of aggregate quantities to the length of the interval as the length of the interval approaches zero). The magnitude T then has the dimension of quantity per unit time. The product of P and T then has the dimension of dollars per unit time.

Because the right-hand side is intended to summarize a continuing process, a flow of physical goods and services, the physical item transferred (good, service, or security) is treated as if it disappeared from economic circulation once transferred. If, for example, a single item, say, a house, were transferred three times in the course of the time interval

for which PT is measured, it would enter into T as three houses for that time interval. Further, only those physical items that enter into transactions are explicitly included in T. The houses that exist but are not bought or sold during the time interval are omitted, though, if they are rented, the rental values of their services will be included in PT and the number of dwelling-unit years per year will be included in T. Clearly, T is a rather special kind of index of quantities: it includes service flows (man-hours, dwelling-unit years, kilowatt hours) but also capital items yielding flows (houses, electric generating plants), weighting each of these capital items in accordance with the number of times it enters into exchanges (its "velocity of circulation" in strict analogy with the "velocity of circulation" of money). Similarly, P is a rather special kind of price index.

The monetary transfer analyzed on the left-hand side of equations (1) and (2) is treated very differently. The money that changes hands is treated as retaining its identity, and all money, whether used in transactions during the time interval in question or not, is explicitly accounted for. Money is treated as a stock, not a flow or a mixture of a flow and a stock. For a single transaction, the breakdown into M and V is trivial: the cash that is transferred is turned over once, or $V = 1$. For all transactions during an interval, we can, in principle, classify the existing stock of dollars of money according as each dollar entered into 0, 1, 2, . . . transactions, that is, according as each dollar "turned over" 0, 1, 2, . . . times. The weighted average of these numbers of turnover, weighted by the number of dollars that turned over that number of times, is the conceptual equivalent of V. The dimensions of M are dollars; of V, number of turnovers per unit time; so, of the product, dollars per unit time.[2]

Equation (2) differs from equation (1) by dividing payments into two categories: those effected by the transfer of hand-to-hand currency (including coin) and those effected by the transfer of deposits. In equation (2) M stands solely for the volume of currency and V for

[2] A common criticism of the quantity equation is that, while it takes account of the velocity of circulation of money, it does not take account of the velocity of circulation of goods. As the preceding two paragraphs make clear, while this criticism is not literally valid, it has a real point. The velocity of circulation of money is explicit; the velocity of circulation of goods is implicit. It might well make the right-hand side of equations (1) and (2) more meaningful to make it the sum of two components—one, the total value of transactions involving continuing flows, the other, the value of transfers of existing items of wealth—and to express the second component as a price times a velocity times a stock. In effect, the shift to the income version of the equation resolves the issue by completely neglecting transfers of existing items of wealth.

the velocity of currency, M' for the volume of deposits and V' for the velocity of deposits.

One reason for the emphasis on this particular division was the persistent dispute about whether the term "money" should include only currency or deposits as well (Friedman and Schwartz 1970, chap. 2). Another reason was the direct availability of figures on $M'V'$ from bank records of clearings or of debits to deposit accounts. These make it possible to calculate V' in a way that it is not possible to calculate V.[3]

Equations (1) and (2), like the other quantity equations I shall discuss, are intended to be identities—a special application of double-entry bookkeeping, with each transaction simultaneously recorded on both sides of the equation. However, as with the national income identities with which we are all familiar, when the two sides, or the separate elements on the two sides, are estimated from independent sources of data, many differences between the two sides emerge (Mitchell 1927, pp. 128–39). This has been less obvious for the quantity equations than for the national income identities—with their standard entry "statistical discrepancy"—because of the difficulty of calculating V directly. As a result, V in equation (1) or V and V' in equation (2) have generally been calculated as the numbers having the property that they render the equations correct. These calculated numbers therefore embody the whole of the counterpart to the "statistical discrepancy."

Just as the left-hand side of equation (1) can be divided into several components, as in equation (2), so also can the right-hand side. The emphasis on transactions reflected in this version of the quantity equation suggests dividing total transactions into categories of payments for which payment periods or practices differ: for example, into capital transactions, purchases of final goods and services, purchases of intermediate goods, payments for the use of resources, perhaps separated into wage and salary payments and other payments. The observed value of V might well be a function of the distribution of total payments among categories. Alternatively, if the quantity equation is interpreted not as an identity but as a functional relation expressing desired velocity as a function of other variables, the distribution of payments may well be an important set of variables.

b) *The Income Form of the Quantity Equation*

Despite the large amount of empirical work done on the transactions equations, notably by Irving Fisher and Carl Snyder (Fisher 1911, pp. 280–318; Fisher 1919; Snyder 1934), the ambiguities of the concepts of "transactions" and the "general price level"—particularly those arising from the mixture of current and capital transactions—were never satisfactorily resolved. The more recent development of national or social accounting has stressed income transactions rather than gross transactions and has explicitly and satisfactorily dealt with the conceptual and statistical problems of distinguishing between changes in prices and changes in quantities. As a result, the quantity equation has more recently tended to be expressed in terms of income rather than of transactions. Let $Y =$ nominal national income, $P =$ the price index implicit in estimating national income at constant prices, and $y =$ national income in constant prices, so that

$$Y = Py. \tag{3}$$

Let M represent, as before, the stock of money; but define V as the average number of times per unit time that the money stock is used in making *income* transactions (that is, payments for final productive services or, alternatively, for final goods and services) rather than all transactions. We can then write the quantity equation in income form as

$$MV = Py, \tag{4}$$

or, if it is desired to distinguish currency from deposit transactions, as

$$MV + M'V' = Py. \tag{5}$$

Although the symbols P, V, and V' are used both in equations (4) and (5) and in equations (1) and (2), they stand for different concepts in each pair of equations.

Equations (4) and (5) are both conceptually and empirically more satisfactory than equations (1) and (2). However, they have the disadvantage that they completely neglect both the ratio of intermediate to final transactions and transactions in existing capital assets.

In the transactions version of the quantity equation, each intermediate transaction—that is, purchase by one enterprise from another—is included at the total value of the transaction, so that the value of wheat, for example, is included once when it is sold by the farmer to the mill, a second time when the mill sells flour to the baker, a third time when the baker sells bread to the grocer, a fourth time when the grocer sells bread to the consumer. In the income version, only the net value added by each of these transactions is included. To put it differently, in the transactions version, the elementary event is an isolated exchange of a physical item for money—an actual, clearly observable event. In the

[3] For an extremely ingenious indirect calculation of V, not only for currency as a whole but for particular denominations of currency, see Laurent (1969).

income version, the elementary event is a hypothetical event that can be inferred from observation but is not directly observable. It is a complete series of transactions involving the exchange of productive services for final goods, via a sequence of money payments, with all the intermediate transactions in this income circuit netted out. The total value of all transactions is therefore a multiple of the value of income transactions only.

For a given flow of productive services or, alternatively, of final products (two of the multiple faces of income), the volume of transactions will clearly be affected by vertical integration or disintegration of enterprises, which reduces or increases the number of transactions involved in a single income circuit, or by technological changes that lengthen or shorten the process of transforming productive services into final products. The volume of income will not be thus affected.

Similarly, the transactions version includes the purchase of an existing asset—a house or a piece of land or a share of equity stock—precisely on a par with an intermediate or final transaction. The income version excludes such transactions completely.

Are these differences an advantage or disadvantage of the income version? That clearly depends on what it is that determines the amount of money people want to hold. Do changes of the kind considered in the preceding paragraphs, changes that alter the ratio of intermediate and capital transactions to income, also alter in the same direction and by the same proportion the amount of money people want to hold? Or do they tend to leave this amount unaltered? Or do they have a more complex effect?

Clearly, the transactions and income versions of the quantity theory involve very different conceptions of the role of money. For the transactions version, the most important thing about money is that it is transferred. For the income version, the most important thing is that it is held. This difference is even more obvious from the Cambridge cash-balances version of the quantity equation. Indeed, the income version can perhaps best be regarded as a way station between the Fisher and the Cambridge versions.

c) *Cambridge Cash-Balances Approach*

The essential feature of a money economy is that it enables the act of purchase to be separated from the act of sale. An individual who has something to exchange need not seek out the double coincidence—someone who both wants what he has and offers in exchange what he wants. He need only find someone who wants what he has, sell it to him for general purchasing power, and then find someone who has what he wants and buy it with general purchasing power.

In order for the act of purchase to be separated from the act of sale, there must be something which everybody will accept in exchange as "general purchasing power"—this is the aspect of money emphasized in the transactions approach. But also there must be something which can serve as a temporary abode of purchasing power in the interim between sale and purchase. This is the aspect of money emphasized in the cash-balances approach.

How much money will people or enterprises want to hold for this purpose? As a first approximation, it has generally been supposed that the amount bears some relation to income, on the assumption that this affects the volume of potential purchases for which the individual or enterprise wishes to hold a temporary abode of purchasing power. We can therefore write

$$M = kPy, \tag{6}$$

where M, P, and y are defined as in equation (4), and k is the ratio of money stock to income—either the observed ratio so calculated as to make equation (6) an identity, or the "desired" ratio so that M is the "desired" amount of money, which need not be equal to the actual amount. In either case, k is numerically equal to the reciprocal of the V in equation (4), the V in one case being interpreted as measured velocity and in the other as desired velocity.

Although equation (6) is simply a mathematical transformation of equation (4), it brings out much more sharply the difference between the aspects of money stressed by the transactions approach and those stressed by the cash-balances approach. This difference makes different definitions of money seem natural and leads to emphasis being placed on different variables and analytical techniques.

The transactions approach makes it natural to define money in terms of whatever serves as the medium of exchange in discharging obligations. By stressing the function of money as a temporary abode of purchasing power, the cash-balances approach makes it seem entirely appropriate to include also such stores of value as demand and time deposits not transferable by check, although this approach clearly does not require their inclusion (Friedman and Schwartz 1970, chap. 3).

Similarly, the transactions approach leads to stress being placed on such variables as payments practices, the financial and economic arrangements for effecting transactions, and the speed of communication

and transportation as it affects the time required to make a payment—essentially, that is, to emphasis on the mechanical aspects of the payments process. The cash-balances approach, on the other hand, leads to stress being placed on variables affecting the usefulness of money as an asset: the costs and returns from holding money instead of other assets, the uncertainty of the future, and so on—essentially, that is, to emphasis on the role of cash in a portfolio.

Of course, neither approach enforces the exclusion of the variables stressed by the other—and the more sophisticated economists who have used them have had broader conceptions than the particular approach they adopted. The portfolio aspects enter into the costs of effecting transactions and hence affect the most efficient payment arrangements; the mechanical aspects enter into the returns from holding cash and hence affect the usefulness of cash in a portfolio.

Finally, with regard to analytical techniques, the cash-balances approach fits in much more readily with the general Marshallian demand-supply apparatus than does the transactions approach. Equation (6) can be regarded as a demand function for money, with P and y on the right-hand side being two of the variables on which demand for money depends, and with k symbolizing all the other variables, so that k is to be regarded not as a numerical constant but as itself a function of still other variables. For completion, the analysis requires another equation showing the supply of money as a function of other variables. The price level or the level of nominal income is then the resultant of the interaction of the demand and supply functions.

The quantity theory in its cash-balances version thus suggests organizing an analysis of monetary phenomena in terms of (1) the factors determining the nominal quantity of money to be held—the conditions determining supply—and (2) the factors determining the real quantity of money the community wishes to hold—the conditions determining demand.

3. Supply of Money in Nominal Units

The factors determining the nominal quantity of money available to be held depend critically on the monetary system. For systems like those which have prevailed in the United States and in the United Kingdom during the past century, they can usefully be analyzed under the three main headings that we have termed the proximate determinants of the money stock: (1) the amount of high-powered money—for any one country this is determined through the balance of payments under an international commodity standard, by the monetary authorities, under a fiduciary standard; (2) the ratio of bank deposits to bank holdings of high-powered money—this is determined by the banking system subject to whatever requirements are imposed on them by law or the monetary authorities; and (3) the ratio of the public's deposits to its currency holdings—this is determined by the public (Friedman and Schwartz 1963*b*, pp. 776–98; Cagan 1965).

4. The Demand for Money

J. M. Keynes's liquidity preference analysis (discussed further in section 5, below) reinforced the shift of emphasis from the transactions version of the quantity equation to the cash-balances version—a shift of emphasis from mechanical aspects of the payments process to the qualities of money as an asset. Keynes's analysis, though strictly in the Cambridge cash-balances tradition, was much more explicit in stressing the role of money as one among many assets, and of interest rates as the relevant cost of holding money.

More recent work has gone still further in this direction, treating the demand for money as part of capital or wealth theory, concerned with the composition of the balance sheet or portfolio of assets.

From this point of view, it is important to distinguish between ultimate wealth holders, to whom money is one form in which they choose to hold their wealth, and enterprises, to whom money is a producer's good like machinery or inventories (Friedman 1956).

a) Demand by Ultimate Wealth Holders

For ultimate wealth holders, the demand for money, in real terms, may be expected to be a function primarily of the following variables:

i) *Total wealth.*—This is the analogue of the budget constraint in the usual theory of consumer choice. It is the total that must be divided among various forms of assets. In practice, estimates of total wealth are seldom available. Instead, income may serve as an index of wealth. However, it should be recognized that income as measured by statisticians may be a defective index of wealth because it is subject to erratic year-to-year fluctuations, and a longer-term concept, like the concept of permanent income developed in connection with the theory of consumption, may be more useful (Friedman 1957, 1959; Brunner and Meltzer 1963; Meltzer 1963).

The emphasis on income as a surrogate for wealth, rather than as a measure of the "work" to be done by money, is conceptually perhaps the basic difference between more recent work and the earlier versions of the quantity theory.

ii) *The division of wealth between human and nonhuman forms.*—The major asset of most wealth holders is their personal earning capacity, but the conversion of human into nonhuman wealth or the reverse is subject to narrow limits because of institutional constraints. It can be done by using current earnings to purchase nonhuman wealth or by using nonhuman wealth to finance the acquisition of skills but not by purchase or sale and to only a limited extent by borrowing on the collateral of earning power. Hence, the fraction of total wealth that is in the form of nonhuman wealth may be an additional important variable.

iii) *The expected rates of return on money and other assets.*—This is the analogue of the prices of a commodity and its substitutes and complements in the usual theory of consumer demand. The nominal rate of return on money may be zero, as it generally is on currency, or negative, as it sometimes is on demand deposits subject to net service charges, or positive, as it sometimes is on demand deposits on which interest is paid and generally is on time deposits. The nominal rate of return on other assets consists of two parts: first, any currently paid yield or cost, such as interest on bonds, dividends on equities, and storage costs on physical assets, and, second, changes in their nominal prices. The second part will, of course, be especially important under conditions of inflation or deflation.

iv) *Other variables determining the utility attached to the services rendered by money relative to those rendered by other assets—in Keynesian terminology, determining the value attached to liquidity proper.*—One such variable may be one already considered—namely, real wealth or income, since the services rendered by money may, in principle, be regarded by wealth holders as a "necessity," like bread, the consumption of which increases less than in proportion to any increase in income, or as a "luxury," like recreation, the consumption of which increases more than in proportion.

Another variable that is likely to be important empirically is the degree of economic stability expected to prevail in the future. Wealth holders are likely to attach considerably more value to liquidity when they expect economic conditions to be unstable than when they expect them to be highly stable. This variable is likely to be difficult to express quantitatively even though the direction of change may be clear from

qualitative information. For example, the outbreak of war clearly produces expectations of instability, which is one reason why war is often accompanied by a notable increase in real balances—that is, a notable decline in velocity.

Still another variable may be the volume of capital transfers relative to income—of trading in existing capital goods by ultimate wealth holders. The higher the turnover of capital assets, the higher the fraction of total assets people may find it useful to hold as cash. This variable corresponds to the class of transactions neglected in going from the transactions version of the quantity equation to the income version.

We can symbolize this analysis in terms of the following demand function for money for an individual wealth holder:

$$\frac{M}{P} = f\left(y, w; r_m, r_b, r_e, \frac{1}{P}\frac{dP}{dt}; u\right), \tag{7}$$

where M, P, and y have the same meaning as in equation (6) except that they relate to a single wealth holder; w is the fraction of wealth in nonhuman form (or, alternatively, the fraction of income derived from property); r_m is the expected nominal rate of return on money; r_b is the expected nominal rate of return on fixed-value securities, including expected changes in their prices; r_e is the expected nominal rate of return on equities, including expected changes in their prices; $(1/P)(dP/dt)$ is the expected rate of change of prices of goods and hence the expected nominal rate of return on real assets; and u is a portmanteau symbol standing for whatever variables other than income may affect the utility attached to the services of money. Each of the four rates of return stands, of course, for a set of rates of return, and for some purposes it may be important to classify assets still more finely —for example, to distinguish currency from deposits, long-term from short-term fixed-value securities, risky from relatively safe equities, and one kind of physical assets from another.[4]

The usual problems of aggregation arise in passing from equation (7) to a corresponding equation for the economy as a whole—in particular, they arise from the possibility that the amount of money demanded may depend on the distribution among individuals of such variables as y and w and not merely on their aggregate or average value. If we neglect these distributional effects, equation (7) can be

[4] Under some assumed conditions, the four rates of return may not be independent. For example, in a special case considered in Friedman (1956, pp. 9–10),

$$r_b = r_e + (1/P)(dP/dt).$$

regarded as applying to the community as a whole, with M and y referring to per capita money holdings and per capita real income, respectively, and w to the fraction of aggregate wealth in nonhuman form.

The major problems that arise in practice in applying equation (7) are the precise definitions of y and w, the estimation of *expected* rates of return as contrasted with actual rates of return, and the quantitative specification of the variables designated by u.

b) *Demand by Business Enterprises*

Business enterprises are not subject to a constraint comparable with that imposed by the total wealth of the ultimate wealth holder. The total amount of capital embodied in productive assets, including money, is a variable that can be determined by an enterprise to maximize returns, since it can acquire additional capital through the capital market. Hence, there is no reason on this ground to include total wealth, or y as a surrogate for total wealth, as a variable in the business demand function for money.

It may, however, be desirable to include a somewhat similar variable defining the "scale" of the enterprise on different grounds—namely, as an index of the productive value of different quantities of money to the enterprise. This is more nearly in line with the earlier transactions approach emphasizing the "work" to be done by money. It is by no means clear what the appropriate variable is: total transactions, net value added, net income, total capital in nonmoney form, or net worth. The lack of availability of data has meant that much less empirical work has been done on the business demand for money than on an aggregate demand curve encompassing both ultimate wealth holders and business enterprises. As a result there are as yet only faint indications about the best variable to use.

The division of wealth between human and nonhuman form has no special relevance to business enterprises, since they are likely to buy the services of both forms on the market.

Rates of return on money and on alternative assets are, of course, highly relevant to business enterprises. These rates determine the net cost to them of holding the money balances. However, the particular rates that are relevant may be quite different from those that are relevant for ultimate wealth holders. For example, rates charged by banks on loans are of minor importance for wealth holders yet may be extremely important for businesses, since bank loans may be a way in which they can acquire the capital embodied in money balances.

The counterpart for business enterprises of the variable u in equation (7) is the set of variables other than scale affecting the productivity of money balances. At least one of these—namely, expectations about economic stability—is likely to be common to business enterprises and ultimate wealth holders.

With these interpretations of the variables, equation (7), with w excluded, can be regarded as symbolizing the business demand for money and, as it stands, symbolizing aggregate demand for money, although with even more serious qualifications about the ambiguities introduced by aggregation.

5. The Keynesian Challenge to the Quantity Theory

The income-expenditure analysis developed by John Maynard Keynes in his *General Theory* (Keynes 1936) offered an alternative approach to the interpretation of changes in nominal income that emphasized the relation between nominal income and investment or autonomous expenditures rather than the relation between money income and the stock of money.

Keynes's basic challenge to the reigning theory can be summarized in three propositions that he set forth:

1. As a purely *theoretical* matter, there need not exist, even if all prices are flexible, a *long-run equilibrium* position characterized by "full employment" of resources.

2. As an *empirical* matter, prices can be regarded as rigid—an institutional datum—for *short-run economic fluctuations;* that is, for such fluctuations, the distinction between real and nominal magnitudes that is at the heart of the quantity theory is of no importance.

3. The demand function for money has a particular empirical form —corresponding to absolute liquidity preference—that makes velocity highly unstable much of the time, so that changes in the quantity of money would, in the main, simply produce changes in V in the opposite direction. This proposition is critical for both propositions (1) and (2), though the reasons for absolute liquidity preference are different in the long run and in the short run. Absolute liquidity preference at an interest rate approaching zero is a necessary though not a sufficient condition for proposition (1). Absolute liquidity preference at the "conventional" interest rate explains why Keynes regarded the quantity equation, though perfectly valid as an identity, as largely useless for policy or for predicting short-run fluctuations in nominal and real income (identical by proposition [2]). In its place, Keynes put the income identity supplemented by a stable propensity to consume.

a) Long-Run Equilibrium

The first proposition can be treated summarily because it has been demonstrated to be false. Keynes's error consisted in neglecting the role of wealth in the consumption function—or, stated differently, in neglecting the existence of a desired stock of wealth as a goal motivating savings.[5] All sorts of frictions and rigidities may interfere with the attainment of a hypothetical long-run equilibrium position at full employment; dynamic changes in technology, resources, and social and economic institutions may continually change the characteristics of that equilibrium position; but there is no fundamental "flaw in the price system" that makes unemployment the natural outcome of a fully operative market mechanism.[6]

b) Short-Run Price Rigidity[7]

Alfred Marshall's distinction among market equilibrium, short-period equilibrium, and long-period equilibrium was a device for analyzing

[5] Keynes, of course, verbally recognized this point, but it was not incorporated in his formal model of the economy. Its key role was pointed out first by Haberler (1941, pp. 242, 389, 403, 491–503) and subsequently by Pigou (1947), Tobin (1947), Patinkin (1951), and Johnson (1961).

[6] This proposition played a large role in gaining for Keynes the adherence of many noneconomists, particularly the large band of reformers, social critics, and radicals who were persuaded that there was something fundamentally wrong with the capitalist "system." There is a long history of attempts, some highly sophisticated, to demonstrate that there is a "flaw in the price system" (the title of one such attempt [Martin 1924]), attempts going back at least to Malthus. In modern times, one of the most popular and persistent is the "social credit" doctrine of Major C. H. Douglas, which even spawned a political party in Canada that captured control (in 1935) of the government of one of the Canadian provinces (Alberta) and attempted to implement some of Major Douglas's doctrines. This policy ran into legal obstacles and had to be abandoned. The successor party now (1969) controls Alberta and British Columbia. But, prior to Keynes, these attempts had been made primarily by persons outside of the mainstream of the economics profession, and professional economists had little trouble in demonstrating their theoretical flaws and inadequacies.

Keynes's attempt was therefore greeted with enthusiasm. It came from a professional economist of the very highest repute, regarded, and properly so, by his fellow economists as one of the great economists of all time. The analytical system was sophisticated and complex, yet, once mastered, appeared highly mechanical and capable of yielding far-reaching and important conclusions with a minimum of input; and these conclusions were, besides, highly congenial to the opponents of the market system.

Needless to say, the demonstration that this proposition of Keynes's is false, and even the acceptance of this demonstration by economists who regard themselves as disciples of the Keynes of *The General Theory,* has not prevented the noneconomist opponents of the market system from continuing to believe that Keynes proved the proposition, and continuing to cite his authority for it.

[7] We are indebted to a brilliant book by Leijonhufvud (1968) for a full appreciation of the importance of this proposition in Keynes's system. This subsection and the one that follows, on the liquidity preference function, owe much to Leijonhufvud's penetrating analysis.

the dynamic adjustment in a particular market to a change in demand or supply. This device had two key characteristics. One, the less important for our purposes, is that it replaced the continuous process by a series of discrete steps—comparable with approximating a continuous function by a set of straight-line segments. The second is the assumption that prices adjust more rapidly than quantities, indeed, so rapidly that the price adjustment can be regarded as instantaneous. An increase in demand (a shift to the right of the long-run demand curve) will produce a new market equilibrium involving a higher price but the same quantity. The higher price will, in the short run, encourage existing producers to produce more with their existing plants, thus raising quantity and bringing prices back down toward their original level, and, in the long run, attract new producers and encourage existing producers to expand their plants, still further raising quantities and lowering prices. Throughout the process, it takes time for output to adjust but no time for prices to do so. This assumption has no effect on the final equilibrium position, but it is vital for the path to equilibrium.

This Marshallian assumption about the price of a particular product became widely accepted and tended to be carried over unthinkingly to the price level in analyzing the dynamic adjustment to a change in the demand for or supply of money. As noted above, the Cambridge cash-balances equation lends itself to a demand-supply interpretation along Marshallian lines (Pigou 1917). So interpreted, a change in the nominal quantity of money (a once-for-all shift in the supply schedule) will require a change in one or more of the variables on the right-hand side of equation (6)—k, or P, or y—in order to reconcile demand and supply. In the final full equilibrium, the adjustment will, in general, be entirely in P, since the change in the nominal quantity of money need not alter any of the "real" factors on which k and y ultimately depend.[8] As in the Marshallian case, the final position is not affected by relative speeds of adjustment.

There is nothing in the logic of the quantity theory that specifies the dynamic path of adjustment, nothing that requires the whole adjustment to take place through P rather than through k or y. It was widely recog-

[8] The "in general" is inserted to warn the reader that this is a complex question, requiring for a full analysis a much more careful statement of just how the quantity of money is increased. However, these more sophisticated issues are not relevant to the point under discussion and so are bypassed.

nized that the adjustment during what Fisher, for example, called "transition periods" would in practice be partly in k and in y as well as in P. Yet this recognition was not incorporated in formal theoretical analysis. The formal analysis simply took over Marshall's assumption. In this sense, the quantity theorists can be validly criticized for having "assumed" price flexibility—just as Keynes can be validly criticized for "assuming" that consumption is independent of wealth, even though he recognized in his asides that wealth has an effect on consumption.

Keynes was a true Marshallian in method. He followed Marshall in taking the demand-supply analysis as his framework. He followed Marshall in replacing the continuous adjustment by a series of discrete steps and so analyzing a dynamic process in terms of a series of shifts between static equilibrium positions. Even his steps were essentially Marshall's, his short-run being distinguished from his long-run by the fixity of the aggregate capital stock. However, he tended to merge the market period and the short-run period, and, true to his own misleading dictum, "in the long run we are all dead," he concentrated almost exclusively on the short run.

Keynes also followed Marshall in assuming that one variable adjusted so quickly that the adjustment could be regarded as instantaneous, while the other variable adjusted slowly. Where he deviated from Marshall, and it was a momentous deviation, was in reversing the roles assigned to price and quantity. He assumed that, at least for changes in aggregate demand, quantity was the variable that adjusted rapidly, while price was the variable that adjusted slowly,[9] at least in a downward direction. Keynes embodied this assumption in his formal model by expressing all variables in wage units, so that his formal analysis—aside from a few passing references to a situation of "true" inflation—dealt with "real" magnitudes, not "nominal" magnitudes (Keynes 1936, pp. 119, 301, 303). He rationalized the assumption in terms of wage rigidity arising partly from money illusion, partly from the strength of trade unions. And, at a still deeper level, he rationalized wage rigidity by proposition (1): under conditions when there was no full-employment equilibrium, there was also no equilibrium nominal price level; something had to be brought in from outside to fix the price level; it might as well be institutional wage rigidity. Put differently, flexible nominal wages under such circumstances had no economic function to perform; hence they might as well be made rigid.

[9] I have referred to "quantity," not "output," because I conjecture that Keynes, if pressed to distinguish the market from the short-run period, would have done so by regarding quantity available to purchase as adjusting rapidly in the market period largely through changes in inventories, and in the short-run period through changes in output.

However rationalized, the basic reason for the assumption was undoubtedly the lack of concordance between observed phenomena and the implications of a literal application of Marshall's assumption to aggregate magnitudes. Such a literal application implied that economic fluctuations would take the form wholly of fluctuations in prices with continuous full employment of men and resources. Clearly, this did not correspond to experience. If anything, at least in the decade and a half between the end of World War I and the writing of *The General Theory*, economic fluctuations were manifested to a greater degree in output and employment than in prices. It therefore seemed highly plausible that, at least for aggregate phenomena, relative speeds of adjustment were just the reverse of those assumed by Marshall.[10]

Keynes explored this penetrating insight by carrying it to the extreme: all adjustment in quantity, none in price. He qualified this statement by assuming it to apply only to conditions of underemployment. At "full" employment, he shifted to the quantity-theory model and asserted that all adjustment would be in price—he designated this a situation of "true inflation." However, Keynes paid no more than lip service to this possibility, and his disciples have done the same; so it does not misrepresent the body of his analysis largely to neglect the qualification.

Given this assumption, a change in the nominal quantity of money means a change in the real quantity of money. In equation (6) we can divide through by P, making the left-hand side the real quantity of money. A change in the (nominal and real) quantity of money will then be matched by a change in k or in y.

Nothing up to this point seems to prevent Keynes from having a purely monetary theory of economic fluctuations, with changes in M being reflected entirely in y. However, this conflicted with Keynes's interpretation of the facts of the Great Depression, which he regarded, I believe erroneously, as showing that expansive monetary policy was ineffective in stemming a decline (Friedman 1967). Hence, he was inclined to interpret changes in M as being reflected in k rather more

[10] I do not mean to suggest that Marshall's assumption is always the best one for particular markets. On the contrary, one of the significant advances in recent years in relative price theory is the development of more sophisticated price adjustment models that allow the rates of adjustment of both price and quantity to vary continuously between instantaneous and very slow adjustment. However, these developments are not directly relevant to the present discussion, although they partly inspire section 12 below.

than in y. This is where his proposition (3) about liquidity preference enters in.

Indeed, in the most extreme, and I am tempted to say purest, form of his analysis, Keynes supposes that the whole of the adjustment will be in k. And, interestingly enough, this result can also be regarded as a direct consequence of his assumption about the relative speed of adjustment of price and quantity. For k is not a numerical constant but a function of other variables. It embodies liquidity preference. In Keynes's system, the main variable it depends on is the interest rate. This too is a price. Hence, it was natural for Keynes to regard it as slow to adjust, and to take, as the variable which responds, the real quantity of money people desire to hold.

If changes in M do not produce changes in y, what does? Keynes's answer is the need to reconcile the amount some people want to spend to add to the stock of productive capital with the amount the community wants to save to add to its stock of wealth. Hence Keynes puts at the center of his analysis the distinction between consumption and saving, or more fundamentally, between spending linked closely to current income and spending that is largely independent of current income.

As a result of both experience and further theoretical analysis, there is hardly an economist today who accepts Keynes's conclusion about the strictly passive character of k, or the accompanying conclusion that money (in the sense of the quantity of money) does not matter, or who will explicitly assert that P is "really" an institutional datum that will be completely unaffected even in short periods by changes in M (Friedman 1968, 1970b).

Yet Keynes's assumption about the relative speed of adjustment of price and quantity is still a key to the difference in approach and analysis between those economists who regard themselves as Keynesians and those who do not. Whatever the first group may say in their asides and in their qualifications, they treat the price level as an institutional datum in their formal theoretical analysis. They continue to regard changes in the nominal quantity of money as equivalent to changes in the real quantity of money and hence as having to be reflected in k and y. And they continue to regard the initial effect as being on k. The difference is that they no longer regard interest rates as institutional data, as Keynes in considerable measure did. Instead, they regard the change in k as requiring a change in interest rates which in turn produces a change in y. Hence, they attribute more significance to changes in the quantity of money than Keynes and his disciples did in the first decade or so after the appearance of *The General Theory*.

A striking illustration is provided in a recent Cowles Foundation Monograph, edited by Donald Hester and James Tobin, on *Financial Markets and Economic Activity* (Hester and Tobin 1967). A key essay in that book presents a comparative static analysis of the general equilibrium adjustment of stocks of assets. Yet the distinction between nominal and real magnitudes is not even discussed. The entire analysis is valid only on the implicit assumption that nominal prices of goods and services are completely rigid, although interest rates and real magnitudes are flexible.[11]

The National Bureau series of monetary studies illustrates the other side of the coin—the approach of those of us who do not regard ourselves as Keynesians. Many of the questions discussed in these monographs would not have appeared to be open questions, and large parts of them would never have been written, had we, implicitly or explicitly, accepted Keynes's assumption that prices are an institutional datum.

c) *Absolute Liquidity Preference*

Keynes gave a highly specific form to equation (6) or (7). The quantity of money demanded, he argued, could be treated as if it were divided

[11] See Tobin and Brainard (1967). A specific example documenting this statement is that Tobin and Brainard explicitly assume that central banks can determine the ratio of currency (or high-powered money) to total wealth including real assets (Hester and Tobin 1967, pp. 61–62). If prices are flexible, the central bank can determine only nominal magnitudes, not such a real ratio.

Other papers in Monograph 21, notably the paper by Brainard, "Financial Institutions and a Theory of Monetary Control" (Brainard 1967), make the same implicit assumptions. The word "prices" does not appear in the cumulative subject index of this monograph and of two companion volumes, Monographs 19 and 20.

Still another more recent example is a paper by the same authors, "Pitfalls in Financial Model Building" (Tobin and Brainard 1968), in which they present a simulation of a "fictitious economy of our construction." In this economy, the replacement value of physical assets is used as the numeraire of the system, and all prices are expressed relative to the replacement value. The result is that the system—intended to illuminate the problems of monetary analysis—takes the absolute price level as determined outside the system. The Central Bank is implicitly assumed to be able to determine the *real* and not merely the *nominal* volume of bank reserves.

Another striking example is Gramley and Chase (1965). In this article, the assumption about price rigidity is explicit and presented as if it were only a tentative assumption made for convenience of analysis. Yet the empirical significance Gramley and Chase attach to their results belies this profession.

See also the econometric study by Goldfeld (1966), which concentrates on real forms of the functions estimated because of "the superiority of the deflated version" (p. 166).

Evidence for a somewhat earlier period is provided by Holzman and Bronfenbrenner (1963). Theories of inflation stemming from the Keynesian approach stress institutional, not monetary, factors.

into two parts, one part, M_1, "held to satisfy the transactions- and precautionary-motives," the other, M_2, "held to satisfy the speculative-motive" (Keynes 1936, p. 199). He regarded M_1 as a roughly constant fraction of income. He regarded the (short-run) demand for M_2 as arising from *"uncertainty* as to the future of the rate of interest" and the amount demanded as depending on the relation between current rates of interest and the rates of interest expected to prevail in the future (Keynes 1936, p. 168; italics in original). Keynes, of course, emphasized that there was a whole complex of interest rates. However, for simplicity, he spoke in terms of "the rate of interest," usually meaning by that the rate on long-term securities that involved minimal risks of default—for example, government bonds. The key distinction to Keynes was between short-term and long-term securities, not between securities fixed in nominal value and those that were not. The latter distinction was rendered irrelevant by his assumption that prices were rigid.

The distinction between short-term and long-term securities was important to Keynes because it corresponded to differences in risk of capital gain or loss as a result of changes in interest rates. For short-term securities, changes in interest rates would have little effect. For long-term securities, the effect is important. Leijonhufvud has argued, and we believe correctly, that Keynes used the term "money" as referring not only to currency and deposits narrowly defined but to the whole range of short-term assets that provided "liquidity" in the sense of security against capital loss arising from changes in interest rates.[12] Needless to say, Keynes also regarded other kinds of risks, such as risks of default, as highly relevant, but, consistent with his proposition (2), he almost entirely disregarded risks arising from changes in the price level of goods and services (Leijonhufvud 1968, chap. 2).

It is therefore somewhat misleading to regard Keynes, as most of the literature does, as distinguishing between "money" and "bonds." Nonetheless, we shall continue to follow current practice and use that terminology. One justification for doing so is that Keynes did treat the short-term assets he labeled "money" as yielding no interest return. (It is well to recall that he was writing at a time when short-term interest rates were extremely low both absolutely and relative to long-term rates. His procedure would seem highly unrealistic today.)

[12] In this respect, the Radcliffe Committee is faithful to Keynes in treating "liquidity" broadly defined as the relevant monetary aggregate rather than "money" narrowly defined.

To formalize Keynes's analysis in terms of the symbols we have used so far, we can write his demand function as

$$\frac{M}{P} = \frac{M_1}{P} + \frac{M_2}{P} = k_1 y + f(r - r^*, r^*), \tag{8}$$

where r is the current rate of interest, r^* is the rate of interest expected to prevail, and k_1, the analogue to the inverse of income velocity of circulation of money, is treated as determined by payment practices and hence as a constant at least in the short run.[13] The current interest rate, r, is an observed magnitude. Hence it will be the same for all holders of money, if, like Keynes, we abstract from the existence of a complex of interest rates. The expected rate, r^*, is not observable. It may differ from one holder to another and, for each holder separately, is to be interpreted as the mean value of a probability distribution, not as a single value anticipated with certainty. For an aggregate function, r^* should strictly speaking be interpreted as a vector, not a number. Though I have introduced P into the equation for consistency with my earlier equations, Keynes omitted it because of his proposition (2), which meant that P, or, more precisely, the wage rate, was taken to be a constant.

In a "given state of expectations," that is, for a given value of r^*, the higher is the current rate of interest, the lower will be the amount of money people would want to hold for speculative motives. The cost of holding money instead of securities would be greater in two ways: first, a larger amount of current earnings would be sacrificed; second, it would be more likely that interest rates would fall, and hence security prices rise, and so a larger amount of capital gains would be sacrificed.

Although expectations are given great prominence in developing the liquidity function expressing the demand for M_2, Keynes and his followers generally did not explicitly introduce them, as I have done, into that function. For the most part, Keynes and his followers in practice treated the amount of M_2 demanded simply as a function of the current interest rate, the emphasis on expectations serving only as a reason for their attribution of instability to the liquidity function.[14]

The reason for this omission is their concentration on the short-run demand function. For that function, they regarded r^* as fixed, so that the speculative demand was a function of r alone. I have introduced

[13] Later writers in this tradition have argued that k_1 too should be regarded as a function of interest rates. See Baumol (1952), and Tobin (1956). However, this issue is not relevant to the present discussion.

[14] A notable exception is Tobin (1958, pp. 65–86).