**RECORD OF SOCIETY OF ACTUARIES
1979 VOL. 5 NO. 4**

RECENT ADVANCES IN PREDICTION THEORY

Moderator: ROBERT J. JOHANSEN. Panelists: WILLIAM E. WECKER, CRAIG F. ANSLEY. Discussant: IRWIN T. VANDERHOOF*

In Professor Wecker's paper, statistical methods are developed to predict "path properties" of a Time Series. Examples of path properties are turning points or minima.

Professor Ansley's paper shows how and to what extent Box-Jenkins Time Series methods can be modified to predict two or more time series simultaneously. The emphasis will be on workable techniques that are already being successfully used in practice. Several practical examples will be presented.

Following the presentation of the two papers by the participants from the American Statistical Association, the Discussant will present a discussion of the two papers from the actuary's viewpoint. Discussion from the floor will follow.

MR. ROBERT J. JOHANSEN: This is the first of three sessions arranged by the American Statistical Association (ASA), and I would like to tell you how the ASA sponsorship came about. As the Society's representative to ASA and to the Committee of Presidents of Statistical Societies, I attended several annual meetings of ASA. These meetings covered a number of topics (follow-up studies, demographic presentations, statistical techniques, papers on economic subjects) of interest or useful to the actuarial profession. Conversely, some of the ASA papers referred to actuarial and other techniques for life tables, follow-up studies, and exposure calculations. Because a joint sharing of results and techniques, would seem to be helpful to actuaries and statisticians, I have been talking with both the American Statistical Association and the Society's program committee to see what could be done. While the logistics of a joint meeting are discouraging, the possibility of each group providing sessions at the other's meetings, is feasible. The three sessions at this meeting are the first exchange and the Society has been invited to participate in the 1980 ASA Annual Meeting in Houston. What will happen in the future depends on how useful such interchanges are to the members of both organizations. If you have any comments, please write to the Society's Office or let me know.

Our first speaker, Dr. William E. Wecker, was responsible for the American Statistical Association effort in organizing the three sessions on Economics; he was at the time Program Chairman for the Business and Economic Statistics Section of the 1979 ASA Annual Meeting. Dr. Wecker has a Masters in Operations Research and a Ph.D. in Statistics and Management Science, both from the

*Dr. Wecker, not a member of the Society, is Associate Professor of Statistics and Management Science at the University of Chicago Graduate School of Business, Chicago, Ill.

University of Michigan. Now in academia, he had been a consultant to Ford Motor, Esso Research and the Xerox Education Group. He has authored a number of papers ranging from random walk and Markov processes to multiple time series.

Our second speaker is Dr. Craig Ansley, Assistant Professor of Statistics and Econometrics of the University of Chicago Graduate School of Business. While he is wearing a statistical hat at this session, Dr. Ansley is a Fellow of the Institute of Actuaries and an Associate of the Society and he has a Ph.D. in Business Administration from Michigan. He has been a consulting actuary in New Zealand and in London mainly on pensions and life insurance, and still does some consulting primarily in the casualty field. Currently Dr. Ansley's research interests cover statistical forecasting and he has published a number of papers.

The Discussant at this session is Irwin T. Vanderhoof. Irv is well known to most of you for his papers and discussions in the area of investment theory and analysis. An FSA, he is Senior Vice-President, Office of Corporate Development and Finance at The Equitable in New York. He also has an attachment to academia, serving as Adjunct Associate Professor with the College of Insurance. He is an Associate of both the Institute of Actuaries and the Casualty Society and is a member of the New York Society of Security Analysts.

DR. WILLIAM E. WECKER*: The title of my talk is "How to Predict the Timing and the Extent of the Next Recession", but the emphasis is really not on Gross National Product (GNP) but on statistical methodology. The procedures that I shall discuss will be applied to the real seasonally adjusted quarterly GNP to predict the occurrence and extent of the recession that began in 1974. But I hope that in our discussion today it may occur to you that there are other series that would be of interest to you to which these same techniques could apply. The main technical point is that the event that I am trying to predict (whether it is a turning point which marks the beginning of a recession, or the minimum of the series which indicates the depth of a recession) is a property of the path of the sequence.

The term "turning point" and certainly the word "recession" are often used by economists, but statisticans have not given them much attention. I reviewed several important reference works in the area of time series predictions without finding the concept of a turning point. None of these reference works does what I am about to do here, so it seems the technical aspects of this problem have been overlooked.

Now let me show you why predicting turning points is a problem by presenting some examples. If you make predictions of the future for the GNP series on the basis of data through 1968 using a Box-Jenkins type technology or from

*Additional details of Dr. Wecker's presentation may be found in his paper "Predicting the Turning Points of a Time Series" published in the University of Chicago Journal of Business, 1979, vol.52, no. 1.

an econometric model you will find these predictions do not have any turning points in them (see fig. 1). Thus one naively might assert that the prediction is stating that there will be no turning points, but that would be an incorrect interpretation. Consider the extreme case of a random walk. Most of us are aware that the way one predicts random walks is to take a point prediction equal to the last value (see fig. 2). Thus the prediction shows no turning points although the series has many. This is the difficulty: that path properties of predictions are generally quite different from the path properties of actual data. If you are interested in the prediction of some path property it is not correct to just make point estimates of the future and then examine the path properties of those estimates.

I will not tell you how to predict turning points yet. First I will show one more way to do it wrong. In the previous examples I made predictions successively further in the future from the same timepoint. Making successive one-quarter ahead predictions that you keep updating does not work either. As an example consider the GNP series from 1972 through 1976. If you compare the actual values and the sequentially updated one-quarter ahead prediction of GNP, you see that the turning points in the actual data occur before the turning points in the prediction series (see fig. 3). This illustrates what one sometimes hears that such and such model is very good for predicting GNP, but it has one problem, that it lags on the turning points. I can show analytically that this is a property of an optimal minimum mean square forecast. In other words this is not a fluke. It is simply a fact that this approach to prediction will not be good in forecasting turning points.

So how can we get a reasonable prediction of a turning point? Well we start by thinking a little more clearly about the problem. Let X(t) where t = 1,2,3, etc. denote the GNP series or whatever sequence it is you are trying to extrapolate. Also let Z(t) take on a value of either zero or one depending on whether or not there is a turning point at time t. With this bit of notation, we can distinguish clearly the usual prediction problem from the turning point prediction problem. The usual prediction problem is: given a sequence of X's, predict future values of that same sequence. The solution to this problem is to get a probability distribution of the future values conditioned on the past values. The turning point prediction problem is: given the same data predict the Z's where the Z's have to be defined in terms of the X's. The solution to the turning point problem will also be a predictive distribution but it will be a distribution of the Z's conditioned on the X's.

Earlier I said that there was not really anything in the statistical literature about turning point prediction, but occasionally one does run into a paper in the economics literature on the subject. Frequently, these papers predict future Z's based on past Z's. They look at how many turning points there were and look at the regularity in that turning point sequence and use that to extrapolate. This is inefficient; it loses information. You do not want to condition on the past Z's, you want to condition on the past X's. The other way to do it wrong is the way we have discussed, to get a probability distribution for the future X's and not the Z's.

It is important to understand that in general Z is an indicator that can be defined to pick out any path phenomenon of interest. If you are interested in a minimum instead of turning points, you just define Z to reflect that. In the example we are presently considering, Z represents turning points but we still must give a careful definition of turning point. In a way that has been the failing of the previous literature, the term "turning point" or "recession" is used without giving it a careful definition. Paul Samuelson has said, two successive declines in quarterly real GNP would be a good servicable definition of a recession. It is a handy and simple definition on which to base Z but any other unambiguous definition you prefer could be used. I will define Z(t) as having a value of one at the second decline of any two successive quarterly declines following a period of increase. For other t, Z(t) = 0.

In outline form then, here is the procedure to use to predict turning points. Given the GNP data, use Box-Jenkins times series method, or any other procedure, to get the solution to the usual prediction problem which is a predictive distribution of the future X's. Then transform the entire joint distribution of the sequence of X's into a predictive distribution of the Z's. For some definitions of Z, the transformation part is trivial. For example, for linear transformations, just take the same linear transformation of the point estimates. But for non-linear transformations like a maximum or a minimum or a turning point, the transformation can be rather difficult to do analytically, and so I used a fairly straightforward numerical procedure on the GNP series.

Looking now at the particular example that I worked out more closely, I pretend that I know GNP up to the first quarter in 1972 only, and then use a time series Box-Jenkins type extrapolation to get the joint distribution of the future values. Then I do the transformation and see if I can pick up the turning point in 1974. The display that I use to report these predictions is a distribution of the time until the next turning point, so that the vertical axis indicates quarters into the future (see fig. 4). By the definition of a turning point that I have given this is a legitimate distribution. There will be eventually a turning point in this series with probability one. It turns out that most of the probability is down around two years or three years in the future, which does agree with when the turning point actually occurred. There is almost no probability of a turning point in the first three quarters, which would be the rest of 1972. In contrast if I base my predictions on data through the second quarter of 1973, the probability of a turning point shifts to be mostly within the next six or eight quarters (see fig. 5). Finally if we go to one quarter before the turning point, when in fact, we would have already observed one decrease in GNP and need only one more decrease to declare a recession according to the definition we adopted, there is almost 100 percent probability that this will occur in the next quarter, as it actually did (see fig. 6).

Now consider another path property. Once you know the recession is on, you might ask how bad is it going to be? A more precise way of phrasing this question is to ask what the minimum of the future values of X will be during the next two years. You must limit your search for the minimum to two years or some finite period, because otherwise it is not a well posed problem and the answer will be minus infinity. As before I get a predictive distribution on the X's. Now my transformation is not to Z, but to a variable M which is the minimum of those future values of GNP over the next two years. In this example the probability distribution of the minimum has a mean of 1.0583 trillion (see fig. 7). The actual value turned out to be 1.0585, but that is a fluke. The answer is really the entire distribution and the fact that the actual came right on the mean is just good fortune.

DR. CRAIG F. ANSLEY*: The subject of my talk is the extension of Box-Jenkins prediction methods to the problem of predicting more than one time series at once. Box-Jenkins prediction procedures were invented during the 1960's. Box and Jenkins wrote a book in 1970 that is now very famous summarizing these methods. Most of the attention in that book was directed towards the problem of predicting a time series given only its past. However, that is a little naive if we really believe that perhaps one set of circumstances influences another. Recently people have spent a lot of time working on the extension of these methods to the problem of predicting two, perhaps three, or even more series simultaneously. This is already done to a certain extent by all of the commercial econometric models. They are in fact predicting several hundred series simultaneously. The difference is that although the theoretic structure in these models is quite sophisticated, the statistical methodology is really not sophisticated at all. We are really starting from the other end where we try to model the data extremely accurately, but so far the number of series that can be tackled at once is fairly small.

Before I discuss multiple time series, I will go quickly through the Box-Jenkins prediction method for univariate processes. A few technical terms are required. The term "stationary process" refers to a time series X(t) where t = 1,2,3, etc. whose expected values are the same for each t. Also the covariance (which is nothing more than a measure of linear relationship between two variables) of X(t) and X(t + s) depends only on the separation s between the two variables and not on their absolute position in the series. For instance this means the relationship between the first and second values in a stationary time series is the same as the relationship between the 100th and 101st values. Furthermore the condition that the covariances depend only on the separation implies, when the separation is zero, that the variances around the mean value are constant. In less technical terms, we can say that the general statistical behavior of a stationary time series does not change as time goes on. In other words, if you look at one piece of series, it will look the same as another piece of series later on, even though the individual values are different. Now immediately you realize that almost no time series fit this description. So the first step of Box-Jenkins procedure deals with how to transform a non-stationary time series, for instance one with a trend, into a time series which is stationary.

*Additional details of Dr. Ansley's presentation are contained in his paper "Predicting Multiple Time Series" available from him.

The next important part of the Box-Jenkins procedure is to find a good way of describing relationships along the series. One method is just to work directly with these covariances between different values. However since covariances depend on scale we convert everything to correlations which are independent of scale by dividing by the variances. The autocorrelation function is the sequence of correlations between X(t) and X(t-k) as k=1,2,3, etc. That is one way of looking at relationships in a time series. There is another way that is equally useful, called the partial autocorrelation function. It is the sequence of correlations between X(t) and X(t-k) after allowing for the explanatory effects of the k-1 intervening values X(t-1) through X(t-k+1). You can think of it as the incremental value of another piece of information. Now in fact these two functions contain exactly the same information; they are just two different ways of summarizing the internal statistical relationships.

The reason I have introduced these two functions is that they are used in a Box-Jenkins procedure to decide the type of model that will suitably fit the stationary series. One of the types of models is something called a moving average model of order q, denoted MA(q). It is basically just a linear combination of independent random variables a(t), a(t-1), through a(t-q) where the coefficient of a(t) is one. A moving average model of order q has the first q autocorrelations non-zero and the others are zero. On the other hand the partial autocorrelations approach zero only gradually. Thus, if the stationary series X(t) that we wish to predict has autocorrelation and partial autocorrelation functions that fit these patterns, an MA(q) model is suggested. Another type of model used by Box and Jenkins is an autoregressive model of order p, denoted AR(p). This is like a regression equation in which the time series is regressed on itself. In particular an AR(p) model for X(t) would be a linear combination of X(t-1) through X(t-p) plus an additional random variable a(t). Such a process has autocorrelations that die out gradually, while only the first p partial autocorrelations are non-zero.

Now we can take a quick look overall at Box-Jenkins procedure. Most series are non-stationary and so in almost all cases we will have to do something to reduce the series to at least approximate stationarity. One method that usually works fairly well, especially on economic data, is to take logarithms of the series values and then differences. Having reduced the data to stationarity, the next step in the Box-Jenkins procedure is to choose the type of model and its order that will be used to fit the series, whether it will be a moving average process, an autoregressive process, or a mixture. The autocorrelation and partial autocorrelation functions are used to make this identification. Once a tentative model choice has been made it remains to estimate the coefficients. However, I do not want to talk about estimation because it would take forever to go into all the details. There are some procedures in Box and Jenkins' book and many people have written about them since. Finally, because it is possible that we might have picked the wrong model, we calculate the a's and estimate their autocorrelations. Since these residual a's are assumed to be independent, their autocorrelations should be zero. Thus if there are large non-zero values in the residual autocorrelation function, the model is inadequate and we must return to the identification stage and choose a different model.

To illustrate the Box-Jenkins method for univariate series, I have two examples to show you. One series is the seasonally adjusted average hourly wage rate from January 1972 through March 1979 (see fig. 8) published in the Survey of Current Business each month. The other series is the Consumer Price Index (CPI) over the same period (see fig. 9) and from the same source. You can see both series increase with a roughly exponential appearance, as you would expect. The CPI series has a little more of a wavy pattern, but it still can be described reasonably well as exponential behavior. Since both series have trends in them, they are not stationary. If I take logarithms of this data the hourly wage rate series looks almost like a straight line and the CPI series at least is spread evenly about a straight line trend. By taking differences the two series become approximately stationary, i.e. the statistical behavior seems to be roughly the same over any part of the series. (see figs. 10 and 11).

To save some time, for the next step of the Box-Jenkins procedure, which is to fit a model to the stationary data series that would create it, I shall consider only the CPI series. From here on I restrict myself to the period January 1972 through March 1978 and save the last twelve months to test my predictions. The autocorrelation function and the partial autocorrelation function are theoretical numbers so we can never tell exactly what they are and we have to rely on statistical estimation. For the autocorrelation function that amounts to nothing more than just calculating the numerical correlation for the data points separated by one time period, then the correlation between data points separated by two time periods, and so on. The result (see fig. 12) is quite a strong relationship from one time period to the next, especially for the first few months apart. Knowing how the Consumer Price Index is put together, this is to be expected. Many items are sampled only once every three months. Thus some numbers are kept in the index for three months and therefore there should be fairly strong relationships for up to three months in this series. I also estimated the partial autocorrelation function and find that only the first two values exceed the limits that you would expect them to be within 95% of the time if the values were really zero (see fig. 12). Thus an AR(2) model seems to be a reasonable first approximation for this series because its first two partial autocorrelations are non-zero, its remaining partial autocorrelations are zero and its autocorrelations die away slowly.

As I indicated earlier, some of the estimation procedures can be very complicated, but the one I used in this case simply arises as a by-product of calculating the autocorrelation function. The .0058 in the resulting model (see fig. 14) is the mean value of the differenced log series and corresponds to the average annual rate of increase in the CPI of 7.2% over the period January 1972 through March 1978. Having fit an AR(2) model to the data, I need to test it to see if it is adequate. To do this I solve the model formula for the residual $a(t)$ and estimate its autocorrelation function (see fig. 13). I find none of the correlations exceed the approximate 95% limit, which is consistent with the a's being serially independent as required by theory. Thus I am reasonably satisfied that this model will describe my data fairly accurately.

However the acid test is whether the model forecasts well. Starting at March 1978 I made 12 forecasts for the CPI (see fig. 15). The first was one month ahead, the second two months ahead and so on. Then I assumed I was not in March 1978 but April 1978 and I made another set of forecasts, which appear on the second diagonal. This time, because I only actually had data through March 1979 I had only 11 values to test my forecasts. Then I started in May 1978 and so on. Looking at the mean square errors of the predictions is one way to summarize. You can see that the size of the error tended to increase, naturally, the further ahead I tried to predict.

Finally we get to the subject of my talk, multiple time series. Before we were just looking at the simple series X(t). Now I want to look at m of them at once, which I can consider as a column vector whose entries are time series. The stationarity conditions which we had before can be generalized and each of the m individual time series has to be detrended as the first step in the Box-Jenkins procedure.

The covariance terms are much more complicated now, because we are considering not just the covariance between one time period and another, but also the relationship of the covariance between one series and another series. In general we look at the covariance between say the i-th series at time t and the j-th at time t+s. Instead of simply having covariances depend on time we have some going across series as well, so we have interseries relationships as well as intertemporal relationships. As before we can substitute correlations for covariances. Correlations really have the same information, but since they are scale free, it makes it a little bit easier for us to get familiar with the numbers. The autocorrelation function now becomes a sequence of matrices. For instance in the s-th matrix the row 1, column 1, entry is the correlation between the first series at time t, and the first series at time t+s. The row 1, column 2 entry is the correlation between the first series at time t and the second series at time t+s. So you can see that there are altogether m^2 possible combinations of relationships with the separation of lag s. We can also define a partial autocorrelation matrix, but I will not try to do anything more now because the idea of partial correlations for multiple time series is really quite complicated. For one thing it is not unique; there are many different definitions.

The models for multiple time series look much the same as the models we had in the univariate case. The difference is that each of the X's and the a's is a vector and the coefficients are no longer single numbers but matrices of numbers. That complicates the estimation procedure quite a bit, because for instance with only two series, we will now need four coefficients for a first order model instead of one coefficient. The best way of illustrating this perhaps is just to look at an example. The example that I shall use is the same as I used before: wages at time t and Consumer Price Index at time t. Previously I forecast each one of these independently without taking any notice of what was happening to the second series. Now I shall predict them as if they were a 2x1 vector and look at the joint properties.

The 2x2 autocorrelation matrices can be computed fairly easily (see fig. 16). The top left hand values are the same correlations between the wage series at time t and the wage series at time t+s that we looked at in the univariate case. In the same way the bottom right hand corners are the correlations between the Consumer Price Index at time t and the Consumer Price Index at time t+s that we have already seen. The off diagonal entries, in the bottom left and the top right, describe the interrelationships between the two series, and this is what is new. As before, the correlation matrices will drop to zero for a moving average process but not for an autoregressive process. Since quite a few correlation matrix entries exceed 95% confidence limits, it does not look like a moving average model. Next I computed the partial autocorrelation matrices, from which chi-square statistics can be deduced. I will not have time to tell you how I do that. The chi-square statistics should be close to zero if the underlying partial correlation is zero and will be big otherwise. We find that the chi-square values exceed the 95% critial value at lag 1 and lag 2 only (see fig. 17). Because we have a model for which the autocorrelations die away slowly and only the first two partial autocorrelations appear to be non-zero it looks like an AR(2) model will fit the data reasonably well.

The two matrices of estimated coefficients for my model can be computed as by-products of the partial autocorrelation function. The non-zero off diagonal value in the upper right hand corner of the first 2x2 coefficient matrix (see fig. 18) tells us that the hourly wage rate is predicted partly not only by past hourly wages but also is predicted partly by past values of the Consumer Price Index. This is about what you would expect, because the wage rates react to past movements in the Consumer Price Index. On the other hand, the CPI is being influenced not only by its own past but also, through the non-zero value in the lower left hand corner, by the past of the wage series. This model then is telling us exactly how the wage price spiral works. Note that it only happens at lag one; at lag two the off diagonal values in the coefficient matrix were not originally zero but they were very close to zero so I just left them out.

Next I looked at the 2x1 column vector of a's. The values of these residual vectors should be serially uncorrelated, so I worked out their autocorrelation matrices in the same way I did for the original series (see fig. 19). Since none of the values in these matrices is very big I find that my first try at the model is satisfactory, that there is no usable information left in the residual series.

Finally I did some forecasting with this model using the same successive origin dates as in the univariate forecasts (see fig. 20). By comparing the root mean square errors of these predictions with the average errors resulting earlier from my predictions using univariate Box-Jenkins technology, it is evident that the bivariate model is significantly better than the two separate univariate models. This improvement is greater for the wage series than for the Consumer Price Index, because the amount of feedback in the joint model is greater for hourly wages.

MR. IRWIN T. VANDERHOOF: My discussion of these papers will be in three pieces. First, some general comments on what I think of the papers. Second, some general comments on what I think of stochastic predictive techniques in general. And third the way in which I believe the contents of these papers can be used in the one area that is of real interest to me, viz. how to make money.

My general comments on the papers are that they are useful and very interesting. Professor Wecker's paper is sort of off on one side of the Box-Jenkins procedure, in the sense that his contribution should not be restricted to time series techniques. There are other predictive techniques for which this particular method of forming a distribution of the expected results and a transformation of those distributions could apply. Professor Ansley's contribution is sort of on the other side of the Box-Jenkins techniques because he is essentially allowing for a deterministic element to enter into the pure time series method in that two time series are considered that react on each other. Thus it is possible to take the point of view that one time series is a causal factor in the development of the other.

Now I would like to talk about time series techniques and more generally about stochastic forecasting as a methodological approach. I condemn such an approach. I have strong reservations about the usefulness of any such approach and the desirability of embarking on this kind of a method of solving problems. Professor Wecker made reference to Samuelson, a Nobel Prize winner, saying that a decline in GNP in two successive quarters is a good way to decide when you had a recession. I will cite another winner of some prizes, Einstein, who said, "God does not play dice with the universe." This is the basic argument against time series forecasting techniques: that they make certain assumptions about the underlying nature of the processes we are studying without ever trying to find any kind of rational explanation for them. Professor Ansley talked about the basic underlying philosophy of Box-Jenkins. It says that the particular elements of the universe we are studying are subject to shock or innovation. We do not know what those innovations are, but we assume based on past experience that the mean is zero and the variance is constant. However, that is making some fairly big assumptions about the world, particularly when you start by saying you do not understand and you are not trying to understand what these innovations are or how they affect the world.

I think these techniques have the twin disadvantages of being very seductive and quite effective. They are seductive in that the mathematical techniques that have evolved are interesting, exciting and fun to try to analyze and understand. They are effective, perhaps more so than deterministic methods. Since they are effective they lead people away from attempting deterministic model building solutions for the problem of prediction. This is a disadvantage because without a deterministic model there is no theory by which to attempt to influence the course of events. For instance, if we have a wonderful predictive method for GNP, for CPI and for hourly wage rates that is based upon pure time series analysis we are discouraged from trying to build models, perhaps involving the money supply or dependency ratios, that would tell us how to affect the GNP and other facets of the economy, in ways that might allow us to alter the course of the economy for the better. As another example, consider that the Mayas in their temples, with high class time series techniques might have been able to predict the movement of the stars and eclipses with great accuracy; and yet that would never have led to the law of gravity which is, in fact, more valuable than just correct prediction based purely on observation and experience.

Now I would like to discuss the way in which I think these techniques can be used to make money. Professor Wecker's contribution might be applicable to some purely insurance functions, perhaps loss ratios on health insurance. Sometimes we see loss ratios on loss of time products rising slowly and declining slowly. It would be very interesting to have some method of determining where those turning points would fall, because if we think the loss ratio on loss of time insurance simply will go up forever we better get out of the business. If we have some idea where it is going to peak or reach a trough, we have some idea about when we should try to emphasize or de-emphasize the product. Professor Ansley's results might relate to questions about productivity of agents. To some extent the number of agents we employ determines our sales of insurance. On the other hand, sales of insurance are closely related to disposable income in the economy and if the agents sell more they can stay in business. We therefore have cross effects. So in these two specific applications, one for Professor Wecker and one for Professor Ansley, there seem to be possibilities of applying these techniques and making money.

DR. WECKER: I want to make a brief response to the discussant. I think that a bad impression may have been left here. A distinction has been made between a law such as a gravitational law, a sort of knowledge about the universe, and the sort of information that is contained in these statistical models. I do not see that distinction. For instance take Craig's wage rate series, transformed by logging and differencing. There is a statistical regularity to that series, the left side looks much like the right side and there is evidence of a constant mean. That is real; that is not some statistical fabrication. The data are real and the properties are real. We are representing what we actually see in the world with these models. In that sense I would not distinguish between laws of gravitation and facts about wage rates.

DR. ANSLEY: I want to add my response too. First, all of the examples as to why it is nice to have deterministic rather than stochastic models were addressed to physics. This is kind of unfortunate, especially when we talk about the law of gravity, because after all the latest theory on gravity is entirely probabilistic. So in fact if we have a law of gravity it is a statistical law, not a deterministic law. The same applies to most modern particle physics and quantum mechanics. Secondly it is a pity for an actuary to condemn stochastic models, seeing that your profession and indeed mine is based entirely on a large stochastic model, namely that of life contingencies. It makes no sense at all to reject stochastic models.

MR. VANDERHOOF: Let me respond first to the second of the two comments made. The examples I gave were not only from physics. I also talked about the policy value of having some idea of the relationship between actions by the federal government through the Federal Reserve on the economy and what actually happens in the economy. It is not very good to be able to predict the CPI, if you believe, as many people do, that a determinant factor of the CPI is the money supplies controlled by the Federal Reserve. As for the regularity of the wage rate series being a real fact of the world, how is it a real fact of the world? Only after you have taken the logarithms of the data and after you have taken the differences. I can find a way to transform any series of data that will eventually produce a regularity if I massage it enough. What Professor Ansley has done is perfectly fine, but suppose that something happens, for instance oil goes to $50 a barrel. The trend would change and what is now a regularity ceases to be a regularity. Without knowing the effect of oil on the economy, there is no guarantee that the trend and this particular way of massaging the data will continue to work.

MR. JAMES A. TILLEY: I want to comment on Mr. Vanderhoof's remarks. Having spent many years of my life working in physics I can tell you the problems there are difficult, but they are at least pure. The particular example he mentioned is one where Einstein himself did not have much success. I subscribe to Mr. Vanderhoof's philosophy of trying to find out as much as one can. Any technique that circumvents that is not good. Nevertheless, I share the concerns that were expressed by both Professors in answering his objection and I agree with them. Predicting various things in the economy is tough and if you can learn something by looking at a time series and the data contained in it by itself, or in the way it correlates with other time series, then it is a useful exercise. That does not mean that one should not aspire to higher things. That their techniques do not explain how the economy works, is not a valid argument against developing and using such techniques.

I also have a question for each of the Professors. Professor Ansley, how do your techniques differ from what is commonly known as Box-Jenkins bivariate transfer function techniques?

DR. ANSLEY: The difference is basically that transfer function models regard only one of the series as being stochastic, the other is regarded as being non-stochastic.

MR. TILLEY: Professor Wecker, does your technique for predicting turning points take into account the fact that if you have used a Box-Jenkins model, the model parameters may not have been precisely determined because you only had a part of the series to look at to estimate them?

DR. WECKER: It is customary in applied time series work to ignore the fact that there is uncertainty about the parameters. The usual justification is that the order of uncertainty is not very great relative to other uncertainties, but the facts are that it is hard from a technical point of view to take this uncertainty into account. So the first answer to your question is that I am assuming that the parameters are correct even though that is not really true. The second answer is that everybody else does it too, and the third answer is that the procedure that I discussed today would apply without change if you did account for the uncertainty in the parameters.

MR. HARRISON GIVENS: Professor Wecker, how good is your procedure at predicting when the stock market will turn one way or another?

DR. WECKER: My procedure would be optimal! But given the random walk character of the stock market, there would be no money to be made by this optimal prediction. If you accept this martingale description of stock price sequence, then no one else will have a procedure that will make money either.

MR. GIVENS: Professor Ansley, if you had taken the same series at a different time, how stationary are the coefficients?

DR. ANSLEY: Well you never can be certain with those two series, but there is no evidence that there is any non-stationarity.

MR. JAMES G. BRIDGEMAN: I could not help comparing the two talks. Could one improve the prediction of turning points by doing a two series analysis?

DR. WECKER: There was nothing special about using a single time series method to make the turning point prediction. A multiple time series approach might be used to advantage, but I stuck with univariate time series because it is sort of an easy canned everyday procedure, whereas even just two time series is rather exotic by modern day applications.

MR. STEVEN M. MARTIN: My question is for Professor Wecker. You demonstrated that your technique worked very well on the example for predicting that 1974 downturn. Have you applied the technique to prior downturns and were you as successful?

DR. WECKER: The only example that I computed was the one that I showed here today. That is because I am in the business of producing the technique and not really concerned with GNP. I am completely certain that this is a statistically sound procedure and it will work correctly in any period.

MR. STEPHEN L. WHITE: Professor Wecker, your method involves a transformation of the measured X's to the new variables Z. The only specific transformation you showed, that for predicting when the series would turn, let Z assume only the values of 0 or 1. Is this restriction necessary, or do your methods also work for continuous transformations? It would appear you must have used a continuous function for your later prediction of the minimum value of the GNP.

DR. WECKER: The method can use any transformation. Letting the Z's be either 0 or 1 is a standard choice for a model where there are exactly two alternatives.

MR. WHITE: Early in your remarks, you said it is incorrect to project the X's beyond the present and then look at path properties of the projected series. However, later you did make some projection of the X's to find the Z's. Is there a contradiction?

DR. WECKER: There is no contradiction. My procedure is to compute the entire predictive distribution of future X's given past X's, and then transform that distribution to obtain the distribution of future Z's (given past X's). At no time is it correct to transform point predictions of the X's. Just recall the elementary result that in general $E[f(X)] \neq f(E[X])$ to see why.

MR. BRIDGEMAN: Professor Wecker, one of the kinds of yes/no questions that some actuaries are interested in is the ruin question for a risk enterprise. Is there anything in what you have done that might be applied to get better answers than we have right now for this problem?

DR. WECKER: Well I do not know what answers you have right now. However, the probability that some particular sequence will fall below some particular value can be computed in the way that I showed. The main point here is clarity of thinking. When you want to produce a prediction, do it with care. Define the random variable you are interested in and do not just hastily do something ad hoc that seems right, because you may not get the right answer.

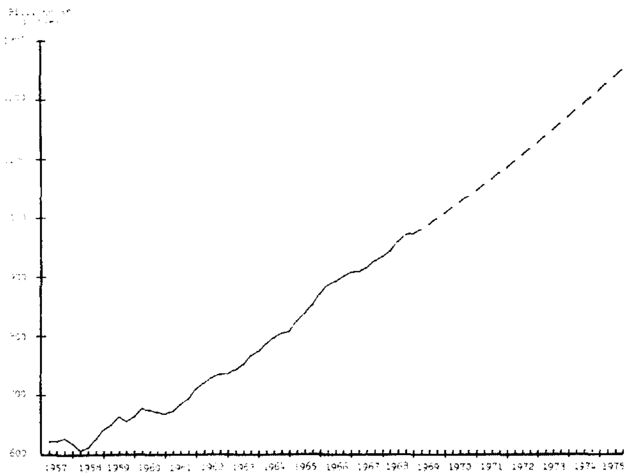


Figure 1. - Quarterly, seasonally adjusted real GNP with predictions.

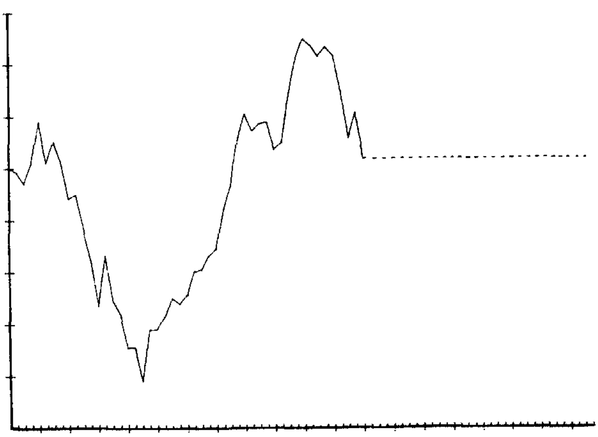


Figure 2. - An artificially generated random walk with optimal (mean squared error) predictions.

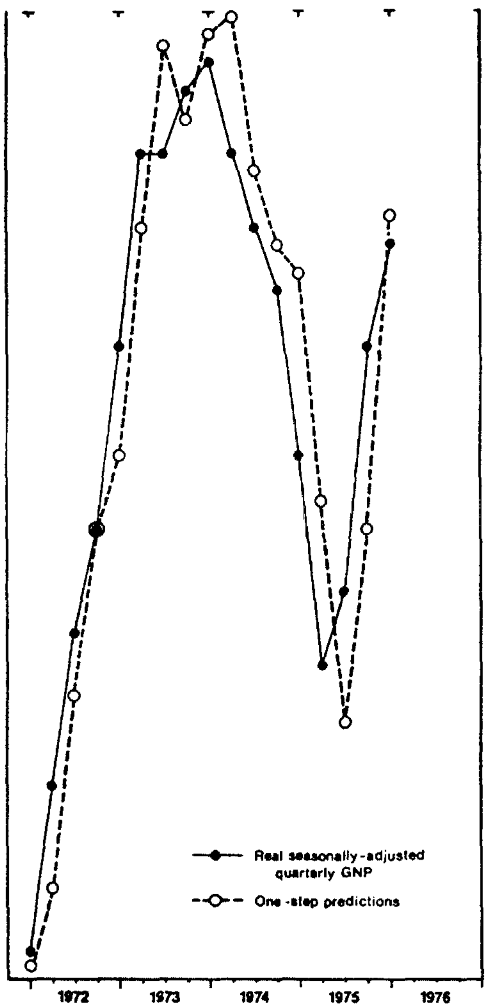


Figure 3. —Real, seasonally adjusted quarterly GNP and one-step predictions.

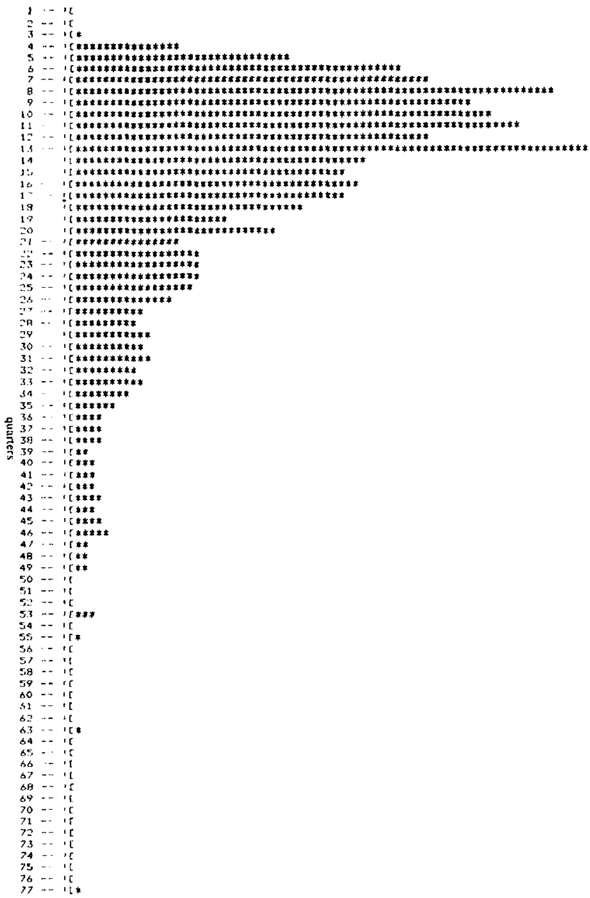


Figure 4.—Distribution of time until the next turning point—prediction origin 1972:1.

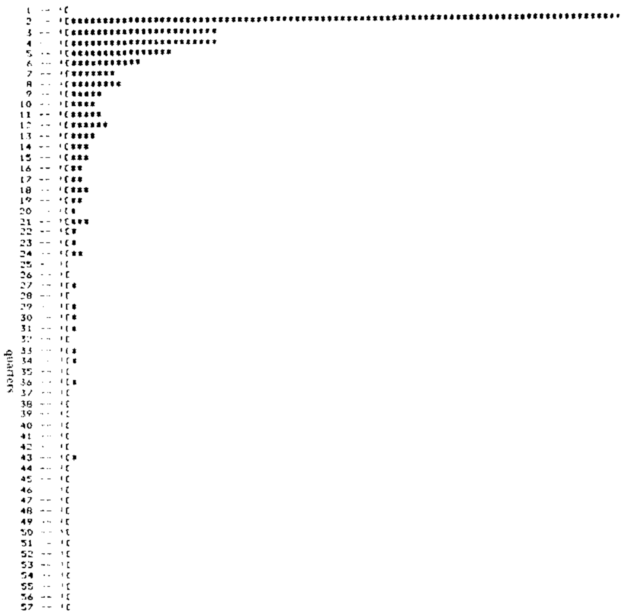


Figure 5. - Distribution of time until the next turning point-prediction origin 1973:2.

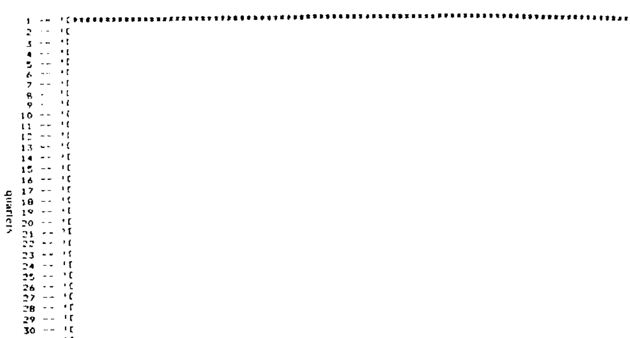


Figure 6. - Distribution of time until the next turning point-prediction origin 1974:1.

```
          930.09   --  ![
          933.23   --  ![
          936.36   --  ![
          939.49   --  ![*
          942.62   --  ![
          945.75   --  ![*
          948.88   --  ![
          952.01   --  ![
          955.14   --  ![
          958.27   --  ![
          961.40   --  ![**
          964.53   --  ![**
          967.66   --  ![*
          970.79   --  ![*
          973.92   --  ![**
          977.05   --  ![**
          980.18   --  ![***
          983.31   --  ![*****
          986.44   --  ![**
          989.57   --  ![*******
          992.70   --  ![****
          995.83   --  ![*****
          993.96   --  ![*****
         1002.09   --  ![*******
         1005.22   --  ![****
         1008.35   --  ![*****
         1011.48   --  ![*****
         1014.61   --  ![********
 R       1017.74   --  ![********
 e       1020.87   --  ![**************
 a       1024.00   --  ![*******
 l       1027.13   --  ![*********
         1030.26   --  ![*********
 G       1033.39   --  ![********
 N       1036.52   --  ![******
 P       1039.65   --  ![***************
         1042.78   --  ![*************
         1045.91   --  ![********
         1049.04   --  ![***********
         1052.17   --  ![***************
         1055.30   --  ![**************
         1058.43   --  ![***************
         1061.56   --  ![***************
         1064.69   --  ![**********
         1067.82   --  ![***************
         1070.95   --  ![****************
         1074.08   --  ![******************
         1077.21   --  ![****************
         1080.34   --  ![******************
         1083.47   --  ![**********
         1086.60   --  ![*******************
         1089.73   --  ![*********************
         1092.86   --  ![****************
         1095.99   --  ![********************
         1099.12   --  ![**************
         1102.25   --  ![****************
         1105.38   --  ![***************
         1108.51   --  ![*************
         1111.64   --  ![******
         1114.77   --  ![**
         1117.90   --  ![
```

Figure 7. —The distribution of the minimum value of GNP (in billions of dollars) over the next eight quarters—prediction origin 1974:2.

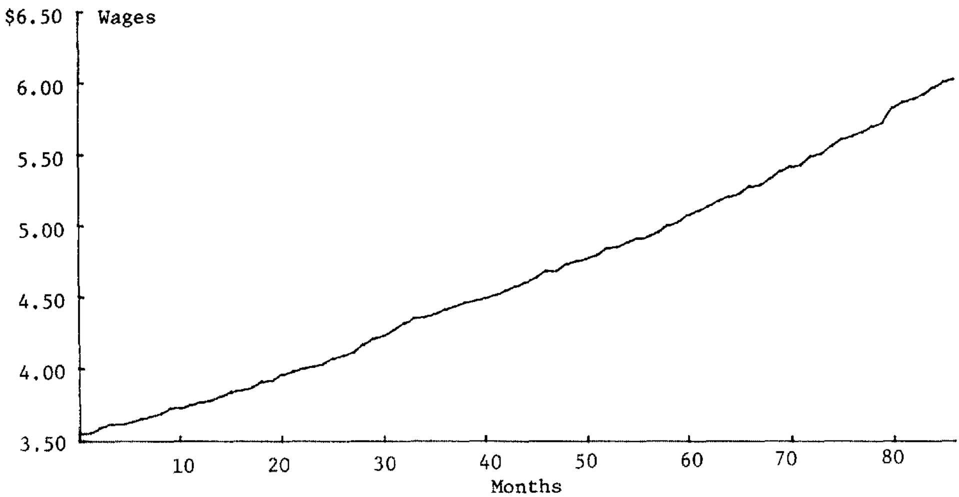


Figure 8. Average hourly wages, non-farm, non-supervisory workers January 1972 - March 1979 (seasonally adjusted) Source: Survey of Current Business

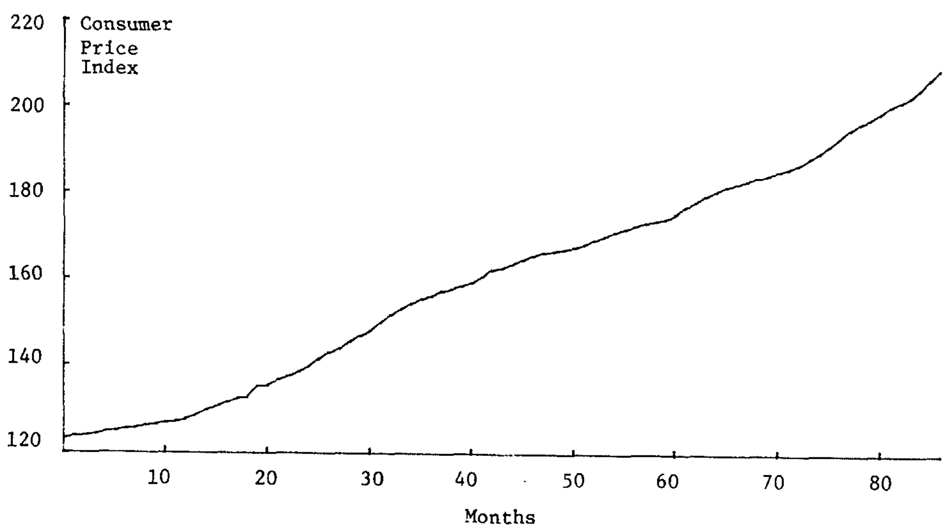


Figure 9. Consumer Price Index January 1972 - March 1979 (1970 = 100) Source: Survey of Current Business

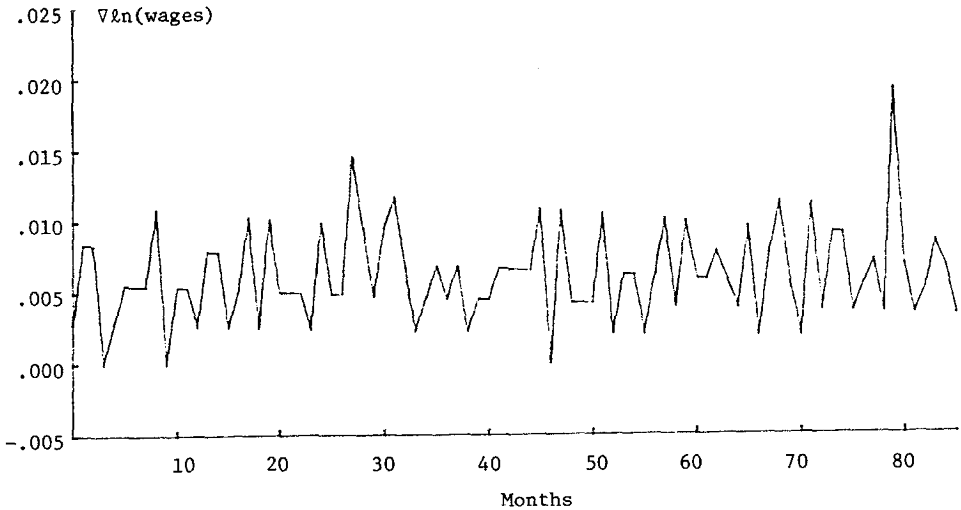


Figure 10. Differences of logarithms of hourly wages February 1972 - March 1979

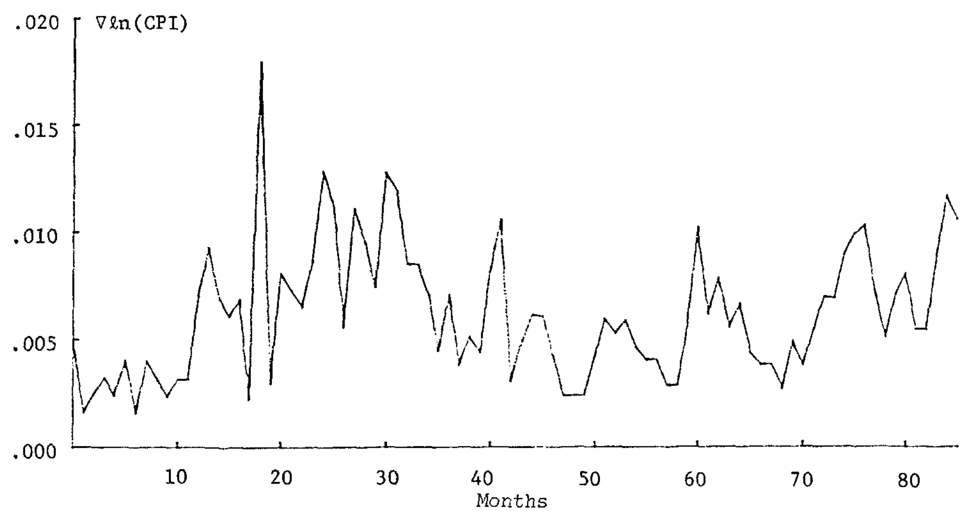
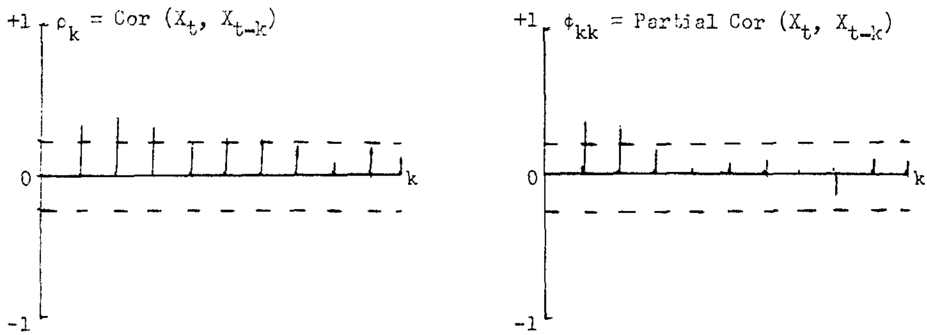
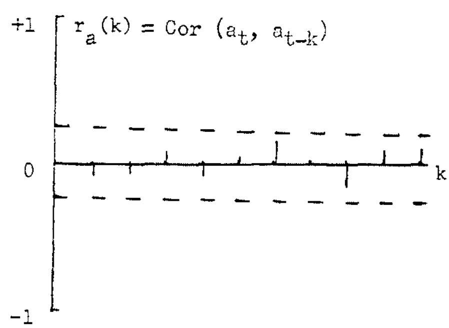


Figure 11. Differences of logarithms of Consumer Price Index February 1972 - March 1979.



– – – – Approximate 95% limits

Figure 12. Consumer Price Index autocorrelations and partial autocorrelations estimated for period January 1972 through March 1978. (Log difference series.)



– – – – Approximate 95% limits

Figure 13. Consumer Price Index model residual autocorrelations.

$$C_t = \text{Consumer Price Index time } t$$

$$X_{2t} = \nabla \ell n C_t = \ell n C_t / C_{t-1}$$

$$(X_{2t} - .0058) = .23(X_{2,t-1} - .0058) + .33(X_{2,t-2} - .0058) + a_{2t}$$

$$\text{Var}(a_{2t}) = .82.10^{-5}$$

Figure 14. - Consumer Price Index AR(2) model.

Date	Actual	Steps Ahead 1	2	3	4	5	6	7	8	9	10	11	12
4-78	191.4	-0.5											
5-78	193.3	-0.6	-1.2										
6-78	195.5	-0.5	-1.2	-2.0									
7-78	196.7	0.2	-0.4	-1.4	-2.2								
8-78	197.7	0.5	0.8	-0.1	-1.1	-2.1							
9-78	199.1	-0.2	0.4	0.8	-0.1	-1.3	-2.3						
10-78	200.7	-0.4	-0.7	0.1	0.5	-0.5	-1.7	-2.7					
11-78	201.8	0.3	-0.3	-0.6	0.3	0.7	-0.3	-1.6	-2.7				
12-78	202.9	0.2	0.5	-0.2	-0.5	0.4	0.9	-0.2	-1.5	-2.6			
1-79	204.7	-0.7	-0.4	0.1	-0.8	-1.2	-0.1	0.3	-0.8	-2.1	-3.2		
2-79	207.1	-1.1	-1.9	-1.6	-1.1	-2.0	-2.4	-1.3	-0.8	-2.0	-3.3	-4.5	
3-79	209.3	-0.5	-1.9	-2.9	-2.6	-2.1	-3.0	-3 4	-2.3	-1.8	-3.0	-4.3	-5.5
Root Mean Square		0.53	1.05	1.34	1.30	1.45	1.85	1.97	1.79	2.15	3.17	4.40	5.50

Figure 15. - Forecast errors for univariate AR(2) model for Consumer Price Index.

Lag	Autocorrelation Matrix		Indicator Matrix	
0	1.00	.11		.
	.11	1.00	.	
1	-.27	.23	-	.
	.22	.35	.	+
2	-.06	.02	.	.
	.02	.41	.	+
3	.24	.16	+	.
	.10	.33	.	+
4	-.05	.15	.	.
	.13	.23	.	.
5	-.16	-.08	.	.
	.08	.25	.	+
6	.25	.16	+	.
	.32	.25	+	+
7	-.22	.07	.	.
	-.05	.18	.	.
8	.01	-.08	.	.
	-.01	.07	.	.
9	-.01	.08	.	.
	-.15	.19	.	.
10	-.08	-.08	.	.
	.09	.14	.	.

Approximate 95% limits $\pm$.23 for purely random model.

. matrix element within limits

+ matrix element above limits

- matrix element below limits

Figure 16. - Bivariate autocorrelation matrices.

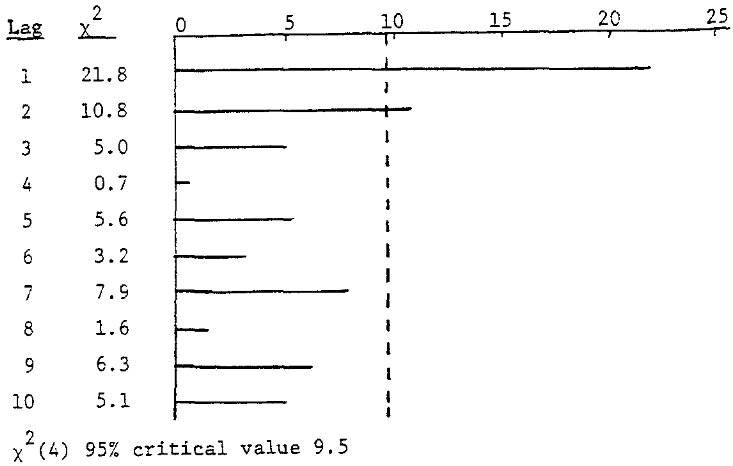


Figure 17. - Bivariate partial autocorrelations: chi-square statistics.

X_{1t} = $\nabla \ell n$(hourly wages)

X_{2t} = $\nabla \ell n$(consumers price index)

$$\begin{bmatrix} X_{1t}-.0060 \\ X_{2t}-.0058 \end{bmatrix} = \begin{bmatrix} -.38 & .32 \\ .11 & .24 \end{bmatrix} \begin{bmatrix} X_{1,t-1}-.0060 \\ X_{2,t-1}-.0058 \end{bmatrix} + \begin{bmatrix} -.19 & 0 \\ 0 & .25 \end{bmatrix} \begin{bmatrix} X_{1,t-2}-.0060 \\ X_{2,t-2}-.0058 \end{bmatrix} + \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix}$$

$$\mathrm{Cov} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix} = \begin{bmatrix} .82.10^{-5} & .79.10^{-6} \\ .79.10^{-6} & .79.10^{-5} \end{bmatrix}$$

Figure 18. - Bivariate AR(2) model.

Lag	Autocorrelation Matrix		Indicator Matrix	
0	1.00	.11		.
	.11	1.00	.	
1	.04	.00	.	.
	.00	-.08	.	.
2	.06	-.06	.	.
	-.08	-.06	.	.
3	.08	.09	.	.
	.08	.12	.	.
4	-.04	.06	.	.
	.12	-.08	.	.
5	-.23	-.04	.	.
	-.15	.03	.	.
6	.08	-.08	.	.
	.16	.20	.	.
7	-.22	-.14	.	.
	.08	.00	.	.
8	-.02	-.06	.	.
	-.07	-.16	.	.
9	-.06	-.14	.	.
	.12	.12	.	.
10	-.01	.16	.	.
	-.14	.10	.	.

Approximate 95% limits $\pm .24$ for uncorrelated residuals.

. matrix element within limits
+ matrix element above limits
- matrix element below limits

Figure 19. - Bivariate residual autocorrelation matrices.

Date	Actual	Steps Ahead 1	2	3	4	5	6	7	8	9	10	11	12
4-78	191.4	-0.4											
5-78	193.3	-0.5	-1.1										
6-78	195.3	-0.6	-1.2	-1.9									
7-78	196.7	0.2	-0.5	-1.3	-2.1								
8-78	197.7	0.5	0.7	-0.1	-1.0	-2.0							
9-78	199.1	-0.3	0.4	0.7	-0.2	-1.2	-2.2						
10-78	200.7	-0.1	-0.7	0.1	0.5	-0.6	-1.6	-2.6					
11-78	201.8	0.3	-0.1	-0.7	0.3	0.7	-0.4	-1.5	-2.6				
12-78	202.9	0.1	0.5	0.1	-0.6	0.4	0.8	-0.3	-1.4	-2.5			
1-79	204.7	-0.7	-0.5	0.0	-0.4	-1.3	-0.1	0.3	-0.9	-2.1	-3.1		
2-79	207.1	-1.0	-1.9	-1.7	-1.2	-1.6	-2.5	-1.3	-0.9	-2.1	-3.3	-4.4	
3-79	209.3	-0.5	-1.8	-2.9	-2.7	-2.1	-2.6	-3.5	-2.3	-1.8	-3.1	-4.3	-5.4
Root Mean Square		0.50	1.02	1.33	1.29	1.37	1.74	1.96	1.77	2.14	3.17	4.35	5.40

Figure 20. - Forecast errors for bivariate AR(2) model--Consumer Price Index.