A Comment

by Albert Ando

In rereading papers presented at this conference and sorting out my recollection of the discussion at the conference, there appear to be two messages conveyed by authors of the conference papers to econometricians in general. The first is an exposition of a variety of methods of time series analysis, which many of us found very useful. There is no dispute on the implication of this exposition: that is, econometricians should be more careful and systematic in analyzing the residual term of behavioral equations; and I have nothing further to say on this point.

The second message, which is never completely explicit in any of the four papers but comes through very strongly when one reads through all four papers, seems more controversial. It is a recommendation that the whole thrust of empirical work in macroeconomics be redirected. Instead of constructing macroeconomic models based as much as possible on macroeconomic theory, making sure that sufficient identifying conditions are imposed on them, and then estimating parameters by traditional methods, it is suggested that we look for empirical regularities among time series and attempt to identify "causal" relations in Granger's sense. It may be that none of the authors literally meant to advocate this position, but this is the message that comes through. For instance, Granger and Newbold, in the introduction and conclusion of their paper, speak of combining the best features of both the time series method and the traditional econometric methodology; yet in their concrete, eight-step suggested procedure for building a model outlined in section 4 of their paper, there is no reference to *a priori* restrictions that economic theory or other considerations may impose on the model.

Even after reading through the brilliant essay by Sims on causality, I am still quite unclear how to interpret the test of causality in Granger's sense. Suppose that we are working with the definition of causality in the sense of H. A. Simon, and one of the identifying conditions of the model is critical in specifying the causal relationship in the model. Then the test of this identifying condition, assuming that the model is overidenti-

fied, has a very specific interpretation as a test of one of the specifications of the behavioral hypothesis. Sims' test of Granger causal ordering, on the other hand, is a test of specification of cross covariance function between two variables, and the relation between such a test and a specific behavioral hypothesis is by no means obvious. Thus, when Sims announced that, according to his tests, money causes income but not the other way around, my instinctive reaction was that either the test or the definition of causality must be defective. During the period covered by the data used in his test, the monetary authority controlled the supply of money some of the time, the short-term interest rate some of the time, and some other items, such as free reserves, at still other times. The "residual" of the relationship between money and income must then contain, among other things, rates of interest and the effects of per capita real income; and by assuming that such a pregnant residual is an *ARIMA* process, the test is addressing itself to only the least useful part of the structure as far as the determination of behavioral relationships is concerned. It is a mildly curious fact that the covariance function between income and money for a particular past period in history happens to satisfy the properties required by Granger causal ordering. We know, however, that the true relation between these two variables is a more complex one, involving other critical variables, and that the pairwise test of Granger causal ordering tells us nothing about this underlying behavioral relationship. Hence, the observed Granger ordering of these variables in past data says nothing whatever about the possibility of the future response of income to any change in money, or the response of money to income.

From the point of view of a more traditional econometric approach to the problem, the difficulties of the time series analysis as it has been applied to macroeconomic problems in practice is that before the statistical analysis begins, the researchers do not seem to ask why they are carrying out their analysis.† The result is the difficulty in interpreting statistical results in terms of well-perceived behavioral hypotheses, even in the best of works of this type, as my comment on Sims' work above suggests. Since the statistical test is not specifically tied to a specific behavioral hypothesis, well-meaning graduate students learning about the test of Granger's causal ordering are justified in thinking that many other tests that are in some sense "similar" to the one carried out by Sims must be as meaningful as the original work of Sims. Suppose now that such a graduate student selects 50 or so variables in addition to money which the student views as possibly having causal relationships with income,

†Sargent and Sims attempt to justify this deemphasis of economic theory by saying that macroeconomic theory does not have much foundation in any case. It would make this comment much too long for me to engage in a debate on this larger topic. But I feel compelled to point out that the Phillips curve, or the determination of the general price level, is the least suitable example to debate of this issue because the determination of the general price level is on an entirely different theoretical foundation from the determination of relative prices and quantities, both in macroeconomics and in general equilibrium analysis.

carries out the test between each one of them and income, and finds that seven of them "cause" income but are not caused by income, four of them are caused by income but do not cause income, and two of them cause income and also are caused by income.

Now the intriguing question at this point is the probability that through sampling errors, the above test result is accepted on a basis of a finite sample when, in population, there is no causal relationship between any of these 50 variables and income. This, of course, depends on the level of significance of the test chosen and on the power of the test. Unfortunately, none of the authors at this conference showed us how to compute powers of the proposed test of the Granger causality, so I am unable to answer my own question.

I am sure authors of the papers at this conference will protest that they know what I am suggesting is data mining and that they had no intention of advocating data mining. But I am afraid that their intention is not relevant because any statistical analysis which is not firmly related to a well articulated behavioral hypothesis concerning the underlying data generating process is data mining. Now I am not necessarily against data mining under all circumstances, and I cheerfully admit to having done data mining from time to time quite profitably. What I most seriously object to is data mining by persons who are not aware that their analyses are data mining operations. I may be permitted to indicate the seriousness of danger by way of quoting an earlier, very obscure result of mine.

Suppose that a student has one variable y to explain and a set of variables x_1 to x_p that are considered as potentially capable of contributing to the explanation of the variance of y, and a finite sample of size n on y and x's.

Let r_{0i} represent the sample correlation coefficient between y and x_i. Corresponding population values will be denoted by ρ_{0i}. The student does not have any behavioral hypothesis underlying this set of data and proceeds to compute all r_{0i}'s, $i=1,2,\ldots p$, and to accept all x_i's having sample correlation coefficients "significantly different from zero" as potential explanatory variables of y. Now anyone can see that this procedure involves a serious bias. Even if all ρ_{0i}, $i=1,2,\ldots,p$, are exactly zero, for a finite sample, r_{0i}'s will have distribution with finite variance, so that the probability that some r_{0i} is different from zero by more than some finite distance is not zero. The student should have computed — under the assumption that all ρ_{0i} are zero and given that there are p of x_i's and that the sample size is n — the expected number of sample r_{0i}'s that are different from zero by the distance defined by the significance test and then compared that sample against this number. Perhaps better still, the expected value $r^*(p,n) = E[r(p,n)]$ of

$$r(p,n) \equiv \max_{1 \leq i \leq p} |r_{0i}|$$

could have been computed under the assumption that ρ_{0i} is zero for all i and that y and all x_i's are pure white noise series, generated by, say, standardized normal distribution. Then only those x's whose correlation with y is "significantly" larger than $r^*(p,n)$ can be considered as potential explanatory variables. This expected value is extremely difficult to obtain analytically, and some years ago, I computed them numerically using 50 artificially generated samples for a variety of values of p and n. I reproduce below a few representative values of r^*.† It should be noted that these values represent a finite sample bias and that the attention to power functions of the test will not alleviate this problem.

Values of r^* for selected values of p and n.‡

p \ n	30	50	.70	100
20	.39	.31	.25	.22
30	.42	.33	.28	.23
50	.45	.37	.29	.24

The analogy between this procedure and the search for statistically significant causal ordering in Granger's sense without specification of a behavioral hypothesis to guide the search is obvious. I am sure authors of the papers at this conference will consider my description through this analogy of their recommendation on how to do econometric research to be grossly distorted and unfair. But the central point remains. These authors appear to say that by relying on techniques of the time series analysis, we need to specify the behavioral hypothesis underlying economic data much less than we do when working with the more traditional econometric approach. In my judgment, it is not only wrong but it is dangerous even to appear to make such a claim.

I believe that econometricians have a good deal to learn from techniques developed in the time series analysis, but these techniques are no more or less than one more tool that econometricians should know and make use of when appropriate. These techniques are not capable of replacing economic theory and all other econometric methods that have been developed and are being utilized by econometricians of today. To claim more for these time series techniques is a serious disservice to the techniques themselves.

†See Ando and Kaufman [2].

‡ Source: See [2].