Scott Freeman
Department of Economics
University of Texas, Austin

Money and Output: Correlation or Causality?

A puzzle

At the center of monetary economics is a puzzling observation: innovations (unexpected surges) in the nation's total supply of money are historically correlated with innovations in real output. Standard monetary economics can easily explain why an increase in the number of dollars will increase the prices of goods and thus nominal output, the dollar value of the economy's production.

But why should the number of nearly fictitious items called dollars be linked to the amount of real goods produced by workers and machines? Can dollars make workers more intelligent or reduce the breakdown of machinery?

The question is of interest to policymakers as well as academics. The Federal Reserve System has a great deal of control over the money supply of the United States. If changes in the money supply truly cause changes in real output, the Fed can stimulate (or hold back) the nation's real production by manipulating the money supply. A moderation of the business cycle might then be within the powers of the Federal Reserve. If, instead, changes in the money supply have no effect on real output, then efforts to stimulate the economy through money supply increases will result only in inflation.

That changes in the money supply cause the changes in real output is only one way to interpret the observed correlation between the two. It might also be that the output changes cause the money supply to change or that both the money supply and output are reacting to innovations in some third factor. This last possibility is the focus of this article.

Friedman and Schwartz in their historical studies (1963a and 1963b) brought the positive correlation between the money supply and real output to the attention of the profession. Positive innovations in the total stock of money were found to occur during expansions in real economic activity and negative innovations during recessions. In addition, Sims (1972) found that the innovations in the money stock preceded the innovations in output. Past innovations in output were of no help in predicting changes in the money stock, but past innovations in the money stock could help predict changes in output.[1] This finding led some in the profession to believe that the money stock innovations caused the output innovations.[2]

However, the usefulness of the money/output correlation as a guide for monetary policy is challenged by the observation that while innovations in the total money stock are strongly correlated with output, innovations in the *monetary base*—the money distributed by the Federal Reserve—are not. Cagan (1965), Sims (1972), and King and Plosser (1984) all find that innovations in output are more strongly linked to innovations in *inside money*—that part of the money stock

[1] *When interest rates were included in the studies with money and output, it was found by Sims (1980) and by Litterman and Weiss (1985) that their innovations, not those of money, helped predict output. Stock and Watson (1989) find that detrending the money data restores much of the predictive content of monetary data. See Stock and Watson for a summary of recent investigations into the money/output correlation.*

[2] *In opposition to this belief, Tobin (1970) argued that the precedence of the changes in money did not imply that money changes caused the output changes. His reasoning, that the Federal Reserve might be forward-looking, differs from the reasoning to be presented here.*

consisting of deposits at banks—than to innovations in the money issued by the Federal Reserve, or *outside money*. This observation is the key to the explanation of the money/output puzzle that I will now propose.

An explanation

Some definitions. To study the links between money and output, it is essential to identify carefully the components of each. First, a definition of real output is needed. In any period t+1, real output (GNP_{t+1}) is an increasing function $F(...)$ of its inputs. For simplicity, I limit the list to the two most obvious, current labor (L_{t+1}) and previously created capital (K_t). An important but realistic assumption is that new investment in capital cannot produce goods instantly—factories take time to build. I also assume that the productivity of these inputs is subject to random shocks, which I represent with a parameter x_{t+1}. We can now represent real output in the following way:

$$(1) \qquad GNP_{t+1} = x_{t+1}F(L_{t+1}, K_t).$$

Capital may come from two sources, direct investment by individuals ($\hat{K}_t$) and investment financed by bank loans (H_t), implying

$$(2) \qquad GNP_{t+1} = x_{t+1}F(L_{t+1}, \hat{K}_t + H_t).$$

The money supply in some period t (MS_t) is defined as the total dollar value of assets readily used in making purchases—currency and liquid (checkable) deposits at banks and similar financial intermediaries,[3] which can be written

(3) $MS_t \equiv$ *nominal currency* + *nominal deposits.*

Currency consists solely of notes and coins issued by the Federal Reserve. Deposits are more complex; they are not simply held as cash in the vaults of banks. A fraction of deposits is held as noninterest-bearing reserves in bank vaults and at the Federal Reserve; the rest is lent to businesses and home buyers for the interest it generates. In this way, bank deposits are (partially) backed by capital, be it the investments of businesses or housing. Let us therefore call the interest-bearing assets of banks *intermediated capital*, capital that is invested from funds gathered by banks. This leads to the equation

(4) *nominal deposits* = *nominal reserves* + *nominal intermediated capital.*

Together, equations 3 and 4 imply that

(5) $MS_t \equiv$ *nominal currency* + *nominal reserves* + *nominal intermediated capital.*

By law, currency and reserves can only be held in the form of noninterest-bearing *fiat money*, money created by the Federal Reserve. Therefore, these two terms are combined in equation 3 to get

$$(6) \quad MS_t \equiv \text{nominal fiat money} + \text{nominal intermediated capital} = \text{nominal fiat money} \times \left(1 + \frac{\text{nominal intermediated capital}}{\text{nominal fiat money}}\right).$$

Equation 6 reveals that the total money supply equals the stock of fiat money times an expression I will refer to as the *money multiplier*. The money multiplier is 1 plus the ratio of intermediated capital to fiat money.[4] The stock of fiat money, often called the monetary base—or high-powered money—is always known, as it is chosen by the Federal Reserve. The money multiplier then reveals the total money supply for any given value of the monetary base. Consider some examples. Suppose that people hold currency but no deposits. The money multiplier is then simply 1, implying that the total money supply is simply the supply of fiat money, all of it used as currency.

[3] *Each measure of money calculated by the Federal Reserve includes both currency and some deposits. The measures differ by the variety of deposits included, but that need not concern us here. Most of the empirical studies looked at the most restrictive definition, called M1.*

[4] *If there were no holdings of currency, the money multiplier would simplify to the inverse of the reserve-to-deposit ratio, which is the money multiplier often taught in introductory textbooks.*

As the use of deposits expands, there is now some money backed by intermediated capital in addition to the money backed by the monetary base, so that for any given level of the monetary base, there is more total money.

Let me introduce some notation to make equation 6 more readable and useful. Let M_t denote the nominal stock of fiat money, the monetary base, and let Q_t denote the real demand for fiat money, the number of goods that people will give up to get the fiat money balances they desire. Recall that H_t denotes the real stock of intermediated capital. A nominal value of a variable is the product of the price level (p_t)—an average of the prices of goods—and its real value, the goods that can be purchased by the nominal value. This implies that

(7) *nominal intermediated capital* = $p_t H_t$, and

(8) *nominal fiat money* $\equiv M_t = p_t Q_t$.

If one substitutes these expressions into equation 6 and divides the top and bottom of the fraction by the price level, the money multiplier is expressed in real terms, which will help link it to real output. This we do in equation 9:

$$(9) \qquad MS_t = M_t\left(1+\frac{p_t H_t}{p_t Q_t}\right) = M_t\left(1+\frac{H_t}{Q_t}\right).$$

Notice that there are two possible sources of fluctuations in the total money supply—changes in the monetary base and changes in the money multiplier. Changes in the monetary base are caused by actions of the Federal Reserve. Changes in the money multiplier, however, can occur if changes occur in the ratio of intermediated capital to fiat money, a ratio affected by a private decision—how much money to hold in currency and deposits. The more people favor deposits relative to currency, the greater the money multiplier and the total money stock; that is, for each unbacked dollar of the monetary base (fiat money), there are more dollars backed by bank loans and more money in total when the ratio of deposits to currency increases.

What might influence an individual's choice of deposits or currency? It is reasonable to expect that the choice will be influenced by the person's relative costs and rates of return. Currency is generally the more convenient of the two, at least for small transactions; it takes less time and effort to use than checks drawn on bank deposits. Deposits, however, offer interest and thus a better rate of return. Most of us hold both deposits and currency, balancing their relative costs and returns. It follows that if the costs of one relative to the other increase, less will be held of that form of money. Similarly, if the rate of return of one should rise, more of that form of money will be held.[5]

Banks also face a choice of how to allocate the deposits they receive between interest-bearing assets and reserves. Reserves, whether held as vault cash or as deposits with the Federal Reserve, represent funds that banks can use to make payments at little cost, but they pay no interest. Therefore, the higher the rate of return available to banks, the less banks will want to keep in reserves and the more they will want to invest in interest-bearing assets. In recent decades of high interest rates, banks have tended to hold as reserves only the minimum required by the Federal Reserve. In earlier decades, when interest rates were lower, notably during the Great Depression, banks often held more reserves than required.

The money/output correlation. The empirical studies of Cagan and the others listed above found that changes in output are correlated with changes in the money multiplier. Let us examine an example of how this correlation might come about.[6]

[5] *While most of us may not pay much attention to the currency balances we hold, this decision is not trivial for those handling large amounts of currency on which the foregone interest may be substantial.*

[6] *The explanation of the money/output correlation presented here is taken from Freeman and Huffman (1991), which drew on ideas from Sargent and Wallace (1982) and Freeman (1986). The example presented is only one of many possible causes of a money/output correlation. It is presented as an illustration, not as an assertion that it is the only possible cause. See Tobin (1970) and McCallum (1983) for alternative explanations.*

Suppose some random event[7] leads people to believe that capital will be less productive (and thus offer to pay a lower rate of return) in the next period. A likely reaction of investors will be to invest less in the current period. Not only will direct investment ($\hat{K}_t$) fall, but so will investment financed through bank loans (H_t). The effect on real output is clear: it will fall in the next period as the result of both the lower productivity of capital and the lower investment. In addition, the flow of savings from investment to other assets will drive down the real interest rate of all assets.

What is the effect on the money supply, $MS_t = M_t(1+\frac{H_t}{Q_t})$? The low return on capital does nothing to change the monetary base, M_t, but what about the money multiplier, $(1+\frac{H_t}{Q_t})$? Faced with a reduction in the loans they can make (H_t), banks must either hold more reserves or accept fewer deposits by lowering the rate of return offered on deposits, thereby encouraging people to use more currency. The demand for fiat money (Q_t) is therefore increased through some combination of an increased demand for reserves and for currency. As we saw above, a decrease in the ratio of bank loans to fiat money represents a decrease in the money multiplier, $(1+\frac{H_t}{Q_t})$ and thus a decrease in the total money supply (MS_t).

Finally, what is the effect on the price level, p_t? The price level can be determined from equation 8, which tell us that the nominal stock of fiat money must equal its demand in nominal terms:

$$M_t = p_t Q_t$$

$$\text{or} \quad p_t = \frac{M_t}{Q_t}. \tag{10}$$

The increased demand (increase in Q_t) for an unchanged stock of fiat money will raise the value of the money; that is, it will lower the price level. These effects are summarized in Figure 1.

The pattern predicted by this analysis fits the data puzzled over in the introduction. Changes in the total nominal money stock are correlated with changes in real output. Moreover, the two move in the same direction, with the change in money preceding the change in output. While I chose only one particular shock to serve as an illustration, the money/output correlation is far more general. Look again at the equations defining output and the money stock:

$$GNP_{t+1} = x_{t+1}F(L_{t+1}, \hat{K}_t + H_t), \text{ and} \tag{2}$$

$$MS_t = M_t(1+\frac{H_t}{Q_t}). \tag{9}$$

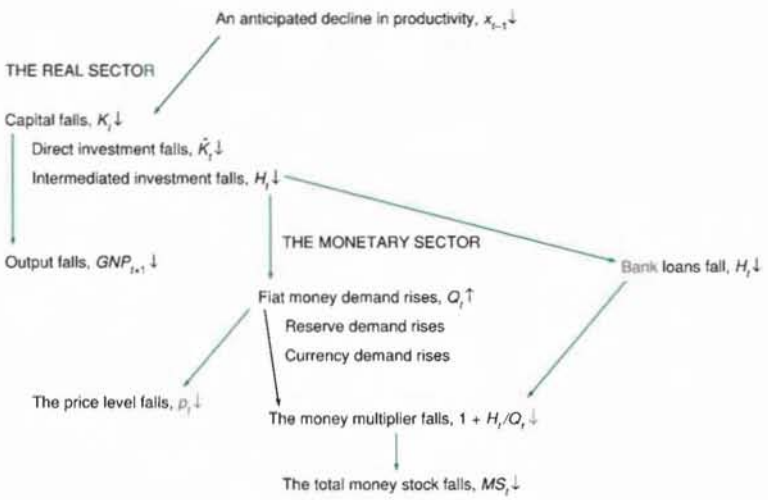
Equations 2 and 9 reveal that both output and the money stock are increasing functions of H_t, which represents bank loans or intermediated capital. Whatever causes bank loans to increase results in an increase in the money supply and, with a delay, in real output (other things being equal).

Active monetary policy. The Federal Reserve has considerable control over the total money supply. Given the observed positive correlation between money and output, can the Federal Reserve stimulate the economy by expanding the stock of money?

Let's try this out. The most direct way to increase the money stock is to print more fiat money. Suppose, therefore, that in some period t, the monetary authority doubles the stock of fiat money, distributing the new money to people in proportion to their holdings of money so that no income is redistributed. Will this bring about a change in real output? No. This is a purely nominal change, a change in units. Prices will double, but no real decision will be affected. People will choose to hold the same real value of deposits, capital, reserves, and currency as before because no rate of return is changed by this one-time expansion of the number of dollars. In particular, the nominal change will not induce any change in real investment plans or real output. A policy of printing fiat money whenever inside money contracts can stabilize the total money supply and the price level but not real output.

[7] *Examples include such diverse events as political or military threats to oil supplies, summer droughts affecting the fall harvest, technological changes, or even unfounded fears of investors, as emphasized in the Keynesian tradition.*

Figure 1
The Model's Chain of Causes and Effects



Green pieces of paper cannot substitute for the real capital that banking provides.[8]

Lessons

The above analysis (or "model") is of interest not simply because it explains the money/output correlation. Any clever economist can come up with a model to explain some single fact. This particular explanation is especially intriguing because it not only explains the money/output correlation but in so doing explains a phenomenon not generally addressed by traditional models of money—that inside money is more tightly linked to output than is the money distributed by the Federal Reserve.

I cannot claim, however, that this simple model represents absolute truth or captures every complexity of the behavior of money and output. Nevertheless, this model illustrates two ways in which the monetary economists and policymakers may have been misled by the observed money/output correlation. First, the model illustrates why a correlation observed between two variables does not imply that one caused the other. Second, it shows how measures of money may mislead the analysis of monetary policy by lumping together two very different types of money. Let us examine these in turn.

Correlation or causality? Although money may be historically correlated with real output, we see from the illustration above that this does not imply that the changes in the money supply cause the changes in output. In the example studied, when money and output both fall, both are reacting to the anticipated decline in the productivity of capital. As investors anticipate a reduction in the return from capital, both direct and intermediated investment fall, which reduces output. The reduction in intermediated investment (bank loans) implies a reduction in the money multiplier and the total money stock. The money stock reacts first because a switch to currency and reserves can be accomplished instantaneously, while a reduction in investment will take a period to reduce output because of the delay between the act of investment and the output it produces.

[8] *An interesting feature of the model in Freeman and Huffman (1991) and the related model of Lacker (1988) is that anticipated future inflation can stimulate output by inducing people to switch from currency to deposits. The effect is a small one, however, since currency holdings are small relative to the nation's capital stock.*

(In many ways this is similar to stock market drops that occur before recessions. The stock market can respond instantly to a drop in anticipated profits, but it takes time for the related drop in investment to show up as a drop in output.)

This example demonstrates a common pitfall in econometric work involving the total money supply. The money supply is an endogenous variable: it reacts to other changes in the economy. Therefore, observed correlations between money and some other endogenous variable (such as output or interest rates) may result from the reactions of both to some economic event. The observed correlation and even precedence of money innovations to output innovations in no way implies that money innovations cause the output innovations.

It is easy to understand that statistical evidence that any two variables move together does not prove that changes in either one causes the changes in the other; a third variable could be the source of both changes. The simple intuition underlying this example is that even evidence that one of the variables changes first does not prove that the first one to change causes the change in the other. It may again be a third factor that causes the other changes, but one variable reacts before the other. Therefore, as Cooley and LeRoy (1985) and Leamer (1985) have argued, the direction of a causal relation cannot be established by the purely statistical device of establishing which variable changes first.

The quantity theory. Finally, the example presented here warns against exclusively focusing on monetary aggregates that treat inside and outside money as if they were entirely the same. Inside money represents deposits invested through banks into capital projects. In this way there is a direct link between inside money and the real economy. In contrast, outside money (the monetary base, or fiat money created by the Federal Reserve) represents merely unbacked pieces of paper with no direct link to real production. It is not surprising, therefore, that these two forms of money have very different links to output in the data.

The quantity theory of money is so named as a statement that the total quantity of money, not its composition, matters. This may be true for the provision of transaction services that money provides. However, both the data and the theory described here indicate that the links of money to real output are very different for inside and outside money. When measures of money fail to distinguish between the two, correlations between inside money and output appear only as correlations between total money and output. Observing the correlation between total money and output, the Federal Reserve quite naturally might hope that the (outside) money the Fed prints will increase output. These hopes will be disappointed if output is only related to inside money. If the money/output correlation is to be studied as a guide to monetary policy, only the links between output and Federal Reserve actions should be examined.

References

Cagan, Philip (1965), *Determinants and Effects of Changes in the U.S. Money Stock, 1875–1960* (New York: National Bureau of Economic Research).

Cooley, Thomas, and Stephen LeRoy (1985), "Atheoretical Macroeconomics—A Critique," *Journal of Monetary Economics* 16: 283–308.

Freeman, Scott (1986), "Inside Money, Monetary Contractions, and Welfare," *Canadian Journal of Economics* 19: 87–98.

——— and Gregory Huffman (1991), "Inside Money, Output, and Causality," *International Economic Review* 32 (August): 645–67.

Friedman, Milton, and Anna Schwartz (1963a), "Money and Business Cycles," *Review of Economics and Statistics* 45 (February): 32–64.

——— and ——— (1963b), *A Monetary History of the United States* (Princeton, N.J.: Princeton University Press).

King, Robert, and Charles Plosser (1984), "Money, Credit, and Prices in a Real Business Cycle," *American Economic Review* 74 (June): 363–80.

Lacker, Jeffrey (1988), "Inside Money and Real Output," *Economic Letters* 28: 9–14.

Leamer, Edward (1985), "Vector Autoregressions for Causal Inference?" in Karl Brunner and Allan Meltzer, *Carnegie–Rochester Conference Series on Public Policy* 22: 255–303.

Litterman, Robert, and Laurence Weiss (1985), "Money, Real Interest Rates, and Output: A Reinterpretation of U.S. Postwar Data," *Econometrica* 53 (January): 129–56.

McCallum, Bennett (1983), "A Reconsideration of Sims' Evidence Concerning Monetarism," *Economics Letters* 13: 167–71.

Sargent, Thomas, and Neil Wallace (1982), "The Real–Bills Doctrine vs. The Quantity Theory: A Reconciliation," *Journal of Political Economy* 90 (December): 1212–236.

Sims, Christopher (1972), "Money, Income, and Causality," *American Economic Review* 62 (September): 540–52.

——— (1980), "Comparison of Interwar and Postwar Cycles: Monetarism Reconsidered," *American Economic Review* 70 (May): 250–57.

Stock, James, and Mark Watson (1989), "Interpreting the Evidence on Money–Income Causality," *Journal of Econometrics* 40 (January): 161–81.

Tobin, James (1970), "Money and Income: Post Hoc Ergo Propter Hoc?" *Quarterly Journal of Economics* 84 (May): 328–29.