Money, Income, and Causality

Christopher A. Sims

The American Economic Review, Volume 62, Issue 4 (Sep., 1972), 540-552.

Stable URL:
http://links.jstor.org/sici?sici=0002-8282%28197209%2962%3A4%3C540%3AMIAC%3E2.0.CO%3B2-%23

Money, Income, and Causality

By CHRISTOPHER A. SIMS*

This study has two purposes. One is to examine the substantive question: Is there statistical evidence that money is "exogenous" in some sense in the money-income relationship? The other is to display in a simple example some time-series methodology not now in wide use. The main methodological novelty is the use of a direct test for the existence of unidirectional causality. This test is of wide importance, since most efficient estimation techniques for distributed lags are invalid unless causality is unidirectional in the sense of this paper. Also, the paper illustrates the estimation of long lag distributions without the imposition of the usual restrictions requiring the shape of the distribution to be rational or polynomial.

The main empirical finding is that the hypothesis that causality is unidirectional from money to income agrees with the postwar *U.S.* data, whereas the hypothesis that causality is unidirectional from income to money is rejected. It follows that the practice of making causal interpretations of distributed lag regressions of income on money is not invalidated (on the basis of this evidence) by the existence of "feedback" from income to money.

* Associate professor of economics, University of Minnesota. Work for this paper was carried out during my tenure as a research fellow at the National Bureau of Economic Research, and a more extended paper on this topic may appear as a *NBER* publication. Numerous members of the *NBER* staff provided support at various stages of the research. Special thanks are due to Philip Cagan, John Hause, Milton Friedman, the Columbia Monetary Economics Workshop, and a seminar at the Cowles Foundation, whose objections and advice have sharpened the paper's argument. Josephine Su carried out the computational work. H. I. Forman drew the charts.

I. The Causal Ordering Question for Money and Income

It has long been known that money stock and current dollar measures of economic activity are positively correlated. There is, further, evidence that money or its rate of change tends to "lead" income in some sense.[1] A body of macro-economic theory, the "Quantity Theory," explains these empirical observations as reflecting a causal relation running from money to income. However, it is widely recognized that no degree of positive association between money and income can by itself prove that variation in money causes variation in income. Money might equally well react passively and very reliably to fluctuations in income. Historically observed timing relations between turning points have also for some time been recognized not to be conclusive evidence for causal ordering. James Tobin and William Brainard and Tobin provide explicit examples of the possibilities for noncorrespondence between causal ordering and temporal ordering of turning points. People in close connection with the details of monetary policy know that some components of the money supply react passively to cyclical developments in the economy. Frank DeLeeuw and John Kalchbrenner, for example, argue that the monetary base (currency plus total reserves) is not properly treated as an exogenous variable in a regression equation because of the known dependence be-

[1] See Milton Friedman and Anna Schwartz (1963b), Friedman (1961), (1964) reprinted in chs. 10–12 of Friedman (1969).

tween certain of its components and cyclical factors.

Phillip Cagan uses an analysis of the details of money-supply determination to argue convincingly that the long-run relation between money supply and the price level cannot be due primarily to feedback from prices to money. His application of the same analytical technique to cyclical relations of money with income measures fails to yield a firm conclusion, however.

Friedman and Schwartz have argued on the basis of historical analysis that major depressions have been caused by autonomous movements in money stock.[2]

The issues between the monetarists and the skeptics are not easily defined on the basis of the literature cited in the preceding paragraphs. Probably few of the skeptics would deny *any* causal influence of money on income. But, on the other hand, leading exponents of the monetarist approach seem ready to admit that there is "clear evidence of the influence of business change on the quantity of money,"[3] at least for the mild cycles which have characterized the postwar United States.

Now if the consensus view that there is some influence of business conditions on money is correct, if this influence is of significant magnitude, and if current dollar *GNP* is a good index of business conditions,[4] then distributed lag regressions treating money as strictly exogenous are not causal relations. Since such regressions are now treated as causal relations by some economists, it is important to test the assumption of causal priority on which they rest.

As will be shown below, there is a natural analogue in a dynamic system to Wold's "causal chain" form for a static econometric model.[5] This analogue turns out to be exactly a model in which causation is unidirectional according to the criterion developed by C. W. J. Granger. But Wold's form is in general not testable in a static context; any multivariate set of data with a specified list of endogenous variables can be fit by a recursive model. The dynamic analogue is, however, easily testable: If and only if causality runs one way from current and past values of some list of exogenous variables to a given endogenous variable, then in a regression of the endogenous variable on past, current, and future values of the exogenous variables, the future values of the exogenous variables should have zero coefficients.

Application of this test to a two-variable system in a monetary aggregate and current dollar *GNP* with quarterly data shows clearly that causality does not run one way from *GNP* to money. The evidence agrees quite well with a null hypothesis that causality runs entirely from money to *GNP*, without feedback.

II. The Meaning of the Results

Before giving a rigorous explanation of the notion of causal direction and the detailed description of statistical results, it is worthwhile to consider in a nontechnical way what the results do and do not prove. That the test applied in this paper shows no feedback from y to x is a necessary condition for it to be reasonable to interpret a distributed lag regression of y on current and past x as a causal relation or to apply any of the common estimation methods involving use of lagged dependent variables or corrections for serial cor-

[2] See Friedman and Schwartz (1963b), p. 217–18 as reprinted in Friedman (1969).

[3] The quoted phrase is from Milton Friedman's introduction to Cagan, p. xxvi, and summarizes one of Cagan's main results.

[4] As I will argue below, it may be that the one-dimensional current dollar *GNP* index is so inadequate a measure of those aspects of business conditions which influence money supply that there is no feedback from current dollar *GNP* to money despite the existence of feedback from business conditions to money.

[5] See Edmond Malinvaud, p. 511 ff., for a description of causal chain models.

relation. Hence the most conservative way to state the results for money and income is that they show it to be unreasonable to interpret a least squares lag distribution for money on *GNP* as a causal relation, and that they provide no grounds for asserting that distributed lag regressions of *GNP* on money do not yield estimates of a causal relation. It is natural, and I believe appropriate, to phrase the result more positively: the data verify the null hypothesis that distributed lag regressions of *GNP* on money have a causal interpretation. However, it is possible to concoct models in which a money on *GNP* regression does not yield a causal relation and yet this paper's test would not detect feedback.

The test will fail to detect within-quarter feedback of a certain type. The "innovation" in the stochastic process x_t is that part of x_t which cannot be predicted from x_t's own past (i.e., the residual in a regression of x_t on its own past). If x_t and y_t are connected by two causal relations—one from x to y involving a distributed lag, and the other from y to x but with only the current innovation in y_t on the right-hand side—then the test used in this paper will not detect the y to x feedback.[6]

Where the data show negligible serial correlation, this failing of the test becomes important. For then y and x are their own innovations and one expects that causal relations may be purely contemporaneous. In the general case, with serially correlated data, the failing is not likely to be important. It can result in false conclusions only where there is a certain sort of exact relation between the lag distributions defining the causal structure and the autocorrelation functions of the error terms. With one important class of exceptions, there is seldom reason to suppose any relation at all between the causal structure and the properties of the error terms.

The exception arises for models in which some elements of optimal control enter. If one of the two relations in a bivariate system is chosen optimally, then the innovations in the variables become structural elements of the system. This fact is important for money and income, since it is easy to imagine that money may have been controlled to influence or to conform to income. It can be shown that in a bivariate system with optimal control of one variable, there will be in general two-way causality by the Granger criterion. The only exception is that if the information lag in the control process is just one period and if the criterion for control is minimal variance in, say, y, then causality will spuriously appear to run from y to x.[7] But then the only way optimal control would be likely to hide income-to-money feedback would be if income were controlled to hold down variance in money. This seems farfetched.

The fact that this paper finds no evidence of feedback from *GNP* to money is not direct evidence on the structure of money-supply determination. All that is necessary to allow interpretation of the money on *GNP* distributed lags as causal relations is the hypothesis that in this particular historical sample (1947–69), the determinants of money supply showed no *consistent* pattern of influence by *GNP*. Thus it would be enough if, for example, money supply were influenced quite differently by real and price components of *GNP* movements, so long as actual *GNP* movements were not dominated by one

[6] One elementary consequence is that it is possible for the test to show no feedback in either direction, despite the existence of well-defined lag distributions in both x on y and y on x regressions. This is the case where all the relation between y and x consists of contemporaneous correlation of their innovations.

[7] Proving this in any generality would require stretching the length and increasing the technical level of the paper. I expect to take up this point at greater length in a subsequent paper.

component or the other. Alternatively, a consistent pattern of feedback from *GNP* to money could have been swamped in this sample period by extraneous influences on money. The situation is analogous to that in a supply and demand estimation problem, where we have evidence that in a particular sample elements other than price dominated supply. Such evidence proves that in the sample the price-quantity relation traces the demand curve, but it does not in itself prove anything about the supply curve. Thus one can imagine that if heightened awareness of the importance of monetary policy makes money respond more consistently to the business cycle, single-equation estimates of the money-to-*GNP* relation will become unreliable.

Finally, we ought to consider whether the bivariate model underlying this paper could be mimicking a more complicated model with a different causal structure. The method of identifying causal direction employed here does rest on a sophisticated version of the *post hoc ergo propter hoc* principle. However, the method is not easily fooled. Simple linear structures with reversed causality like the one put forth by Tobin cannot be constructed to give apparent money-to-*GNP* causality. Complicated structures like that put forward by Brainard and Tobin in which both *GNP* and money are endogenous will except under very special assumptions yield a bivariate reduced form showing bidirectional causality. The special assumptions required to make endogenous money appear exogenous in a bivariate system must make money essentially identical to a truly exogenous variable. Thus, if money has in the sample been passively and quickly adjusted to match the animal spirits of bankers and businessmen, and if animal spirits is a truly exogenous variable affecting *GNP* with a distributed lag, then money might falsely appear to cause *GNP*. However, if there is substantial random error in the correspondence between animal spirits and money and that error has a pattern of serial correlation different from that of animal spirits itself, then the bivariate relation between money and *GNP* will appear to show bidirectional causality.[8]

An assumption that future values of money or income cause current values of the other, via economic actors' having forecasts of the future better than could be obtained from current and past money and *GNP*, will affect the apparent direction of causality. However, the effect is much more likely to make a truly unidirectional structure appear bidirectional than vice versa. For example, it is easy to see that if current money supply is determined in part by extraneous knowledge of *GNP* for several future quarters, past money could spuriously appear to affect current *GNP*. However, it is difficult to imagine in such a situation why past *GNP* and all the variation in future *GNP* which can be predicted from past *GNP* should *not* affect money. Without such an artificial assumption, one cannot explain a one-sided lag distribution of *GNP* on money by a "reversed-causation-with-accurate-anticipations" model.

III. Testing for the Direction of Causality[9]

In a single, static sample, the "direction of causation" connecting two related groups of variables is ordinarily not identified. That is, one can construct many different models of causal influence all of which are consistent with a given pattern

[8] This point is not obvious, but to prove it would, as in the case of the previous point about optimal control, overextend the paper. The technically sophisticated reader may easily verify the proposition for himself.

[9] It is my impression that many of the results in this section, even where they have not previously been given formal expression, are widely understood. For example, H. Akaike clearly understands that a two-sided transfer function implies the existence of feedback.

of covariances amongst the variables. If one is willing to identify causal ordering with Wold's causal chain form for a multivariate model, and if enough identifying restrictions are available in addition to those specifying the causal chain form, one can test a particular causal ordering as a set of overidentifying restrictions. The conditions allowing such a test are seldom met in practice, however.

Granger has given a definition of a testable kind of causal ordering based on the notion that absence of correlation between *past* values of one variable X and that part of another variable Y which cannot be predicted from Y's own past implies absence of causal influence from X to Y. More precisely, the time-series Y is said to "cause" X relative to the universe U (U is a vector time-series including X and Y as components) if, and only if, predictions of $X(\mathrm{t})$ based on $U(s)$ for all $s<\mathrm{t}$ are better than predictions based on all components of $U(s)$ except $Y(s)$ for all $s<\mathrm{t}$.

We will give content to Granger's definitions by assuming all time-series to be jointly covariance-stationary, by considering only linear predictors, and by taking expected squared forecast error as our criterion for predictive accuracy.

Consider the jointly covariance-stationary pair of stochastic processes X and Y. If X and Y are jointly purely linearly indeterministic (linearly regular in the terminology of Yu. S. Rozanov), then we can write

$$\begin{aligned} X(\mathrm{t}) &= a^*u(\mathrm{t}) + b^*v(\mathrm{t}) \\ Y(\mathrm{t}) &= c^*u(\mathrm{t}) + d^*v(\mathrm{t}) \end{aligned} \tag{1}$$

where u and v are mutually uncorrelated white noise[10] processes with unit variance, a, b, c, and d all vanish for $\mathrm{t}<0$, and the notation

$$g^*f(\mathrm{t}) = \sum_{s=-\infty}^{\infty} g(s)f(\mathrm{t}-s)$$

The expression (1) is the moving average representation of the vector process $\left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right]$ and is unique up to multiplication by a unitary matrix.[11]

A useful result, not proved by Granger, is

THEOREM 1: *Y does not cause X in Granger's definition if, and only if, a or b can be chosen identically 0.*[12]

This result gives us another intuitive handle on Granger causality. If causality is from X to Y only, then of the two orthogonal white noises which make up X and Y, one is X itself "whitened" and the other is the error in predicting Y from current and past X, whitened. (A whitened variable is one which has been passed through a linear filter to make it a white noise.)

Granger has shown that if there is an autoregressive representation, given by

$$B^*\begin{bmatrix} X \\ Y \end{bmatrix}(\mathrm{t}) = \begin{bmatrix} u \\ v \end{bmatrix}(\mathrm{t}), \tag{2}$$

$B(\mathrm{t})=0$ for $\mathrm{t}<0$, u, v defined by (1), then the absence of causality running from Y to X is equivalent to the upper right-hand element of B being zero. That is, causality runs only from X to Y if past Y does not influence current X. From this point it is not hard to show:

THEOREM 2: *When $\left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right]$ has an autore-*

[10] A "white noise" is a serially uncorrelated process.

[11] Actually, the statement that (1) is the moving average representation of $\left[\begin{smallmatrix} X \\ Y \end{smallmatrix}\right]$ is a condition for uniqueness. There will be forms of (1) for which a, b, c, and d are all 0 for $\mathrm{t}<0$ and u and v are white noises but do not yield moving average representations. These forms of (1) will not be unitary transformations of the moving average representation and can be distinguished from the true moving average representation by the fact that in a true moving average representation $a(0)u(\mathrm{t})+b(0)v(\mathrm{t})$ is the limiting forecast error in forecasting $X(\mathrm{t})$ from all past X and Y.

[12] Proofs of both theorems appear in the Appendix.

gressive representation, Y can be expressed as a distributed lag function of current and past X with a residual which is not correlated with any values of X, past or future, if, and only if, Y does not cause X in Granger's sense.

We can always estimate a regression of Y on current and past X. But only in the special case where causality runs from X to Y can we expect that no future values of X would enter the regression if we allowed them. Hence, we have a practical statistical test for unidirectional causality: Regress Y on past and future values of X, taking account by generalized least squares or prefiltering of the serial correlation in $w(\mathrm{t})$. Then if causality runs from X to Y only, future values of X in the regression should have coefficients insignificantly different from zero, as a group.

An implication of Theorem 2 is that many commonly applied distributed lag estimation techniques are valid only if causality runs one way from independent to dependent variable. The condition that the independent variable X be "strictly exogenous," central to most statistical theory on time-series regression, is exactly the Theorem 2 condition that $X(\mathrm{t})$ be uncorrelated with the residual $U(s)$ for any t, s. For example, quasi differencing to eliminate serial correlation in residuals will produce inconsistent estimates without the one-way causality condition; and the "Koyck transformation" which is invoked to allow interpretation of regressions with autoregressive terms as estimates of infinite lag distributions depends on one-way causality. Hence in principle a large proportion of econometric studies involving distributed lags should include a preliminary test for direction of causality.

Remarks on Distributed Lag Methodology

Especially in a study of this kind, where we wish to make fairly precise use of F-tests on groups of coefficients, it is important that the assumption of serially uncorrelated residuals be approximately accurate. Therefore all variables used in regressions were measured as natural *logs* and prefiltered using the filter $1-1.5L+.5625L^2$; i.e., each logged variable $x(\mathrm{t})$ was replaced by $x(\mathrm{t})-1.5x(\mathrm{t}-1)+.5625x(\mathrm{t}-2)$. This filter approximately flattens the spectral density of most economic time-series, and the hope was that regression residuals would be very nearly white noise with this prefiltering.

Two problems are raised by this prefiltering. First, if the filter has failed to produce white noise residuals, it is quite unlikely to fail by leaving substantial positive first-order serial correlation. Durbin-Watson statistics are therefore of little use in testing for lack of serial correlation, and tests based on the spectral density of the residuals were used instead. Second, as I pointed out in an earlier paper (1970), prefiltering may produce a perverse effect on approximation error when lag distributions are subject to prior "smoothness" restrictions. Therefore, no Koyck, Almon, or rational lag restrictions were imposed a priori, and the length of the estimated lag distributions was kept generous.

In applying the F-tests for causal direction suggested in the previous section, one should bear in mind that the absolute size of the coefficients is important regardless of the F value. It is a truism too often ignored that coefficients which are "large" from the economic point of view should not be casually set to zero no matter how statistically "insignificant" they are. Thus, the fact that future values of the independent variable have coefficients insignificantly different from zero only shows that unidirectional causality is possible. If the estimated coefficients on future values are as large or larger than those on past values, bidirectional causality may be very important in practice, despite insignificant F's. Moreover, small coefficients on future values of the independent variable may sometimes be safely ignored even when they are statistically significant. This is especially true in the light of my observation (1971) that nonzero coefficients on future values may be generated in discrete-time data from a "one-sided" continuous-time distributed lag.[13]

All the data used in the regressions presented in this paper were seasonally adjusted at the source. This creates potential problems of a sort which has not been widely recognized heretofore. Most seasonal adjustment procedures in common use allow for a seasonal pattern which shifts slowly over time, and the rate at which the seasonal pattern is taken to shift varies from one series to another. It can be shown[14] that in distributed lag regressions relating two variables which have been deseasonalized by procedures with different assumed rates of shift in the seasonal pattern, spurious "seasonal" variation is likely to appear in the estimated lag distribution. The lag distributions estimated in this paper are long enough and free enough in form that bias from this source should be obvious wherever it is important (and it is important in one regression). However, it would be better to start from undeseasonalized data, being sure that both variables in the relation are deseasonalized in the same way. A check along these lines, using frequency-domain procedures, was carried out for this paper and is mentioned in the discussion of results below.

[13] The definition of causality given in the previous section generalized easily to continuous time. One simply reinterprets (1) as a continuous-time relation, and "Y does not cause X" still corresponds to "b identically zero."

[14] I showed this in an earlier mimeographed version of this paper. A separate short paper on this topic is in preparation.

IV. Time Domain Regression Results

The data used cover the period 1947–69, quarterly. Money was measured both as monetary base (MB)—currency plus reserves adjusted for changes in reserve requirements—and as $M1$—currency plus demand deposits. Figures for MB were taken from the series prepared by the Federal Reserve Bank of St. Louis and supplied to the National Bureau of Economic Research data bank. Results were similar for $M1$ and MB, so we sometimes use M or money to refer to both $M1$ and MB in what follows.

Regressions of the *log* of GNP (in current dollars) on future and lagged *log* M were significant, as were the reversed regressions of *log* M on future and lagged *log* GNP. (See Table 1.) Table 2 reports tests for homogeneity between the pre-1958 and post-1958 sections of the sample. No significant differences between the subsamples appeared in the regressions. Future values of GNP were highly significant in explaining the M dependent variable, but future values of M were not significant in explaining the GNP dependent variable. (See Table 3.) The largest individual coefficients in each GNP on M regression occur on past lags,

Table 1—Summary of *OLS* Regressions[a]

	F for Independent Variables	$\bar{R}^2$	Standard Error of Estimate	Degrees of Freedom
$GNP=f(M1$, 8 past lags)	1.89*	0.7927	0.01018	64
$GNP=f(M1$, 4 future, 8 past lags)	1.37	0.7840	0.01040	60
$GNP=f(MB$, 8 past lags)	2.24**	0.8004	0.00999	64
$GNP=f(MB$, 4 future, 8 past lags)	1.61	0.7924	0.01019	60
$M1=f(GNP$, 4 future, 8 past lags)	11.25**	0.8385	0.00403	60
$MB=f(GNP$, 4 future, 8 past lags)	5.89**	0.8735	0.00410	60

* Significant at 0.10 level.
** Significant at 0.05 level.
[a] All regressions were fit to the period 1949III-1968IV. $M1$ is currency plus demand deposits. MB is monetary base as prepared by the Federal Reserve Bank of St. Louis. The F-tests shown are for the null hypothesis that all right-hand side variables except trend and seasonal dummies had zero coefficients. See also notes to Table 4.

TABLE 2—F'S FOR COMPARISONS OF SUBPERIODS 1948III–1957III vs. 1957IV–1968IV[a]

Regression Equation	F	Degrees of Freedom
$GNP=f(M1$, 8 past lags)	1.44	(14, 50)
$GNP=f(MB$, 8 past lags)	0.64	(14, 50)
$M1=g(GNP$, 4 future, 8 past lags)	0.88	(18, 46)
$MB=f(GNP$, 4 future, 8 past lags)	1.01	(18, 46)

[a] Tests are for the null hypothesis that all coefficients (including trend and seasonals) remained the same in both subsamples.

TABLE 3—F-TESTS ON FOUR FUTURE QUARTERS' COEFFICIENTS[a]

Regression Equation	F
GNP on $M1$	0.36
GNP on MB	0.39
$M1$ on GNP	4.29**
MB on GNP	5.89**

** Significant at 0.05 level

[a] All tests apply to regressions run over the full sample and are assumed distributed as $F(4, 60)$.

and the estimated shapes for those regressions appear broadly reasonable on the assumption that coefficients on future lags are small and coefficients on past lags are nonzero and fairly smooth. (See Table 4 and Figures 1 and 2.)

These results allow firm rejection of the hypothesis that money is purely passive, responding to GNP without influencing it. They are consistent with the hypothesis that GNP is purely passive, responding to M according to a stable distributed lag but not influencing M.

But let us note a few statistical *caveats.* Though the estimated distribution looks like what we expect from a one-sided true distribution, the standard errors on the future coefficients are relatively high. These results are just what a unidirectional causality believer would expect, but they are not such as to necessarily force a believer in bidirectional causality to change his mind. Also, seasonality problems are

TABLE 4—LAG DISTRIBUTIONS FROM TIME-DOMAIN REGRESSIONS[a]

Coefficient on lag of:	GNP on MB past only	GNP on MB with future	MB on GNP	GNP on $M1$ past only	GNP on $M1$ with future	$M1$ on GNP
−4		−0.65	.162		−.300	.050
−3		.290	−.013		.120	.117
−2		−.088	.105		.126	.069
−1		−.110	.179		.105	.125
0	.603	.532	.171	.570	.484	.181
1	.593	.507	.015	.370	.412	.089
2	.509	.515	.052	−.034	−.017	.116
3	−.029	.080	.264	.543	.582	.107
4	−.011	.023	.107	−.242	−.363	.027
5	−.865	−.822	−.009	−.178	−.147	.027
6	−.037	−.053	.016	−.180	−.136	.025
7	−.296	−.282	.147	−.157	−.139	.123
8	.072	.039	.130	−.326	−.405	.112
Standard errors of coefficients:						
Largest s.e.	.313	.338	.052	.293	.318	.051
Smallest s.e.	.272	.276	.045	.274	.294	.044
Sum of coefficients	.540	—	—	.365	—	—
Standard error of sum	.442	—	—	.523	—	—

[a] Regressions were on *logs* of variables, prefiltered as explained in the text. Each regression included, in addition to the leading and lagging values of the independent variable for which coefficients are shown, a constant term, a linear trend term, and three seasonal dummies. Trends were in all cases significant. Seasonal dummies were insignificant. (The data were seasonally adjusted.)

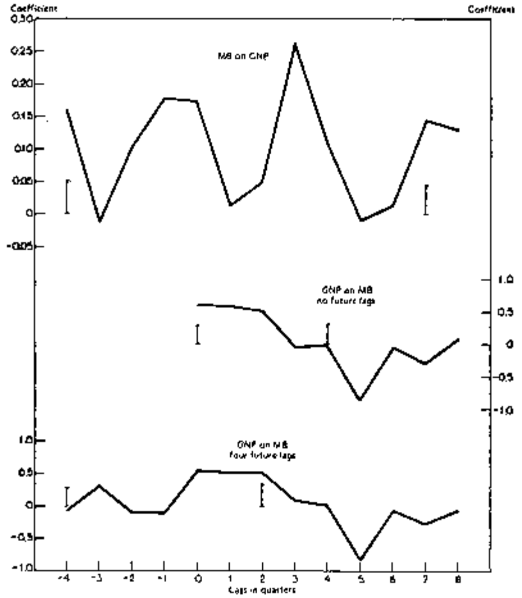

FIGURE 1. LAG DISTRIBUTIONS FOR MB AND GNP
Note: Smallest and largest standard errors are displayed as vertical lines above or below corresponding coefficients.

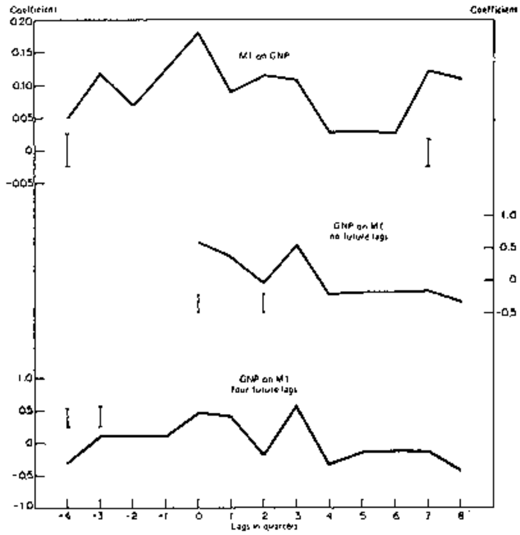

FIGURE 2. LAG DISTRIBUTIONS FOR $M1$ AND GNP
Note: Smallest and largest standard errors are displayed as vertical lines above or below corresponding coefficients.

clearly present in the MB on GNP regression. Seasonality effects appear to be less of a problem with $M1$ than with MB.

DeLeeuw and Kalchbrenner have argued, in attacking the "reduced form" money vs. GNP regressions put out by the St. Louis Fed, that the monetary base is not truly exogenous. We have discussed above the substance of that argument. Suffice it to say here that they claim that one could make the monetary base more "exogenous" by extracting from it borrowed reserves and (possibly) cash in hands of the public. Attempts to use these adjusted MB series (one of them is actually unborrowed reserves) failed, in the sense that relations were less significant statistically and GNP on adjusted MB regressions did not show one-sided lag distributions.

The same regression equations used for GNP and M were estimated also with GNP replaced by the GNP deflator ($PGNP$) and then by real GNP ($RGNP$) with MB the money variable. Quantity theory even in its modern guise does not claim to have firm implications about the way income changes divide into real and price components, but it seemed useful to examine the possibility that monetary variables would predict the components separately as well as their product. Standard errors of the (logarithmic) equations regressing $RGNP$ on MB were slightly larger than corresponding standard errors for current dollar GNP. Values of coefficients and F-statistics were much the same with $RGNP$ as dependent variable as with GNP the dependent variable. Future lags were again highly significant for MB on GNP regressions and highly insignificant for the reversed relation. However, with $RGNP$, current plus eight past lagged values of MB were not as a group significantly different from zero at the .10 level. With $PGNP$, standard errors of estimate were small, but almost every F-test failed to attain significance, including the test on future lags in the *MB* on *PGNP* relation.

V. Tests for Serial Correlation in Residuals

Durbin-Watson statistics for all reported regressions are close to two. This is to be expected because of the prefiltering. The test on the cumulated periodogram of the residuals, described by James Durbin, yields results in the indeterminate range for each regression.[15] The test on the cumulated periodogram is in principle capable of detecting departures from serial independence even when there is no first-order serial correlation, and in this sense is a stronger test than the Durbin-Watson for the case at hand.

The central difficulty here, though, is that a total of 17 of the available 78 degrees of freedom have been used up in the regression, so that the easily-computed bounds tests leave a wide range of indeterminancy. An alternative to the bounds tests is to use the likelihood ratio test for the null hypothesis that the periodogram of the residuals has constant expectation across a number of intervals. This test is described in E. J. Hannan (1960), p. 98.[16] In application to regression residuals this test is justified only when the number of observations is much larger than the number of independent variables, which is clearly not the case here. The statistics reported in Table 5 would be distributed as chi-square with 7 degrees of freedom if asymptotic results applied, but the true

TABLE 5—LIKELIHOOD-RATIO TESTS FOR WHITE NOISE RESIDUALS[a]

GNP on *MB*	*GNP* on *M*1	*MB* on *GNP*	*M*1 on *GNP*
13.02	19.01	11.04	12.64

Note: .05 significance level for chi-squared with 7 degrees of freedom, 14.1

[a] The statistics shown are each distributed asymptotically as chi-square with 7 degrees of freedom on the null hypothesis of white noise residuals. As noted in the text, the asymptotic distribution is probably not a good approximation to the true distribution here. For the *GNP* on *M* equations, residuals were taken from the form with no future lags. For the *M* on *GNP* equations, residuals were taken from the form including future lags.

significance levels of the test will be higher than the nominal ones. Even at nominal significance levels, though, only the residuals from the regression of *GNP* on *M*1 are significantly "nonwhite" at a 5 percent level.

The conclusion from this list of approximate or inconclusive tests can only be that there is room for doubt about the accuracy of the *F*-tests on regression coefficients.

As a check on the least squares results, these same regressions were estimated also using a frequency-domain procedure, Hannan's (1963) "inefficient" procedure.[17] This procedure has some disadvantages relative to least squares, but it has the two advantages that 1) it makes it computationally simple to estimate the variance-covariance structure of the residuals and use the estimate in constructing tests on the estimated regression coefficients and 2) it makes it easy to deseasonalize raw data directly. Not all the tests for significance of groups of coefficients came out

[15] The test carried out was actually based on cumulation of the periodogram over 128 equally spaced points, instead of over the 39 harmonic frequencies as would be appropriate to get Durbin's test. This difference is, however, demonstrably asymptotically negligible (as sample size increases Durbin's test converges in distribution to any test based on more points than half the sample size) and seems unlikely to have been very important even at this particular sample size.

[16] Hannan's description includes Bartlett's small-sample correction to the likelihood ratio test. The results reported in Table 5 do not include the Bartlett correction, since it was small.

[17] The theory of these estimates has been extended in Hannan (1967) and Wahba. It is worthwhile noting that Wahba's proof that the Hannan inefficient estimates are "approximately" least squares estimates is not a proof that the Hannan inefficient estimates have the same asymptotic distribution as least squares, and their asymptotic distributions are in fact different.

the same way at the same significance levels in the frequency-domain estimates, but the general agreement with the least squares results was so close that there is no point in reproducing the frequency-domain results here.[18] Raw data for the monetary base was not readily available, but frequency-domain estimates using raw data on $M1$ and GNP, symmetrically deseasonalized, gave results very similar to those obtained with least squares on published deseasonalized data.

VI. The Form of the Lag Distribution

The lag distribution estimated here to relate GNP to M has only a loosely determined form because of the lack of prior restrictions on its shape. Still, it is worthwhile noting that it agrees in general shape with many previous estimates, and that it can be given an economic explanation. The distribution is positive at first, then becomes mostly negative beyond the fourth lag. The initial positive coefficients sum to a number greater than one, though the sum of all the coefficients is less than one. (Note, though, that the standard error on the sum of coefficients is very large. See Table 4.) The pattern of a short-run elasticity exceeding unity and a long-run elasticity below unity agrees with the theoretical speculations of Friedman (1969), pp. 138–39, concerning the effects of a demand for money dependent on permanent rather than on current income. However, note that the contemporaneous quarter response is less than unitary, and that negative response does not set in for several quarters. To explain this, one must either introduce an averaging procedure into the other side of the equation, making "permanent money" depend on permanent income, or one must introduce the possibility of transactional frictions which keep the economy off its demand curve for money in the short run. At least the latter of these elements is not novel. Alan Walters pointed out that over short enough time intervals people are likely to be off their demand curves. It seems only natural that, since individuals' money balances always fluctuate over short periods due to random timing of transactions, it should take time for changes in money balances to affect individuals' spending behavior.

[18] The frequency-domain results were presented and discussed in an earlier mimeographed version of this paper.

VII. Conclusion

The main conclusions of the paper were summarized in the introduction. I repeat them more briefly here: In time-series regression it is possible to test the assumption that the right-hand side variable is exogenous; thus the choice of "direction of regression" need not be made entirely on a priori grounds. Application of this test to aggregate quarterly data on $U.S.$ GNP and money stock variables shows that one clearly should not estimate a demand for money relation from these data, treating GNP as exogenous with money on the left-hand side; no evidence appears to contradict the common assumption that money can be treated as exogenous in a regression of GNP on current and past money.

Appendix

THEOREM 1: *Y does not cause X in Granger's definition if, and only if, in the moving average representation*

$$\text{(A)} \qquad \begin{bmatrix} X(\text{t}) \\ Y(\text{t}) \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} * \begin{bmatrix} u \\ v \end{bmatrix} (\text{t}),$$

a or b can be chosen to be identically zero.

PROOF:

Following Rozanov we introduce the notation $H_z(\text{t})$ to stand for the completion under the quadratic mean norm of the linear space of random variables spanned by $z(s)$ for $s \leq \text{t}$. Suppose b is zero. Clearly $X(\text{t})$ then lies in $H_u(t)$. By the definition of a moving average (*m.a.*) representation, $H_{X,Y}(t)$ is identical to $H_{u,v}(t)$. But it follows from Rozanov's "Remarks" on pages 62–63 that if $H_u(t)$ and $H_X(t)$ are not identical, then with b zero the identity of $H_{X,Y}(t)$ and $H_{u,v}(t)$ fails. Therefore, $H_u(t)$ and $H_X(t)$ are identical. But then the projection of $X(t)$ on $H_{X,Y}(t-1)$ is in $H_X(t-1)$, which is to say that given past X, past Y does not help in predicting current X. One side of the double implication is proved.

In Granger's definition, Y not causing X is the same thing as the projection of $X(t+1)$ on $H_{X,Y}(t)$ lying in $H_X(t)$. Assuming this condition holds, define $u(t)$ as the difference between $X(t)$ and the projection of $X(t)$ on $H_X(t-1)$. Define $w(t)$ as the difference between $Y(t)$ and the projection of $Y(t)$ on $H_{X,Y}(t-1)$. Finally, define $v(t)$ as that part of $w(t)$ orthogonal to $u(t)$ (i.e., the residual in a regression of $w(t)$ on $u(t)$). By definition, $u(t)$ and $w(t)$ and therefore, $v(t)$ are uncorrelated with past values of each other. Also, $u(t)$ and $v(t)$ are contemporaneously uncorrelated and $H_{u,v}(t)$ is identical to $H_{X,Y}(t)$. Expressing $X(t)$ and $Y(t)$ in terms of the coordinates $u(s)$, $s \leq t$, will give us a moving average representation of the form (A).

THEOREM 2: *When* $\begin{bmatrix} X \\ Y \end{bmatrix}$ *has an autoregressive representation, Y can be expressed as a distributed lag function of current and past X with a residual which is not correlated with any $X(s)$, past or future if, and only if, Y does not cause X in Granger's sense.*

PROOF:

Suppose Y can be expressed as a distributed lag on X with a residual $w(t)$ independent of $X(s)$ for all s. Let $u(t)$ be the fundamental white noise process in the moving average representation of $X(t)$ alone and $v(t)$ be the fundamental white noise process in the *m.a.* representation of $w(t)$ alone. Write the assumed distributed lag relation

$$\text{(B)} \qquad Y(t) = \mu^* X(t) + w(t)$$

Then clearly

$$\text{(C)} \qquad Y(t) = \mu^* a^* u(t) + d^* v(t),$$

where a^*u and d^*v are the *m.a.* representations of X and w, respectively. The equation (C) together with the *m.a.* representation of X are clearly in the form (A) with $b \equiv 0$. Now we need only verify that u and v are jointly fundamental for X and Y, and for this we need only show that $H_{X,Y}(t)$ includes $H_{u,v}(t)$. $H_u(t)$ is in $H_X(t)$ by definition. $H_v(t)$ is in $H_w(t)$ which is in turn (by inspection of (B)) in $H_{X,Y}(t)$. One side of the double implication is proved. Suppose that we have the autoregressive representation

$$\text{(D)} \qquad \begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} * \begin{bmatrix} X \\ Y \end{bmatrix}(t) = \begin{bmatrix} u \\ v \end{bmatrix}(t)$$

and that the *m.a.* representation has the form (A) with $b \equiv 0$. Let G be the matrix on the right-hand side of (A) and H be the matrix on the left-hand side of (D). Then almost everywhere $\tilde{G}^{-1} = \tilde{H}$. (The tilde denotes a Fourier transformation.) Since $\tilde{G}$ can be written in triangular form, $\tilde{H}$ (and thus H) can be written triangular also. But then we can substitute the first equation of (D) into the second euqation of (A) to obtain

$$\text{(E)} \qquad Y(t) = c^* \alpha^* X(t) + d^* v(t) x$$

Equation (E) has the desired properties, since X can be expressed entirely in terms of u and v is uncorrelated with u.

REFERENCES

H. Akaike, "Some Problems in the Application of the Cross-Spectral Method," in B. Harris, ed., *Advanced Seminar on Spectral Analysis of Time Series*, New York 1967.

W. Brainard, and J. Tobin, "Pitfalls of Financial Model Building," *Amer. Econ. Rev. Proc.*, May 1968, *58*, 99–122.

P. Cagan, *Determinants and Effects of Changes in the Stock of Money*, Nat. Bur. Econ. Res. *Stud. Business Cycles*, No. 13, New York 1965.

F. DeLeeuw, and J. Kalchbrenner, "Monetary and Fiscal Actions: A Test of Their Relative Stability—Comment," *Fed. Reserve Bank St. Louis Rev.*, Apr. 1969, *51*, 6–11.

J. Durbin, "Tests for Serial Correlation in Regression Analysis Based on the Periodogram of Least Squares Residuals," *Biometrika*, Mar. 1969, *56*, 1–16.

M. Friedman, "The Monetary Studies of the National Bureau," Nat. Bur. Econ. Res. *Annual Report* 1964; reprinted in Friedman (1969).

———, "The Lag in the Effect of Monetary Policy," *J. Polit. Econ.*, Oct. 1961, *69*, 447–66; reprinted in Friedman (1969).

———, *The Optimum Quantity of Money and Other Essays*, Chicago 1969.

——— and A. Schwartz, (1963a) *Monetary History of the United States 1867–1960*, Nat. Bur. Econ. Res. *Stud. Business Cycles*, No. 12, Princeton 1963.

——— and ———, (1963b) "Money and Business Cycles," *Rev. Econ. Statist.*, Feb. 1963, supp., *45*, 32–64; reprinted in Friedman (1969).

C. W. J. Granger, "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, July 1969, *37*, 424–38.

E. J. Hannan, *Time Series Analysis*, London 1969.

———, 'Regression for Time Series," in M. Rosenblatt, ed., *Time Series Analysis*, New York 1963.

———, "Estimating a Lagged Regression Relation," *Biometrika*, 1967, *54*, 409–18.

E. Malinvaud, *Statistical Methods of Econometrics*, Chicago 1965.

Y. S. Razanov, *Stationary Random Processes*, San Francisco 1967.

C. A. Sims, "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," *J. Amer. Statist. Ass.*, Mar. 1972, *67*, 169–75.

———, "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," mimeo 1970.

———, "Discrete Approximation to Continuous-Time Distributed Lags in Econometrics," *Econometrica*, May 1971, *39*, 545–63.

J. Tobin, "Money and Income: Post Hoc Ergo Propter Hoc?" *Quart. J. Econ.*, May 1970, *84*, 301–17.

G. Wahba, "Estimation of the Coefficients in a Multi-Dimensional Distributed Lag Model," *Econometrica*, July 1969, *37*, 398–407.

A. Walters, "Professor Friedman on the Demand for Money," *J. Polit. Econ.*, Oct. 1965, *73*, 545–55.