

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Monetary Trends in the United States and United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867-1975

Volume Author/Editor: Milton Friedman and Anna J. Schwartz

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-26409-2

Volume URL: <http://www.nber.org/books/frie82-2>

Publication Date: 1982

Chapter Title: The General Theoretical Framework

Chapter Author: Milton Friedman, Anna J. Schwartz

Chapter URL: <http://www.nber.org/chapters/c11402>

Chapter pages in book: (p. 16 - 72)

The General Theoretical Framework

As is so pungently stated in the quotation from de Morgan that we use as the epigraph of this book, every empirical study rests on a theoretical framework,¹ on a set of tentative hypotheses that the evidence is designed to test or to adumbrate. It may help the reader if we set out explicitly the general theoretical framework that underlies this and our earlier volumes.²

The words *general* and *framework* are included in this statement of the purpose of this chapter to make clear its limits. The chapter does not present a fully developed theory that has as implications all the empirical regularities that those of us studying monetary phenomena have isolated. A fully developed theory is much to be desired, but it would require an entire treatise. It is not something that can readily be done in an introductory chapter of an empirical study concerned with only one aspect of monetary relations. Further, the chapter makes no attempt to present a comprehensive doctrinal history of the development of either the quantity theory or Keynesian theory. References to earlier writers are simply expository devices to illuminate analytical points. Finally, this chapter

1. The first five sections of this chapter draw heavily on Milton Friedman, "Money: Quantity Theory," *International Encyclopedia of the Social Sciences* (New York: Macmillan and Free Press, 1968).

2. Several reviewers of *A Monetary History* criticized us for not making explicit the theoretical framework employed in that book. This chapter is largely a response to that criticism. See John M. Culbertson, "United States Monetary History: Its Implications for Monetary Theory," *National Banking Review* 1 (March 1964): 372-75; Allan H. Meltzer, "Monetary Theory and Monetary History," *Schweizerische Zeitschrift für Volkswirtschaft und Statistik* 101, no. 4 (1965): 409-22; James Tobin, "The Monetary Interpretation of History," *American Economic Review* 55 (June 1965): 464-85.

supplements, rather than replaces, others of our writings on issues in monetary theory.³

Our theoretical 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 mid-1930s, it was generally supported by serious students of economics, those whom today we would term professional economists, and rejected by laymen. However, the success of the Keynesian revolution led to its rejection by many, perhaps most, professional economists. Only in the past two decades has it experienced a revival so that it once again commands the adherence of many professional economists. Its initial acceptance, its rejection, and its recent revival have all been grounded basically on judgments about empirical regularities.

2.1. 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 to designate money—talents, shekels, pounds, francs, liras, drachmas, dollars, and so on. The real quantity of money is the quantity expressed in terms of the volume of goods and services 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 done implicitly 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 whose components 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 real quantity of money can be expressed in terms of the number of weeks of the household’s average level of consumption it could finance with its money balances or, alternatively, in terms of the number of weeks of its average income to which its

3. These caveats are occasioned by the reaction to an earlier version of this chapter published separately. See Robert J. Gordon, ed., *Milton Friedman’s Monetary Framework: A Debate with His Critics* (University of Chicago Press, 1974).

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 its money balances are 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. For example, the ratio of the annual transactions of the community to its stock of money is the "transactions velocity of circulation of money," since it gives the number of times the stock of money would have to "turn over" in a year to accomplish all transactions. Similarly, the ratio of annual income to the stock of money is termed "income velocity." In every case, the real quantity of money or velocity is calculated 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 first, that what ultimately matters to holders of money is the real quantity rather than the nominal quantity they hold and, second, that there is a fairly definite real quantity of money that people wish to hold in 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 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 customary or 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 what 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.

2.2 Quantity Equations

The tautology embodied in the quantity equation is a useful device for classifying the variables stressed in the quantity theory. The quantity equation has taken different forms as quantity theorists have stressed different variables.

2.2.1 Transactions Equation

The most famous version of the quantity equation is doubtless the transactions version popularized by Irving Fisher:⁴

$$(1) \quad MV = PT,$$

or

$$(2) \quad MV + M'V' = PT.$$

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

4. See Irving Fisher, *The Purchasing Power of Money* (New York: Macmillan, 1911; rev. ed. 1920; 2d rev. ed. 1922; reprinted New York: Kelly, 1963), pp. 24–54.

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 to a *point* in time by the usual limiting process of letting the interval 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 , 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 interval. Further, only those physical items that enter into transactions are explicitly included in T . 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 are included in PT and the number of dwelling-unit years per year are included in T . Clearly, T is a rather special kind of index of quantities: it includes service flows (man-hours) and also physical capital items yielding flows (houses, electric-generating plants) and securities representing such capital items as well as such intangible capital items as “goodwill.” Each of the capital items or securities is weighted by the number of times it enters into transactions (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 as 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 of time, 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.⁵

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 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.⁶ Another reason was the direct availability of data on $M'V'$ from bank records of clearings or of debits to deposit accounts. These data make it possible to calculate V' in a way that is not possible for V .⁷

Equations (1) and (2), like the other quantity equations we 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 them emerge.⁸ This statistical defect 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) and V and V' in equation (2) have generally been calculated as the numbers

5. A common criticism of the quantity equation is that it takes account of the velocity of circulation of money but not of the velocity of circulation of goods. As the preceding two paragraphs make clear, this criticism, though not literally valid, makes 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.

6. See Milton Friedman and Anna J. Schwartz, *Monetary Statistics of the United States* (New York: Columbia University Press for NBER, 1970), chap. 2.

7. For an extremely ingenious indirect calculation of V , not only for currency as a whole but for particular denominations of currency, see Robert Laurent, “Currency Transfers by Denominations,” Ph.D. diss., University of Chicago, 1969.

8. See Wesley C. Mitchell, *Business Cycles* (New York: NBER, 1927), pp. 128–39.

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, and 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.

2.2.2 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,⁹ the ambiguities of the concepts of “transactions” and the “general price level”—particularly those arising from the mixture of current and capital transactions—have never been satisfactorily resolved. The more recent development of national or social accounting has stressed income transactions rather than gross transactions and has explicitly if not wholly satisfactorily dealt with the conceptual and statistical problems involved in 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 transactions rather than of gross transactions. Let Y = nominal national income, P = the price index implicit in estimating national income at constant prices, N = the number of persons in the population, y = per capita national income in constant prices, and $y' = Ny$ = national income at constant prices, so that

$$(3) \quad Y = PNy = Py'.$$

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

9. See Irving Fisher, *Purchasing Power of Money*, pp. 280–318; Irving Fisher, “Money, Prices, Credit and Banking,” *American Economic Review* 9 (June 1919): 407–9; and Carl Snyder, “On the Statistical Relation of Trade, Credit and Prices,” *Review of International Statistical Institute* 2 (October 1934): 278–91.

$$(4) \quad MV = PNy = Py',$$

or, if we desire to distinguish currency from deposit transactions, as

$$(5) \quad MV + M'V' = PNy.$$

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 conceptually and empirically more satisfactory than equations (1) and (2). However, they have the disadvantage that they completely neglect 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 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 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.

2.2.3 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.

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

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

$$(6) \quad M = kPNy = kPy',$$

where M , N , P , y , 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 being interpreted in one case 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 aspect of money stressed by the transactions approach and that stressed by the cash-balances approach. This difference makes different defini-

tions of money seem natural and leads to placing emphasis 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 in addition such stores of value as demand and time deposits not transferable by check, although the cash-balances approach clearly does not require their inclusion.¹⁰

Similarly, the transactions approach leads to emphasis 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 aspect of the payments process. The cash-balances approach, on the other hand, leads to emphasis 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 have had broader conceptions than the particular approach they adopted. Portfolio considerations enter into the costs of effecting transactions and hence affect the most efficient payment arrangements; mechanical considerations 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 , N , and y on the right-hand side being three of the variables on which demand for money depends and 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; (2) the factors determining the real quantity of money the community wishes to hold—the conditions determining de-

10. Friedman and Schwartz, *Monetary Statistics*, chap. 3.

mand; and (3) the reconciliation of demand with supply—the conditions determining how changes in demand or supply work themselves out through prices and quantities.

2.2.4 The Transmission Mechanism: Money to Income, Prices, Output

A frequent criticism of the quantity theory and the quantity equations concerns the third of these items—the mechanism whereby a change in the nominal quantity of money is transmitted to prices and quantities. The criticism is that the transmission mechanism is not specified, that the proponents of the quantity theory rely on a black box connecting the input—the nominal quantity of money—and the output—effects on prices and quantities.

On one level this criticism is not justified; on another it points to an important element in the unfinished agenda of research—an element to which some of the later chapters of this book make, we trust, a contribution.

The criticism is not justified insofar as it implies that there is a fundamental difference between the adjustment mechanism implicit or explicit in the quantity equation and in a demand-supply analysis for a particular product—shoes, or copper, or haircuts. In both cases the demand function for the community as a whole is an aggregation of demand functions for individual consumer or producer units, and the separate demand functions are determined by the tastes and opportunities of the units. In both cases, supply functions depend on production possibilities, institutional arrangements for organizing production, and the conditions of supply of resources. In both cases a shift in supply or in demand introduces a discrepancy between the amounts demanded and supplied *at the preexisting price*. In both cases any discrepancy can be eliminated only by either a price change or some alternative rationing mechanism, explicit or implicit.

On this level, two features of the demand-supply adjustment for money have concealed the parallelism. One is that demand-supply analysis for particular products typically deals with flows—number of pairs of shoes or number of haircuts per year—whereas the quantity equations deal with the stock of money at a point in time. In this respect the analogy is with the demand for, say, land, which, like money, derives its value from the flow of services it renders but has a purchase price and not merely a rental value. The second is the widespread tendency to confuse “money” and “credit,” which has produced misunderstanding about the relevant price variable. The “price” of money is the quantity of goods and services that must be given up to acquire a unit of money—the inverse of the price level. This is the price that is analogous to the price of land or of copper or of haircuts. The “price” of money is not the interest rate, which is the “price” of credit. The interest rate connects stocks with flows—the rental

value of land with the price of land, the value of the service flow from a unit of money with the price of money. Of course, the interest rate may affect the quantity of money demanded—just as it may affect the quantity of land demanded—but so may a host of other variables.

On a more sophisticated level, the criticism about the transmission mechanism applies equally to money and to other goods and services. In all cases it is desirable to go beyond equality of demand and supply as defining a stationary equilibrium position and examine the variables that affect the quantities demanded and supplied and the dynamic temporal process whereby actual or potential discrepancies are eliminated. For money, an examination of the variables affecting demand and supply has been carried very far—farther than for most other economic goods or services, as sections 2.3 and 2.4 and the references contained therein, indicate. But for both money and most other goods and services, there is as yet no satisfactory and widely accepted description, in precise quantifiable terms, of the dynamic temporal process of adjustment—though in recent decades much research has been devoted to this question. It remains a challenging subject for research. Section 2.6 discusses a particular hypothesis about the adjustment mechanism, and chapters 8 and 9 explore some of the issues empirically.

2.2.5 The International Transmission Mechanism

From its very earliest days, the quantity theory was intimately connected with the analysis of the adjustment mechanism in international trade. A commodity standard, in which money was specie or its equivalent, was taken as the norm. Under such a standard, the supply of money in any one country is determined by the links between that country and other countries that use the same commodity as money. Under such a standard, the same theory explains links among money and prices and nominal income in various parts of a single country—money, prices, and nominal income in Illinois and money, prices, and nominal income in the rest of the United States—and the corresponding links among various countries. The differences between interregional adjustment and international adjustment are empirical: greater mobility of people, goods, and capital among regions than among countries, and hence a more rapid adjustment.

The specie-flow mechanism developed by Hume and elaborated by Ricardo and his successors analyzed the links among countries primarily in terms of the effect of a disequilibrium stock of money on prices and thereby the balance of payments. “Too” high a money stock in country A tends to make prices in A high relative to prices in the rest of the world, encouraging imports and discouraging exports. The resulting deficit in the balance of trade will be met by shipments of specie, which reduces the quantity of money in country A and raises it in the rest of the world. The

changes in the quantity of money will tend to reduce prices in country A and raise them in the rest of the world, correcting the original disequilibrium. The process will continue until price levels in all countries are at a level at which balances of payments are in equilibrium (which may, of course, mean a continuing movement of specie, for example, from gold- or silver-producing countries to non-gold- or silver-producing countries, or between countries growing at different secular rates).

Another strand of the classical analysis has recently been revived under the title "the monetary theory of the balance of payments." This theory is logically equivalent to the specie-flow mechanism except that it makes different assumptions about the speed of adjustment of the several variables. The specie-flow mechanism implicitly assumes that prices are slow to adjust and do so only in response to changes in the quantity of money produced by specie flows. However, there can be only a single price for goods traded internationally if the markets are efficient and transportation costs are neglected. Speculation tends to assure this result for the prices of traded goods expressed in a common currency. Internally, competition between traded and nontraded goods tends to keep their relative prices in a relation determined by relative costs. If these adjustments are rapid, prices will always be in equilibrium among countries ("the law of one price"). If the money stock is not distributed among countries in such a way as to be consistent with these prices, the excess demands and supplies of money will lead to specie flows. If the quantity of money in a country is "too" low, domestic nominal demand will not be adequate to absorb a total value of domestic goods plus imports equal to the value of domestic output. Export of the excess will produce a balance of payments surplus for that country, which will raise the quantity of money. Specie flows are still the adjusting mechanism, but they are produced not by discrepancies in prices but by differences between demand for output in nominal terms and the supply of output at world prices. Putative, not actual, price differences are the spur to adjustment. This description is a highly over simplified picture, primarily because it omits the important role assigned to short- and long-term capital flows by all theorists—those who stress the specie-flow mechanism and even more those who stress the single-price mechanism.¹¹

In practice, few countries have had pure commodity standards. Most have had a mixture of commodity and fiduciary standards. Changes in the

11. For a fuller discussion, see: Jacob A. Frenkel, "Adjustment Mechanisms and the Monetary Approach to the Balance of Payments," in *Recent Issues in International Monetary Economics*, ed. E. Claassen and P. Salin (Amsterdam: North-Holland, 1976); Jacob A. Frenkel and Harry G. Johnson, "The Monetary Approach to the Balance of Payments: Essential Concepts and Historical Origins," in *The Monetary Approach to the Balance of Payments*, ed. J. Frenkel and H. Johnson (Toronto: University of Toronto Press 1976), pp. 21-45.

fiduciary component of the stock of money can replace specie flows as a means of adjusting the quantity of money.

The situation is more complex for countries that do not share a unified currency, that is, a currency in which only the name assigned to a unit of currency differs among countries. Changes in the rates of exchange between national currencies then serve to keep prices in various countries in the appropriate relation when expressed in a common currency. Exchange rate adjustments replace specie flows or changes in the quantity of domestically created money. And exchange rate changes too may be produced by actual or putative price differences or by short- or long-term capital flows.

And, of course, there are all kinds of mixtures of commodity and fiduciary standards. The most important in recent decades have been currencies linked by rates of exchange fixed, at least temporarily, by governments rather than by the commodity content of the different currencies. Though superficially similar to a unified currency, such fixed rates are fundamentally different since they contain no automatic mechanism for equilibrating adjustment. The adjustments often have taken the form of direct controls over foreign exchange transactions, subsidies to exports, and obstacles to imports, sometimes giving rise to an implicit multiple exchange rate system, sometimes effected through an explicit multiple rate system; government borrowing to finance balance-of-payment deficits, or governmental lending to offset surpluses; and, ultimately, exchange rate adjustments.¹²

For the purposes of this chapter, we shall neglect these complications and proceed throughout as if we are dealing with a closed economy or, equivalently, with a set of national economies using a unified currency. We shall return to the problem of the international transmission mechanism on an empirical level in chapter 7, where we consider the interrelations between the United States and the United Kingdom during the century that our study covers.

2.2.6 First-Round Effects

Another frequent criticism of the quantity equations is that they neglect any effect the source of change in the quantity of money may have on the outcome—in Tobin's words, the question is whether "the genesis of new money makes a difference," in particular, whether "an increase in the quantity of money has the same effect whether it is issued to purchase goods or to purchase bonds."¹³

12. M. Connolly and D. Taylor, "Adjustment to Devaluation with Money and Non-traded Goods," *Journal of International Economics* 6 (August 1976): 289–98; and idem, "Testing the Monetary Approach to Devaluation in Developing Countries," *Journal of Political Economy* 84, part 1 (August 1976): 849–59.

13. James Tobin, in Gordon, *Milton Friedman's Monetary Framework*, p. 87.

This criticism too is invalid on a purely theoretical level, but it raises an important question for research. On a theoretical level, there is no difficulty in allowing for the source of the change in the quantity of money by including the appropriate variables in the demand (or supply) function (e.g., the ratio of interest-bearing government debt to total wealth).

On an empirical level, the basic issue is ancient—whether the “first-round effect” of a change in the quantity of money largely determines the ultimate effect. As John Stuart Mill put a view very much like Tobin’s in 1844, “The issues of a *Government* paper, even when not permanent, will raise prices; because Governments usually issue their paper in purchases for consumption. If issued to pay off a portion of the national debt, we believe they would have no effect.”¹⁴

Ludwig von Mises in his theory of the cycle implicitly accepted a similar empirical judgment. For example, Lionel Robbins, in his Misesian analysis of the Great Depression, says, “In normal times, expansion and contraction of the money supply comes, not *via* the printing press and government decree, but *via* an expansion of credit through the banks. . . . This involves a mode of diffusion of new money radically different from the case we have just examined—a mode of diffusion which may have important effects.”¹⁵

Of course, Mill, von Mises, and Tobin are right that the way the quantity of money is increased will affect the outcome in some measure or other. If one group of individuals receives the money on the first round, they will likely use it for different purposes than another group of individuals. If the newly printed money is spent on the first round for goods and services, it adds directly at that point to the demand for such goods and services, whereas if it is spent on purchasing debt it has no such immediate effect on the demand for goods and services. Effects on the demand for goods and services come later as the initial recipients of the “new” money themselves dispose of it. Clearly, also, as the “new” money spreads through the economy, any first-round effects will tend to be dissipated. The “new” money will be merged with the old and will be distributed in much the same way.

One way to characterize the Keynesian approach (sec. 2.5) is that it gives almost exclusive importance to the first-round effect. This leads to emphasis primarily on flows of spending rather than on stocks of assets. Similarly, one way to characterize the quantity-theory approach is to say that it gives almost no importance to first-round effects.

The empirical question is how important the first-round effects are compared with the ultimate effects. Theory cannot answer that question.

14. See John Stuart Mill, “Review of Books by Thomas Tooke and R. Torrens,” *Westminster Review* 41 (June 1844): 579–93; quotation, p. 589.

15. See Lionel Robbins, *The Great Depression* (London: Macmillan, 1934), pp. 35–36.

The answer depends on how different are the reactions of the recipients of cash via alternative routes, on how rapidly a larger money stock is distributed through the economy, on how long it stays at each point in the economy, on how much the demand for money depends on the structure of government liabilities, and so on. Casual empiricism yields no decisive answer. Tobin can say, "The monetization of commercial loans . . . seems to me to be alchemy of much deeper significance than semi-monetization of Treasury bills."¹⁶ But we could answer, "True, but remember that the transactions velocity of money may well be twenty-five to thirty or more times a year, to judge from the turnover of bank deposits. So the first round covers at most a two-week period, whereas the money continues circulating indefinitely." Maybe the first-round effect is so strong that it dominates later effects; but maybe it is highly transitory. We shall have to examine empirical evidence systematically to find out.

Despite repeated assertions by various authors that the first-round effect is significant in this sense, none, so far as we know, has presented any systematic empirical evidence to support that assertion. The apparently similar response of income to changes in the quantity of money at widely separated dates in different countries and under diverse monetary systems seems to us to establish something of a presumption that the first-round effect is not highly significant. More recently, several empirical studies designed explicitly to test the importance of the first-round effect have supported this presumption.¹⁷

Perhaps other studies will reverse this tentative conclusion. In any event, the importance of the first-round effect will be provided by empirical evidence, not by argumentation or theory.¹⁸

16. See James Tobin, "The Monetary Interpretation of History," *American Economic Review* 55 (June 1965): 464-85; quotation, p. 467.

17. Cagan investigated the first-round effect on interest rates. He was able to identify the existence of such an effect, but it was of minor quantitative importance. Auerbach found no evidence of a first-round effect on nominal income of the division of the change in the quantity of money between high-powered money and bank credit, or the division of high-powered money between financing current government expenses and debt redemption. Bordo, in a thesis dealing with the pre-World War I period for the United States, found at best very limited traces of the first-round effect. See Robert Auerbach, "The Income Effects of the Government Deficit," Ph.D. diss., University of Chicago, 1969; Phillip Cagan, *The Channels of Monetary Effects on Interest Rates* (New York: NBER, 1972); Michael Bordo, "The Effects of the Sources of Change in the Money Supply on the Level of Economic Activity," Ph.D. diss., University of Chicago, 1972.

18. In a more recent article, "Monetary Policies and the Economy: The Transmission Mechanism," James Tobin, on the basis of his approach, which stresses the role in the transmission process of the credit markets and of the ratio of the market value to the replacement value of physical capital, again concludes that "the effects of an expansion of monetary aggregates depends on how it is brought about," and that "inside money is . . . more powerful stuff than outside money." The only empirical evidence cited refers to the influence of the ratio of the market to the replacement value of investment. However, there

2.3 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 that 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—determined for a country that has a fiduciary standard by the monetary authorities, for any one country under an international commodity standard through the balance of payments; (2) the ratio of bank deposits to bank holdings of high-powered money—determined by the banking system subject to any requirements that are imposed on them by law or the monetary authorities; and (3) the ratio of the public's deposits to its currency holdings—determined by the public subject to any controls on interest rates imposed by law or the monetary authorities.¹⁹

These factors determine the *nominal*, but not the *real* quantity of money. The *real* quantity of money is determined by the interaction between the nominal quantity supplied and the *real* quantity demanded and, in our view, ultimately by demand rather than supply. In the process, changes in demand for real balances have feedback effects on the variables determining the nominal quantity supplied, and changes in nominal supply have feedback effects on the variables determining the real quantity demanded. In our judgment, these feedback effects are for the most part relatively minor, so that the *nominal* supply can generally be regarded as determined by a set of variables distinct from those that affect the quantity of *real* balances demanded. In this sense the nominal quantity can be regarded as determined primarily by supply; the real quantity, by demand. Whether or not this judgment is correct, any discussion of the interrelation between demand and supply that neglects

is many a slip between that empirical result and his quoted conclusions, if those conclusions are regarded as referring to the effect of changes in monetary aggregates on nominal income, prices, and output. Tobin does not refer to any empirical evidence to support such an interpretation of his conclusions. His thoughtful and sophisticated analysis is concerned almost wholly with the effect of credit institutions on the composition of output and the structure of interest rates. *Southern Economic Journal* 44 (January 1978): 421–31; quotations, p. 431.

19. High-powered money consists of specie and obligations of the monetary authorities that are used either as currency by the public or as reserves by the banks. The monetary authorities are the central bank and the Treasury. See Friedman and Schwartz, *A Monetary History*, pp. 776–98 and note 59, p. 50, for use of the term high-powered money as early as the 1930s by Federal Reserve research personnel; for analysis of the proximate determinants, see also Phillip Cagan, *Determinants and Effects of Changes in the Stock of Money, 1875–1960* (New York: Columbia University Press for NBER, 1965).

the distinction between the nominal and the real quantity of money is necessarily incomplete and misleading.²⁰

The preceding paragraph is another way of stating that part of our description of the quantity theory that asserts that "substantial changes in the supply of nominal balances can and frequently do occur independently of any change in demand." If this generalization were not valid, that is, if (a) the quantity of money supplied were a function of the same variables as the quantity demanded and (b) the supply function were as stable over time and place as the demand function, observed data on the quantity of money, nominal and real, and on the variables affecting the quantities of money supplied and demanded, would simply record random perturbations about the intersection of the stable demand and supply functions. A function calculated from such data could not then be regarded as an estimate of a demand function—in the jargon of econometrics, the demand function would not have been identified.²¹

20. This paragraph was stimulated by the criticism of the analysis of the Great Depression in our *Monetary History* by Peter Temin, *Did Monetary Forces Cause the Great Depression?* (New York: Norton, 1976), esp. pp. 14–27. Temin's analysis is basically flawed by his failure to make this distinction.

Consider, for example, one statement from his discussion of our analysis, in which we have inserted in brackets words to make clear the confusion between nominal and real in Temin's analysis: "There is nothing in the narrative in Chapter 7 of the *Monetary History* to refute the following story: Income [real or nominal?] and production fell from 1929 to 1933 for nonmonetary reasons. Since the demand for [real] money [balances] is a function of [real] income, the demand for [real] money [balances] fell also. To equilibrate the money market [i.e., credit market], either interest rates, the [nominal or real?] stock of money, or both, had to fall. And since the [nominal] supply of money was partly a function of the [nominal] interest rate, this movement down along the supply curve of [nominal] money meant a decrease in both [interest rates and the (nominal or real?) stock of money]" (ibid., p. 27).

Omit the bracketed words, and the statement seems eminently reasonable—but only because of the implicit identification of nominal and real magnitudes, and the confusion between the interest rate, which is the price of credit, and the inverse of the price level, which is the price of money. According to his story, what explains the sharp decline in prices? If production fell for independent reasons, and this led to a movement down the nominal supply curve of money, so that demand and supply were continuously equated, why should prices have fallen? How does a decline in the demand for real balances produce a decline in nominal supply? Would it not instead, in the first instance, lead to upward pressure on prices, so that on Temin's alternative story, prices should have risen rather than fallen?

Temin has succumbed to the Keynesian assumption that the price level is an institutional datum discussed in greater detail in section 2.5 below.

For a fuller examination of Temin's analysis, see Arthur E. Gandolfi and James Lothian, "Did Monetary Forces Cause the Great Depression?" *Journal of Money, Credit and Banking* 9 (November 1977), 679–91; and Anna J. Schwartz, "Understanding 1929–1933," in *The Great Depression Revisited*, ed. Karl Brunner (Boston: Martinus Nijhoff, 1981).

21. For one of the earliest and still pertinent discussions of this point, see E. J. Working, "What Do Statistical 'Demand Curves' Show?" *Quarterly Journal of Economics* 41 (Febru-

Our rejection of points *a* and *b* is of course an empirical finding, not something that can be justified by theoretical considerations alone. With respect to point *a*, theory suggests many possible links between the quantity of money supplied and both real income and interest rates. Changes in real income affect imports and exports, both directly and by altering domestic prices for given monetary growth. Under international financial arrangements embodying fixed exchange rates, these effects disturb the balance of payments unless offset by the appropriate changes in the quantity of money. Under a fiduciary or fiat standard embodying flexible exchange rates, changes in real income may affect systematically the behavior of the monetary authorities. Under both standards, changes in real income affect the reserve ratios desired by banks and the currency-deposit ratios desired by money holders; changes in interest rates similarly alter international capital flows, the behavior of monetary authorities, and reserve and currency ratios. However, these numerous effects by no means all run in the same direction. For example, under an international commodity standard, a higher rate of real growth adds directly to imports, which tends to reduce the rate of monetary growth. On the other hand, higher real growth has a downward effect on prices, which raises exports, tending to increase the rate of monetary growth; and higher real growth may produce either capital outflows or capital inflows. Changes in rates of interest similarly have effects in both directions. Moreover, time lags enter on both the demand and the supply side, and there is no reason to expect them to be the same. Hence, even if the same named variables were to enter importantly into the demand and the supply functions, the economically relevant variables might differ because differently dated.

With respect to point *b*, theoretical considerations suggest that the supply function depends on the financial structure—for example, will be different for a commodity standard and a fiduciary standard. Financial institutions have undergone major changes in the century our data cover and have differed in important respects between the United States and the United Kingdom, introducing changes in supply. However, many elements in the financial structure remained the same throughout the period and have been common to both countries. There is no way in principle to judge whether the changes over time and the differences between countries had major or minor effects on the supply function.

The findings of chapters 6 and 7 are indirect empirical evidence that neither point *a* nor point *b* can be accepted. A much larger body of

ary 1927): 212–35. A classic statement of the problem is given in T. C. Koopmans, "Identification Problems in Economic Model Construction," in *Studies in Econometric Method*, ed. W. C. Hood and T. C. Koopmans, Cowles Commission Monograph no. 14 (New York: Wiley, 1953). For a more elaborate statement, see Franklin M. Fisher, *The Identification Problem in Econometrics* (New York: McGraw-Hill, 1966).

evidence justifying the same conclusion is contained in our *Monetary History*, in Phillip Cagan's companion volume, *Determinants and Effects of Changes in the Stock of Money, 1875-1960*, and in studies of the supply of money by other scholars.²²

These studies have concluded that neither interest rates nor real income have a consistent and sizable influence on the nominal quantity of money supplied. The same result is implicit in the evidence we have summarized elsewhere that supports the conclusion that the cyclical relation between money and income reflects primarily an influence running from money to income, which dominates the reflex influence running the other way.²³

This prior body of evidence explains why we do not in this book explore systematically the supply function of money. We shall for the most part take it for granted that the nominal quantity of money available to be held is largely independent of the variables entering into the demand function—that, in the jargon of econometrics, it can be treated as an exogenous variable entering into the determination of such endogenous vari-

22. For the United States, in addition to Cagan, see Albert E. Burger, *The Money Supply Process* (Belmont, Calif.: Wadsworth, 1971), and Idem, "Money Stock Control," in *Controlling Monetary Aggregate II: The Implementation*, pp. 33-55, Conference Series no. 9 (Boston: Federal Reserve Bank of Boston, Feb. 1973); David I. Fand, "Some Implications of Money Supply Analysis," *American Economic Review Papers and Proceedings* 57 (May 1967): 380-400; J. R. Zecher, "An Evaluation of Four Econometric Models of the Financial Sector," *Federal Reserve Bank of Cleveland, Economic Papers* Dissertation Series no. 1, (January 1970); Robert H. Rasche, "A Review of Empirical Studies of the Money Supply Mechanism," *Federal Reserve Bank of Saint Louis Review* 54 (July 1972): 11-19.

For the United Kingdom, see A. R. Nobay, "A Model of the United Kingdom Monetary Authorities' Behaviour 1959-1969," paper presented at the February 1972 Money Study Group Conference; D. Fisher, "The Instruments of Monetary Policy and the Generalized Tradeoff Function for Britain, 1955-1968," *Manchester School of Economics and Social Studies* 38 (September 1970): 209-22; R. L. Crouch, "A Model of the United Kingdom Monetary Sector," *Econometrica* 35 (July-October 1967): 398-418. See also Harry G. Johnson and associates, eds., *Readings in British Monetary Economics*, (Oxford: Clarendon Press, 1972), sec. 3, "The Supply of Money," pp. 203-77; and Harold Black, "The Relative Importance of Determinants of the Money Supply: The British Case," *Journal of Monetary Economics* 1 (April 1975): 251-64.

23. "The Monetary Studies of the National Bureau," in *The National Bureau Enters Its Forty-Fifth Year* (44th Annual Report, 1964), pp. 7-25 (reprinted in Milton Friedman, *The Optimum Quantity of Money and Other Essays* [Chicago: Aldine, 1969], pp. 261-84).

See also Donald P. Tucker, "Dynamic Income Adjustment to Money-Supply Changes," *American Economic Review* 56 (June 1966): 433-49; Christopher A. Sims, "Money, Income, and Causality," *American Economic Review* 62 (September 1972): 540-52; E. L. Felge and D. K. Pearce, "The Casual Relationship between Money and Income: Some Caveats for Time Series Analysis," *Review of Economics and Statistics* 61 (November 1979): 521-33; David Laidler, "Monetarism: An Interpretation and an Assessment," University of Western Ontario Centre for the Study of International Economic Relations, Working Paper no. 8010 (July 1980).

ables as nominal income, prices, interest rates, and real income.²⁴ However, nothing essential would be altered if the nominal quantity of money supplied were expressed as a function of other variables, comparable to the demand function, provided that, insofar as the same variables enter the supply and demand functions, the functional relation between them and the quantity of money is different. For example, the supply function is frequently written as:

$$(6a) \quad M^s = h(R, Y),$$

where R is an interest rate or set of interest rates, either actual or anticipated, and Y is nominal income or NP_y . In the special case of M^s strictly exogenous, the supply function reduces to

$$(6b) \quad M^s = M_o.$$

The simple quantity theory then specifies that

$$(6c) \quad M^D = M^s,$$

where M^D is defined by equation (6), and M^s by either equation (6a) or equation (6b), and where, in the long run, the variable that equates demand and supply is the price level, though, in the short run, in so-called transition periods, other variables may also be affected.

24. Of course, that does not mean that the nominal quantity of money is not an endogenous variable from a different point of view. It simply means that the variables determining it are largely independent of the variables we are seeking to analyze. To put the matter differently, there is no "first cause." Whatever is taken for granted at one level of analysis itself requires explanation at a different level. The quantity of money is what it is at any time because antecedent circumstances have made it that amount rather than something else. These antecedent circumstances are a valid subject for examination. In such an examination the quantity of money would be treated as endogenous, and some other variables affecting it as exogenous—variables such as the balance of payments, the identity of the members of the Federal Open Market Committee, the operating procedures of the New York Federal Reserve Bank, and so on in endless variety. At a still deeper level of analysis, these other variables would be treated as endogenous, and so on in infinite regress.

A basic scientific problem is how to carve up a broad question into narrower sectors for investigation. The desideratum is to have sectors that are orthogonal to one another, in the sense that there is a minimum of interaction between them, so that the analysis of each can proceed independently. In terms of "endogeneity" and "exogeneity," this means that variables that are treated as exogenous for one sector should be determined in another sector by variables other than those regarded as endogenous in the first sector. The "recursive" systems analyzed extensively by Herman Wold are a particular example of systems satisfying this requirement. See Herman Wold, "Statistical Estimation of Economic Relationships," *Econometrica* 17, suppl. (July 1949): 1-22; also Herman Wold and R. H. Strotz, "Recursive vs. Non-Recursive Systems: An Attempt at Synthesis," *Econometrica* 28 (April 1960): 417-27.

2.4 The Demand for Money

J. M. Keynes's liquidity preference analysis (discussed further in sec. 2.5) reinforced the 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.²⁵

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.²⁶

2.4.1 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:

1. *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

25. Much attention has been devoted in the past decade or so to the so-called micro-foundations of money (see Robert J. Barro and Stanley Fischer, "Recent Developments in Monetary Theory," *Journal of Monetary Economics* 2, April 1976: 151-55). The aim has been to provide a deeper theoretical underpinning for the kind of demand functions we develop in this section (or an alternative to such functions) in terms of a general equilibrium analysis of individual utility maximizing choices. The aim is admirable but, like Walrasian general equilibrium analysis for the most part, we suspect that the return will be primarily in improving our "analytical filing box" rather than in generating substantive hypotheses about economic phenomena (see M. Friedman, "Leon Walras and His Economic System" *American Economic Review* 45 [December 1955]: 900-909).

Much of the work along this line is summarized and an extensive bibliography is provided in *Models of Monetary Economies*, Federal Reserve Bank of Minneapolis, 1980, containing the proceedings of a conference held in December 1978.

One particular theoretical construction for which its authors made extravagant claims is the attempt to base the theory of money on an overlapping generations model (see papers by Wallace; and by Cass, Okuno and Zilcha in *ibid.*). We share the view expressed by James Tobin in his cogent comments on these papers (*ibid.*, pp. 83-90) that, as it has been developed so far at least, this model is *not* "the key to the theory of money." On the contrary, in our view this ingenious and subtle model abstracts from what we regard as the essential role of money. What is left may be of interest in other contexts but not for the theory of money.

26. See Milton Friedman, "The Quantity Theory of Money—A Restatement," in *Studies in the Quantity Theory of Money*, ed. M. Friedman (Chicago: University of Chicago Press, 1956); reprinted in *The Optimum Quantity of Money* (Chicago: Aldine, 1969).

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.²⁷

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 the more recent analyses of the demand for money and the earlier versions of the quantity theory.

2. *The division of wealth between human and nonhuman forms.* The major asset of most wealth holders is personal earning capacity. However, 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 of human wealth 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.

3. *The expected rates of return on money and other assets.* These rates of return are the counterparts to 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, a change in the nominal price of the asset. The second part will, of course, be especially important under conditions of inflation or deflation.

4. *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

27. See M. Friedman, *A Theory of the Consumption Function* (Princeton: Princeton University Press for NBER, 1957); idem "The Demand for Money: Some Theoretical and Empirical Results," *Journal of Political Economy* 67 (August 1959): 327–51, reprinted as Occasional Paper no. 68 (New York: NBER), and in *The Optimum Quantity of Money* (Chicago: Aldine, 1969); Karl Brunner and Allan H. Meltzer, "Predicting Velocity: Implications for Theory and Policy," *Journal of Finance* 18 (May 1963): 319–54; Allan H. Meltzer, "The Demand for Money: The Evidence from the Time Series," *Journal of Political Economy* 71 (June 1963): 219–46.

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 war is often accompanied by a notable increase in real balances—that is, a notable decline in velocity.

The rate of inflation enters under item 3 as a factor affecting the cost of holding various assets, particularly currency. The variability of inflation enters here, as a major factor affecting the usefulness of money balances. Empirically, variability of inflation tends to increase with the level of inflation, reinforcing the negative effect of higher inflation on the quantity of money demanded.

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:

$$(7) \quad M/P = f(y, w; R_M^*, R_B^*, R_E^*, g_P^*; u),$$

where M , P , and y have the same meaning as in equation (6) except that they relate to a single wealth holder (for whom $y = y'$); w is the fraction of wealth in nonhuman form (or, alternatively, the fraction of income derived from property): an asterisk denotes an expected value, so 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; $g_P^* = (1/P) (dP/dt)^*$ is the expected rate of change of prices of goods and hence the expected nominal rate of return on physical assets in addition to any direct income they yield (or storage costs they impose);²⁸ and u is a portmanteau symbol standing for whatever variables other than income may affect the utility

28. See also the discussion in section 6.6.3. We use the term physical assets in contrast to nominal assets to refer to all sources of permanent income, whether they are tangible assets,

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.²⁹

The usual problems of aggregation arise in passing from equation (7) to a corresponding equation for the economy as a whole—in particular, 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 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 .

2.4.2 Demand by Business Enterprises

Business enterprises are not subject to a constraint comparable to 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 an enterprise can determine 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, on different grounds, a somewhat similar variable defining the “scale” of the enterprise—namely, as an index of the productive value of different quantities of money to the enterprise, in line with the earlier transactions approach emphasizing the “work” to be done by money. It is by no means clear

such as factories, buildings, or the like; or intangible assets such as goodwill or the productive capacities of human beings.

We shall use g to refer to the percentage rate of change of the variable designated by a subscript.

29. Under some assumed conditions, the four rates of return may not be independent. For example, in a special case considered in Friedman, “The Quantity Theory of Money—a Restatement,” pp. 9–10,

$$R_B = R_E.$$

Note that R_E is here defined differently than r_e was in the source here cited; r_e there referred to the real, not nominal, return on equities.

what the appropriate variable is: for example, total transactions, net value added, net income, total capital in nonmoney form, or net worth. The lack 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, the rates banks charge 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 subset of such variables—namely, expectations about economic stability and the variability of inflation—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.

2.5 The Keynesian Challenge to the Quantity Theory

The income-expenditure analysis developed by John Maynard Keynes 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 nominal income and the stock of money.³⁰

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

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

30. *The General Theory of Employment, Interest and Money* (London: Macmillan, 1936).

2. As an *empirical* matter, prices can be regarded as rigid—an institutional datum—for *short-run economic fluctuations*; that is, the distinction between real and nominal magnitudes that is at the heart of the quantity theory is not important for such fluctuations.

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, in the main, changes in the quantity of money frequently produce offsetting changes in V . This proposition is critical for the other two, 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.

2.5.1 Long-Run Equilibrium

Though this book is about monetary trends, and hence the first proposition about long-run equilibrium is particularly relevant, that 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.³¹ 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 a natural outcome of a fully operative market mechanism.

This proposition played a large role in gaining for Keynes the adherence of many noneconomists, particularly the large band of reformers,

31. 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 Gottfried Haberler, *Prosperity and Depression* 3d ed. (Geneva: League of Nations, 1941), pp. 242, 389, 403, 491–503; and subsequently by Arthur C. Pigou, “Economic Progress in a Stable Environment,” *Economica*, n.s., 14 (August 1947): 180–88; James Tobin, “Money Wage Rates and Employment,” in *The New Economics*, ed. Seymour Harris (New York: Knopf, 1947); Don Patinkin, “Price Flexibility and Full Employment,” in *Readings in Monetary Theory*, ed. F. A. Lutz and L. W. Mints (Homewood, Ill.: Irwin, 1951), a revised version of an article that appeared in *American Economic Review* 38 (September 1948): 543–64; Harry G. Johnson, “The General Theory after Twenty-Five Years,” *American Economic Association Papers and Proceedings* 51 (May 1961): 1–17.

social critics, and radicals who were persuaded that there was something fundamentally wrong with the capitalist "system." There is a long history, going back at least to Malthus, of attempts, some highly sophisticated, to demonstrate that there is a "flaw in the price system."³² In modern times, one of the most popular and persistent attempts is the "social credit" doctrine of Major C. H. Douglas, which even spawned a political party in Canada that in 1935 captured control of the government of the Canadian province of 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 controlled Alberta until 1971, when it gave way to the Progressive Conservative Party; it controlled British Columbia for most of the period from 1952 to the present (1980). However, while retaining the name the successor party rejected the basic social-credit doctrine. Before Keynes these attempts had been made primarily by persons outside the mainstream of the economics profession, and professional economists had little trouble 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.

2.5.2 Short-Run Price Rigidity

Alfred Marshall's distinction among market equilibrium, short-period equilibrium, and long-period equilibrium was a device for analyzing the dynamic adjustment in a particular market to a change in demand or supply.³³ This device had two key characteristics. One, the less important

32. The title of one such attempt by P. W. Martin, *The Flaw in the Price System* (London: King, 1924).

33. We are indebted to a brilliant book by Axel Leijonhufvud, *On Keynesian Economics and the Economics of Keynes* (London: Oxford University Press, 1968), for a full appreciation of the importance of this proposition in the Keynesian system. This subsection and the one that follows, on the liquidity preference function, owe much to Leijonhufvud's penetrating analysis.

for our purposes, is that it replaced the continuous process by a series of discrete steps—comparable to 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. In the long run, it will 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.³⁴ 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 N , 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 , N , and y ultimately depend.³⁵ 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 initial adjustment (Marshall’s market equilibrium) to take place through P rather than through k or y (it clearly is unlikely to affect N in any short period). It was widely recognized that the adjustment during what Fisher, for example, called “transition periods” would in practice be partly in k and 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,

34. Pigou, “Economic Progress.”

35. 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.

even though he recognized in his asides that wealth has an effect on consumption.³⁶

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

36. In an article, "On the Short-Run Non-Neutrality of Money in the Quantity Theory," *Banca Nazionale del Lavoro Quarterly Review* 100 (March 1972): 3-22, Don Patinkin cites evidence that he regards as decisively contradicting the interpretation that Fisher and quantity theorists "simply took over Marshall's assumption" that "prices adjust more rapidly than quantities" (as contended in Friedman, "A Theoretical Framework for Monetary Analysis," *Journal of Political Economy* 78 [March/April 1970]: 207-8, and Gordon, *Milton Friedman's Monetary Framework*, p. 17). Yet we regard the evidence Patinkin cites as strikingly confirming our interpretation. One sample of his evidence will do:

"The sequence of effects visualized by Fisher" after an increase in the quantity of money is as follows:

"1. Prices rise.

"2. Velocities of circulation (V and V') increase; the rate of interest rises, but not sufficiently.

"3. Profits increase, loans expand, and the Q 's [i.e., the real volume of trade] increase.

"4. Deposit currency (M') expands relatively to money (M).

"5. Prices continue to rise; that is, phenomenon No. 1 is repeated. Then No. 2 is repeated, and so on."

Is not Fisher's sequence precisely the counterpart for the aggregate to Marshall's analysis for a particular product summarized in the second paragraph before the one to which this footnote is attached?

Further proof is that just before listing the five steps that Patinkin quotes, Fisher states that "an increase in currency cannot, even temporarily, very greatly increase trade. . . . almost the entire effect of an increase of deposits must be seen in a change of prices" (Fisher, *Purchasing Power of Money*, pp. 62-63).

Consider how a Keynesian would describe the effects of an increase in the quantity of money. It would go:

1. Interest rates fall.

2. Investment increases.

3. Output and real income increases.

4. Consumption increases.

It is not clear when he would come to the statement "prices rise," but it would surely be late in his list. Moreover, his step 1 implies that velocity falls, but he would be most unlikely ever to refer to that phenomenon.

Is this not precisely the contrast that we draw between the quantity theorists and the Keynesians when we say that Keynes "deviated from Marshall . . . in reversing the roles assigned to price and quantity"? References in Patinkin's article to statements by Pigou, Keynes, Robertson, Lavington, and Chicago economists, all equally strike us as confirming our interpretation.

Patinkin also criticizes our assertion that "this recognition was not incorporated in formal theoretical analysis," asserting, "The facts of the case, however, are quite different," and giving as evidence that "Fisher wrote incomparably more on his monetary proposals for mitigating the cyclical problems of the 'transition period' than on the long-run proportionality of prices to money. This concentration on short-run analysis was even more true for the policy-oriented Chicago quantity-theory school of the 1930s and 1940s."

However, there can be a great difference between what is implied by or contained in a formal theory, what proponents of the theory may believe it implies or contains, and what

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.³⁷

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, at least downward.³⁸ Keynes embodied this assumption in his formal model by expressing all variables in wage

they write about. Of course Fisher, the Chicago monetary economists, and the host of other economists who studied business cycles wrote a great deal about short-run movements and constructed many ingenious theories about business cycles that have much to teach us. In particular, Fisher's distinction between nominal and real interest rates, which dates back to some of his earliest writing, remains a seminal and penetrating insight. Yet, so far as we know, none of this voluminous writing and none of these theories provide a formal theoretical extension of the quantity theory to explain the division of changes in nominal income between changes in prices and in output or of changes in the quantity of money between changes in velocity, in prices, and in output, just as none of Keynes's extensive discussion of changes in money-wage rates before the point of full employment provides a formal theoretical analysis of such changes.

37. This famous quoted remark is from Keynes, *A Tract on Monetary Reform* (London: Macmillan, 1923), reprinted as volume 4 of *The Collected Works of John Maynard Keynes* (London: Macmillan, 1971), p. 65. See the full quotation in context in chapter 6, note 11, below.

38. The reference to "quantity," not "output," is based on the 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.

The statement that Keynes assumed prices rigid is an oversimplification, since he distinguished between the price level of products and the wage rate and allowed for a change in the ratio of prices to wages, even before the point of full employment. However, this change in prices in wage-units plays no important role in the aspects of his theory that are relevant to our purposes, so we have simplified our analysis of Keynes's theory by regarding prices as well as wages as rigid—a simplification that has been widely used. (Explicit reference to this simplification should have been made in M. Friedman, "Theoretical Framework for Monetary Analysis." We are indebted to an unpublished paper by Paul Davidson for recognition that the exposition on this point in that source may have been misleading.)

Keynes himself minimized the importance of changes in prices relative to wages, noting that, "This policy [of maintaining the money wage level as a whole as stable as possible] will result in a fair degree of stability in the price-level. . . . Apart from 'administered' or monopoly prices, the price-level will only change in the short period in response to the extent that changes in the volume of employment affect marginal prime costs; whilst in the

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.³⁹ 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 in such circumstances had no economic function to perform; hence nominal wages might as well be made rigid.

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, experience did not correspond. 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.⁴⁰

long period they will only change in response to changes in the cost of production due to new technique and new or increased equipment” (*General Theory*, p. 270).

Leijonhufvud, in response to criticism by Herschel I. Grossman, has retracted his initial position on Keynes’s assumptions, writing, “it is *not* correct to attribute to Keynes a general reversal of the Marshallian ranking of relative price and quantity adjustment velocities. In the ‘shortest run’ for which system behavior can be defined in Keynes’ model, output-prices *must* be treated as perfectly flexible.” See Grossman, “Was Keynes a ‘Keynesian’? A Review Article,” *Journal of Economic Literature* 10 (March 1972): 26–30; Leijonhufvud, “Keynes’ Employment Function,” *History of Political Economy* 6 (1974): 158–70; quotation from p. 169.

Leijonhufvud’s textual exegesis is correct, and relevant to Keynes’s employment function, but it does not alter the role that the reversal of the Marshallian ranking of relative price and quantity adjustments played in Keynes’s theory. In the “shortest run” that Leijonhufvud refers to, the elasticity of demand for labor is high, so that, as Keynes noted, prices in wage-units will be highly stable.

And, whatever may be true for Keynes himself, there is no doubt that his followers who shaped much of economic thinking since *The General Theory* appeared, took product prices as well as wages as determined by forces outside those dealt with in Keynesian theory (see footnotes 43 and 45 below).

39. Keynes, *General Theory*, pp. 119, 301, 303.

40. Marshall’s assumption is clearly not 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 adjust-

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 such a situation one 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 , N , or 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, a purely monetary theory conflicted with Keynes's interpretation of the Great Depression, which he regarded, we believe erroneously, as showing that expansionary monetary policy was ineffective in stemming a decline.⁴¹ Hence he was inclined to interpret changes in M as reflected in k rather more than in y' . This is where his proposition 3 about liquidity preference enters in.

Indeed, in the most extreme, and we are 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 the interest rate as slow to adjust and to take, as the variable that 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 investment, the amount some people want to add to the stock of productive capital, with savings, the amount the community wants to add to its stock of wealth. Hence Keynes puts at the center of his analysis the distinction between investment and consumption, or more fundamentally between spending that is largely independent of current income and spending linked closely to current income.

ment 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 2.6 below.

41. See Milton Friedman, “The Monetary Theory and Policy of Henry Simons,” *Journal of Law and Economics* 10 (October 1967): 1–13; reprinted in *Optimum Quantity of Money*, pp. 81–93.

As a result of both experience and further theoretical analysis, hardly an economist today 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 will explicitly assert that P is "really" an institutional datum that will be completely unaffected, even in short periods, by changes in M .⁴²

Yet Keynes's assumption about the relative speed of adjustment of price and quantity remains 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 that 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 several decades after the appearance of *The General Theory*.

The statement that Keynes and his followers "treat the price level as an institutional datum in their formal theoretical analysis" does not mean they assert that prices and wages are in fact constant, or even that in their empirical work they do not introduce relations designed to predict the movements of prices and wages. Treating the price level or the wage level as an institutional datum, or, as Keynes did, as the "numeraire," is not equivalent to asserting that wages or prices are constant. It means, rather, that the theory in question has nothing to say about what determines the wage level; that the forces determining the wage level are forces abstracted from in the theory. This assumption is reflected in the kind of ad hoc relations Keynesians introduce into their empirical work to predict prices and wages.⁴³

42. Milton Friedman, "Money: Quantity Theory," *International Encyclopedia of the Social Sciences; The Counter-Revolution in Monetary Theory*, Institute of Economic Affairs for the Wincott Foundation, Occasional Paper 33 (London: Tonbridge, 1970).

43. The price equations generally simply link prices to costs, mainly wages. This equation can be regarded as derivable from Keynes's system. But the wage equations are either purely ad hoc or, insofar as they are derivable from any theoretical system, it is the pre-Keynesian classical system rather than Keynes's. For example, Patinkin (Gordon, *Milton Friedman's Monetary Framework*, p. 128) refers approvingly to Lawrence Klein's comment that "the main reasoning behind this equation is that of the law of supply and demand. Money wage rates move in response to excess demand on the labor market." The "law of supply and demand" is hardly Keynesian! More important, Klein misapplies it. The "classical law," as taken over by Keynes, connects *real-wage rates*, not *money-wage rates*,

It is important to distinguish between the logical implications of a theory and the statements about observable phenomena that a professed adherent of the theory may make. As Keynes says, "We can keep 'at the back of our heads' the necessary reserves and qualifications and the adjustments which we shall have to make later on."⁴⁴ Of course, both the Keynesians and Keynes himself recognize that, as a factual matter, changes in income are partly in prices and partly in output; and, of course, both have instructive ideas and insights about the factors that determine the division in particular cases. But Keynes's formal theory has nothing to say about what determines the absolute price or wage level, though it does have some implications for the behavior of prices relative to wages.⁴⁵

with excess supply or demand. Klein's inclusion of the rate of change of prices in the equation, which Patinkin cites, is a move toward the correct classical inclusion of real wages, but if it went wholly in that direction it would leave money wages and money prices either undetermined or a simple inheritance from past history—which is precisely what we say Keynes's system assumes.

44. Keynes, *General Theory*, p. 297.

45. For a fuller discussion of this point, see Gordon, *Milton Friedman's Monetary Framework*, pp. 79–80, 93–94, 127–29, 143, 155–57, 176–77.

A striking illustration of the Keynesian tendency to treat the price level as an institutional datum is provided in Cowles Foundation Monograph 21, *Financial Markets and Economic Activity* (New York: Wiley 1967). A key essay in that book presents a comparative static analysis of the general equilibrium adjustment of stocks of assets (W. C. Brainard and J. Tobin, "Financial Intermediaries and the Effectiveness of Monetary Controls," *ibid.*, pp. 55–93). 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.

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 (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" (*ibid.*, pp. 94–141), 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 example is a paper by the same authors, "Pitfalls in Financial Model Building" (*American Economic Association Papers and Proceedings* 58 (May 1968): 99–122), 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 Lyle Gramley and Samuel B. Chase, "Time Deposits in Monetary Analysis," *Federal Reserve Bulletin* 51 (October 1965): 1380–1406, reprinted in Karl Brunner, ed., *Targets and Indicators of Monetary Policy* (San Francisco: Chandler, 1969), pp. 219–49. In this article the assumption about price rigidity is explicit and presented

The NBER series of monetary studies, including this volume, 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 those monographs would never have been written, had we, implicitly or explicitly, accepted Keynes's assumption that prices are an institutional datum.

2.5.3 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 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."⁴⁶ 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.⁴⁷ 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 that were 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 a difference in risk of capital gain or loss as a result of a change in the interest rate. The capital value of short-term securities is not much affected by a change in the interest rate; the capital value of long-term securities is. Leijonhufvud

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 Stephen M. Goldfeld, *Commercial Bank Behavior and Economic Activity* (Amsterdam: North-Holland, 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 Franklyn D. Holzman and Martin Bronfenbrenner, "Survey of Inflation Theory," *American Economic Review* 53 (September 1963): 593–661. Theories of inflation stemming from the Keynesian approach stress institutional, not monetary, factors.

An even more striking example is Peter Temin's attack on our interpretation of the Great Depression. His central criticism is marred precisely by the implicit identification of nominal and real magnitudes (see footnote 20 above).

46. *General Theory*, p. 199.

47. *Ibid.*, p. 168; italics in original.

has argued, 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 provide "liquidity" in the sense of security against capital loss arising from a change in the interest rate.⁴⁸ 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 a change in the price level of goods and services.⁴⁹

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.)

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

$$(8) \quad M/P = M_1/P + M_2/P = k_1 y' + f(R - R^*, R^*),$$

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.⁵⁰ 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 we have introduced P into the equation for consistency with our 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.

48. In this respect the Radcliffe Committee was faithful to Keynes in treating "liquidity" broadly defined as the relevant monetary aggregate rather than "money" narrowly defined. (Radcliffe) Committee on the Working of the Monetary System, 1959. *Report*. Cmd. 827.

49. Leijonhufvud, *On Keynesian Economics*, chap. 2.

50. Later writers in this tradition have argued that k_1 too should be regarded as a function of interest rates. See W. J. Baumol, "The Transactions Demand for Cash: An Inventory Theoretic Approach," *Quarterly Journal of Economics* 66 (November 1952): 545-56; James Tobin, "The Interest-Elasticity of Transactions Demand for Cash," *Review of Economics and Statistics* 38 (August 1956): 241-47. However, this issue is not relevant to the present discussion.

In a "given state of expectations," that is, for a given value of R^* , the higher 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 about interest rates are given great prominence in developing the liquidity function expressing the demand for M_2 , Keynes and his followers generally did not explicitly introduce an expected interest rate into that function, as we have done. 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 attributing instability to the liquidity function.⁵¹

The reason for the omission of the expected interest rate 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. We have introduced R^* to distinguish between the different reasons that are implicit in Keynes's analysis for absolute liquidity preference in the short run and the long run.

Keynes's special twist was less expressing the demand function in the general form described by equation (8) than the particular form he gave to the function $f(R - R^*, R^*)$. For given R^* , he believed that this function was highly elastic at $R = R^*$, the degree of elasticity at an observed numerical value of R depending on how homogeneous the expectations of different holders of money are and how firmly they are held.⁵² Let there be a substantial body of holders of money who have the same expectation and let them hold that expectation firmly, and the function f would become perfectly elastic at that current interest rate. Money and bonds would become perfect substitutes; liquidity preference would become absolute. The monetary authorities would find it impossible to change the interest rate because speculators holding these firm expectations would frustrate them.

An attempt by the monetary authorities to increase the amount of money by buying bonds tends to raise bond prices and lower the rate of return. Even the slightest lowering would, Keynes argued, lead specula-

51. A notable exception is James Tobin, "Liquidity Preference as Behavior towards Risk," *Review of Economic Studies* 25 (February 1958): 65-86.

52. Tobin, "Liquidity Preference," presents an excellent and illuminating analysis of this case. Because he assumes that shifts into or out of securities involve commitments for a finite period equal to the unit of time in terms of which the interest rate is expressed, his critical value is not $R = R^*$ but $R = R^*/(1 + R^*)$, current income on the securities compensating for an expected capital loss.

tors with firm expectations to absorb the additional money balances and sell any bonds demanded by the initial holders of the additional money. The result would simply be that the community as a whole would be willing to hold the increased quantity of money at an essentially unchanged interest rate; k would be higher and V lower. Conversely, an attempt by the monetary authorities to decrease the amount of money by selling bonds would tend to raise the rate of interest, and even the slightest rise would induce the speculators to absorb the bonds offered.⁵³

Or, again, suppose there is an increase in nominal income for whatever reason. That will require an increase in M_1 , which can come out of M_2 without any further effects. Conversely, any decline in M_1 can be added to M_2 without any further effects. The conclusion is that *in circumstances of absolute liquidity preference*, income can change without a change in M or in interest rates and M can change without a change in income or in interest rates. The holders of money are in metastable equilibrium, like a tumbler on its side on a flat surface; they will be satisfied with whatever the amount of money happens to be.

For the long-run demand schedule, the reason for absolute liquidity preference is different. In long-run equilibrium, R must equal R^* , so $f(R - R^*, R^*)$ reduces to a function of R^* alone. Let there be a deficiency of investment opportunities, the kind of situation envisaged in Keynes's proposition 1, so that R^* becomes very low. The lower the rate, the lower the returns from capital assets other than money—whether these be bonds, equities, or physical assets (recall that because of the assumption that the price level is rigid, Keynes did not regard the distinction among these assets as important). Accordingly, the lower R^* , the lower the cost of holding money. At a sufficiently low, yet finite rate, the extra return from holding nonmoney assets would only just compensate for the extra risks involved. Hence at that rate liquidity preference would be absolute. The “market rate” of interest could not be indefinitely low; a bottom limit was set by the widespread desire to substitute money for other assets at low interest rates.

This conclusion was a key element in Keynes's proposition 1. One way to summarize his argument for that proposition is in terms of a possible conflict between the “market” and the “equilibrium” rate of interest. If investment opportunities were sparse, yet the public's desire to save were strong, the “equilibrium” rate of interest, he argued, might have to be very low or even negative to equate investment and saving. But there was a floor to the “market rate” set by liquidity preference. If this floor exceeded the “equilibrium rate,” he argued, there was a conflict that

53. In Keynes's analysis, the result would be the same if the amount of money were increased or decreased by operations that added to or subtracted from total wealth, rather than by substituting one form of wealth for another, because he assumed that wealth had no direct effect on spending.

could be resolved only by unemployment that frustrated the public's thriftiness. The fallacy in this argument is that the introduction of money not only introduces a floor to the "market rate"; it also sets a floor to the "equilibrium rate." And, in the long run, the two floors are identical. This is the essence of the so-called Pigou effect.⁵⁴

Neither Keynes himself nor most of his followers distinguished as sharply as we have between the short-run and long-run liquidity traps. They tended to merge the two and, in line with the general emphasis on the short run, to stress the elasticity of the demand for money with respect to current, not expected, interest rates.⁵⁵

Keynes regarded absolute liquidity preference as a strictly "limiting case" of which, though it "might become practically important in future," he knew "of no example . . . hitherto." However, he treated velocity as if in practice its behavior frequently approximated that which would prevail in this limiting case.⁵⁶

Economists no longer explicitly avow absolute liquidity preference. The failure of repeated attempts by central banks to peg interest rates at low levels has made that proposition untenable. No Keynesian can anymore say, as Keynes did in the sentence immediately following that quoted in the preceding paragraph, "Indeed, owing to the unwillingness of most monetary authorities to deal boldly in debts of long term, there has not been much opportunity for a test [of absolute liquidity preference]."⁵⁷ Yet, like absolutely rigid prices, absolute liquidity preference still plays an important role in the theorizing of many an economist. It is implicit in the tendency to regard k or velocity as passively adjusting to changes in the quantity of money. It is explicit in the tendency to regard the demand for money as "highly" elastic with respect to interest rates.

Consider again equation (6). Let there be a change in M . Economists in the Keynesian tradition continue, as we noted earlier, to regard P as an institutional datum and so unaffected. They must therefore regard the change in M as affecting k or N or y . With absolute liquidity preference, k can absorb the impact without any change in the interest rate. Since they take the interest rate as the only link between monetary change and real income, the whole of the change would then be absorbed in k with no

54. See Milton Friedman, *Price Theory* (Chicago: Aldine, 1962), pp. 262–63; (2d ed., Chicago: Aldine, 1976), pp. 313–15.

55. Tobin makes an explicit distinction of this kind, though not in connection with a liquidity trap as such.

56. *General Theory*, p. 207.

In his criticism of an earlier version of this chapter, Patinkin objects to the key role we assign to absolute liquidity preference in our interpretation of Keynes, citing as evidence solely the quotation in the prior sentence of the text. Friedman's reply cites thirteen quotations from *The General Theory* supporting our interpretation. Gordon, *Milton Friedman's Monetary Framework*, pp. 129–30, 168–70, 175–76.

57. *General Theory*, p. 207.

effect on N or y . If liquidity preference is not absolute, k can change only through a change in the interest rate. But a change in the interest rate affects Ny through investment spending. The more elastic is the demand for money, the less interest rates will have to change. The more inelastic are investment spending and saving with respect to the interest rate, the less will any given change in the interest rate affect y' . Hence the tendency for these economists to regard k as absorbing the main impact of changes in M means that implicitly or explicitly they regard the demand for money as highly elastic with respect to the interest rate and investment spending and saving as highly inelastic.

The tendency on the part of many economists to assume implicitly that prices are an institutional datum and that the demand for money is highly elastic with respect to the interest rate underlies some of the criticisms that have been directed against our earlier work and that of some of our associates. We have been interpreted, wrongly, we believe, as saying that k is completely independent of interest rates.⁵⁸ In that case, changes in M need not be reflected at all in k . If, also, P is taken as an institutional datum, all of the effect will be on y' . This is the implicit source of the criticism leveled against us, that we regard the quantity of money as determining the level of economic activity. Not only, say our critics, do we believe that money matters, we believe that money is all that matters.⁵⁹ If P is not regarded as an institutional datum, and we have not so regarded it, then even if we supposed k to be completely insensitive to interest rates and to anything else that might be affected by changes in M (such as the rate of change in P or in y') and so to be an absolute constant, aside from random disturbances, something other than the quantity of money would have to be brought into the analysis to explain how much of the change in M would be reflected in P and how much in y' (see sec. 2.6).

We have always tried to qualify our statements about the importance of changes in M by referring to their effect on *nominal* income. But this qualification appeared meaningless to economists who implicitly identified nominal with real magnitudes. Hence they have misunderstood our conclusions.

We have accepted the quantity-theory presumption and have thought it supported by the evidence we examined, that changes in the quantity of money as such have a negligible effect *in the long run* on real income, so that nonmonetary forces are "all that matter" for changes in real income over the decades and money "does not matter." On the other hand, we

58. See M. Friedman, "Interest Rates and the Demand for Money," *Journal of Law and Economics* 9 (October 1966): 71-85; reprinted in M. Friedman, *Optimum Quantity of Money*, pp. 141-55.

59. See Arthur M. Okun, "Money and Business Cycles: A Comment," *Review of Economics and Statistics* 45 suppl. (1), part 2 (February 1963): 72-77; Tobin, "Monetary Interpretation of History," p. 481.

have regarded the quantity of money, plus the other variables (including real income itself) that affect k as essentially "all that matter" for the long-run determination of nominal income. The price level is then a joint outcome of the monetary forces determining nominal income and the real forces determining real income.⁶⁰

For shorter periods of time, we have argued that changes in M will be reflected in all variables on the right-hand side of equation (6): k , P , N , and y . But we have argued that the effect on k is empirically not to absorb the change in M , as the Keynesian analysis implies, but often to reinforce it, changes in M and k frequently affecting income in the same rather than opposite directions. Hence we have emphasized that changes in M are a major factor, though even then not the only factor, accounting for short-run changes in both nominal income and the real level of activity (y'). We regard the description of our position as "money is all that matters for changes in nominal income and for short-run changes in real income" as an exaggeration, but one that gives the right flavor of our conclusions. We regard the statement that "money is all that matters," period, as a basic misrepresentation of our conclusions.⁶¹

Another, more subtle difference between the approach of economists in the Keynesian tradition and the approach we have adopted has also contributed to much misunderstanding. This difference is in the transmission mechanism that is assumed to connect a change in the quantity of money with a change in total nominal income (= total spending). The Keynesians regard a change in the quantity of money as affecting in the first instance "the" interest rate, interpreted as a market rate on a fairly narrow class of financial liabilities. They regard spending as affected only "indirectly" as the changed interest rate alters the profitability and amount of investment spending, again interpreted fairly narrowly, and as investment spending, through the multiplier, affects total spending. Hence the emphasis they give in their analysis to the interest elasticities of the demand for money and of investment spending. We, on the other hand, stress a much broader and more "direct" impact on spending, saying, as in section 2.1, that individuals seeking "to dispose of what they regard as their excess money balances . . . will try to pay out a larger sum

60. See Milton Friedman, "The Supply of Money and Changes in Prices and Output," in United States Congress, Joint Economic Committee, Compendium, *The Relationship of Prices to Economic Stability and Growth* (1952), pp. 242-46; reprinted in M. Friedman, *Optimum Quantity of Money*, pp. 171-87; M. Friedman and A. J. Schwartz, *A Monetary History of the United States, 1867-1960* (Princeton: Princeton University Press for NBER, 1963), p. 695.

61. Friedman, "Supply of Money and Changes in Prices and Output," pp. 246-51; Friedman and Schwartz, *Monetary History*, pp. 678, 695; idem, "Money and Business Cycles," *Review of Economics and Statistics*, 45, suppl. (1), part 2 (February 1963): 38-39, 45-46, 55-64; reprinted in Friedman, *Optimum Quantity of Money*, pp. 189-235.

for the purchase of securities, goods, and services, for the repayment of debts, and as gifts than they are receiving from the corresponding sources.”⁶²

The two approaches can be readily reconciled on a formal level. The transmission mechanism we have stressed can be described as operating “through” the balance sheet and “through” changes in interest rates. The attempt by holders of money to restore or attain a desired balance sheet after an unexpected increase in the quantity of money will tend to raise the prices of assets and reduce interest rates, which will encourage spending to produce new assets and also spending on current services rather than on purchasing existing assets. This is how an initial effect on balance sheets gets translated into an effect on income and spending.

The difference between us and the Keynesians is less in the nature of the process than in the range of assets considered. The Keynesians tend to concentrate on a narrow range of marketable assets and recorded interest rates. We insist that a far wider range of assets and interest rates must be taken into account—such assets as durable and semi-durable consumer goods, structures, and other real property. As a result, we regard the market rates stressed by the Keynesians as only a small part of the total spectrum of rates that are relevant.⁶³

This difference in the assumed transmission mechanism is largely a by-product of the different assumptions about price. The rejection of absolute liquidity preference forced Keynes’s followers to let the interest rate be flexible. This chink in the key assumption that prices are an institutional datum was minimized by interpreting the “interest rate” narrowly, and market institutions made it easy to do so. After all, it is most unusual to quote the “interest rate” implicit in the sales and rental prices of houses and automobiles, let alone furniture, household appliances, clothes, and so on. Hence the prices of these items continued to be regarded as an institutional datum, which forced the transmission process to go through an extremely narrow channel. On our side there was no such inhibition. Since we regarded prices as flexible, though not “perfectly” flexible, it was natural for us to interpret the transmission mechanism in terms of relative price adjustments over a broad area rather than in terms of narrowly defined interest rates.

62. We have put “indirectly” and “direct” in quotes because this distinction, tirelessly repeated, is purely semantic and has no substantive content. What is regarded as “indirect” or “direct” depends simply on the theoretical structure that is found most convenient. For example, start with the quantity theory equations, and the effect of a change in the quantity of money on desired spending is “direct,” the effect on interest rates “indirect,” since it will be described as arising via the change in desired spending (as in the quotation to which this note is attached). Start with the Keynesian structure and the situation is reversed: the effect on interest rates is “direct,” the effect on desired spending is “indirect.”

63. See Milton Friedman, “The Lag in Effect of Monetary Policy,” *Journal of Political Economy* 69 (October 1961): 461–63; Milton Friedman and David Meiselman, “The Relative Stability of Monetary Velocity and the Investment Multiplier in the United States,

2.6 The Adjustment Process

In an earlier publication preliminary to this chapter, we outlined the elements common to simple quantity theory and Keynesian models, noting that these common elements form an incomplete system with one equation missing and that the key difference between the two theories is the assumption adopted to fill the gap. For the simple quantity theory, the assumption is that aggregate real income is determined outside the system; for the Keynesian theory, the assumption is that the nominal wage (and hence price) level is determined outside the system. We also sketched a third possibility, the assumption that the elasticity of demand for real balances with respect to real income is unity plus the twin assumptions that speculators determine the interest rate in accord with firmly held anticipations, and that the difference between the permanent real interest rate and the secular growth of output can be taken as a constant for short-period fluctuations. We called the third possibility a theory of nominal income, since it defines only the path of nominal income, not of prices and output separately.

Though we regard the third approach as distinctly superior to the other two, all three have the basic defect that they say nothing about the factors that determine the proportions in which a change in nominal income will, in the short run, be divided between price change and output change. In addition, the simple quantity and Keynesian approaches have nothing to say about the adjustment process and leave little room for anticipations to play a role. The monetary theory of nominal income is less unsatisfactory in these respects but shares with the other two the absence of a satisfactory link between short-run change and long-run adjustment.

To remedy the defects common to all three theories, the key is a theory that will explain (a) the short-run division of a change in nominal income between prices and output; (b) the short-run adjustment of nominal income to a change in autonomous variables; and (c) the transition between the short-run situation and a long-run equilibrium.⁶⁴

The central idea we shall use in sketching the direction in which such a theory might be developed is the distinction between actual and anticipated magnitudes or, to use a terminology that need not be identical but that we shall treat for this purpose as if it is, between measured and permanent magnitudes. At a long-run equilibrium position, all anticipa-

1897-1958," in *Stabilization Policies* ed. Commission on Money and Credit (Englewood Cliffs, N.J.: Prentice-Hall, 1963), pp. 217-22; Friedman and Schwartz, "Money and Business Cycles," pp. 59-63; M. Friedman, *Counter-Revolution in Monetary Theory*, pp. 24-25; Karl Brunner, "The 'Monetarist Revolution' in Monetary Theory," *Weltwirtschaftliches Archiv* 105, no. 1 (1970): 3-5.

64. Still other parts of the theoretical framework are developed more fully in the course of the analysis of specific issues in later chapters of this book.

tions are realized, so that actual and anticipated magnitudes, or measured and permanent magnitudes, are equal.⁶⁵

We shall regard long-run equilibrium as determined by the Walrasian equations of general equilibrium, which determine the real variables, plus the quantity theory, which, for the given real variables, determines the price level.

We shall regard short-run equilibrium as determined by an adjustment process in which the rate of adjustment in a variable is a function of the discrepancy between the measured and the anticipated values of that variable or its rate of change, as well as, perhaps, of other variables or their rates of change. Finally, we shall let at least some anticipated variables be determined by a feedback process from past observed values.

2.6.1 Division of a Change in Nominal Income between Prices and Output

It seems plausible that the division of a change in nominal income between prices and output depends on two major factors: anticipations about the behavior of prices—this is the inertia factor stressed by Keynes—and the current level of output or employment compared with the full-employment (permanent) level of output or employment—this is the supply-demand response stressed by quantity theorists. We can express this in general form as:

$$(9) \quad g_P = f[g_Y, g_P^*, g_{y'}^*, y', y'^*]$$

$$(10) \quad g_{y'} = j[g_Y, g_P^*, g_{y'}^*, y', y'^*],$$

where an asterisk attached to a variable denotes the anticipated value of that variable and where the form of equations (9) and (10) must be consistent with the identity

$$(11) \quad Y = P y',$$

so that only one of equations (9) and (10) is independent.

To illustrate, a specific linearized version of equations (9) and (10) might be

$$(12) \quad g_P = g_P^* + \eta(g_Y - g_Y^*) + \xi(\log y' - \log y'^*);$$

$$(13) \quad g_{y'} = g_{y'}^* + (1 - \eta)(g_Y - g_Y^*) - \xi(\log y' - \log y'^*).$$

65. Note that the equality of actual and anticipated magnitudes is a necessary but not a sufficient condition for a long-run equilibrium position. In principle, actual and anticipated magnitudes could be equal along an adjustment path between one equilibrium position and another. The corresponding proposition is more complicated for measured and permanent magnitudes and depends on the precise definition of these terms. However, since we shall be considering a special case in which the stated condition is treated as both necessary and sufficient for long-run equilibrium, these complications will be bypassed.

The sum of equations (12) and (13) is exactly the logarithm of equation (11), differentiated with respect to time, provided the anticipated variables also satisfy a corresponding identity,⁶⁶ so the equations satisfy the specified conditions.

The extreme quantity theory assumption that all the change in income is in prices, and that output is always at its permanent level, is obtained by setting $\eta = 1$ and $\xi = \infty$. An infinite value of ξ corresponds to "perfectly flexible prices" and assures that $y' = y'^*$. The unit value of η assures that prices absorb any change in nominal income, so that real income grows at its long-term rate of growth.⁶⁷

The extreme Keynesian assumption, that all the change in income is in output, so long as there is unemployment, and all in prices, once there is full employment, is obtained by setting $g_P^* = 0$, and $\eta = \xi = 0$ for $y' < y'^*$, and then shifting to the quantity theory specification of $\eta = 1$, $\xi = \infty$ for $y' \geq y'^*$. The zero value of g_P^* assures that anticipations are for stable prices and, combined with the zero values of η and ξ , that $g_P = 0$. It would be somewhat more general, and perhaps more consistent with the spirit rather than the letter of Keynes's analysis, and even more that of his modern followers, to let g_P^* differ from zero while keeping $\eta = \xi = 0$ for $y' < y'^*$. This would introduce the kind of price rigidity relevant to Keynes's short-period analysis, yet it could be regarded as capturing the phenomenon that his modern followers have emphasized as cost-push inflation.⁶⁸

Equations (12) and (13) do not by themselves specify the path of prices or output beginning with any initial position. In addition, we need to know how anticipated values are formed. Presumably anticipations are affected by the course of events so that, in response to a disturbance that produces a discrepancy between actual and anticipated values of the variables, there is a feedback effect that brings the actual and anticipated variables together again (see below). If this feedback process proceeds rapidly, then the transitory adjustments defined by equations (12) and (13) are of little significance. The relevant analysis is the analysis that connects the asterisked variables.

Chapter 9 explores empirically the adjustment mechanism both to evaluate the relative importance of anticipations and rate of capacity

66. This also explains why g_y^* does not appear explicitly in equation (12), or g_z in equation (13), as they do in equations (9) and (10). They are implicitly included in g_y^* .

67. With ξ infinity, and $\log y' = \log y'^*$, the final expression in equations (12) and (13) is $\infty \cdot 0$, or technically indeterminate. The product can be taken to be zero in general, except possibly for a few isolated points at which $\log y'$ deviates from $\log y'^*$, a deviation closed instantaneously by infinite rates of change in $\log P$ and $\log y'$.

68. The simple monetary theory of nominal income developed in Gordon, *Milton Friedman's Monetary Framework*, pp. 34–48, is of course consistent with these equations in their general form since it does not specify anything about the division of a change in nominal income between prices and output.

utilization and to examine the time path of adjustment. In the process, we develop approximations to these abstract differential equations that can be estimated empirically (see especially sec. 9.10).

2.6.2 Short-Run Adjustment of Nominal Income

For monetary theory, the key question is the process of adjustment to a discrepancy between the nominal quantity of money demanded and the nominal quantity supplied. Such a discrepancy could arise from either a change in the supply of money (a shift in the supply function) or a change in the demand for money (a shift in the demand function). The key insight of the quantity-theory approach is that such a discrepancy will be manifested primarily in attempted spending, and through that route in the rate of change in nominal income. Put differently, money holders cannot determine the nominal quantity of money (though their reactions may introduce feedback effects that will affect the nominal quantity of money), but they can make velocity anything they wish.

What, on this view, will cause the rate of change in nominal income to depart from its permanent value? Anything that produces a discrepancy between the nominal quantity of money demanded and the quantity supplied, or between the two rates of change of money demanded and money supplied. In general form

$$(14) \quad g_Y = f [g_Y^*, g_{MS}, g_{MD}, M^S, M^D],$$

where M^S refers to money supplied, M^D refers to money demanded, and the two symbols are used to indicate that the two are not necessarily equal. That is, equation (21) replaces the adjustment equation (6c), $M^D = M^S$, common to all the simple models.

To illustrate, a particular linearized version of equation (14) would be

$$(15) \quad g_Y = g_Y^* + \Psi(g_{MS} - g_{MD}) + \phi(\log M^S - \log M^D).$$

Unlike equations (12) and (13), the two final adjustment terms on the right-hand side do not explicitly include any asterisked magnitudes. But implicitly they do. The amount of money demanded will depend on anticipated or permanent income and prices as well as on the anticipated rate of change in prices.⁶⁹

69. The three simple models considered in Gordon, *Milton Friedman's Monetary Framework*, pp. 34–46 all require setting $\Phi = \infty$ in our equation (15) to assure that $M^S = M^D$. However, once this is done, the rest of the equation provides no information on the adjustment process, since the final term, which is then of the form $\infty \cdot 0$ is indeterminate. Hence, even though $M^S = M^D$ implies that

$$(a) \quad g_{MS} = g_{MD}$$

so that the second term on the right-hand side of equation (15) is zero for any finite value of Ψ , it does not follow that

In its general form, equation (15) allows for changes in both supply of money and demand for money. It also implicitly allows for the forces emphasized by Keynes, shifts in investment or other autonomous expenditures, through the effect of such changes on M^S and M^D . For example, an autonomous rise in investment demand will tend to raise interest rates. The rise in interest rates will tend to reduce M^D , introducing a discrepancy in one or both of the bracketed expressions on the right-hand side of equation (15), which will cause g_Y to exceed g_Y^* .

Chapter 8 explores empirically, for time units spanning a phase, the adjustment of nominal income to current and prior monetary change, developing empirically manageable approximations to equation (14). See especially sec. 8.4.

2.6.3 Money Demand and Supply Functions

To complete the theory of the adjustment process, it is necessary to specify the functions connecting M^D and M^S with other variables in the system, and also to provide relations determining any additional variables—such as interest rates—entering into these functions. Sections 2.3 and 2.4 discuss the demand and supply functions for money that we regard as relevant for this purpose, so only a few brief supplementary comments are required for present purposes.

First, for reasons discussed in section 2.3, we have taken M^S itself as an autonomous variable in much of our empirical work and have not incorporated in the analysis any feedback from other adjustments.

$$(b) \quad g_Y = g_Y^*$$

The requirement (a) leads to the equation

$$(c) \quad g_Y = g_M$$

for the simple quantity theory, since, with real income and the interest rate fixed, the quantity of money demanded is proportional to prices and hence to nominal income. This equation says that a change in money supply is reflected immediately and proportionately in nominal income.

For the simple Keynesian theory, equation (a) leads, from the equation for the LM curve (equation 22 in Gordon, p. 33) to

$$(d) \quad g_M = \left[\frac{\partial \log \ell}{\partial \log Y} + \frac{\partial \log \ell}{\partial R} \frac{dR}{d \log Y} \right] g_Y$$

where $dR/d \log Y$ is to be calculated from the equation for the IS curve (equation [21] in Gordon, p. 33). In the special case of absolute liquidity preference $\partial \log \ell / \partial R = \infty$; in the special case of completely inelastic investment and saving functions, $dR/d \log Y = \infty$. In either of these cases, equation (d) implies that g_M finite, $g_Y = 0$; that is, a change in the supply of money has no influence on nominal income. In the more general case, equation (d) says that a change in money supply is reflected immediately, but not necessarily proportionately, in nominal income.

For the monetary theory of nominal income, equation (a) implies equation (41), in Gordon, p. 42, which allows for a delayed adjustment of permanent income to measured income, but not for any discrepancy between M^S and M^D .

Second, the function specifying M^D might in principle include a transitory component. That is, the theory here sketched is entirely consistent with distinguishing between a short-run and a long-run demand for money, as some writers have done.⁷⁰

Chapter 6 explores this and other issues empirically. We there find it possible to estimate a single demand equation fitting all our data: that is, a single demand equation for the United States and the United Kingdom for a century, confirming in a rather remarkable way the initial insight of the quantity theory approach on the stability of the demand for money.

2.6.4 Determination of Interest Rates

Given that interest rates enter into the demand function for money (equation 7) and also, possibly, into the supply function (equation 6a), a complete model must specify the factors determining them. Our long-run model determines their permanent values. So what is needed is an analysis of the adjustment process for interest rates comparable with that for prices and nominal income discussed above—provided, as seems reasonable, that measured as well as permanent values of interest rates enter into the money demand and supply functions.

The pure theory of this adjustment process is outlined in the initial section of chapter 10. The components of the adjustment include an initial liquidity and loanable funds effect, a subsequent income effect, and a still more delayed price anticipation effect. The rest of chapter 10 explores these adjustments empirically, giving special attention to the adjustment process via the anticipated rate of price change incorporated in the monetary theory of nominal income.

In some of our empirical work, particularly in chapters 6 through 9, we have treated interest rates as exogenous.

2.6.5 Determination of Anticipated Values

The transition between the short-run adjustment process and long-run equilibrium is produced by a revision of anticipated values in response to measured values in such a way that, for a stable system, a single disturbance sets up discrepancies that are in the course of time eliminated. To put this in general terms, we must have

$$(16) \quad g_P^*(t) = f [g_P(T)]$$

$$(17) \quad g_Y^*(t) = h [g_Y(T)]$$

70. See H. Robert Heller, "The Demand for Money: The Evidence from the Short-Run Data," *Quarterly Journal of Economics* 79 (May 1965): 291–30; Gregory C. Chow, "On the Long-Run and Short-Run Demand for Money," *Journal of Political Economy* 74 (April 1966): 111–31; H. König, "Demand Function, Short-Run and Long-Run Function, and the Distributed Lag," *Zeitschrift für die Gesamte Staatswissenschaft* (February 1968): 124 ff.; J. Carr and M. R. Darby, "The Role of Money Supply Shocks in the Short-Run Demand for Money," *Journal of Monetary Economics* 8 (September 1981): 183–99.

$$(18) \quad y'^*(t) = j [y'(T)]$$

$$(19) \quad P^*(t) = k [P(T)],$$

where t stands for a particular point in time and T for a vector of all dates before t .

A disturbance of long-term equilibrium, let us say, introduces discrepancies in the two final terms in parentheses on the right-hand side of equation (15). These discrepancies cause the rate of change in nominal income to deviate from its permanent value, which through equations (12) and (13) produce similar discrepancies between the rates of price and output change and their permanent values. These may in turn reenter equation (15), but whether they do or not, through equations (16)–(19) they produce revisions in the anticipated values that, sooner or later and perhaps after a cyclical reaction process, eliminate the discrepancies between measured and permanent values.

These anticipation equations are in one sense very general, in another, very special. They require that anticipations be determined entirely by the history of the particular variable in question, not by other history or other currently observed phenomena. They thereby deny any “autonomous” role to anticipations.

One response to this potential defect has been the theory of rational expectations that has recently received much attention.⁷¹ This theory asserts that economic agents should be treated as if their anticipations fully incorporate both currently available information about the state of the world and a correct theory of the interrelationships among the variables. Anticipations formed in this way will on the average tend to be correct (a statement whose simplicity conceals fundamental problems of interpretation, as we point out in sec. 10.7). The theory of rational expectations has been extremely fruitful on an analytic level but as yet is in a preliminary stage as a source of empirically testable hypotheses about the formation of expectations.

In our own empirical work, we have relied primarily on expectation models of the general type described by equations (16)–(19) and on simple adaptive expectations models, in which an anticipated value is revised at a rate proportional to the discrepancy between the actual and anticipated value. However, this area is attracting much research attention, so rapid progress in the development of specific models can be expected.⁷²

71. Robert J. Shiller, “Rational Expectations and the Dynamic Structure of Macroeconomic Models: A Critical Review,” *Journal of Monetary Economics* 4 (January 1978): 1–44; J. J. Sijben, *Rational Expectations and Monetary Policy* (Germantown, Md.: Sijthoff and Noordhoff, 1980).

72. See Brian Kantor, “Rational Expectations and Economic Thought,” *Journal of Economic Literature* 17 (December 1979): 1422–41. Studies that use a weighted average of past values to obtain expected values have been criticized in the rational expectations

One subtle problem in this kind of a structure, in which the absence of a discrepancy between actual and anticipated values defines long-period equilibrium, is to assure that the feedback relations defined by equations (16)–(19), as well as the other functions, are consistent with the expanded system of Walrasian equations that specify the long-term equilibrium values. At least some values are implicitly determined in two ways: by a feedback relation such as equations (16)–(19) and by the system of long-run equilibrium equations. The problem is to assure that at long-run equilibrium these two determinations do not conflict.

2.7 An Illustration

It may help to clarify the general nature of this theoretical approach if we apply it to a hypothetical monetary disturbance.⁷³

Let us start with a situation of full equilibrium with stable prices and full employment and with output growing at, say, 3 percent per year. For simplicity, assume that the income elasticity of demand for money is unity, so that the quantity of money is also growing at the rate of 3 percent per year. Assume also that money is wholly noninterest-bearing fiat money and that its quantity can be taken as autonomous.

Assume that there is a shift at time $t = t_0$ in the rate of growth of the quantity of money from 3 percent per year to, say, 8 percent per year and that this new rate of growth is maintained indefinitely. Chart 2.1 shows the time path of the money stock before and after time t_0 . The lines are not drawn strictly to scale. For emphasis, they exaggerate the difference in the slopes of the lines before and after t_0 .

2.7.1 Long-Run Equilibrium

Let us first ask what the long-run equilibrium solution will be. Clearly, after full adjustment, nominal income will be rising at 8 percent per year. If, for the moment, we neglect any effect of this monetary change on real output and the rate of growth of output, prices would be rising at 5 percent per year. It might therefore seem as if the equilibrium path of nominal income would duplicate that of the quantity of money in chart 2.1 (redrawn as the solid plus dashed lines in chart 2.2). But this is not the

literature. Benjamin Friedman has defended distributed lags not just as an acceptable procedure but an optimal means of forecasting. See his "Optimal Expectations and the Extreme Information Assumptions of 'Rational Expectations' Macromodels," *Journal of Monetary Economics* 5 (January 1979): 23–42. Karl Brunner, Alex Cukierman, and Allan H. Meltzer also defend adaptive expectations as rational in "Stagflation, Persistent Unemployment, and the Permanence of Economic Shocks," *Journal of Monetary Economics* 6 (October 1980): 467–92.

73. For an application to quarterly data for 1952–70, see Dean Taylor, "Friedman's Dynamic Models: Empirical Tests," *Journal of Monetary Economics* 2 (November 1976): 531–38.

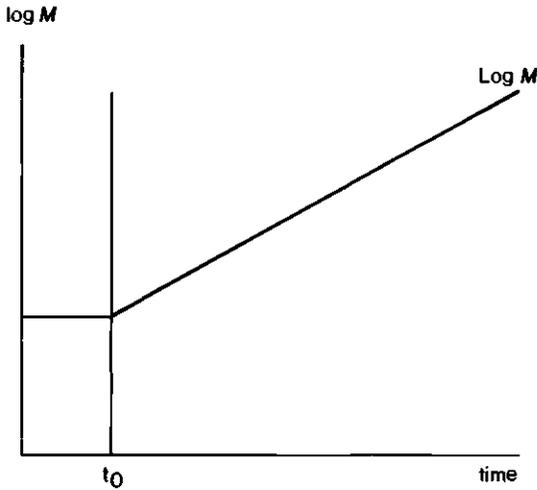


Chart 2.1 Time path of money stock before and after time t_0 .

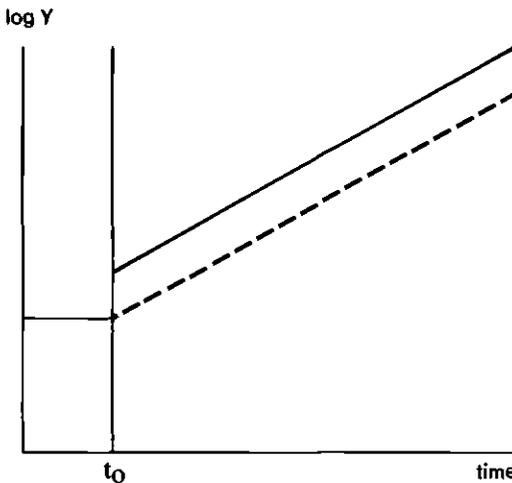


Chart 2.2 Equilibrium path of nominal income before and after time t_0 .

case. With prices rising at the rate of 5 percent per year and, at equilibrium, with this price rise fully anticipated by everyone, it is now more costly to hold money. As a result, equation (7) would indicate a decline in the real quantity of money demanded relative to income, that is, a rise in desired velocity. This rise would be achieved by a rise in nominal income above that required to match the rise in the nominal quantity of money. The equilibrium path of nominal income would be like the solid line in chart 2.2 rather than the dashed line.

If equilibrium real output and the rate of growth of real output were unaffected by the monetary change, as we have so far assumed, the

equilibrium path of prices would be the same as that of nominal income, except that it would have a slope of 3 percent per year less, to allow for the growth in real income. However, equilibrium real output will not be unaffected by this monetary change. The exact effect depends on just how real output is measured, in particular whether it includes or excludes the nonpecuniary services of money. If it includes them, as in principle it should, then the level of real output will be lower after the monetary change than before. It will be lower for two reasons: first, the higher cost of holding cash balances will lead producers to substitute other resources for cash, which will lower productive efficiency; second, the flow of nonpecuniary services from money will be reduced.⁷⁴ For both reasons, the price level of output will have to rise more than nominal income—a solid line and a dashed line like those for nominal income in chart 2.2 would be farther apart vertically for prices of final products than for nominal income.

It is harder to be precise about the equilibrium rate of growth, since that depends on the particular growth model. What is clear is that the aggregate stock of nonhuman capital, including money, will be lower relative to human capital, but that the aggregate stock of physical (non-money) capital will be higher, so that the real yield on capital will be lower. The nominal interest rate (the R_B of equation 7) will equal this real yield plus the rate of change in prices, so it will be higher. If these changes have any effect on the rate of growth of real output, they will tend to reduce it, so that the equilibrium price level of final products not only will be higher relative to its initial value than the equilibrium level of nominal income, but also may rise more rapidly.⁷⁵ For simplicity, we shall neglect this possibility and assume that the equilibrium rate of rise in prices is 5 percent per year.

2.7.2 The Adjustment Process

So much for the equilibrium position. What of the adjustment process? This description of the equilibrium position already tells us one thing about the adjustment process. To produce the shift in the equilibrium path of nominal income from the dashed to the solid line, nominal income and prices must rise over some period at a faster rate than the final equilibrium rate—at a faster rate than 8 percent per year for nominal

74. See M. Friedman, *Optimum Quantity of Money*, pp. 14–15.

75. See Jerome L. Stein, "Money and Capacity Growth," *Journal of Political Economy* 74 (October 1966): 451–65; Harry G. Johnson, "The Neo-Classical One-Sector Growth Model: A Geometrical Exposition and Extension to a Monetary Economy," in *Essays in Monetary Economics* (London: Allen and Unwin, 1967); idem, "Neutrality of Money in Growth Models: A Reply," *Economica*, n.s., 34 (February 1967): 73–74; Alvin Marty, "The Optimal Rate of Growth of Money," *Journal of Political Economy*, 76, part 2 (July/August 1968): 860–73.

income and 5 percent per year for prices. There must, that is, be a cyclical reaction, an overshooting, in the rate of change in nominal income and prices, though not necessarily in their levels.

How will this adjustment process be reflected in our theoretical sketch of the adjustment process? The shift in g_{MS} at time t_0 from 3 percent to 8 percent introduces a discrepancy of positive sign into the second term on the right-hand side of equation (15), while initially leaving the third term unchanged. As a result, g_Y will increase, exceeding g_Y^* , which, viewed in this transitional process as an anticipated value rather than as a long-run equilibrium value, is unchanged from the prior long-run equilibrium value. How rapidly the rate of growth of nominal income rises depends partly on the value of Ψ , the coefficient indicating speed of adjustment, and partly on the demand function for money. If the latter depends only on anticipated values [that is, if all the variables in equation (7) have asterisks], g_{MD} will initially be unchanged, so everything will depend on Ψ , which might have any value, from zero, meaning no adjustment, to a value higher than unity, meaning that nominal income would rise initially by more than 8 percent per year.⁷⁶

Whatever the rate of rise in nominal income, it will be divided into a rise in prices and a rise in output, in accordance with equations (12) and (13). If η is less than unity, both real output and prices will start rising, their relative rates depending on the size of η .

The rising prices and nominal income will start affecting anticipated rates of change, through equations (16)–(19), feeding back into equations (15) and (12) and (13).

All of this is so at time t_0 , with no effect on the levels of any of the variables. As the process continues, however, the levels start being affected. In equation (15), $\log M^S$ comes to exceed $\log M^D$, so the final term of equation (15) adds to the upward pressure on g_Y , making for a speeding up in the expansion of nominal income. In equations (12) and (13), $\log y'$ comes to exceed $\log y'^*$, thus increasing the fraction of income increase absorbed by prices and reducing the fraction absorbed by output. The changed levels of y' and P feed into equations (18) and (19) and so start altering y'^* and P^* .

The changes in all of the variables now start affecting the demand functions for money, both directly, as these variables enter the demand functions, and indirectly, as they affect other variables, such as interest rates, that in turn enter the demand functions. As a result g_{MD} and $\log M^D$ in equation (15) start to change. The process will, of course, finally be completed when the relevant measured variables are all equal to their permanent counterparts and these are equal to the long-run equilibrium values discussed above.

76. The model briefly sketched in the final two paragraphs of M. Friedman, "Demand for Money," implicitly has an initial value of Ψ that is much higher than unity.

It is impossible to carry much further this verbal statement of the solution of an incompletely specified system of simultaneous differential equations. The precise adjustment path depends on how the missing elements of the system are specified and on the numerical values of the parameters, but perhaps this sketch suffices to give the flavor of the kind of adjustment process they generate and to indicate why the process is necessarily cyclical.

What is the reflection in these equations of the point made in the first paragraph of this subsection, namely, that g_Y and g_P must, during the transition, average higher than their final long-term equilibrium values? Consider equation (15). Suppose that over a period the *average* value of g_Y and g_P had been 8 percent per year and 5 percent per year, respectively. Suppose the anticipation functions (16)–(19) were such that this was fully reflected in anticipated values. Then, as we have seen, though M^S would have risen at the rate of 8 percent per year, M^D would not have; so the final term in equation (15) would not be zero, even though the middle term on the right-hand side might be. Hence, g_Y would exceed g_Y^* , which by assumption is at its long-run equilibrium value; so full equilibrium would not have been attained.

Chart 2.3 summarizes various possible adjustment paths of g_Y consistent with the theory sketched. The one common feature of all of them is that the area above the 8 percent line must exceed the area below. In principle, of course, still other paths are possible. For example, it is conceptually possible for the adjustment to be explosive rather than damped. Restricting ourselves to damped paths is an empirical judgment.

2.8 Conclusion

The climate of professional opinion has changed greatly since the first draft of this chapter was written. Many issues about which controversy then raged now seem outdated. Defending the quantity theory approach and adopting the kind of framework outlined in this chapter no longer seem idiosyncratic and reactionary. On the contrary, they are more nearly in the mainstream, though still not without vigorous critics. Theoretical controversy today is less about Keynes and the classics than about rational expectations and “supply-side” economics, or about the microfoundations of macroeconomics.

This change in the climate of theoretical opinion has not been produced by the persuasiveness or lack thereof of the arguments adduced by economic theorists. Just as the emergence of the Keynesian revolution was a reaction to the brute facts of depression, so the resurgence of the quantity theory (renamed undescriptively “monetarism”) and the rejection of simple Keynesianism have been a reaction to the emergence of inflation and stagflation. Theoretical analysis has an essential role to play

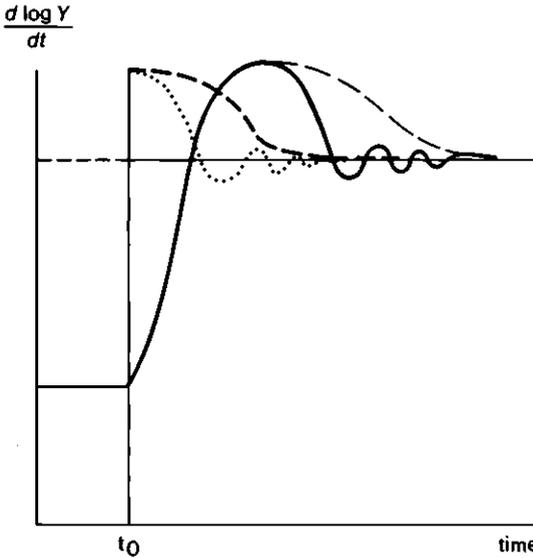


Chart 2.3 Possible adjustment paths of rate of change in nominal income.

in guiding and organizing research, in interpreting empirical evidence, in providing compact ways to summarize masses of generalizations, and in avoiding errors. But in our opinion the basic differences among economists are not theoretical, but empirical.

The controversy about the role of money in economic affairs that raged for so long and has by no means died down even yet reflects different implicit and explicit answers to empirical questions such as those considered in the later chapters of this book: Is the demand function for money stable? What variables are most important in determining the quantity of money demanded? How elastic is the response of the quantity of money demanded to interest rates (chap. 6)? What are the channels whereby changes in one country influence another, and how important are those influences (chap. 7)? When changes in the demand or supply produce discrepancies between the quantity of money the public holds and the quantity it desires to hold, how rapidly do these discrepancies tend to be eliminated? Is the impact on nominal income asymptotic or cyclical (chap. 8)? What about the separate adjustment of prices and output? Do the reactions of prices depend more on price anticipations or on the level of output relative to capacity? How are anticipations formed (chap. 9)? What about the effects on interest rates? Does the distinction between “nominal” and “real” interest rates play a key role? Or is that purely a theoretical construct of little practical importance (chap. 10)?

The reason differences about such empirical questions have been able to persevere is that adjustment to monetary disturbances takes a long time and affects many economic magnitudes (chaps. 8 and 9). If adjustment were swift, immediate, and mechanical, as some earlier quantity theorists may have believed or, more likely, as was attributed to them by their critics, the role of money would be clearly and sharply etched even in the imperfect figures that have been available. But if the adjustment is slow, delayed, and sophisticated, then crude evidence may be misleading, and a more subtle examination of the record may be needed to disentangle what is systematic from what is random and erratic. That, not the elaboration of the theory, is the primary aim of this book as well as of the other monetary studies of the NBER.