

The Box-Jenkins Forecasting Technique

Joseph George Caldwell, PhD

January 1971
(Reformatted September 2006)

© 1971, 2006 Joseph George Caldwell. All Rights Reserved.

Posted at Internet websites <http://www.foundationwebsite.org> and <http://www.foundation.bw> .

Note: This document (*The Box-Jenkins Forecasting Technique*), posted at <http://www.foundationwebsite.org/BoxJenkins.htm> , presents a nontechnical description of the Box-Jenkins methodology. For a technical description of the Box-Jenkins approach, see the document, *TIMES Box-Jenkins Forecasting System*, posted at <http://www.foundationwebsite.org/TIMESVol1TechnicalBackground.pdf> . A set of briefing slides describing mathematical forecasting using the Box-Jenkins methodology is posted at http://www.MathematicalForecasting_Box-Jenkins.pdf . A computer program that can be used to develop a broad class of Box-Jenkins models is posted at the Foundation website, <http://www.foundationwebsite.org> (6 February 2009).

Contents

I. Introduction	1
II. Previous Techniques	2
A. Intuitive Methods	3
1. Smoothing	3
2. Curve Fitting	3
3. Some Classical Times Series Models	4
B. Forecasters Derived from Tested Mathematical Models	5
1. The Model Building Process	5
2. Classes of Models	6
3. Econometric Models	6
4. Stochastic Models	7
III. The Box-Jenkins Method	8
A. Stochastic Box-Jenkins Models	8
1. Estimation of the Model	8
2. Derivation of the Optimal Forecaster	9
3. Simulation Using a Stochastic Box-Jenkins Model	10
B. Stochastic-Dynamic Box-Jenkins Models	10
1. General Considerations	11
2. Identification of Stochastic-Dynamic Model Structure	12
3. Process Control Applications	13
IV. Applications	13
A. Purely Stochastic Models	13
B. Stochastic-Dynamic Models	14
V. Some Comparisons	15
VI. Development Effort	16
VII. Summary	17
References	19
Endnotes	20

The Box-Jenkins Forecasting Technique

Joseph George Caldwell, PhD

I. Introduction

The operations of most businesses are continually affected by events beyond the control of the management. Some of these events are unanticipated or of such importance that they require special management consideration to determine the response to the new situation. For example, new legislation, a large fire, or the introduction of a significant new competitive product, are examples of changes in the environment which will require custom-tailored analysis by management. On the other hand, there are in the course of normal operations a tremendous number of events that cannot be controlled or predicted in detail, but whose occurrence is nevertheless anticipated. For example, product demand, raw material costs, and interest rates are continually changing. There is uncertainty associated with the exact magnitude of the variation in these items, but such variation represents a normal part of the environment. So long as the environment remains essentially unchanged, this variation is expected, and the response to this variation should be automatic, not requiring unusual management analysis. The effectiveness of management response to this variation often depends to a great degree on the extent to which it can reduce the uncertainty about the magnitude and direction of the variation. For example, if a company can improve its ability to predict product demand, it can likely lower its inventory requirements or improve its scheduling efficiency. While it is often desirable to predict long-term changes in the environment, a very basic problem in the routine business operation is that of short-term forecasting: predicting variation in the near future, under the assumption that the essential nature of the variability will continue as in the past (or in some other specified fashion). Management's solution to problems of this sort represent its standard response to the variability problem; its analysis and response to long-term changes in the environment will no doubt require special attention.

The term "near future" used above may refer to several days, months, or years into the future, depending on the time frame of the application. The number of days, months, or years for which we are potentially able to significantly reduce the variability will, of course, depend on the nature of the process we are studying. For example, in forecasting monthly product sales, we may be able to forecast well only a few months ahead if the product is nonseasonal, or a year or two ahead if the sales exhibit strong annual seasonal behavior.

Since short-term forecasting problems are concerned with prediction of variation under the assumption (explicit or not) that the essential nature of the process will continue, it is appropriate to try to determine fixed rules to predict the near future from the recent past. We refer to any forecasting procedure that is a specified function of observed data as a mathematical forecaster. Quite a number of short-term forecasting procedures have been developed in the past. Depending on the nature of the problem, these procedures have ranged from very simple smoothing to the use of formulas based on elaborate econometric models. For example, if a company is forecasting monthly sales of hundreds of products as part of its inventory control system, forecasting procedures (or "forecasters" for short) having low data and computational requirements are appropriate. Alternatively, the same company might use a detailed econometric model and sample survey data to obtain short-term forecasts of the highest possible accuracy for the cost of a key raw material or for total quarterly sales.

In any event, no matter what type of forecaster is appropriate, it is desirable to obtain the best possible forecast for the amount of effort expended. Many short-term forecasters in wide use fall far short of providing the maximum accuracy possible, given the information on which they are based and the computational efforts they require. If little or no historical data are available, then it is quite possible for the errors of judgment in choosing a short-term forecaster to be smaller than the random errors associated with estimating a forecaster from data. If extensive historical data are available, however, it becomes possible to derive a mathematical forecaster based on this data which can predict better than a mathematical forecaster that is chosen by judgment. Since the business situations in which short-term forecasting is of concern generally continue over long periods of time, or for a number of similar items, such data are often available.

This article describes a new technique, developed by Profs. G.E.P. Box and G.M. Jenkins, that enables the development from historical data of forecasters that have both high accuracy and low computational requirements. The technique may be applied to quickly determine forecasters that are as uncomplicated in form as the simple smoothing methods, or that involve a number of economic variables. In either case, use of technique enables efficient utilization of other predictive information contained in the data and offers assurance of obtaining the highest forecasting accuracy possible in terms of the variables on which the forecast is based.

Although the Box-Jenkins model first appeared in book form (Reference 2) in 1967, the business forecasting community seems still largely unaware of the potential of the method. This situation is perhaps understandable since published applications of the technique have appeared, to this author's knowledge, only in technical journals (e.g., Reference 9). The situation has improved somewhat with the publication of Box and Jenkins' book (Reference 1) in late 1970. A recent *Harvard Business Review* article (Reference 4) includes the Box-Jenkins technique in a discussion of the problem of selecting the appropriate forecasting techniques for various applications. There is no question, however, that business awareness of the Box-Jenkins forecasting method is much less than the importance of forecasting problems would warrant. This article is intended to help increase this awareness.

In addition to describing the Box-Jenkins technique, this article briefly describes some previous forecasting techniques and cites some comparisons between the forecasting ability of the old and new techniques. The article avoids mathematical detail, but does involve discussion of certain technical concepts considered fundamental to the problem of developing good forecasters.

II. Previous Techniques

Two approaches have been used in the past to develop forecasters from past observations. On the one hand, forecasters have been developed which possess intuitively satisfying properties. Alternatively, they have been developed from tested mathematical models of the process under study. These two methods represent fundamentally different philosophical approaches to the forecasting problem, and, as we shall see later, they differ considerably in their forecasting abilities.

A. Intuitive Methods

1. Smoothing

A procedure which has been used widely in the past to produce forecasts of future observations is smoothing. Smoothing methods represent attempts to determine some sort of "average" value around which the observations appear to be fluctuating. Two examples of smoothing procedures are the moving average method and exponential smoothing. In the method of moving averages, the forecast value is computed as the average of a fixed number of preceding observations. The number of observations to be included in the average is usually determined arbitrarily to compromise between the "responsiveness" and the "stability" of the forecaster. (Alternatively, the number of observations included might be chosen to remove some sort of periodic behavior, such as a monthly cycle in weekly data. In this case, the moving average is in fact a forecast of the mean level over the next month, rather than a forecast of the next week's level.)

With exponential smoothing, the forecast value is a weighted average of the preceding observations, where the weights decrease with the age of the past observations. In exponential smoothing, there are one or more parameters (constants) which determine the rate of decrease of the weights. These parameters, called smoothing constants, are determined either arbitrarily or by some formal estimation procedure such as the method of least squares.

The smoothing methods described above are easy to describe and straightforward to implement. As their simplicity might suggest, they can generally be improved upon as methods for forecasting. We shall now describe some forecasting methods that are much more elaborate than, but offer little if any improvement over, smoothing methods.

2. Curve Fitting

A graph of the history of any process generally contains characteristic "patterns" which appear over and over. The tendency to extrapolate such patterns is often hard to resist, a number of forecasting methods have been based on the premise that such extrapolation is a reasonable thing to do.

Consider for the moment the graph of the process known as a "random walk." At each successive point in time, there is an equal likelihood that the process will move upward or downward. In the short term, stock prices often exhibit such behavior. Although quadratic-like curves often occur throughout the graph, any attempt to forecast by extrapolating these curves will result in forecast errors of larger average magnitude than the "random walk" forecaster that forecasts the next period's value to be the same as the current value. Reference 1 contains an interesting example of an elaborate method of exponential smoothing derived from fitting curves to a process that was later shown to be a random walk. Not surprisingly, the "curve fitting" method produced forecast errors of substantially larger magnitude than those produced by the "random walk" forecaster.

3. Some Classical Times Series Models

A series of observations taken over various points in some time interval is technically called a *time series*. During the past several decades, a considerable amount of statistical theory has been developed to analyze time series. As with all techniques of statistical analysis, the conclusions of time series analysis are critically dependent on the assumptions underlying the analysis. One of the tragedies of modern "scientific" investigation is that the computational procedures of statistics have often been uncritically applied to data, the underlying assumptions ignored, and false conclusions drawn from this so-called "statistical" analysis. An example of this situation is the application of regression analysis to historical data to derive an equation which is then (wrongly) purported to indicate what will happen if the regressor variables are varied. With respect to the forecasting problem, a similar misapplication of statistical procedures has been carried out. As in the case of regression analysis, the inappropriateness has been obscured by the complexity of the estimation procedures.

Statistical time series analysis provides us, among other things, with good procedures for estimating parameters of various types of models. The structure of the models often corresponds to physically meaningful behavior. For example, in structural analysis, the physics of a situation may dictate a model which is trigonometric in nature. In this situation, a model involving sine and cosine terms would be reasonable, and good statistical estimation procedures (e.g., multiple regression analysis) are available for estimating the model parameters. The use of such models for describing economic time series, however, is not necessarily appropriate simply because the time series data exhibit periodic behavior. Nevertheless, such models have been used in the past to describe, for example, a seasonal time series. Along this same vein of fitting an arbitrarily chosen model to data, a number of "seasonal adjustments" could be estimated from the data. If the postulated model structure were appropriate, then use of the corresponding estimation procedure would be reasonable. In any event, the statistical adequacy of the model must be determined before accepting any such model as a basis for forecasting.

Thus we see that the statistical estimation procedures can be uncritically applied to fit a particular type of model to time series data, and this fitted model then used as a basis for forecasting. The two key points that are often overlooked are that the particular model structure chosen may be inappropriate and that the fitted model may fail to be an adequate representation of the process. These points appear to be overlooked because of the elaborate statistical procedures that are used to derive estimates of the parameters of the model. Of course, if the underlying model assumptions are satisfied, then these estimation procedures -- often developed out of sophisticated theoretical analysis -- do provide good estimates of the model parameters and a good fitted model. As we shall see later, however, there is a vast difference between "model fitting" and "model building" -- the process of determining an adequate mathematical representation of the process generating the time series data. This process involves subjecting a fitted model to a variety of diagnostic checks of its adequacy to represent the process being modeled. A competent time series analysis, of course, includes this latter process. This section does not take issue with the methods of statistical time series analysis -- they are valid. The trouble arises with the uncritical application of an arbitrarily selected technique of time series analysis -- parameter estimation (model fitting) -- and subsequently neglecting to subject the fitted model to various tests of adequacy.

The fact is that many classical time series models were derived for physical situations (as in astronomical, electrical, and mechanical fields) in which underlying deterministic components (such as sine and cosine terms) are obscured by simple random noise, and the various estimation procedures were developed to estimate these components. Economic time series, however, are essentially stochastic rather than deterministic in nature, and for this reason, many classical time series models are simply inappropriate. The use of these essentially deterministic time series models (also referred to as "unobserved component" models) has been remarkably widespread, apparently a carryover resulting from their successful (and appropriate) application in many physical situations for more than a century. It appears, however, that the questionable use of such models in economic applications is finally drawing to an end (Reference 6).

The previous paragraphs have illustrated a number of procedures for fitting a model to data. In view of the large amount of effort that has been expended using fitted models as a basis for forecasting, it is indeed unfortunate that the fact that a particular model is efficiently fitted to data is no assurance that it is an appropriate basis for forecasting.

B. Forecasters Derived from Tested Mathematical Models

1. The Model Building Process

The reason for referring to the techniques described above as "intuitive" is no doubt clear by now -- there is no substantiation that the model underlying the forecaster is an adequate representation of the process under study. Whatever properties the forecaster might possess relate to the model to which it corresponds, but these properties don't mean very much if there is little assurance that the model is a good representation of the actual process. To determine a good short-term forecaster for a process, we need a good model of the short-term behavior of the process. We shall now turn our attention to forecasters based on models which are good representations (in the short term) of the processes they represent.

The process of developing a good model for a process is called *model building*. Model building involves four basic steps. (See Reference 1 for a detailed discussion of model building.) First, we must have available a class of models which is capable of exhibiting the essential characteristics of the process under study. Second, a preliminary analysis of the problem under study will suggest a tentative subclass of models which is reasonable to entertain. Third, observed data are used to "fit" one of these models, i.e., to estimate the parameters of the model. Fourth, the fitted model is subjected to a number of diagnostic checks to test whether the model is an adequate representation of the process under study. If the tests are not satisfied, the tentative model is modified. Steps 3 and 4 are then repeated until the tests are satisfied, i.e., until we have an adequate model. In summary, model building involves: (1) selection of a general model class; (2) model identification (tentative model selection); model fitting (estimation of parameters); (4) diagnostic checking (tests of model adequacy) and model modification; (5) repeat steps 2, 3, and 4 until an adequate model is found. Thus model building is an iterative process involving much more than simply fitting a model to data, the basis for intuitive forecasting. Good procedures for estimating parameters are easy to specify in terms of formulas and, understandably, there has been a great deal of model fitting done in the past, with the subsequent use of intuitive forecasters. Diagnostic checking and model modification

requires critical statistical analysis, rather than the straightforward application of standard statistical estimation formulas, and in fact represents the bulk of the human effort required in model building; (the computations required to fit tentative models and compute the statistics required to test model adequacy can be performed by computers). Thus we see that model building involves a wide range of time series analysis techniques, not just those associated with parameter estimation.

2. Classes of Models

The first step in model building is to choose a class of time series models with which to work. At the extremes, there are two basic approaches. On the one hand, we may attempt to develop a *dynamic*, or *causal* model of the process under study. In such a model we attempt to relate the behavior of the variable of primary interest to the behavior of other variables. The variables to tentatively include in the model, and the tentative model structure, are suggested by economic theory, and the model is referred to as an *econometric* model. As an alternative to the econometric model, we may attempt to develop a purely *stochastic*, or *empirical* model of the process. With the approach we strive simply to develop a model which exhibits the same essential characteristics as the process under study, without attempting to identify the casual nature of the relationships between the various relevant interacting variables.

The terms used above are not perfectly descriptive. For example, a "dynamic" model will always possess a simple *stochastic* term to describe the variation unexplained by the dynamic structure. Furthermore, a "casual" economic model will undoubtedly employ *empirically* derived relationships between the variables. Intermediate between the two extremes of the dynamic and the stochastic models is the class of *stochastic-dynamic* models, which contain both casual components and nontrivial stochastic components.

In forecasting with a stochastic model of a process, we are in essence attempting to predict the next few moves of the process based generally on all of the past behavior of the process, and in particular, on the recent history of the process. The model derived from the data describes how the process behaves, and the forecaster predicts the near future of the time series from the recent past, based on the stochastic behavior of the process as characterized by the model. With an econometric model, the forecaster predicts the near future from recent past, based on the economic relationships characterized in the model.

The problem of choosing a model class involves balancing accuracy requirements against the costs associated with developing and implementing the forecaster. For example, an elaborate econometric model would clearly be inappropriate for forecasting the short-term sales of each of thousands of items of an inventory system. Reference 4 includes a discussion of the problem in choosing an appropriate model class.

3. Econometric Models

To construct an econometric time series model, we identify the variables that are considered to have an effect on the variable of interest, and then pose a tentative structure for the model. In most cases a linear or linearized model relating the variables is considered, and standard

regression analysis is used to estimate the model parameters (regression coefficients). The estimation is, however, not always straightforward. For example, suppose that we wish to forecast sales, and it is known that sales are related to price, and price in turn to sales. Then a simultaneous system of equations is necessary to describe the system, the usual regression estimates are inappropriate, and some other method, such as "two-stage" least squares, must be used to estimate the parameters.

Once the estimation has been completed, i.e., we have a fitted model, then it is necessary to subject the model to various diagnostic checks. These checks often involve the model "residuals." A "residual" is the difference, or "error," between an actual observation and the value predicted by the model. In an econometric model, one of the usual underlying assumptions is that the residuals are unrelated to each other in a certain statistical sense. In technical terms, the residuals are not *autocorrelated*. If, for a fitted model, the residuals are autocorrelated, then the current tentative model must be modified and a new tentative model entertained. The new tentative model might involve either a somewhat different dynamic structure or include some new variable. Alternatively, if the residuals are relatively small in magnitude, their autocorrelation can be taken into account in the estimation procedure, without changing the dynamic model structure. If the latter approach is followed, then we in fact have a simple example of a stochastic-dynamic model.

In developing an econometric model (or any other model for that matter) it can be unwise to include a very large number of variables in the model, particularly if the inclusion is based solely on empirically observed relationships. With a large enough number of variables, it would not be surprising to find a model that seemed generally adequate and yet proved, for several reasons, to be a poor basis for forecasting. First, the number of parameters may simply be so large compared to the number of observations that it becomes difficult to perform sensitive tests of model adequacy. Second, by increasing the number of variables in a model we increase the risk of discovery, by chance, an apparent relationship of the variable we wish to forecast to some other totally unrelated variable. This, of course, is a principal drawback associated with using regression analysis to simply "fit" a forecasting model to data. An additional problem associated with a many-variable forecasting model is that in using it we implicitly assume that the relationship between all the variables of the model will continue in the future as in the past. The larger the number of variables involved, the less reasonable such an assumption becomes. In general, if a large number of parameters seems necessary, it is reasonable to suspect that the identified model structure is inappropriate.

4. Stochastic Models

In order to develop a stochastic time series model to represent a process, it is necessary to have a flexible class of models available. We have already mentioned a few classical time series models above and noted their applicability to essentially deterministic situations. Two other models that have been used and that are essentially stochastic in nature are the (finite) autoregressive model and the (finite) moving average model. In the autoregressive model, the current observation is represented as a weighted average of previous observations, plus a random term that is uncorrelated with the random terms of other observations. In the moving average model, the current observation is represented as a weighted average of uncorrelated random terms.¹

In general, the preceding models have not proved sufficiently general to model arbitrary economic time series with a reasonable number of parameters, and this fact probably accounts in part for the limited use of stochastic models (of stochastic-dynamic models) for model building. This situation has now been remedied, for the models investigated by Box and Jenkins possess the capability to efficiently represent a tremendous variety of economic time series.

This section has described in brief detail the major categories of techniques that have been used in the past to develop forecasting models. We shall now turn our attention to a description of the Box-Jenkins forecasting method.

III. The Box-Jenkins Method

A. Stochastic Box-Jenkins Models

1. Estimation of the Model

As the preceding section has suggested, not a great deal of forecasting has been done using tested stochastic or stochastic-dynamic time series models. The essential reason for this situation is that a suitable class of stochastic models had not been identified as possessing the flexibility necessary to represent efficiently (i.e., using a reasonable number of parameters) the tremendous variety of characteristics of economic time series. This situation no longer holds, for Box and Jenkins have thoroughly investigated a class of models that prove to be quite satisfactory for both the stochastic and stochastic-dynamic situations. We shall first describe the purely stochastic models. Technically, these models are called autoregressive integrated moving average (ARIMA) models, or simply Box-Jenkins models for short.

The Box-Jenkins models can be used to represent processes that are stationary or nonstationary. A stationary process is one whose statistical properties are the same over time; in particular, such a time series fluctuates around a fixed mean value. Examples of nonstationary time series include series which include changes in level, trends, changes in trends, or seasonal behavior.

The purely stochastic Box-Jenkins model is remarkably simple in form. The current observation is represented by a linear combination (weighted average) of previous observations, plus an error term associated with the current observation, plus a linear combination of error terms associated with previous observations. The error terms have zero mean, constant variance, and are uncorrelated with each other. The portion of the model involving the observations is called the *autoregressive* part of the model, and the portion involving the error terms is called the *moving average* part of the model.²

The problem of building a stochastic Box-Jenkins model of a process in essence involves determining the number of terms in the autoregressive and moving average parts of the model,

and determination of values for the parameters associated with those parts. By statistical analysis of the time series data, it is possible to choose from the full class of autoregressive-integrated moving average models, subclasses of models having a specific structure appropriate to the particular time series under examination. By determining such a reasonable structure for the model, the number of parameters to be estimated in the model can be substantially reduced. This parameter reduction is quite important, for "nonlinear" statistical estimation procedures are generally required to fit a tentative Box-Jenkins model.

After a tentative Box-Jenkins model has been fitted, it is subjected to various diagnostic checks to test its adequacy as a stochastic representation of the process under study. If the model is found to be inadequate, analysis of the model residuals suggests ways to modify the model structure to obtain a new tentative model which will likely do an improved job of representing the process. The basic statistic for assisting identification of a reasonable structure for a new tentative model is the autocorrelation function of the residuals of the current model. For testing model adequacy, the power spectrum of the residuals can also be used as an alternative to the autocorrelation function (see Reference 7). This process of fitting a tentative model, testing it, and determining a new tentative model, is repeated until a model is found which does an adequate job of representing the process. (References 1 and 3 contain detailed descriptions of this model building process.)

2. Derivation of the Optimal Forecaster

After a model has been determined that, according to the various tests to which it is subjected, is considered to be an adequate representation of the time series under examination, we are in a position to derive a forecaster for the time series. In order to determine a forecaster from a model, it is necessary to specify a criterion which the forecaster satisfies. Ideally, the criterion should take the "cost" associated with forecast errors into account. We would then like to derive the forecaster for which the expected cost associated with forecast errors is minimized. If the cost function is not apparent, a reasonable approach is to determine the (linear) forecaster that has the least forecast error variance, or mean squared error of prediction. That is, for each specified time in the future we wish to determine the forecast that has minimum mean squared error at the point. The minimum mean squared error forecaster is usually referred to as the "optimal" forecaster. The optimal forecaster corresponding to a Box-Jenkins model turns out to be the expected (mean) value of the process, conditional on the past observations. For example, the optimal forecast one time period into the future is the expected value of the process at that point in time, given that the past values of the time series are as observed.

Computation of the optimal forecast is quite easy for a stochastic Box-Jenkins model. To compute the lead-one forecast, for example, we simply substitute, into the formula defining the model, the observed (past) values and the optimal estimates of the past error terms. These error terms can be estimated recursively: they are simply the past forecast errors. To compute the optimal forecast beyond lead one, we substitute forecasts for future observations and zeros for future error terms. Thus we see that the computational effort required to forecast using a Box-Jenkins model is on the same order as that required by the simple smoothing techniques.

Computation of tolerance (probability) limits around the forecasts is straightforward for a Box-Jenkins model. If we have a stationary process, the distance between these limits will gradually

increase for forecasts further and further into future time, to a fixed value proportional to the variance of the process. For a nonstationary process, the variance of the process is undefined, and the tolerance interval grows substantially wider as we forecast further into the future.

It is of interest to note that a forecaster derived from a stochastic model is an adaptive forecaster. Adaptivity is, of course, a very desirable property for a forecaster to possess and a necessary one for a forecaster of a nonstationary process. (The various smoothing methods -- exponential smoothing, moving average -- have been so widely used because, in addition to requiring few computations, they are adaptive in nature.) A very important feature of the Box-Jenkins models is their ability to efficiently represent quite general nonstationary processes. By its very definition, a stochastic model embodies the changing nature of the process, and the corresponding forecaster is in essence a description of how the future is likely to turn out, given the recent past behavior of the process. Suppose, for example, that a process appears to have "trends" that change slope over time, or that it exhibits pseudo-periodic behavior with a stochastically varying phase and amplitude. If such a process is modeled using, for example, trends, trend adjustments, seasonals, and seasonal adjustments, then the parameters of the model will probably have to be updated rather frequently; i.e., the forecaster cannot adapt automatically to the changing situation, since the model does not represent the process well. The key point here is that we want an adaptive *forecaster* rather than an adaptive *model*. We wish to derive a fixed model and to derive an adaptive forecaster from it. An adequately tested stochastic model of a process meets this requirement. By taking the essential stochastic characteristics of the process properly into account, it produces a forecaster that in essence tells *how to* take the changing features of the process into account.

3. Simulation Using a Stochastic Box-Jenkins Model

Although we are primarily concerned with forecasting in this article, a few words are in order regarding the use of the Box-Jenkins models for simulation. For example, we may be interested in determining the effect of sales variability on a new inventory policy, starting from the current sales position. In such a case, we need a model which exhibits the same statistical properties as the actual sales themselves. An econometric model would be difficult to use in such a situation because, as the model is simulated into the future, we need to substitute values for all of the variables of the model, and the econometric model does not specify the behavior of any of these variables other than the variable(s) of primary interest. A Box-Jenkins model can readily be used in such a situation, for it is in fact a description of the statistical properties of each succeeding observation in terms of the preceding observations. Each simulated observation becomes input to the succeeding simulated observation. It should be remembered, however, that a Box-Jenkins model is concerned essentially with the local (short-term) behavior of a process, since it is through understanding the local behavior that we are able to predict the near future from the recent past. If a long-term simulation model is desired, then special care must be taken to make sure that the model reflects the longer-term properties of the process.

B. Stochastic-Dynamic Box-Jenkins Models

1. General Considerations

As we mentioned earlier, the Box-Jenkins method can be used to develop stochastic-dynamic models, in which the behavior of the variable of primary interest (the endogenous variable, or variable we wish to forecast) is related not only to its past behavior, but to the behavior of other (exogenous) variables as well. The reason for including exogenous variables is obvious: since the model class is expanded, the precision of the forecast may be increased over that corresponding to the pure stochastic model. If an exogenous variable is included in a model to be used for forecasting, however, then the values of that variable must be known or forecast over the forecasting period of interest. Typically, the behavior of the exogenous variable presages that of the variable of primary interest; that is, the current behavior of the variable of primary interest is related to the past behavior of the exogenous variable. Such an exogenous variable is called a *leading indicator*. Obviously, if there is a strong lagged relationship between an exogenous variate and the variable of primary interest, the precision of the forecasts will be considerably enhanced.

It is recalled that a stochastic-dynamic model is a hybrid between a purely econometric model and a purely stochastic model. In it, the behavior of the variable of interest that cannot be explained in terms of the behavior of the exogenous variates is represented by a stochastic portion of the model. As we add additional exogenous variables to the model, the stochastic portion of the model becomes less and less elaborate, and we may approach the "purely dynamic" model in which the stochastic portion of the model is simple uncorrelated random variation. (In a particular situation it may not be feasible to derive such a "purely dynamic" model; the practical limit may be a stochastic-dynamic model with a nontrivial stochastic component.)

Note that once we have introduced an exogenous variate into a model, we might refer to such a stochastic-dynamic model as an econometric model. Rather than simply implying the inclusion of more than one economic variable, however, the term *econometric* usually indicates that the model has a special structure dictated by economic theory, rather than empirically determined. Usually, the stochastic part of the such econometric models is quite simple in form, and the predictive power of the model arises out of the dynamic rather than the stochastic relationships. Hence, while an econometric model is certainly an example of a stochastic-dynamic model (in which the stochastic portion is usually simple in form), we shall generally use the term "stochastic-dynamic" to describe models in which the relationships between the variables are empirically determined.

A multivariable³ Box-Jenkins stochastic-dynamic model may include all of the relevant variables of a so-called econometric model. The Box-Jenkins stochastic-dynamic models simply represent a particular class of empirical models that are capable of efficiently representing a wide variety of processes involving more than one variable. While the structure of these models is quite flexible, the economic nature of a process may suggest a special model structure which could be an even more efficient representation of the process. In a sense a Box-Jenkins stochastic-dynamic model could be viewed as an empirically-derived "econometric" model, in contrast to a causally-derived "econometric" model; (we are using the term "econometric model" to refer to this latter type of model). To the extent possible, of course, mathematical model building should always take advantage of special features or understanding of the real-world

process being modeled. Such understanding forms the fundamental basis for selecting a particular class of models to represent a process.

The performance of a forecaster based on an econometric model is often compared to that of a forecaster based on a simple stochastic model, such as a finite autoregressive scheme. The forecasting ability of the autoregressive forecaster is often taken as a minimum standard of performance for an econometric model (References 5, 8 and 10). Needless to say, a more reasonable comparison would be that of the econometric forecaster to a forecaster based on a tested stochastic model, rather than on an arbitrarily selected and probably inappropriate stochastic model.

2. Identification of Stochastic-Dynamic Model Structure

Often econometric models have been constructed using linear regression analysis. The term "linear" refers, of course, to the nature of the statistical estimation, rather than the functional form of the regression function, which may be highly nonlinear. Computationally, about the most difficult problem that might arise with this approach would be that the residuals (error terms) of the model may be autocorrelated, but the standard regression methods can usually still be used for the analysis after appropriate steps have been taken. There are essentially two reasons for such autocorrelation. First, it may simply not be possible or practical to include a sufficient number of exogenous variables in the model so that the stochastic part of the model would be simple in form (not autocorrelated). Second, the structure of the model may not be representing the relationship between the variables of the model in a very efficient fashion.

By way of analogy, suppose that a one-parameter moving average process best described the stochastic behavior of a time series, but that an autoregressive model was fitted instead to the data. By allowing a sufficient number of parameters in the autoregressive model, a satisfactory fit could be obtained, with residuals having negligible autocorrelation. However, if we restrict the number of autoregressive parameters, we would have neither a very good fit nor negligible autocorrelation in the residuals. Similarly, with the stochastic-dynamic model, it may be quite inefficient to insist that the variable of interest be explained solely in terms of linear combinations of other variables. A much more efficient structure for the model, allowing for a better fit with fewer parameters and less autocorrelation in the residuals, may be possible if we represent the process also in terms of a linear combination of the variable of primary interest.

Thus we see that, if we are to determine stochastic-dynamic models which, in terms of the number of parameters used, are efficient representations of the processes, it is important to identify reasonable model structures to investigate. The practical question that arises is, of course, how to accomplish this identification, which can be rather difficult whenever both the variable of interest and the exogenous variables are autocorrelated. It turns out that this identification is often facilitated through the use of stochastic models for the exogenous variables of the model.

Essentially what is done is to develop a stochastic model for an exogenous variable and then use this model as a "filter" to transform the exogenous variable to a "white noise" series. This same filter is also applied to the variable of primary interest. In electrical engineering terminology, the preceding process is known as "prewhitening." It then turns out that a certain

statistic (the cross correlation function between the prewhitened variables) can be used to rapidly identify a reasonable structure for the part of the model relating to the exogenous variable. Thus, just as the autocorrelation function assists identification of the structure of a stochastic model, the cross correlation function assists identification of the structure of a stochastic-dynamic model.

3. Process Control Applications

This article is primarily concerned with the forecasting applications of the Box-Jenkins method. As noted earlier, the models can be used for simulation as well. Another area of significant application is the field of process control. In control applications we can not only observe the behavior of the "explanatory" variables of the model (called control variables, rather than exogenous variables), but we can control them to produce changes in the variable of primary interest. To control the variable of primary interest in the desired fashion, we need a good mathematical model of the process. As in forecasting applications, the Box-Jenkins models have sufficient flexibility to efficiently model the behavior of a wide variety of physical processes.

In developing a model for control we are in a somewhat different position from the situation in forecasting, in that data can be collected corresponding to forced changes in the control variables. In economic situations we can in general only observe, not manipulate, the exogenous variables. The manner in which the data are collected has a fundamental effect on the use to which the developed model can be put. For example, we cannot use a model to predict what changes in a system will result from our manipulating a control variate, if the model was developed from data in which the system was merely observed, and not interfered with. The manner in which the control variables are manipulated to produce the data from which the model is derived will depend on the application. The subject will not be discussed here but is described in detail in Box and Jenkins' book (Reference 1).

The reader interested in control problems should also consider the time-varying dynamic systems models investigated by R.E. Kalman and R.S. Bucy. The need for a time-varying representation seems less strong for economic processes, however, than for time series arising in physical applications.

IV. Applications

A. Purely Stochastic Models

The obvious application of the Box-Jenkins technique is to develop a purely stochastic model from which to derive an optimal forecaster. This use of the Box-Jenkins method will no doubt receive widespread application, since the model development is rapid, the optimal forecaster is easy to compute, and the data requirements for both model development and forecast computation are low (only data on the variable of primary interest are required). Computation of tolerance limits around the forecasts is also straightforward.

The Box-Jenkins method is particularly suited for development of models of processes exhibiting strong seasonal behavior. Earlier techniques of fitting trigonometric models, seasonal patterns, or seasonal adjustments, often failed to allow for gradual changes in the "shape" of the seasonal "pattern." Often, in fact, the parameters of the model had to be updated in order to take such changes into account. Once again, if the changing nature of the seasonal behavior of the process is one of the basic stochastic properties of the process, use of a Box-Jenkins model of the process will result in a forecaster that adapts automatically to the changing seasonal "shapes."

In addition to the direct application of stochastic models for developing forecasters, we have noted the use of stochastic models in assisting the identification of relationships in stochastic-dynamic models. As we shall soon see, pure stochastic models also play an important role in forecasting with a stochastic-dynamic model.

B. Stochastic-Dynamic Models

As we observed above, it may be possible and advantageous to entertain a Box-Jenkins model with exogenous variates in order to improve the accuracy of the forecasts based on the model. The associated cost lies with the increased effort involved in model construction, and the additional data requirements for model development and forecasting. We shall now examine some problems associated with models which include exogenous variables.

If the time lag between the variable of primary interest and the leading indicator is not very great (that is, the needed values of the exogenous variable are known only for a short time into the forecasting period), then a question arises as to what values should be used as forecasts for the exogenous variates. It turns out that in order to determine the optimal forecasts from such a model we need to use the optimal forecasts for the exogenous variates in the model. That is, we need stochastic models for the exogenous variates of a stochastic-dynamic model in order to determine the optimal forecasts for the variable of primary interest of such a model.

Since the stochastic-dynamic model class becomes a more and more general class as we increase the number of exogenous variables, greater and greater forecasting accuracy becomes possible. However there is a tradeoff: As we increase the number of exogenous variables in the model we increase the data requirements of the model. Also, depending on the length of lag between the variable of primary interest and the exogenous variates, we increase the number of variables for which forecasts are needed, and hence for which we must construct stochastic models. The tradeoff made between forecasting accuracy and difficulty of implementation will be determined by the nature of the cost associated with increased data collection and analysis requirements on the one hand, and decreased precision on the other. Situations can occur, of course, in which virtually no improvement in forecasting accuracy occurs as exogenous variates are added to the model. The forecasting capability is simply transferred from the stochastic portion of the model to the dynamic portion of the model, with little overall improvement. (See, for example, Reference 10.)

It is noted that econometric forecasting models often include "anticipatory" variables. An anticipatory variable is a measure, determined by sample survey, of the expectations of a particular group. For example, a sample of companies in a particular industry might be polled

concerning their expected capital investment over the next quarter, and the results of this survey included as an explanatory variable in the capital investment forecasting model. In some situations (References 5 and 8) the use of anticipatory data alone can provide a better forecast than the best econometric model based on nonanticipatory data. Once again, however, a tradeoff arises between forecasting accuracy and cost for the cost of the sample survey to collect the anticipatory data can be considerable.

V. Some Comparisons

The essential property of the Box-Jenkins method is that it enables rapid development of a forecaster that, with respect to the variables included in the model, has as high accuracy as possible.

The theoretical basis for forecasters derived from tested models offers assurance that the forecaster will have greater accuracy than intuitive forecasters. Such assurance is not enough, however, for it is of interest to have some idea of the magnitude of the difference in actual applications. Similarly, we would like to know the differences in forecasting accuracy corresponding to different model classes, such as the stochastic and the stochastic-dynamic model classes. Questions such as these can be answered only by applying the various techniques to the same problem and comparing the results. Unfortunately since the Box-Jenkins technique is new, there are not a very large number of such comparisons on record. This situation will no doubt change, but for the time being at least the "case histories" are relatively few in number.

Box and Jenkins present a revealing comparison between a forecaster based on a very general curve-fitting method and the optimal forecaster based on a stochastic model. The curve-fitting method was one which fitted a combination of sines, cosines, polynomials and exponential functions to data using a "discounted least-squares" method to estimate model parameters. In the particular example quadratic polynomials provided a good fit, and the corresponding forecaster turned out to be an exponential smoothing procedure. The stochastic model which fit best was an integrated moving average process like the random walk model mentioned earlier. The error variance of the optimal forecaster based on the stochastic model was in general (over 10 periods into the future) one-half that of the curve-fitting forecaster.

Reference 8 describes a comparison of forecasts based on econometric and stochastic models for hog prices and quantities sold in domestic commodity markets. The econometric model represented an attempt to describe the supply-demand relationships between quantity and price, whereas separate stochastic models were used for the two quantities. For forecasting quantity, the stochastic models and econometric models exhibited about the same forecasting accuracy. For forecasting prices, the standard deviation of the forecast errors associated with the econometric model was about 80% of the standard deviation of the forecast errors for the stochastic model. Thus, for prices, a considerable improvement in forecasting accuracy resulted by considering the relationship of the variable of interest to other variables. This comparison is not exactly fair, however, in that a *system* of equations was used to represent the econometric model, whereas price and quantity were modeled separately in the stochastic approach. Had a multivariate Box-Jenkins model been used, it is expected that the econometric and stochastic models would have been quite similar in performance, since the import variables

of the econometric model were simply the two variables under study, namely, price and quantity.

VI. Development Effort

There is little doubt that forecasters developed using the Box-Jenkins technique will possess higher accuracy than other empirical methods utilizing the same data. In addition to considering forecasting accuracy, the decision to employ the Box-Jenkins method must include consideration of the amount of effort required to develop and implement the Box-Jenkins forecaster. As we have already noted, the computational requirements of the Box-Jenkins forecaster are generally no greater and quite possibly less than those of other forecasters utilizing the same data. It remains to consider the data and effort required to develop the Box-Jenkins forecaster.

Since the Box-Jenkins forecaster is developed from a time series analysis, it is of course necessary to have a history of the process for which a forecaster is desired. Let us consider first the case of a univariate stochastic Box-Jenkins model, i.e., a model involving only a single variable. In general, the data requirements for the stochastic Box-Jenkins models are no greater than those for the other models that have been used in the past, such as classical time series models, or smoothing methods in which the smoothing constants are estimated from the past data. In general, 50-150 successive observations of the process would suffice to develop a model. Alternatively, if a common forecasting model is desired for a group of nonseasonal items having "similar" variability, a collection of shorter time series for the items of the group would suffice. For example, a time series of 150 monthly sales of a particular product, or 20 time series of 15 monthly sales for each of 20 products could be used. Of course the variety of tests to which the model developed from the shorter time series could be subjected would be rather limited in this latter case. On the other hand, developing a model from several short series of recent data may be desirable, particularly in working with quarterly or annual data for which a long time series would extend many years into the past. For a seasonal problem, involving for example the prediction of monthly sales in a situation in which there is annual seasonal behavior, the data should cover a number of seasons, possibly as few as five or six, but preferably seven or more. For a situation involving several seasonal components (e.g., prediction of daily interest rates in which weekly, monthly, quarterly or annual seasons might occur), then the preceding data requirement applies to the season of highest period (in this example the annual season). The preceding data requirements are suggested as guidelines only; the nature of the particular situation will delineate the appropriate data requirements. In any event, the more data that are available for analysis, the more precise will be our parameter estimates and tests of model adequacy.

In addition to the data requirements, the Box-Jenkins method requires the application of certain analytical skills. The emphasis here is on the word *analysis*, rather than on the particular procedures involved. Unfortunately, yet understandably, the development of a model requires somewhat more than the ability to compute the least-squares estimate of a parameter, or to test a regression coefficient for significance. The computations required to fit a tentative univariate Box-Jenkins model pose little problem since computer programs are available to perform these computations. The essence of building a Box-Jenkins model involves the use of such a program to develop a tentative model, and the interpretation of the statistics computed by such

a program to either accept the current model as adequate or to suggest a modified tentative model. The efficient development of a Box-Jenkins model thus involves the combination of a critical mind with the computational power of a computer. The actual amount of time required by the analyst to develop a Box-Jenkins model is quite modest. For a typical time series, only a few hours total time are adequate to develop a stochastic model. (In a typical situation, however, computer availability may be limited, and the need for a separate program run for each tentative model would spread this total time over a somewhat longer period of time.) The above time estimates assume knowledge of the Box-Jenkins technique and some experience using it. A company's decision to acquire these skills, as opposed to retaining consultants to develop the model, will depend upon the number and importance of the short-term forecasting problems it encounters.

In short, while the development of a Box-Jenkins forecaster requires a particular set of skills, the amount of human effort required is generally not any greater than that required by some of the more elaborate "curve-fitting" methods that have been employed in the past.

If exogenous variables are included in a Box-Jenkins model, the development is likely to require somewhat more time. In general, development of a Box-Jenkins stochastic-dynamic model including exogenous variables will require no more effort than the development of an econometric model involving the same number of exogenous variables. (However with a stochastic-dynamic Box-Jenkins model, we are in a position to include only a few of the most important explanatory variables, modeling the "remaining" variation with the stochastic component of the Box-Jenkins model.)

VII. Summary

The Box-Jenkins stochastic models represent a flexible class of models that can be used to represent the short-term behavior of a wide class of time series. Stochastic models are useful as a means for developing optimal short-term forecasters solely in terms of the variables of primary interest. In some instances, these stochastic forecasters are about as accurate as those based on elaborate econometric models. This situation would hold to an even greater extent with multivariate stochastic models. The Box-Jenkins stochastic models can be used to provide forecasts for the exogenous variables of an econometric model, to enable determination of the optimal forecast based on the econometric model. They can also aid the identification of a reasonable structure for an econometric model, and can be used to model autocorrelated residuals in an econometric model. Finally, they are especially well-suited to the problem of simulating near future realizations, or outcomes, of a time series.

The Box-Jenkins stochastic-dynamic models include a useful class of models intermediate between the "purely stochastic and the "purely" econometric models. With this class it may be possible to approach the increased precision of an econometric model, without the need for including a large number of exogenous variables in the model. The applications of these models to control problems have been noted.

The important characteristic of the Box-Jenkins method is not, however, that it might produce a forecaster that is as accurate as one based on an econometric model (it probably won't).

Rather, the method is a means for rapidly determining an optimal forecaster in terms of whatever variables are specified.

Box and Jenkins have demonstrated efficient procedures for developing these models from time series data. Using electronic digital computers, these procedures generally involve no more human effort than the procedures for developing many less accurate forecasters.

Because of the recent introduction of the Box-Jenkins method, there is not substantial literature available comparing this method to other methods currently in wide use. It is hoped that this situation will change quickly, as the business community increases its use of the Box-Jenkins method.

References

1. Box, G.E.P., and G.M. Jenkins, *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, 1970.
2. Box, G.E.P., G.M. Jenkins, and D.W. Bacon, "Models for Forecasting Seasonal and Nonseasonal Time Series," in *Spectral Analysis of Time Series*, B. Harris, ed., John Wiley & Sons, Inc., New York, 1967.
3. Box, G.E.P., and D.A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, Vol. 65, No. 332, December 1970, pp. 1509-1526.
4. Chambers, J.C., S.K. Mullick, and D.D. Smith, "How to Choose the Right Forecasting Technique," *Harvard Business Review*, July-August 1971, pp. 45-74.
5. Friend, I., and W. Thomas, "Predictive Ability of Plant and Equipment Anticipations," *Journal of the American Statistical Association*, Vol. 65, No. 330, June 1970, pp. 510-519.
6. Grether, D.M. and M. Nerlove, "Some Properties of 'Optimal' Seasonal Adjustment," *Econometrica*, Vol. 38, No. 5, September 1970, pp. 682-703.
7. Jenkins, G.M. and D.G. Watts, *Spectral Analysis and its Applications*, Holden-Day, San Francisco, 1968.
8. Jorgenson, D.W., J. Hunter and M. Ishag Nadiri, "A Comparison of Alternative Econometric Models of Quarterly Investment Behavior," *Econometrica*, Vol. 38, No. 2, March 1970, pp. 187-224.
9. Leuthold, R.M., A.J.A. MacCormick, A. Schmitz, and D.G. Watts, "Forecasting Daily Hog Prices and Quantities: A Study of Alternative Forecasting Techniques," *Journal of the American Statistical Association*, Vol. 65, No. 329, March 1970, pp. 90-107.
10. Stekler, H.O., "Forecasting with Econometric Models: An Evaluation," *Econometrica*, Vol. 36, No. 3-4, July-October 1968, pp. 437-463.

Endnotes

1. The terminology here may be somewhat misleading. For a moving average process, each observation is a moving average of uncorrelated random terms; for an autoregressive process; each observation is a moving average of past observations. The moving average smoothing procedure described earlier *is not* the appropriate forecaster for a moving average process.

2. In mathematical notation, the purely stochastic Box-Jenkins model is

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} \\ + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}$$

where z_t = observation of time t ; a_t = error at time t ; and the ϕ 's and θ 's are parameters of the model. The a_t 's are a "white noise" sequence, i.e., they have constant mean 0 and variance σ^2 ; and are uncorrelated with each other.

3. In a *multivariate* model, we are directly interested in more than one quantity. For example, we may wish to forecast price *and* quantity, as a vector pair, taking full account of the joint relationship between them. A model used to describe a simple (scalar) variable, such as price alone, is an example of a *univariate* model. We use the term *multivariable* to refer either to a multivariate model or to a univariate model containing exogenous variables. The term "naive forecaster" has sometimes been used to refer to a forecaster that is based on a single-variable model (univariate with no exogenous variables).