Notes on Aggregate Models Based on Time Series

September 5, 1950

by Carl Christ

These notes are intended as suggestions to anyone interested in the construction and testing of aggregate econometric models based on time series data (some of the comments will apply equally to cross section data); in particular they refer to the kind of work represented by Klein's "The Use of Econometric Models as a Guide to Economic Policy" (Econometrica, 15, April 1947, pp. 111-51) [which is more fully expounded in Economic Fluctuations in the United States, 1921-1947 (Cowles Commission Monograph 11, New York, Wiley, 1950)] and Christ's paper in the forthcoming National Bureau volume on the 1949 Conference on Business Cycles Research (which exists in more primitive form as Cowles Commission Discussion Paper, Economics No. 269). The corresponding computations are to be found in the Cowles Commission computation file under Job 288 and Job 333, respectively.

Other suggestions will be found in the Discussion Papers, Economics Nos. 227 (Domar), 241 and 241A (Christ), 253 (Marschak), 259 (Bjerve), and Statistics Nos. 340-341 (Christ), and also in Klein's comments in Christ's paper in the above mentioned National Bureau volume.

These notes are divided into two sections, one on the form of the model, and the other on estimation, prediction, and other operations performed on the model.

A. The Form of the Model.

1. It is often suggested that the process of econometric model-building properly begins with the formal derivation, from theoretical considerations, of equations describing behavior and technical and institutional constraints; and that after the equations have been so derived, they are to be confronted with data so that their parameters may be estimated. This procedure would be the ideal one if our theoretical considerations were numerous enough and specific enough to enable us to derive unique expressions for such things as the consumption function, the demand-for-money equation, etc. The available theories are not able to lead us to unique equations, however (who can say, for example, on theoretical grounds alone, whether the consumption function should depend linearly or quadratically on income, and whether it should depend on past income or past consumption or neither or both?) Since this is the case, it is necessary to use other means in addition to economic theory in order to choose the form of each equation which is to be estimated. A means which suggests itself quite naturally is to choose from among theoretically acceptable forms one which is highly consistent with past observations (i.e., which produces small random-looking residuals when fitted to past data). Of course it would be possible to get an exact fit if a sufficiently complex form were used (e.g., an (n-1)th order polynomial for n observations in two variables), so it will be necessary to sacrifice good fit to simplicity or vice versa.

What I mean to say is that in choosing the form of an equation, it is well in practice to pay attention to how well each of the alternative Theoretically acceptable forms fits the past data, because economic theory leaves a great many possibilities in the "acceptable" category. Experimenting with data and different forms may suggest interesting hypotheses.

The disadvantages of this procedure are several. First, the fitting will probably be done by least squares, which is known to be biased except in special cases.

But if there is a close relationship among several time series, it is likely to result in small residuals from a regression. Second, the data used to test the hypothesis will have been used already in selecting the hypothesis, i.e., the restrictions assumed upon the variables. This means that the standard errors will be spuriously small if they are calculated on the usual assumption that the restrictions are known a priori. Third, there is no guarantee that a form of equation which has been empirically found to fit well for a certain period will continue to do so...there is no guarantee against reading a particular systematic relationship into a short configuration where it does not exist. Still, these are prices worth paying, in the absence of more unequivocal theory.

2. The variable T (essentially tax yield) is not fixed by government policy; it is tax rates which are so fixed. An exogenous variable or variables should be introduced to represent the tax rate structure, and then tax yield should be found as a function of the rate structure and income (and possibly income distribution, though this can probably be assumed constant).

3. The points at which the government impinges on the economy should be enumerated, and if possible an exogenous variable should be included for each such point. These might include tax rate (as mentioned above); interest rate or price of government bonds; consumer credit regulations; foreign credit policy; subsidies to certain industries (e.g. agriculture) or demand for output of certain industries, if the model is broken down into industrial sectors; rent and price controls; policy toward unions (the Wagner and Taft-Hartley Acts had important consequences); public housing; social security and other transfer payment schedules; etc. (see Bjerve's Economics Discussion Paper No. 259).

B. Estimation, Prediction, and Other Operations

1. In the Klein and Christ models referred to on page 1, the exogenous variables G (government expenditures) and T (essentially government revenues) are not

among the predetermined variables used in the limited information estimation. This oversight occurred because G and T do not appear in any stochastic equation of either model. It is a serious oversight, because G and T are two of the most important of all government variables. Some other exogenous variables are also omitted from consideration in the estimation, but they are less important. This situation should be corrected in any future work.

2. It has been suggested by Klein (in Monograph 11, p. 5, and in his comments in the abovementioned National Bureau volume) that better predictions can be made by (a) substituting estimated values of parameters and known (or assumed) values of predetermined variables into the structural equations, and then solving the system of structural equations (including identities) for the jointly dependent variables, than by (b) merely substituting known (or assumed) values of predetermined variables into the reduced-form regressions. The intuitive argument for this is that the structural equations contain a priori restrictions which are used in method (a) but ignored in method (b), so that (a) can be expected to be more efficient.

The results of a few experimental computations to obtain predictions by method (a) are disappointing. Predictions of price level p and disposable income Y, from Klein's model III for 1941 and from my model for 1948, are absurdly different from the corresponding observed values. (Klein's model yields a cubic equation in p, and my model yields a quintic in p, because of nonlinearities.)

The above computations were made on the assumption that the effect of all disturbances is zero. But because of nonlinearities in the models, some of the disturbances appear in the simultaneous solution as squares or cross-products, so that the expected value of their effect is not zero but is a function of their variances and covariances. A recomputation of the 1948 predictions from my model, using estimated variances and covariances where squared or cross-multiplied disturbances appear, yields the same results to two significant figures as do the first computation which assume the effects of all disturbances to be zero.

It would be interesting and useful to investigate the reasons for these disappointing results.

3. In principle it is possible to obtain, from such models as we discuss here, an expression for each endogenous variable in terms of exogenous variables and lagged values of itself, i.e., with no lagged values of other endogenous variables. This would be a difference equation, and could be solved to obtain a time-path for the endogenous variable involved, as a function of disturbances and exogenous variables. For fixed values of the exogenous variables and disturbances, the limiting value, if any, could be found, and also the period (if the solution of the difference equation were oscillatory) and damping factor. The derivatives of these three quantities with respect to exogenous policy variables could also be found. Tinbergen's study of the U. S. 1919-1932 contains some computations of this sort.

4. The Hart-von Neumann test of δ^2/S^2 (ratio of mean square successive difference to variance) for the presence of serial correlation of disturbances (<u>Annals of Mathematical Statistics</u>, 1942) breaks down, as Orcutt and Cochrane show (two papers in <u>Journal of American Statistical Association</u>, 1949); the reason is that the <u>true</u> disturbances are never observed, and so the test must be conducted with the <u>calculated</u> disturbances instead. The distribution of δ^2/S^2 might be susceptible of computation on this basis. Another alternative would be to use a test of runs in the disturbances instead of a test of δ^2/S^2, but there might be a strong cyclical pattern of disturbances which would escape the notice of the runs test, if the pattern were sufficiently saw-toothed (like this ).

5. To catch the above and other non-random features of the disturbances, it would be well to graph them for each possible form of each equation and examine them qualitatively (this suggestion is similar to A. 1. above).