

Forecasting: Does the Box-Jenkins Method Work Better than Regression?

S Nanda

A difficult exercise, forecasting is key to effective management. Models using time series data are frequently used for forecasting likely values of important variables such as supply and demand.

The regression method is the most common, although it involves many critical assumptions that are difficult to satisfy in practice. Efforts to reduce the severity of the assumptions and improve our ability to manipulate data have led to generalized regression and the Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) models.

Nanda tests the forecasts from two models in each method, using data on monthly milk procurement by Amul Dairy from January 1965 to December 1975. The regression method produced better forecasts than the Box-Jenkins method.

S Nanda teaches at the Whittemore School of Business and Economics, University of New Hampshire, Durham, and also works for Elder Care Services Inc., Rowley, MA., USA.

Forecasting is an essential part of the managerial decision making process. If decisions involve marketing or sales planning, demand has to be forecast; for decisions on inventory level, future requirements have to be forecast; and in decisions on new product introduction, the market for the product has to be forecast.

How does one forecast? For a systematic approach one needs a forecasting model. There are many approaches for developing a forecasting model. One approach to forecasting uses time series data. In this paper, we discuss and compare two techniques of forecasting, both using time series data. These are the multiple regression models and the Box-Jenkins' Autoregressive Integrated Moving Average (ARIMA) Models. The two are compared by forecasting milk procurement by Amul Dairy.

Forecasting Models using Time Series Data

A time series consists of observations generated sequentially over time. For example, time series data on milk procurement may consist of actual milk procured every month. Time series data are ordered with respect to time, and successive observations may be dependent. The observed time series is generally referred to as time series realization of an underlying process. The data may indicate that there is a trend over time, i.e. there is a long term behaviour underlying the data. The trend, if observed over a long time, may be increasing, decreasing, or may remain unchanged. There may be cyclical fluctuations, i.e. a pattern of ups and downs over time. In addition, the data may show that the underlying process has periodic fluctuations of constant length, i.e. seasonally.

The purpose of modelling is to capture this underlying process using the observed time series so that one can predict what would be the likely realization at a time point in future. Two methods

that can be used to do so are:

- Regression method
- Box-Jenkins method.

A brief description of these two methods follows.

Regression

The regression method assumes that the characteristic of interest, referred to as the dependent variable, has some association with some independent variables. It attempts to capture this association and uses it to forecast future values. Knowing the values of independent variables, the regression method can be used to predict the average value of the dependent variable. For example, we know that the demand for a product is related to its price. While developing a regression model of demand and price, demand will be referred to as the dependent variable and price as the independent variable. Knowing the likely price at a future date we can forecast what the average demand is likely to be.

In any case, the regression method makes certain assumptions. Two important assumptions— independence of residuals and constant variance of residuals—are discussed next.

Independence of the Residuals. The difference between an observed value of the dependent variable and its estimate using the regression model is called a residual. Regression assumes that the residuals are independent. That is, knowledge of one or more of the residuals is assumed to provide no additional information to assess any other residual. If this assumption is not true then there is what is called autocorrelation among residuals. There should be no autocorrelation. The presence of autocorrelation can be tested using tests such as the Durbin Watson test. The second assumption is regarding the variance of the residuals.

Assumption of Homoscedasticity. Residuals are assumed to have a constant variance. For a given value of the independent variable price, there will be a distribution of residual, namely the difference between the observed demand and the estimated value of demand. This distribution will have a variance. What the constant variance assumption implies is that for all given values of price, the distributions of the residuals will have the same variance. That is the assumption of homoscedasticity.

The heteroscedastic test can be used to check whether this assumption is valid.

Violation of these two assumptions may make the regression estimates meaningless. If one is interested in the distribution of the forecasts and not just the average value, then more assumptions will have to be made.

Multicollinearity. Another factor affecting regression forecasts is known as multicollinearity. When two or more of the independent variables used in a regression model are related, there is said to be multicollinearity. When there is multicollinearity, the estimates of the regression coefficients, or our measure of how the dependent variable is related to the independent variable, can be erroneous. Multicollinearity may be discovered by examining the coefficients of correlation between pairs of independent variables.

In using the regression method one has to ensure the validity of the assumptions mentioned above. This becomes more important when one is dealing with time series data where, for example, the assumption of no autocorrelation may not be true. There are ways to deal with the problem of autocorrelation such as using the generalized least squares. However, the Box-Jenkins' method is considered to be superior as it takes into consideration the problem of autocorrelation directly.

Box-Jenkins Method

The Box-Jenkins method involves three steps (Box and Jenkins, 1970; Pankratz, 1983):

- Identification
- Estimation
- Diagnostic checking.

Identification. The purpose of identification is to understand the pattern in the time series data in terms of stationary. If the data set is not stationary, it is made stationary with respect to mean and variance. If there is seasonality in the data, the length of seasonality is found.

Stationary with respect to mean implies that, when the data are plotted against time, they fluctuate around a constant mean independent of time. Similarly, data being stationary with respect to variance would imply that the variance in the data is invariant with respect to time.

The existence of stationarity in the data is

discovered by examination of the autocorrelation coefficients and partial autocorrelation coefficients.¹ If we find that, for example, autocorrelation coefficients $r_{xt, xt-p}$ computed for $p = 1, 2, 3, \dots$ decreases to zero quickly, it should indicate that the data are stationary in the original form. Otherwise it would suggest that the data are non-stationary. A non-stationary data set can be transformed into a stationary set by suitable differencing. For example, instead of considering the original time series data, one can deal with a series where the first element is the difference between the second element and the first element of the original series, the second element is the difference between the third element and the second element of the original series, and so on.

If the data are stationary with respect to mean and variance, one could then examine the autocorrelation and partial autocorrelation coefficients calculated for the data in order to identify the seasonality.

At the end of the identification step, it would be possible to know the pattern in the data. A tentative model could be formulated as follows:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + e_t - \phi_1 e_{t-1} - \phi_2 e_{t-2} - \dots - \phi_q e_{t-q}$$

where x_1, x_2, \dots is the time series data, e_1, e_2, \dots are the past residuals or errors in forecasting. The first part of the model is the autoregressive part and the second part is called the moving average part. This model is called ARIMA model of order (p,q).

Estimation. Having identified a tentative model, the next step is to estimate the parameters identified in the model. This is done by an iterative process. An initial set of values for the parameters is assumed. Iteratively an optimal set is obtained. The criterion for comparison of different sets of parameter values is the mean square error.

Diagnostic Checking. The third step in the Box-

¹ Partial correlation coefficient between any two variables x_1 and x_2 keeping another variable x_3 constant is the simple correlation coefficient between the two variables x_1 and x_2 , after removing the linear influence of the third variable x_3 . This is done by computing the correlation coefficient of the two residuals obtained from the linear regression of x_1 on x_3 and the linear regression of x_2 on x_3 .

² If we denote the time series as x_1, x_2, \dots, x_n , the autocorrelation of order p , $r_{xt, xt-p}$ is the correlation coefficient between two series x_1, x_2, \dots, x_{n-p} and $x_{p+1}, x_{p+2}, \dots, x_n$.

Jenkins method is diagnostic checking, i.e. checking whether the model represented accurately the underlying process in the given time series. This is done by checking the autocorrelation and partial autocorrelation coefficients of the residuals of the model. Another procedure for diagnostic checking is to try higher order ARIMA models and examine whether or not forecasts improve.

I describe below my test of the two methods to forecast milk supply to Amul.

Forecasting Milk Procurement by Amul Dairy

Amul Dairy is a milk processing/product manufacturing unit located at Anand in the state of Gujarat. Unlike most dairy plants that own their milch animals, Amul Dairy is a union of about 1,000 primary milk producer cooperative societies with over 100,000 producer members.

The procurement of milk by Amul varies over the year. One may ask why milk collected from a set of known members varies. There are reasons for this. Firstly, it is not obligatory for members to sell their milk to the dairy only. Members are free to transact with other agencies purchasing milk in the area. Secondly, and more importantly, milk supplies in most areas show significant seasonal variations. Peak supplies commonly occur during the winter months when they reach a volume normally two to three times that available in the summer, the months of minimal supply. Since demand for milk and milk products is fairly constant all round the year, the seasonal peaks in production/supply of liquid milk are converted to milk products and stored for reconstitution purposes during the lean months.

The seasonally in production and the consequent uncertainty in procurement makes forecasting an important aid to production planning and inventory management. In a typical dairy plant in India, milk constitutes nearly 80 per cent of the cost of production. The ability to accurately forecast milk supply has direct implications for the economic viability of the entire operations of the dairy plant.

The data series used for building the forecasting model is monthly milk receipts by Amul Dairy from January 1965 to December 1975. Only data till December 1974 were used to estimate the model. The remaining 12 observations were used

Table 1: Estimates Using Regression Models

Variable	OLS		GLS	
	Model-1	Model-2	Model-1	Model-2
Proc _(t-1)	0.95	—	0.93	—
Price _(t-1)	—	0.23	—	0.23
D1	51.21	825.56	75.81	745.63
D2	-81.29	645.43	-59.19*	595.67
D3	33.64*	617.09	52.96*	566.06
D4	-86.39	469.11	-67.34	416.74
D5	-88.44	328.54	-72.00	274.75*
D6	-88.83	194.03	-74.92	138.71*
D7	84.57	246.89	96.07	190.04*
D8	176.20	394.12	188.84	355.76*
D9	111.49	475.87	127.03	415.84
D10	176.56	625.16	193.59	563.47
D11	135.33	723.52	155.11	660.07
D12	132.18	813.53	153.87	748.11
RSQ	0.97	0.58	0.95	0.78
CRSQ	0.96	0.54	0.94	0.76
F-Stat	331.00	13.52	195.85	35.15
DW Stat	1.46	0.07	2.06	2.17

D1—D12 are the monthly seasonal dummies

* Insignificant at the 95% confidence limit.

to check the ex-post forecasting accuracy of the model. The price series data used were the price paid by Amul to its suppliers. Figure 1 shows the monthly milk procurement of milk during January 1965 to December 1975.

As can be seen from the graph in Figure 1, the data are not stationary. The average level of procurement has changed. In the periods from January 1965 to March 1969 the average was slightly above 5 million kg per month while the average during March 1969 to December 1975 is around 10 million kg per month.

Regression Method

Several regressions were run using slightly different specifications of the same basic model until one that gave the best fit was reached. An examination of the raw data (Figure 1) reveals a distinct seasonal pattern as well as a trend. Two options were tested to capture seasonality—one was of using seasonal dummies and the other was of using lagged dependent variables (lagged by 12 periods since the data are monthly). Seasonal dummies gave a better goodness of fit.

Two regression models were estimated—one using procurement in the previous month and the other using price in the previous month* as the independent variable. They were:

Model-1: $\text{Proc}_t = f(\text{Proc}_{t-1}, \text{Seasonal Dummies})$

Model-2: $\text{Proc}_t = f(\text{Price}_{t-1}, \text{Seasonal Dummies})$

where Proc_t denotes procurement in period t .

Since there is seasonality in the data, seasonal dummies were introduced to capture seasonality. In Model-1 procurement in any month is a function of procurement in the previous month and the seasonality variable corresponding to that month. In Model-2, procurement price in the previous month is used as the independent variable. No constant term is used in the models since 12 seasonal dummies have been used.

Both models under the ordinary least squares method had autocorrelation problems since the Durbin-Watson (DW) statistic was very low (see Table 1).

The parameters in the two models were reestimated using the Generalized Least-Squares (GLS), which takes into account the existence of autocorrelation. As is evident from the DW statistic in Table 1, GLS was successful in fixing the autocorrelation problem. While Model-1 was autoregressive of order 1, Model-2 was autoregressive of order 2. Model-1, with the lagged dependent variable, performed much better than Model-2, in terms of goodness of fit, in both the ordinary least square estimates (OLS) and generalized least square (GLS) estimates (Table 1). Figure 2 shows actual and forecast milk procurement using OLS. Figure 3 shows actual and forecast milk procurement using GLS.

Box-Jenkins Method

How do the forecasts using the Univariate Box-Jenkins time series (UBJ-ARIMA) models compare with those of regression models?

Identification. The first step in identification is to see whether or not the data series is stationary. Examination of autocorrelation and partial autocorrelation functions indicated the non-stationarity of the data with respect to mean and the presence of seasonality. The heteroscedastic test rejected the null hypothesis of homoscedasticity.

The stationary series in backshift notation was of the form:

$$(1-B)(1-B^{12}) \ln(\text{Proc}_t) = W_t$$

Figure 1. Monthly Milk Procurement by Amul
Jan 1965 to Dec 1975

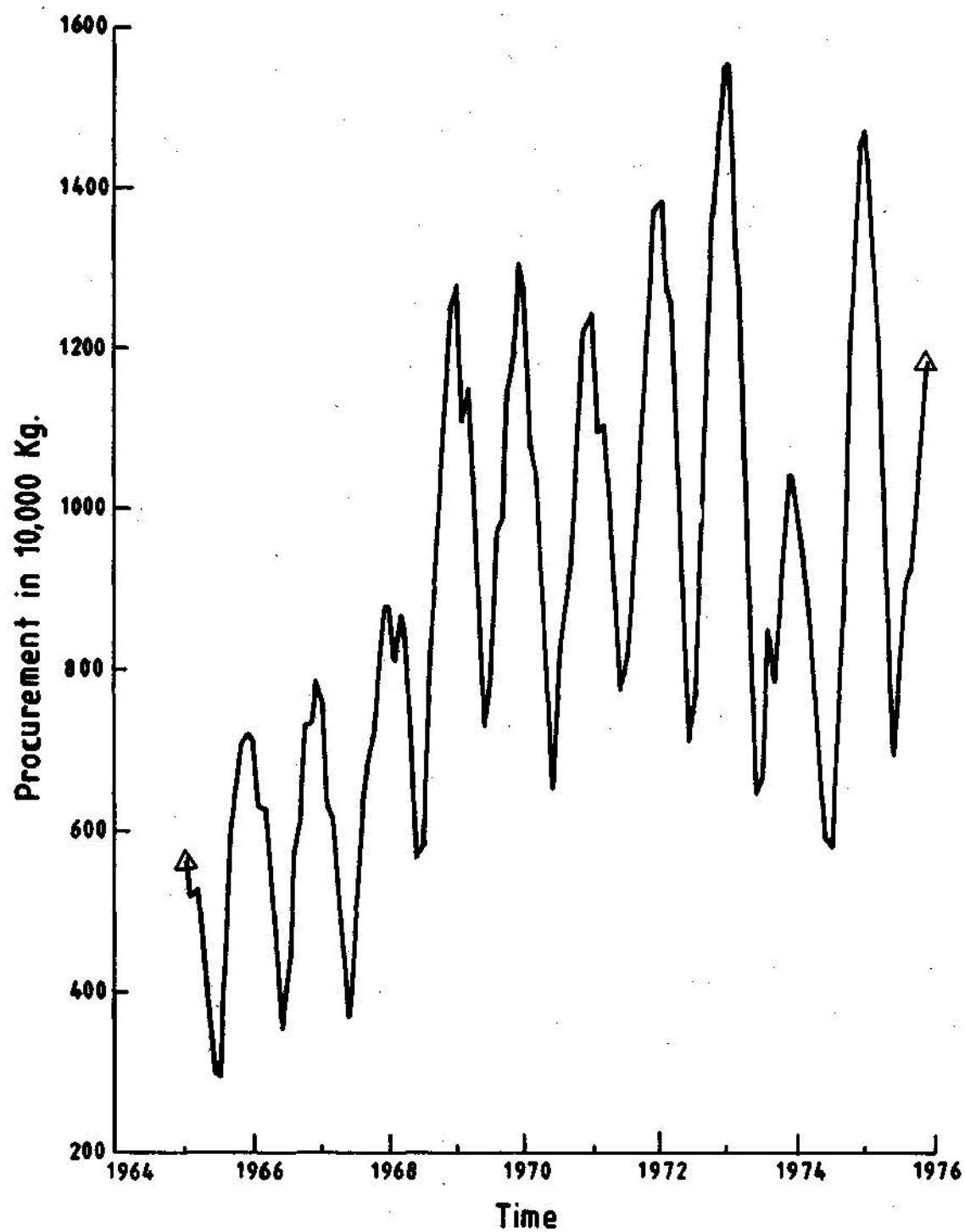


Figure 2 Actual and OLS Forecasts of Milk Procurement
Feb 1965 to Dec 1975

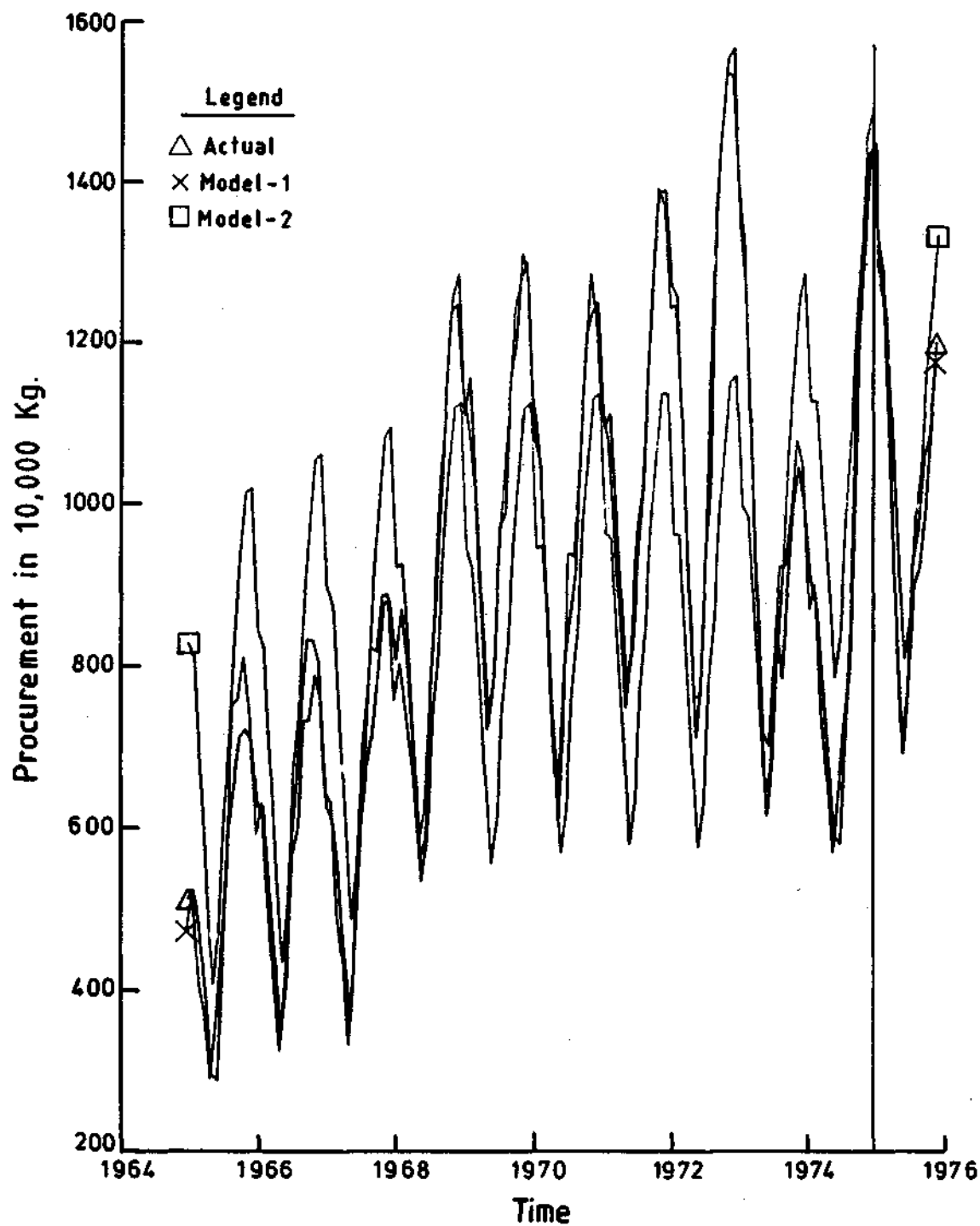
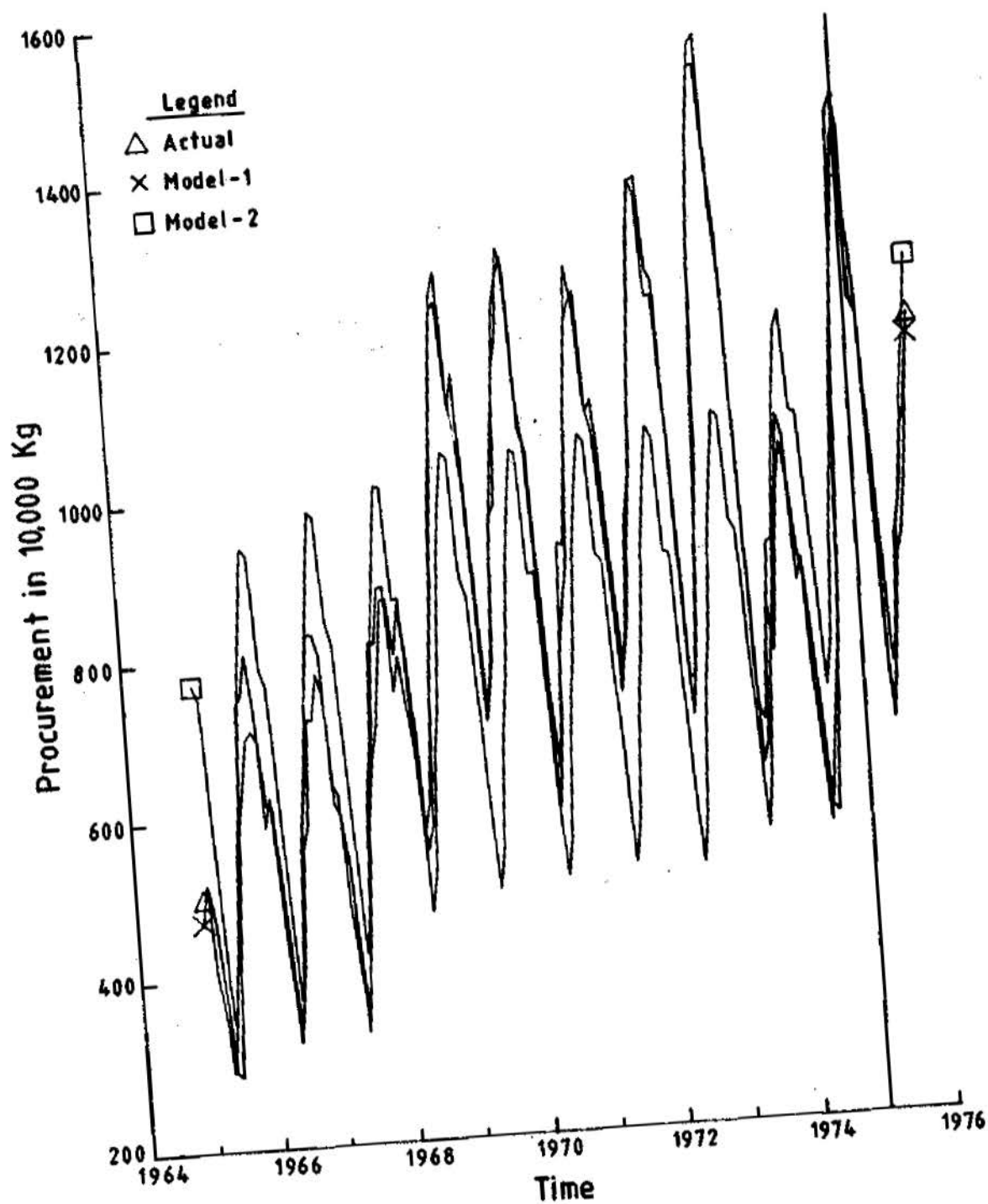


Figure 3. Actual and GLS Forecasts of Milk Procurement
Feb 1965 to Dec 1975



where W_t is the stationary series, $(1-B)$ is the first difference and $(1-B^{12})$ is the twelfth order seasonal difference. Having obtained a stationary series, the autocorrelation function (ACF) and partial autocorrelation functions (PACF) of the stationary series were then used to identify the appropriate ARIMA models. The ACF of the transformed and differenced series indicated a pure seasonal model of moving average of order 12. Similarly, the PACF also indicated a pure seasonal autoregressive model of orders 12 and 24. Thus, the two models identified were:

ARIMA MODEL-1

$$(1-B)(1-B^{12}) \ln(\text{Proc}_t) = (1 - \Delta_1 B^{12}) a_t$$

ARIMA MODEL-2

$$(1-B)(1-B^{12})(1-r_1 B^{12}-r_2 B^{24}) \ln(\text{Proc}_t) = a_t$$

where Δ = MA 12 coefficient
 r_1 = AR 12 coefficient
 r_2 = AR 24 coefficient
 a_t = residual.

Estimation. The values of the parameters were estimated for the identified models.* They are shown in Table 2, which gives a summary of the two ARIMA models.

Table 2: Summary of the two ARIMA Models

	Type	Order	Coefficient	T-Ratio	RMS
Model-1	MA	12	0.887	37.44	0.0032
Model-2	AR	12	-0.882	-14.89	0.0039
	AR	24	-0.654	-10.75	

The constant terms were excluded in estimation, since the series had already been differenced. All the estimates in the two models were significant.

Diagnostic Checking. Checking whether the models identified represented accurately the underlying process of the given time series was done using the ACF and PACF of the residuals. There was no discernible pattern in their ACF and PACF, meaning that the residuals were "white noise." The models represented the given series.

* A list of some software vendors of Box-Jenkins programs is given in Appendix 1.

Appraisal of the Two ARIMA Models

The two ARIMA models satisfy all the criteria for good models. Since the purpose of ARIMA models is forecasting, the most important criterion for the choice of a model is small forecast error. Based on this criterion, ARIMA Model-1 should be preferred to ARIMA Model-2.

Multivariate Time Series Models

One of the problems of UBJ-ARIMA models is that of background noise caused due to trend, seasonality, and outliers. This problem of background noise can be reduced by incorporating appropriate independent variable time series into the model. It has been accepted that a 'fair' multivariate model is always preferred to a 'good' univariate model (McCleary and Hay, 1980).

Evaluation of the Models

How can we test the efficacy of the various models in forecasting?

Two tests were used to test the two models: the Root Mean Square (RMS) per cent error and Theil's U test (Theil, 1966).

The root mean square per cent error of the ex-post forecasts is defined as:

$$\text{RMS per cent error} = \sqrt{\frac{1}{T} \sum_{t=1}^T \left(\frac{\text{Proc}_t^s - \text{Proc}_t^a}{\text{Proc}_t^a} \right)^2}$$

where Proc_t^s is procurement during month t as estimated by the model and Proc_t^a is the actual procurement during month t .

RMS per cent error should be as low as possible.

Theil's inequality coefficient (Pindyck and Rubinfeld, 1981) to evaluate ex-post forecasts is defined as:

$$U = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^T (\text{Proc}_t^s - \text{Proc}_t^a)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^T (\text{Proc}_t^s)^2} + \sqrt{\frac{1}{T} \sum_{t=1}^T (\text{Proc}_t^a)^2}}$$

U always ranges between 0 and 1. A zero value indicates a perfect fit. A value of one for U implies that the predictive performance of the model is as bad as it possibly could be.

Results

How did the various models fare on forecasting milk procurement by Amul Dairy? The performance of the various models on the two tests is given in Table 3.

Table 3: Performance of Various Forecasting Models on Two Tests

Model	RMS Per Cent Error	Theil's U
Regression Method		
OLS		
Model-1	2.20	0.20
Model-2	14.00	0.48
GLS		
Model-1	4.14	0.207
Model-2	7.05	0.35
Box-Jenkin's Method		
Model-1	4.63	0.225
Model-2	5.75	0.317

The reverse causality that exists between price and procurement cannot be captured in single equation models. That is why the univariate models such as Model-1 in the regression method performed better in our example. Model-1 in the Regression method using a lagged dependent variable performed much better than Model-2, which used price as an independent variable. Model-1 in OLS did display autocorrelation in its residuals but not in GLS. UBJ-ARIMA Model-1 performed worse than Model-1 in the OLS regression method. Its performance was closer to Model-1 in the GLS regression method, because GLS adjusts for autocorrelation just as ARIMA looks for stationarity. Low RMS per cent error is only one desirable measure of simulation fit. Another important criterion is how well the model simulates the turning points in the time series data. On this criterion Model-1 in both the regression and the Box-Jenkins methods performed better than the others.

The turning point prediction is critical to good forecasting. Managers should examine all the underlying factors in judging a forecast, especially when they feel that structural changes have taken place or are likely to take place. This should be handled outside of the model chosen as well as in the formulation of the model itself.

In the example, the Box-Jenkins method did

not perform as well as the regression method. This may be partly due to the nature of the data and the nature of the underlying factors. A lagged linear model proved better than the Box-Jenkins method which examines simultaneously various combinations of the underlying factors. Model builders should choose the method most appropriate to the data and the situation.

References

- Box, G E P and Jenkins, G M (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- McCleary, R and Hay, R A (.1980). *Applied Time Series Analysis for the Social Services*. California: Wadsworth.
- Pankratz, Allan, (1983). *Forecasting with Univariate Box-Jenkins Model*. New York: John Wiley.
- Pindyck, R S and Rubinfeld D I (1981). *Econometric Models and Economic Forecasts*. New York: McGraw-Hill.
- Theil, H (1966). *Applied Economic Forecasting*. Amsterdam: North-Holland.

Appendix 1: Software Vendors of Box-Jenkins Programs

Organization	Address	Program Name(s)	Type of Models	Batch/ Conversational
Applied Decisions Systems, Inc.	33 Hayden Avenue Lexington, MA 02173	SIBYL/RUNNER	Univariate and Multivariate	Conversational
Automatic Forecasting Systems, Inc.	P.O. Box 563 Hatboro, PA 19040	PACK System and AUTOBJ	Univariate and Multivariate	Conversational and Batch
Gwilym Jenkins & Partners, Ltd.	1700 Echo Trail Norman, OK 73069	GENISIS	Univariate and Multivariate	Batch
IBM Corporation	Data Processing Division 1133 Westchester Avenue White Plains, NY 10604	APL Forecasting and Time Series Analysis	Univariate and Multivariate	
Charles R. Nelson Associates, Inc.	4921 N.E. 39th St. Seattle, WA	PDQ, et al.	Univariate and Multivariate	Conversational and Batch
SAS Institute, Inc.	Box 8000 Cory, NC	SAS	Univariate and Multivariate	Conversational and Batch
Scientific Computing Associates, Inc.	P.O. Box 625 Dekalb, IL 60115	The SCA System	Univariate and Multivariate	Conversational and Batch
Statistical Laboratory, Iowa State University	C/O Bill Meeker Route 1 Ames, IA 50010	TSERIES	Univariate	Batch
BMDP Statistical Software, Inc.	1964 Westwood Blvd. Suite 202 Los Angeles, CA 90025	BMDP	Univariate Multivariate	Conversational and Batch

Source: John C. Hoff (1983). *A Practical Guide to Box-Jenkins Forecasting*, California : Wadsworth.