COMMENT on Cowles Foundation Paper 49
"A Test of an Econometric Model for the United States, 1921-1947,"
by Carl Christ

MILTON FRIEDMAN, University of Chicago

First, may I congratulate Carl Christ and the Cowles Commission for undertaking to test the predictive value of Klein's econometric model and for the thoroughly objective and scientific manner in which they have performed this task. Economics badly needs work of this kind. It is one of our chief defects that we place all too much emphasis on the derivation of hypotheses and all too little on testing their validity. This distortion of emphasis is frequently unavoidable, resulting from the absence of widely accepted and objective criteria for testing the validity of hypotheses in the social sciences. But this is not the whole story. Because we cannot adequately test the validity of many hypotheses, we have fallen into the habit of not trying to test the validity of hypotheses even when we can do so. We examine evidence, reach a conclusion, set it forth, and rest content, neither asking ourselves what evidence might contradict our hypothesis nor seeking to find out whether it does. Christ and the Cowles Commission have not followed this easy path. They have revised the parts of Klein's econometric model that fit postwar experience least well, then, and this is the important step, have used the revised model to predict an additional year and compared the results with what actually happened. The fact that the results suggest that Klein's experiment was unsuccessful is in some ways less important than the example they set the rest of us to go and do likewise. After all, most experiments are destined to be unsuccessful; the tragic thing is that in economics we so seldom find out that they are.

Klein's model is not the only attempt to construct a system of simultaneous equations to predict short-time changes in important economic
phenomena. Probably the most ambitious was the model constructed by Jan Tinbergen on the basis of United States data for 1919-32. Slightly different in character but in the same general class were the equations computed under the direction of Gardiner C. Means at the National Resources Board and published in *Patterns of Resource Use* (1939). More recently, Colin Clark published in *Econometrica* (April 1949) a rather simpler model for the United States economy. And there are still others, both published and unpublished.

These systems of equations all describe adequately the data from which they were derived; that is, they yield high correlation coefficients and most of the estimated parameters are several times their standard errors. But, as is fairly widely recognized by now, the fact that the equations fit the data from which they were derived is a test primarily of the skill and patience of the analyst; it is not a test of the validity of the equations for any broader body of data.¹ Such a test is provided solely by the consistency of the equations with data not used in their derivation, such as data for periods subsequent to the period analyzed. It is my impression that this test has at various times been applied to Tinbergen's and Means' equations and that neither survived it satisfactorily. Colin Clark's model, as far as I know, has not yet been tested.

Christ accepted tests by Andrew Marshall as a basis for revising Klein's original model. He then proceeded to test his revision of Klein's model in two main ways: first, he tested the internal consistency of the equations by seeing whether errors in predicting 1948 values were larger than might be expected on the basis of the unexplained variation in the sample years; second, he compared the predictions of his econometric model with the predictions of what he terms "naive" models.

As the tests of internal consistency seem to me far less important than the naive model tests, I shall add little to what Christ says about them. I wish to mention only that the choice of the probabilities used in defining the tolerance interval (γ and P in Christ's notation), though not discussed by Christ, is critical. The choice of sufficiently high values for these probabilities will assure the acceptance of almost any equation, no matter how bad the prediction, though of course only at the expense of a high chance of failing to reject the equation when it is 'false'. It is my hunch that, given the size of sample, the values initially chosen by Marshall and used by Christ involve an unduly small risk of rejecting a 'correct' equation compared with the risk of accepting an 'incorrect' equation.

The naive model tests deserve somewhat more analysis. The naive

models should not be taken seriously as techniques for actually making predictions; they are not competing theories of short-time change. Their function is quite different. It is to provide a standard of comparison, to set the zero point, as it were, on the yardstick of comparison. We say that the appropriate test of the validity of a hypothesis is the adequacy with which it predicts data not used in deriving it. But how shall we assess the adequacy of prediction? Obviously we need not require perfect prediction; so the question is when are the errors sufficiently small to regard the predictions as unsuccessful? We cannot judge by the absolute size of the error; on what grounds are we to say that an error of, say, $1$ billion is either small or large? Nor do percentage errors help much, even though they seem intuitively more relevant. An error of $2$ per cent means one thing if the variable being predicted never varies by more than $3$ per cent and quite a different thing if it usually varies by $50$ per cent. Moreover, the percentage error is itself really arbitrary. For example, suppose we know income and seek to predict savings and consumption expenditures. Since consumption expenditures will be something like $10$ times as large as savings, a $20$ per cent error in savings will be approximately a $2$ per cent error in consumption. Which is the appropriate number for judging the adequacy of the prediction? The $2$ per cent or the $20$ per cent error?

If predictions are made for several years (or other units) one simple method of testing the accuracy of the predictions is by the correlation between the predicted and the actual values. This can be computed and compared with the correlations to be expected between chance series, and the prediction judged a success if the correlation is higher than might reasonably be expected from chance alone. But this test, too, has its defects: it is likely to be relatively insensitive for a small number of predicted values; it may require an estimate of the serial correlation among observations if the appropriate sampling distribution is to be used; it is not clear what the appropriate alternative hypotheses are in terms of which the test of significance should be chosen.

The naive models provide an alternative, though related, standard of comparison, which can be used for one year or many years, and which takes account of serial correlation. They are in some sense the 'natural' alternative hypotheses — or 'null' hypotheses — against which to test the hypothesis that the econometric model makes good predictions. The reason can be easily seen. The essential objective behind the derivation of econometric models is to construct an hypothesis of economic change; any econometric model implicitly contains a theory of economic change. Now given the existence of economic change, the crucial question is whether the theory implicit in the econometric model abstracts any of the essential forces responsible for the economic changes that actually occur. Is it better,
that is, than a theory that says there are no forces making for change? Now naive model I, which says the value of each variable next year will be identical with its value this year, is precisely such a theory; it denies, as it were, the existence of any forces making for changes from one year to the next. In the language of the econometric model, it says that the appropriate structure is one in which all the equations contain only a constant term, the rest of the parameters being zero. If the econometric model does no better than this naive model, the implication is that it does not abstract any of the essential forces making for change; that it is of zero value as a theory explaining year to year change.

Of course, there are many varieties of change, and many different objectives may be set for an econometric model or for any other theory of change. The forces that are essential in explaining changes from one year to the next may not be the same as those that are essential in explaining changes over a two-, or a five-, or a twenty-year period. And for each of these types of change there is an appropriate naive model of type I. The fact that an econometric model is rejected for one class of change does not mean that it will be rejected for another; but neither, of course, is there any reason to believe that it will not be.

Change can be differentiated also by criteria other than the period considered. In particular, we frequently distinguish between what is called 'secular' and 'cyclical' change. This is the role of naive model II, which says that the value of each variable next year will differ from its value this year in the same direction and by the same amount as its value this year differed from its value last year. This is a theory of 'pure' secular change, as it were; and it seems to me appropriate if and only if the model being tested has passed naive model test I satisfactorily. In that case, the implication is that the model has successfully abstracted some essential forces making for change, and the question can then be asked whether it has isolated secular forces alone or cyclical forces as well.

Christ's revision of Klein's model does no better than naive model I for the one year for which Christ could make the test, 1948. The econometric model makes larger errors than the naive model for approximately half the variables predicted, and its average error is, if anything, larger than the average error of the naive model.

One is tempted to add that the test is biased in favor of the econometric model because of the way exogenous variables are treated. Christ used the actual values of the exogenous variables for 1948 whereas in making a prediction for a future year it would be necessary to predict the exogenous variables independently. But this is not a valid objection; Christ's procedure is the correct one. The model claims to make only conditional predictions: if the exogenous variables are such and such, the endogenous
variables will be such and such. And the important first question is whether it can make such conditional predictions.

Of course, one swallow does not make a spring; and one must be careful of generalizing too broadly from tests based on predictions for one year. Perhaps if the model were tested for additional years the unfavorable verdict would be reversed; all one can say is that the evidence so far assembled contradicts the hypothesis of short-time economic change implicit in the econometric model. It is highly desirable that additional evidence be accumulated; but meanwhile I shall proceed on the assumption that additional evidence would not reverse Christ’s tentative conclusion.

Christ suggests one qualification to the conclusion that, on the basis of existing evidence, the particular econometric model he tested is worthless. He writes, “Even if such a naive model does predict about as well as our econometric model, our model may still be preferable because it may be able to predict consequences of alternative policy measures and of other exogenous changes, while the naive model cannot.” But this argument is at best misleading, at worst invalid. The naive model can make such predictions too: one can simply assert that a proposed change in policy or in an exogenous variable will have no effect. If this kind of prediction worked as well as the econometric predictions for a change from one year to the next, might it not work as well for policy changes also? Note that the evidence implicitly used in predicting the effect of policy changes by means of the econometric model is derived from year to year changes in the basic data, i.e., from precisely the kind of changes the naive model test suggests the econometric model is incompetent to predict. To put the point in another way, the assertion that the econometric model can be used to predict the consequences of policy changes implicitly assumes that the theory of change implicit in the econometric model abstracts some of the essential forces determining economic change; stated loosely, that the model is an approximation to the ‘correct’ one, and that the parameters are better estimated by giving them the values obtained from the estimated econometric structure than by setting them equal to zero. Now it is precisely these propositions that the naive model test contradicts.

Of course, the policy changes to be predicted may differ in character from the year to year changes that the econometric model failed to predict; and, as Christ suggests, the model may predict the one kind of change even though it does not predict the other. But then it is a pure act of faith to assert that the econometric model can predict the effect of policy changes, and there is no reason for anyone else to share this faith until some evidence for it is presented. Surely, the fact that the model fails to predict one kind of change is reason to have less rather than more faith in its ability to predict a related kind of change.
Granted that this particular experiment in constructing an econometric model must be judged a failure on the basis of present evidence, what implications does this have for future work? One possibility, already mentioned, is that this failure is a freak; that further evidence will show that this model can predict successfully and that one should await such further evidence. Another possibility is that the defects of this model are peculiar to it and not to econometric models of this general kind; that examination of the economic theory implicit in this model, of the detailed shortcomings of individual equations and the like, will permit the construction of an improved model along the same general lines that will work successfully. Neither possibility can be categorically rejected. Like any other prediction, the assertion that it will or will not be realized is a prediction that cannot be made with certainty. My own hunch, however, is that neither possibility will be realized; that additional evidence on this particular model will strengthen rather than reverse the conclusion suggested by the existing evidence and that attempts to proceed now to the construction of additional models along the same general lines will, in due time, be judged failures.

In part, this hunch is simply an extrapolation of experience: as already noted, Klein's model is by no means the first of its general type that has been constructed and tested and so far none has survived the test of ability to predict. But this empirical extrapolation is by itself unsatisfactory. The fundamental premise underlying work in this field is that there is order in the processes of economic change, that sooner or later we shall develop a theory of economic change that does abstract essential elements in the process and does yield valid predictions. When and if such a theory is developed, it will clearly be possible to express it in the form of a system of simultaneous equations of the kind used in the econometric model — mathematics is after all a rather flexible and highly useful language into which practically any economic theory can be translated. Does it not then follow that despite the unsatisfactory results to date, the appropriate procedure is to continue trying one after another of such systems until one that works is discovered?

I think the answer is no. Granted that the final result will be capable of being expressed in the form of a system of simultaneous equations applying to the economy as a whole, it does not follow that the best way to get to that final result is by seeking to set such a system down now. As I am sure those who have tried to do so will agree, we now know so little about the dynamic mechanisms at work that there is enormous arbitrariness in any system set down. Limitations of resources — mental, computational, and statistical — enforce a model that, although complicated enough for our capacities, is yet enormously simple relative to the present state of understanding of the world we seek to explain. Until we can develop a simpler
picture of the world, by an understanding of interrelations within sections of the economy, the construction of a model for the economy as a whole is bound to be almost a complete groping in the dark. The probability that such a process will yield a meaningful result seems to me almost negligible.

The model builders have, of course, recognized this problem. For example, it explains the distinction they make between a model — which is a class of admissible hypotheses — and a structure — which is a single hypothesis. It explains also their emphasis on examining the economic theory implicit in their equations, and on checking the signs of their statistically estimated parameters for 'reasonableness'.

In so far as they think the prospect more hopeful than I do, it is because they assess differently the existing state of our knowledge — they think we have more basis for narrowing the range of admissible hypotheses than I do. On this point, I venture to suggest that they have been misled by failing to distinguish among different kinds of economic theory. We do have a very well developed and, in my view, successful and useful theory of relative prices which tells us a great deal about relationships among different parts of our economic system, about the effects of changes in one part on its position relative to others, about the long-run effects of changes in technology, the resources at our disposal, and the wants of consumers. A theory of short-run changes in the economy as a whole must deal with many of the phenomena that are dealt with in price theory, and thus it is tempting to suppose that price theory substantially reduces the arbitrariness of a system of equations — enables us to narrow substantially the class of admissible hypotheses.

I believe that this is a serious mistake. Our theory of relative prices is almost entirely a static theory — a theory of position, not of movement. It abstracts very largely from just those dynamic phenomena that are our main concern in constructing a theory of economic change. The basic empirical hypothesis on which it rests is that the forces determining relative prices can be considered largely independent of the forces determining absolute prices; and its success is testimony to the validity of this hypothesis.

A theory of change cannot, of course, be constructed completely independently of a theory of relative prices. The two must in some sense be consistent with one another, and thus there is a real point in checking any theory of change to see that it does not have implications for the relations among the parts that are inconsistent with the theory of relative prices. The important point is that the existing theory of relative prices does not really help to narrow appreciably the range of admissible hypotheses about the dynamic forces at work.

Monetary theory, interpreted broadly, has somewhat more to offer. It is
at least concerned with absolute prices. But even monetary theory, in its present state, is less useful than might at first appear. It too has typically been concerned with positions of equilibrium, with comparative statics rather than with dynamics — and this, I may add somewhat dogmatically, applies equally to Keynesian and pre-Keynesian monetary theories.

One cannot, of course, specify in advance what a workable theory of change will look like when it is developed. But I think it is clear that it will have to be concerned very largely with leads and lags, with inter-temporal relations among phenomena, with the mechanism of transmission of impulses — precisely the kind of thing about which neither contemporary price theory nor contemporary monetary theory has much to say.

The direction of work that seems to me to offer most hope for laying a foundation for a workable theory of change is the analysis of parts of the economy in the hope that we can find bits of order here and there and gradually combine these bits into a systematic picture of the whole. In the language of the model builders, I believe our chief hope is to study the sections covered by individual structural equations separately and independently of the rest of the economy.

These remarks obviously have a rather direct bearing on the desultory skirmishing between what have loosely been designated the National Bureau and the Cowles Commission techniques of investigating business cycles. As in so many cases, the difference between the two approaches seems to me much greater in abstract discussions of method than it is likely to prove in actual work. The National Bureau has been laying primary emphasis on seeking to reduce the complexity of phenomena in order to lay a foundation for a theory of change; the Cowles Commission on constructing the theory of change. As the National Bureau succeeds in finding some order, some system, in the separate parts it has isolated for study its investigations will increasingly have to be concerned with combining the parts — putting together the structural equations. As the Cowles Commission finds that its general models for the economy as a whole are unsuccessful, its investigators will increasingly become concerned with studying the individual structural equations, with trying to find some order and system in component parts of the economy. Thus, I predict the actual work of the two groups of investigators will become more and more alike.

LAWRENCE R. KLEIN, National Bureau of Economic Research

Carl Christ has presented a splendid methodological account of a procedure for testing the validity of econometric models, but like many other econometric contributions of recent years it is weak in empirical or sub-
stantive content. I shall argue that his time series data contain an obvious
gross error, that he has not chosen a desirable postwar revision of my
prewar econometric model, and that his forecasting technique is both
wrong and inefficient. Let me make matters quite clear at the outset, I do
not accept any personal responsibility for anything that Christ has done.
I participated to a negligible extent in his work.
The most serious deficiency in Christ's work is in the data he used for
1946-47 to revise my model and bring it up to date. These are critical
observations since they provide the basis for revisions and in samples of
20-25 annual observations can play an important statistical role. In addi-
tion, these data enter as lags in the forecasting for 1948. The series Christ
has constructed show a drop in real aggregate output of more than 10 per
cent from 1946 to 1947; this I do not believe. Every expert whose opinion
I have canvassed concerning this period of experience advances the offhand
guess that real output rose from 1946 to 1947. Christ shows an increase in
total employment of 2.5 million persons from 1946 to 1947, and the Fed-
eral Reserve index of industrial production rose from 170 to 187 in the
same period. Some of the trouble Christ finds with his estimates of the
labor equation may well be traceable to these erroneous cross currents in
employment and output. The same can be said about the estimates of the
production function.
I am not prepared to give a full statement as to the source of his diffi-
culties with the time series data, but it seems plausible to conclude that the
price deflators used to pass from current dollar to constant dollar magni-
tudes are largely responsible. Price controls were lifted in the middle of
1946, and many of the published indexes rose to an excessive degree. I say
excessive because many observers believe that official indexes seriously
understated true prices toward the end of the war and in the early postwar
period. The CIO argued vainly but correctly, I believe, that the BLS cost
of living index was too low during the war. I shall not go into the reasons
since this matter is discussed in other places.
Many of Christ's current dollar estimates could be wrong also because
he used some questionable methods for converting the Department of
Commerce national accounts from the new concepts (post July 1947) to
the old concepts. His aim was apparently to reproduce an extension of
my time series which followed the concepts I used as closely as possible.
This, in itself, was a serious mistake. Looking at the supply-of-data situa-
tion of 1945 and 1946, I tried to do the best I could to get an adequate set

\footnote{Christ, in revising his paper, has written an unsatisfying appendix on the data prob-
lem. An interesting empirical finding of this appendix, however, is that hypothetical
changes in the 1946 observations lead to radically different least-squares estimates of
parameters of important equations.}
of series covering the whole interwar period. I know only too well the deficiencies in my series. Looking at the supply-of-data situation of 1948 and 1949, if I were to have set out upon a revision of my model (Christ's situation), I would first have completely reworked all the series. The national accounts of the Department of Commerce, in their current state, would have been accepted as basic, and everything else would have been adjusted to them, including the pre-1929 data. Instead, Christ adopted the dubious procedure of forming regressions between the old and new concepts of the Department of Commerce during overlapping periods and extrapolated the data on the old concept to recent years from observations of the data on the new concept. Given Christ's questionable objective that he wanted to follow my outmoded time series as closely as possible, he could have done something more satisfactory. He could have accepted the most recent estimates of the Department of Commerce whenever no change of concept was involved, and he could have tried to estimate directly items that account for the change in concept, obtaining more satisfactory estimates of the series based on the old concepts.

In two other specific cases Christ used some data that seem obviously ill chosen. His price index of business capital goods (private producers' nonagricultural plant and equipment) is a weighted average of construction cost indexes and the wholesale price index for metals and metal products. The latter index is a poor substitute for an equipment price index in the postwar period. He then used this price index to extrapolate Fabricant's price index underlying business depreciation charges into the postwar years. It is almost certain that the depreciation deflator did not rise as fast as the price of newly purchased capital goods.

If we want to make a sound judgment about the use of econometric models for predicting some of the main economic magnitudes, we ought to reserve opinion until the most efficient use of the technique with available information has been tested. To forecast in the social sciences is difficult, and it is not likely that we shall get useful results with an inefficient application of any method. Christ's paper represents an inefficient application in many respects, and on the matter of data alone there are numerous things that he must do before he can draw any conclusions. The only really satisfactory approach open to him in the interests of efficiency is to revise all his series to agree with the new data of the Department of Commerce. This is not an easy task and will not appeal, of course, to 'sophisticated' econometricians who are more interested in less tedious research; however, it is imperative. Some day the 'sophisticated' econometrician will learn the trite result that his methods are not very useful when applied to poor data.

If this is too much of a job for Christ, a minimum task may be set out. He must first recalculate his estimates of the current dollar series without
using a regression between the old and new concepts of the Department of Commerce. He should try to estimate directly the magnitudes that reconcile the two series. He should recompute all price deflators more carefully. If he still finds that real output fell from 1946 to 1947, he should revise the model on the basis of the 1947 estimate alone and recompute his extrapolations to 1948. I recommend this step because it seems likely that the main difficulty lies in the price deflators and that these are low in 1946 rather than high in 1947.

The only things for which I assume any responsibility are the construction of the prewar model and the forecasts, from it, for 1946 and 1947. My extrapolation to 1946 (Econometrica, April 1947, p. 134) estimated net national product in 1934 prices to be $121.6 billion. Christ’s figure for the observed value is $115.2 billion. In terms of the customary accuracy involved in economic forecasts, this is not a bad correspondence. It is certainly in the right direction for the postwar situation. My forecast for fiscal 1947 (ibid., p. 133) was $104.5 billion. Christ’s figure for calendar 1947 is $103.3 billion, showing that my fiscal year forecast of real output was undoubtedly near the observed value. Since both my forecasts were made before the events occurred they had to use estimates of the relevant predetermined variables. Some of the estimates were not correct, but that, of course, is the case in any realistic forecasting situation.

The reason for introducing these considerations is to point out that Friedman’s comments on Christ’s paper cannot be accepted. Christ has not shown that econometric models break down as forecasting devices. We have two situations possible. Either Christ’s series are accepted as correct, in which case I have been able to make some satisfactory forecasts from my model, or Christ’s series are deemed incorrect, in which case Friedman cannot draw any substantive conclusions, as yet, from Christ’s work. I am quite sure that the latter possibility is the correct one and that my forecast for 1947 was too low.

To many of us engaged in econometric work, it became obvious in the second half of 1947 that the most serious deficiencies in the existing models lay in the consumption equation and in the group of relations serving to determine absolute prices. During the war, households consumed at a low level in relation to their incomes; i.e., wartime observations on consumption and income lay substantially below the prewar consumption-income relationship. In the first postwar year (or 18 months), consumer spending worked its way back to the prewar relationship. In 1947 and 1948 the high levels of spending far surpassed the old relationship. We have, it appears, returned in 1949 to the neighborhood of the prewar relationship. The consequences of these movements are that the forecasts for 1946 were roughly correct and that those for 1947 and 1948...
were low. Christ has tried to improve the consumption equation by introducing the real stock of cash balances as an additional variable. While there may be some plausibility to this approach, it is evident that it is not adequate. The observed consumption point for 1948 lies well above Christ's equation. In view of the low residual variation reported by Christ for the amended consumption equation in 1947, one might ask whether it would have been possible to know in advance of the 1948 forecast that this equation was going to give a low estimate for 1948. One thing apparent from the statistical estimates of the parameters is that the residual variation from Christ's consumption equation shows high serial correlation. Another shortcoming of his consumption equation is that it has a negative trend. I made two suggestions to Christ at the start of his work that would have improved the consumption equation and reduced the serial correlation of residuals. One change was to introduce lagged consumption as a separate variable in the consumption equation. Sound theoretical justification for this change can be developed. In the revised version of his paper, Christ made some least-squares estimates of a consumption equation of this type and found a remarkable improvement in the extrapolation to 1948. My other suggestion was to split consumption into categories such as durables, nondurables, and services, estimating separate equations for each. My own calculations of the equations for these components show a satisfactory lack of serial correlation in each set of residuals. If relative prices, stocks of consumer durables, and other relevant variables are taken into account, it is quite possible that a more satisfactory explanation of the postwar fluctuations in consumer spending could be obtained.

I find numerous faults in other equations of Christ's model and hold it unfortunate that such unreliable results should have been publicly presented when even Christ admits they contain some striking contradictions to common sense. I can think of many research possibilities that could be investigated to help clear up these faults and feel Christ should have investigated some of them before presenting his model. The production equation is unsatisfactory because a reliable or positive estimate of the marginal productivity of capital has not been found. The link between the prewar and postwar data on the stock of capital is very suspicious. Christ has, as I pointed out above, used a wrong deflator for depreciation in recent years; there is the problem of accounting for the transfer of surplus war property to private hands after the war. I have long insisted that the relevant variable for the production function is the flow of capital services rather than the existing stock of capital but find no real attempt on Christ's part to measure directly the flow of capital services. His indirect measurements, relating use of capital to net investment, strike me as being inadequate. These problems are of some importance, but two other defects of
the production function seem more serious. In the first place, the man-hour concept of employment is far superior to the man-year concept used by Christ. He correctly notes the difficulty in preparing a series on man-hours going back to the early 1920's; however, I feel that some rough estimates should be made from the few sources available merely to see whether the bias in the employment data could have been responsible for the poor estimates of the marginal productivity of capital. My experience with American data shows that the trend influence (technological progress) in the production function has been very rapid, much more rapid than a simple linear function would allow. A quadratic trend would seem to fit an American model much better.

The production function finally selected (3.4) looks, as Christ has pointed out, suspiciously like his labor equation (4.2). An obvious alternative I would propose to avoid this difficulty is to express $N$ as a function of $\frac{pX}{w} \left( \frac{pX}{w} \right)^{-1}$ and $t$. This form is suggested by the theory of profit maximization subject to a Cobb-Douglas production function. Christ should consider this alternative. The other difficulties he encounters in his labor equation analysis can perhaps be explained by the inconsistent data he used on production and employment. I simply cannot believe the fantastic estimates he obtains for (4.0) since this was one of the most stable relationships of the interwar period, showing approximately the same structure for all methods of estimation in all models.

The limited-information estimates of the investment equation (1.0) seem equally implausible. Christ's equation is simply an extension of my own results, but I now lean towards a new formulation I recommended to Christ but which he did not try. I would express aggregate investment as a function of the current and lagged nonwage income originating in the sectors of the economy making the investments and the stock of capital. This is in accordance with the empirical findings in my paper (Part II). It is also possible that we may have written off too hastily the interest elasticity of investment as negligible. While I still do not think that investment for the whole economy is highly interest elastic, there is still some possibility of a small interest effect. I would favor an investment equation using nonwage income (profits before interest), the stock of capital, and bond yields as explanatory variables.

In essence, Christ's revision of the model has been to test the prewar equations against incorrect data of two postwar years, to recompute the parameters of the model with the two later observations, to rewrite the equations with a dubious production function as one of the structural

---

* On the other hand, I find his wage equation (5.0), brought in by the revision, to be quite useful and satisfactory.
equations, and to use real cash balances as a variable in the consumption
equation. As an alternative, I suggest a revision that I feel would be much
more rewarding. First, rework the entire set of data as suggested in the
first part of these comments. Secondly, try to use additional statistical
information such as that provided in quarterly and cross-section data.
Thirdly, revise the equations of the model as follows: The consumption,
investment, production, and labor equations should be treated as sug-
gested above. In case a satisfactory estimate of the production function is
not obtained, the best alternative may be to rewrite the system in such a
way that this equation does not appear explicitly but is imbedded in other
equations. Industrial sectors and other components of variables should
be treated in additional equations in a less aggregative model. New equa-
tions should be introduced to explain corporate savings and imports as
dependent variables. This plan of revision is not simple, but it is the
direction nonsuperficial work must follow. Revisions like these will prove,
I predict, to be much more valuable than any refinements of statistical
methodology. I find Christ’s empirical work disappointing in that it made
practically no attempt to introduce more basic revisions such as these.

After testing and revising my model on the basis of estimated data for
1946 and 1947, Christ extrapolated the revised model to 1948, a year
outside the sample observations. Although the mechanics of his extrapola-
ing procedure are straightforward, I find his technique at serious fault
from an econometric point of view.

An econometric model usually contains as many equations as there are
dependent variables thus enabling one to express the dependent varia-
tives in terms of the predetermined variables, once the structural parameters
are estimated. I find it very curious that Christ has gone to all the trouble
of structural estimation and then has not used the estimated model to
express dependent magnitudes in terms of predetermined variables for
purposes of extrapolation. Table 3 of Christ’s paper contains his calcula-
tions underlying the predictive ability of the model. For the reader’s
benefit some comments are called for on this table. On the basis of a model
involving 10 stochastic equations, Christ makes 13 predictions by one
method and 21 by another. The structure of the model has been seriously
violated, for it is not designed to yield more than 10 predictions. Christ
goes through an elaborate procedure in testing and constructing a model
of essentially 10 equations. He then throws away this information and
makes 13 or 21 forecasts, in the latter case often getting more than one
forecast for the same variable. The mechanics of this procedure are obvi-
ous, but the rationale is surely lacking. For example, equation 11.0 (C) in
Table 2 gives us an estimate of a structural relation showing how the
average interest rate depends on predetermined variables. This equation
extrapolates tolerably well to 1948, predicting the interest rate to be 2.99 per cent; the observed value is 3.08 per cent. Christ estimates, in Table 2, the parameters of a structural equation showing how interest rates are related to predetermined variables, then turns to other equations to predict the interest rate in Table 3. The structural equation gives a much better prediction than the equations used in Table 3. It so happens that by throwing away information, Christ has biased his test of the predictive ability of the econometric model, since the structural equation extrapolates better than either of the naive models; whereas his reduced form predictions in Table 3 are worse than the naive model predictions.

Because of the faulty character of the data used and because of the inadequacy of the production and the demand-for-labor equations, I find it impossible to accept Christ’s revised version of my model. However, I shall present an interesting experiment with the parts of the model that are more or less acceptable. Consider a system composed of CLS equations 1.0, 2.0, 6.5, 7.0, 10.0, 11.0, and definitions 12 and 13. In this model, equations related to the determination of rents in a free market setting are obviously suppressed because of rent controls, a fact Christ neglects for some unknown reason.

As stated previously, my former model was particularly weak in that actual consumption has been far above the consumption equation and there is no satisfactory scheme for the determination of absolute prices. I accept, for the moment, 6.5 (CLS) as the best possible version of the consumption equation until basic research in this area has progressed further, and take the price level as given until a suitable empirical scheme can be developed for incorporating this item into the model as an endogenous variable. Least-squares estimates are used because it is the only type that has been calculated for the particular consumption equation used.

Solving this model for $Y, C, I, D_1, D_2, i$ in terms of the other variables and substituting the observed values of the latter set in 1948 I find the following extrapolations:

$$ Y = \$68 \text{ billion} \quad I = \$1.71 \text{ billion} \quad D_2 = \$1.62 \text{ billion} \quad C = \$73 \text{ billion} \quad D_1 = \$1.90 \text{ billion} \quad i = 2.99 \text{ per cent} $$

This model contains only 6 stochastic equations; hence there are only 6 extrapolations. These estimates are obviously defective as compared with observations, yet proceeding along the lines of Christ’s paper we conclude that in 4 out of 6 cases they are better than those of either naive model.

One year’s test tells us practically nothing in a statistical sense about the merits of econometric techniques. This is as true of my example as of
Christ's paper. Absolutely no scientific conclusions can be drawn until many forecasts have been made under realistic forecasting conditions with efficient methods. My example certainly is not a demonstration of the usefulness of econometric model building; I merely offer it as a challenge to the acceptance of any substantive results in Christ's paper.

In addition to the fact that Christ uses the wrong equations for extrapolating the model beyond the sample points, the entire character of his prediction scheme is so mechanical that it loses much efficiency. It would be convenient if we had arrived at the final situation where forecasting could be reduced to purely mechanical operations, but we are only approaching such a situation, and there are many nonmechanical operations that any sensible forecaster would use together with the econometric model in its present form. For this reason, Christ's extrapolation cannot, in any sense, be considered as optimal, given his facilities. To be more specific, an econometric forecaster should be wary of structural change between the prewar and postwar period. Cross-section data from the Surveys of Consumer Finances and the surveys of investment intentions may throw substantial light on the postwar structure of the consumption and investment equations. Christ did not even consider this material. He could have used these data to check his estimates of consumption and investment or he could have, perhaps, used them to estimate the current structure of the consumption and investment equations to be used together with the other relationships of his model for extrapolation.

The assumption that there is no serial correlation of the disturbance terms of our econometric relationships may not be valid. Correction factors for the estimated values of the endogenous variables may be looked for in the trends in the most recent values of the estimated disturbances.

An alternative to taking into account the serial correlation in the disturbances would be to boost up some of the equations to make them conform with the postwar data as closely as with the prewar data. An objective way would be to introduce a dummy variable that takes on zero values in the prewar period and unit values in the postwar period.

It seems clear, in any case, that a competent forecaster would have used an econometric model on the eve of the prediction period far differently and more efficiently than Christ used his model.

In Appendix C Christ argues that the incorrectness of his data for the postwar years may have affected some of his structural equations but not the predictions made from the model. I find this argument weak in many respects. In the first place, the tests carried out for purposes of revising the model will be affected by a change in data if the estimated parameters are changed. Thus, at the earliest link of a chain of calculations there is weakness. This weak link spoils the entire chain. Secondly, the prediction equa-
tions used in the Appendix are the same as those used in Table 3. I reiterate that these are not the equations we want for purposes of forecasting. There is no easy solution from rough calculations like those of the Appendix. The only scientific way to approach the problem is to rework the entire set of basic data, obtain revisions of the model that stand up better than those offered by Christ (especially the consumption, production, and labor demand equations), and use the most efficient forecasting technique available. When all this has been properly done, we can come to the question of the predictive ability of econometric models.

REPLY BY MR. CHRIST

Lawrence Klein's comments have made it clear that my paper is not a finished piece of work containing well established results, and that its merits, such as they are, lie in the methodological field, illustrating the application of various methods of testing econometric models. I regret that I did not make this clear myself. I regard as the most valuable part of my paper the exposition and illustration of procedures for prediction and testing, rather than the particular model tested or the particular results arrived at, and that is the basis upon which I would like my work to be judged. Perhaps if I had made this clear, Klein would not have found it necessary to reiterate in his comments many of the criticisms and suggestions that already appear (some at his instance) in the paper and its appendices. To the extent that the paper does give the reader the impression that its results are reliable and should be accepted without further investigation, I think Klein is justified in many of his comments.

Klein's comments are divided into several parts, concerning (1) the data, (2) the form of the equations, (3) the number of variables predicted, (4) prediction from the linear reduced-form equations vs. prediction from the structural equations, (5) an experimental calculation of Klein's, (6) the question of mechanical vs. discretionary methods of prediction, and (7) my statement of the probable results of performing the desirable recomputations Klein suggests.

1) I have already essentially accepted most of Klein's comments on the data, as indicated in my Appendix C, with two exceptions worth noting: First, there is his statement that "... as to the source of [my] difficulties with the time series data, ... it seems plausible to conclude that the price deflators used to pass from current dollar to constant dollar magnitudes are largely responsible." I have cited in Appendix C sources that lead me to disagree with this, and to believe that my regression procedure for extending the undeflated series to 1947 is equally responsible. Second,
there is the statement that he "simply cannot believe the fantastic [limited information] estimates" I obtain for the Labor equation (4.0), and that these can perhaps be explained by inadequacies in the data. As I said in my paper, they appear fantastic to me as well, but I think it is clear, from the fact that my least-squares estimates of equation (4.0) are very reasonable and very close to Klein's least-squares and limited information estimates, that the data are not the controlling factor: if they were, my least-squares estimates would have been as absurd as my limited information estimates. I have commented on this matter in Section 12 of my paper.

2) Klein's comments on my choice of equations in the model are relevant to the question of how good the econometric technique is or can be for prediction purposes, because they suggest improvements that can be expected to lead to better results; they are not relevant to the process of testing the predictions of this particular model (though they may help to explain failures). I share most of Klein's criticisms on this point — in fact they are in my paper — except that I do not like to use a quadratic trend in the production function even though it fits the past data well: it is always possible to invent some function of time that fits a given set of data well or even perfectly, but where there are random elements such a function is not a reliable extrapolating device unless there is some substantive reason to believe that it will continue.

3) Klein states that my model "is not designed to yield more than 10 predictions" because it contains only 10 stochastic equations, and that I have made predictions of 13 variables. (Klein might be misunderstood when he says "Christ makes 13 predictions by one method and 21 by another"; the 21 predictions are actually predictions of the same 13 variables, with some duplication due to the possibility of using different sets of restrictions in estimating some of the parameters of the reduced form, as explained in Appendix E.) The three nonstochastic equations are the identities defining disposable income $Y$, private output $X$, and wage rate $w$ (there is a fourth, defining capital stock $K$, but I ignored it because $K$ appears as such nowhere else in the model). It is true that an arbitrary number of new variables could be added, each defined by a new identity, and that by a suitable choice of these new variables it would be possible to change the 'score', i.e., the number of variables predicted successfully by the model, from 6 out of 13 to, say, 16 out of 23, or even to 93 out of 100, without changing the original model in any way. This is what Klein means by saying that the number of variables that can legitimately be predicted by a model cannot exceed the number of stochastic equations in the model. However, there is a certain arbitrariness in deciding, regarding my model for example, which 3 variables should be eliminated by the 3 identities and which 10 should be predicted. Thus, if I had chosen to eliminate $H$, $X$, and $N$
respectively by the 3 identities, the score would have been raised from 6 successes out of 13 variables to 6 out of 10; on the other hand, if I had chosen to eliminate $D_3$, $p$, and either $w$ or $W_1$, the score would have been lowered to 3 out of 10 vis-a-vis naive model I and to 4 out of 10 vis-a-vis naive model II. The identities as they are written suggest that the most natural variables to eliminate are $Y$, $X$, and $w$; had this been done, the score would have become 5 successes out of 10. Since each variable in the model has a real economic meaning and represents an interesting economic magnitude, it seems that there is arbitrariness involved in eliminating 3 by identities, as well as in not eliminating 3 and thereby having more variables to predict than stochastic equations.

4) Klein argues in favor of using structural equations instead of reduced form equations for making predictions (specifically, this means first to substitute known or assumed values of predetermined variables, and estimated values of parameters, into the structural equations; then to solve the resulting system of equations simultaneously for the values of the jointly dependent variables). In support he cites the interest rate, for which limited information estimates of the interest equation (11.0) give a better 1948 prediction than least-squares estimates of the reduced form. Now the interest rate is in a unique position in my model (and in Klein's) because it is already expressed in terms of predetermined variables in equation (11.0), which means that it can be predicted from the structural equations directly without the algebraic operations of simultaneous solution. If one seeks instead to predict price level $p$ or disposable income $Y$ from the structural equations, one finds that, because of the nonlinearities in the model (including the identities), simultaneous solution leads to a quintic equation in $p$ or in $Y$, respectively. (Klein's model leads to a cubic in $p$ or in $Y$, as he does not have the nonlinear identity defining wage rate.) I have not been able to obtain anything except absurd values for $p$ and $Y$ from calculations of this sort, for either my model or Klein's model, but I had not intended to refer to this until I could determine the reason. As matters stand now, it appears on common-sense grounds, as Klein says, that prediction from structural equations will usually be superior to prediction from estimates of linearized reduced form equations, because of using more restrictions, but the few cases I have tried do not bear this out. Further investigation of the reason is called for.

5) Klein has presented an experimental calculation, using a modified model and my data, to see what kind of predictions the structural equations make for 1948 if the model is not required to predict prices, the true 1948 price level being used as if it had been known when the predictions were made. The results are better than mine, as measured by my criterion of the number of cases in which the naive models are bested by the econo-
metric model (though the structural-equation predictions of the two most important variables, Y and C, are quite far off even compared with predictions from my least-squares estimates of the reduced form). Klein's forecasts for 1946 (Econometrica, April 1947), to which he refers in his comments, are likewise based on a model in which the price level is supplied from outside (by an enlightened guess) rather than predicted by the model. At this stage of econometric research it is apparently more accurate to guess at the price level, then use the econometric technique to predict other variables on the basis of that price level than to predict all variables by econometric methods.

6) This brings up an interesting issue. Klein's criticism of my forecasting methods as inefficient and mechanized springs from the fact that he and I are pursuing different objectives. His objective is to make good predictions now with the resources at hand. If this were my objective, I would not rely only on mechanical procedures any more than Klein would; I would try to use judgment in taking account of certain information outside the model, such as recent trends in calculated disturbances, recent cross-section studies, and if predictions were improved by guessing at the price level, I would guess at the price level as well as I could. But my objective is different: it is to make some progress toward evaluating a kind of prediction procedure that is scientific, in that as far as possible it is reproducible and free from discretionary judgments (of course this does not preclude the incorporation of cross-section data into the model). Such evaluation is needed because policy makers cannot be expected to have confidence in scientific forecasting techniques until such techniques are developed to the point where results are reproducible and independent workers can come to the same conclusions.

7) Klein says in his last paragraph that I argue in Appendix C that the difficulties in my postwar data "may have affected some of [my] structural equations but not the predictions made from the model". My words are: "the estimates of the [structural] parameters would be changed, and the 1948 fit of some structural equations would probably be improved, but there is no evidence that the predictions of important variables by the reduced form would be improved" (italics supplied). I agree with Klein's other comments in his last paragraph, apart from the question whether to use the structural equations or the reduced form for predicting, discussed under (4) above, and apart from the question of forecasting technique, discussed under (5) and (6) above.
It may be useful to put together the important results of Christ's Tables 2 and 3, and thereby bring out one or two additional conclusions (see accompanying table).

Column 3 shows the errors in 1948 predictions from Klein's least-squares estimates of his structural equations, which were based on data for 1920-41. Since the predictions are obtained by using actual 1948 values for the jointly determined variables, they are not strictly predictions. Nevertheless, the errors are large. If instead one had assumed that these variables would not change from one year to the next, the average absolute errors that would have been made since 1920 are those given in column 2; the errors that would have been made in 1948 are given in column 6. Comparing columns 3 and 2 we see that in the case of 7 of the 8 'predicted' variables the error from Klein's model is larger, often substantially larger, than the average annual absolute change in the variable. Comparison of columns 3 and 6 yields a similar result. We can infer, therefore, that the information extracted by the model from the prewar data was of very little use in 1948.

The errors in column 4 (Christ's model) are very much smaller than those in column 3. These too are not really forecasts, for the same reason. But the equations yielding these estimates use additional data for 1946 and 1947, whereas Klein used data ending in 1941; also, some changes were made in the actual form of the equations. The result is an improvement, though as judged by column 2 the improvement does not go very far. In the case of five variables the errors from Christ's model (col. 4) are smaller than the average annual changes (col. 2), and in five variables they are larger. The comparison with column 6, where the standard is the actual change 1947-48, is similar. Nevertheless, I think it is clear that Christ's model performed better than Klein's since every one of the errors is reduced substantially. One cannot help but feel, however, that if the addition of two years of data and a revision of some equations makes such a difference the models themselves must rest on a weak foundation. Christ's computations in Appendix C add to that impression. Here a change in the data for several variables for 1946 alone alters the 1948 predictions substantially.

A comparison between columns 4 and 5 is of some interest. Column 5 shows the results of the actual predictions from Christ's model. That is to say, the several variables were predicted from equations utilizing only the so-called predetermined variables. All except two of these predetermined variables are lagged, so that they would presumably be known at the end of 1947. On the other hand, as noted above, column 4 utilizes knowledge
### SUMMARY OF RESULTS OF TESTS

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>ACTUAL VALUE, 1948</th>
<th>AVERAGE ANNUAL ABSOLUTE CHANGE IN ACTUAL VALUES, 1920-41, 1943-48</th>
<th>OBSERVED MINUS PREDICTED VALUES, 1948</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>1941, and 1948 values of related variables (3)</td>
</tr>
<tr>
<td>Investment, bil. 1934 $</td>
<td>1.89</td>
<td>1.02</td>
<td>−2.93</td>
</tr>
<tr>
<td>Inventories, bil. 1934 $</td>
<td>34.3</td>
<td>1.25</td>
<td>−2.52</td>
</tr>
<tr>
<td>Price level, index, 1934:1.0</td>
<td>2.03</td>
<td>.07</td>
<td>−8.90</td>
</tr>
<tr>
<td>Production, bil. 1934 $</td>
<td>90.3</td>
<td>5.03</td>
<td>+7.34</td>
</tr>
<tr>
<td>Employment, mil. man-years</td>
<td>52.1</td>
<td>1.81</td>
<td>+12.26</td>
</tr>
<tr>
<td>Wage bill, bil. current $</td>
<td>114.7</td>
<td>4.93</td>
<td>+7.1</td>
</tr>
<tr>
<td>Wage rate, thous. $ per man-year</td>
<td>2.20</td>
<td>.07</td>
<td>+0.17</td>
</tr>
<tr>
<td>Consumption, bil. 1934 $</td>
<td>82.8</td>
<td>2.86</td>
<td>+12.26</td>
</tr>
<tr>
<td>Income, bil. 1934 $</td>
<td>79.9</td>
<td>4.15</td>
<td>+9.0</td>
</tr>
<tr>
<td>Owned housing, bil. 1934 $</td>
<td>1.86</td>
<td>.23</td>
<td>−3.07</td>
</tr>
<tr>
<td>Rent, index, 1934:1.0</td>
<td>1.24</td>
<td>.05</td>
<td>−0.044</td>
</tr>
<tr>
<td>Rental housing, bil. 1934 $</td>
<td>1.44</td>
<td>.22</td>
<td>−0.59</td>
</tr>
<tr>
<td>Interest rate, per cent</td>
<td>3.08</td>
<td>.38</td>
<td>−0.48</td>
</tr>
</tbody>
</table>

Derivation of columns:

1. Col. 1 of Table 3.
2. Col. 3 of Table 3.
3. Col. 11 of Table 2, using Klein's least-squares estimates of structural equations (KLS).
4. Col. 12 of Table 2, using Christ's least-squares estimates of structural equations (CLS).
5. Col. 8 of Table 3, Christ's least-squares estimates of reduced form equations. The 1948 predictions are made from the 'predetermined variables', of which all except two are lagged 1947 values (see App. D).
6. Col. 10 of Table 3 (Naive model 1).
7. Col. 11 of Table 3 (Naive model II).
of the actual 1948 values of relevant variables. It would seem then that the results in column 4 should be better than those in column 5 since not only is current information being used in column 4 but it is presumably more directly relevant to the predicted variable. Nevertheless, in only 5 of the 10 variables for which a comparison can be made are the errors in column 4 smaller than in column 5; in 4 cases they are larger; one (rent) is ambiguous because of the difference in the number of decimals. It seems then that the variables that were thought, when the model was constructed, to be directly relevant to the ones to be predicted are really not much more relevant than the so-called predetermined variables.

Christ has indicated the results of comparing columns 5 and 6 or 5 and 7. Another way of stating these results is to say that one could get as much information about the 1948 values for the several variables from their own values in 1947 or 1946-47 as one could from knowledge of the 1947 and earlier values of other variables, as used in his model. Put in this way, the results may not seem surprising, though it is clear that if the model were theoretically correct one would expect that a knowledge of the 1947 and earlier values of the variables other than the one being forecast would improve the forecast.

JAN TINBERGEN, Netherlands School of Economics, Rotterdam

1) In Section 3 one should, I think, add as a point in favor of the least-squares method that it requires a theory only for the equation studied and hence prevents the use of erroneous theories for the other equations.

2) It might have been of some use in Section 7 to mention as a possibility Modigliani's approach to the problem of the consumption function.

3) In Section 7 Christ states that he has added no new endogenous variables to the system by his modifications of the consumption function. I think this is true only as far as the statistical testing of his equations is concerned but not for the system looked at from the viewpoint of theory. In other words, the type of movements executed by the system will, of course, be influenced by M as an endogenous variable.