Volume Title: A Theoretical Framework for Monetary Analysis

Volume Author/Editor: Milton Friedman

Volume Publisher: NBER

Volume ISBN: 0-87014-233-X

Volume URL: http://www.nber.org/books/frie71-1

Publication Date: 1971

Chapter Title: Quantity Equations

Chapter Author: Milton Friedman

Chapter URL: http://www.nber.org/chapters/c0912

Chapter pages in book: (p. 3 - 10)

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

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

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

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

2. Quantity Equations

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

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

a) Transactions Equation

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

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

or

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

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

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

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

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

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

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

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

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

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

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

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

b) The Income Form of the Quantity Equation

Despite the large amount of empirical work done on the transactions equations, notably by Irving Fisher and Carl Snyder (Fisher 1911, pp.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

c) Cambridge Cash-Balances Approach

The essential feature of a money economy is that it enables the act of purchase to be separated from the act of sale. An individual who has something to exchange need not seek out the double coincidence—someone who both wants what he has and offers in exchange what he

wants. He need only find someone who wants what he has, sell it to him for general purchasing power, and then find someone who has what he wants and buy it with general purchasing power.

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

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

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

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

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

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

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

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

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

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

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

3. Supply of Money in Nominal Units

The factors determining the nominal quantity of money available to be held depend critically on the monetary system. For systems like those which have prevailed in the United States and in the United Kingdom during the past century, they can usefully be analyzed under the three main headings that we have termed the proximate determinants of the money stock: (1) the amount of high-powered money—for any one country this is determined through the balance of payments under an