

USE OF ARIMA TIME SERIES AND REGRESSORS TO FORECAST THE SALE OF ELECTRICITY

Beatrice Ugiliweneza, University of Louisville, Louisville, KY

ABSTRACT

Objectives: To forecast the sales made by an electric company every month. To show the use of ARIMA and Regression models in forecasting.

Method: The data used to forecast were from the total sale of electricity for each end of the month from 1994 to 2004. The statistical forecasting method used is the ARIMA time series with the regression model.

Results: With a seasonable ARIMA model, a regressor and a dynamic regressor, the model predictions compare with the actual values of sales and hence, the forecast values are reliable.

Conclusions: ARIMA time series are useful models to predict the sales of electricity for this company. From this study, we can conclude that ARIMA and Regression models can be used by other businesses for planning.

INTRODUCTION

Generally, companies focus on the forecasting of sales because they need to plan their expenses and still make a profit. This paper shows how ARIMA time series and Regression models can be used to forecast company sales. A model is used to forecast one year ahead of the total sale of electricity from a sample of ten years available (1994-2004), with the data provided monthly. As a result, the predicted values of the 10 years (1994-2004) compare well enough, with the actual sales values. Thus, the forecast values are reliable. The application can be easily extended to other selling companies. We use these data to demonstrate the use of the SAS Time Series Forecasting System. While the Forecasting System can automate the analysis of data, the investigator needs to interact with the system to find the most effective forecasts. We also show how inflation can be added as a dynamic regressor.

METHOD

The data were collected on a monthly basis, from the total sales of an electric company for ten years (1994-2004). A sample of 128 data points was obtained. The data were provided by the original investigator. The tool used is the SAS Time Series forecasting system. This system is a SAS point-and-click interface that provides automatic model fitting and forecasting as well as interactive model development. The system provides the best fitting model for each time series. We can use system features to identify series behavior, fit candidate forecasting models, and perform diagnostic checks on the fitted models.

To get the SAS Time Series forecasting system, we start with the window shown in Figure 1.

Figure 1. Entry Screen for Time Series Forecasting System

After inputting the data, the Develop model is chosen. After this, the following window (Figure 2) enables us to construct a model by right-clicking in the white area.

Figure 2. Developing the Model

Plots and results will be explored with interactive graphical tools.

The data were stationary and they had a fixed constant for the mean (the average) and the variance. For this reason, the ARIMA time series model was used. Figure 3 shows that the sale of electricity tends to be seasonal. Hence, the seasonal ARIMA model was used.

Figure 3: Plot of total sale of electricity

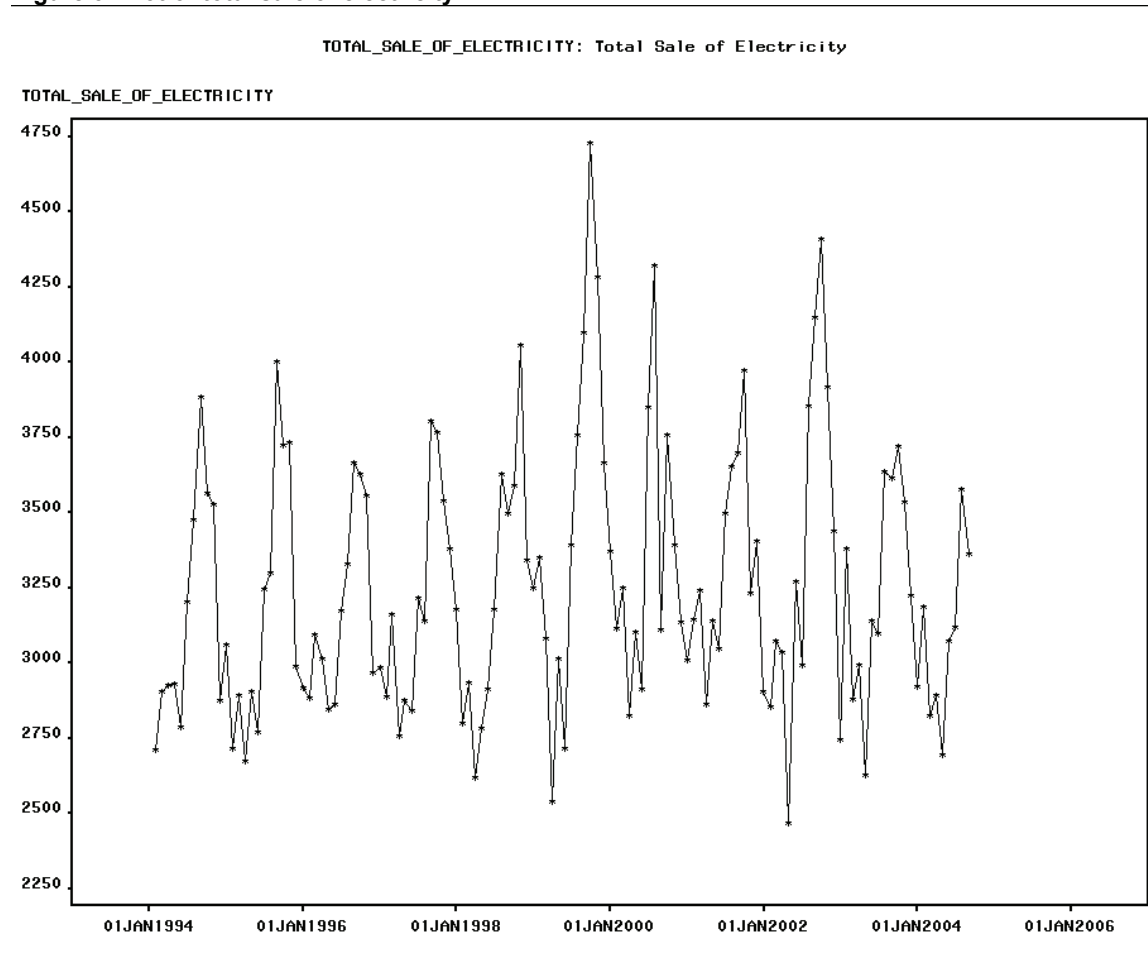
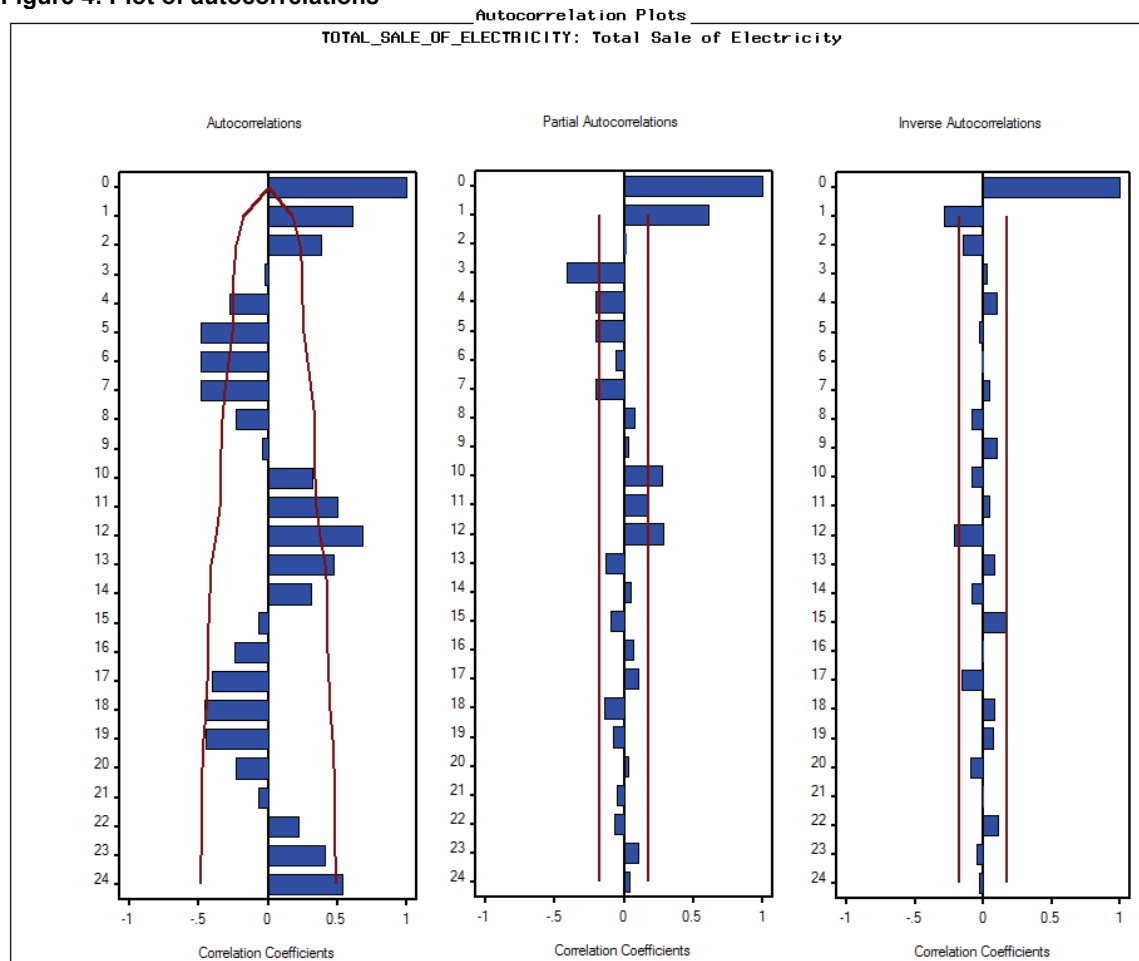


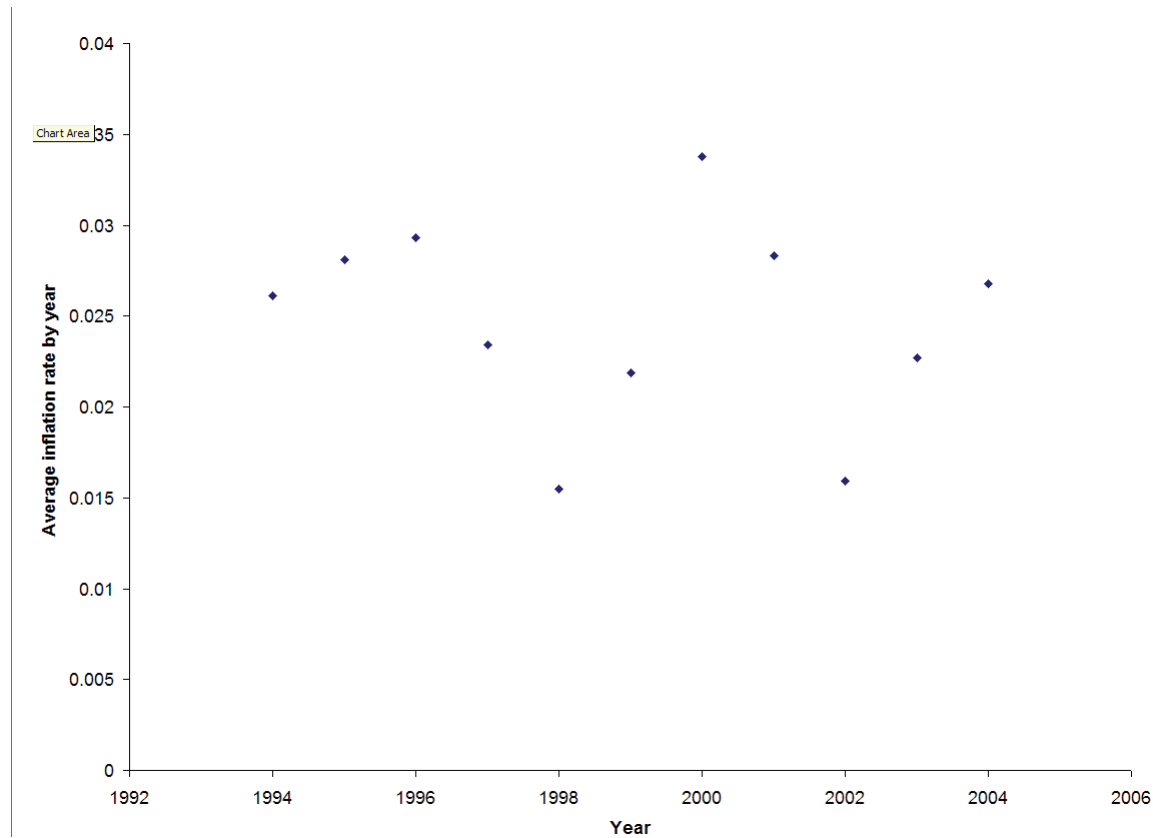
Figure 4 shows that the ACF (Autocorrelations function) is spiked at lag 1 and declines toward zero. It also shows that the PACF (Partial autocorrelation function) is spiked at lag 1 and is zero at lag 2. Considering this, the ARIMA $(1, 0, 0)_s$ was chosen.

Figure 4: Plot of autocorrelations



The sales of electricity should depend on the total electric usage. For this reason, the total electric usage was chosen as a regressor. Moreover, the results of the sales are affected by the inflation rate, which is a variable, non-constant regressor. Thus, the inflation rate was added to the model as a dynamic regressor. With this dynamic regressor, a numerator factor with an order 1 is specified, which means that the inflation rate starts at one month. These inflation rate data were found on the historical inflation data website, inflationdata.com. Figure 5 shows the graph of the mean value of the inflation rate per year for ten years, from 1994 to 2004.

Figure5: Plot of the mean value of the inflation rate by year from 1994 to 2004



Briefly, the model used is:

Total sale of electricity=Inflation rate [N (1))] +Total electric usage +ARIMA (1, 0, 0)₁₂

This model of ARIMA and the regressors was chosen because it yields a better model than the ARIMA alone. In fact, the following tables show that the mean absolute percent error of the ARIMA + Regressors, which is equal to 5.96, is smaller than the mean absolute percent error of the ARIMA alone, which is 6.82.

Table1: Statistics of fit of the use of the model: Total sale of electricity=Inflation rate [N (1))] +Total electric usage + ARIMA (1, 0, 0)₁₂

Statistics of Fit	
TOTAL_SALE_OF_ELECTRICITY: TOTAL_SALE_OF_Electricity Inflation_rate[N(1)] + Total_Electric_Usage + ARIMA(1,0,0)s	
Statistic of Fit	Value
Mean Square Error	67331.3
Root Mean Square Error	259.48284
Mean Absolute Percent Error	5.96220
Mean Absolute Error	197.39628
R-Square	0.633

Table 2: Statistics of fit of the use of the model: ARIMA (1, 0, 0)₁₂

Statistics of Fit	
TOTAL_SALE_OF_ELECTRICITY: TOTAL_SALE_OF_Electricity	
ARIMA(1,0,0) _s	
Statistic of Fit	Value
Mean Square Error	94480.8
Root Mean Square Error	307.37734
Mean Absolute Percent Error	6.82167
Mean Absolute Error	228.42803
R-Square	0.487

RESULTS

The working series contains 128 sales collected in ten years (1994-2004) on a monthly basis and among these, a holdout sample of 60 sales on which the model of forecasting is built. These sales have a mean of \$3250.62 and a standard deviation of 429.13. The values of the sales are in thousands of dollars. The Inflation rate is used as a predictor in the model

[N (1)] + Total electric usage + ARIMA (1, 0, 0)₁₂

We use it to give predicted values for the ten years (1994-2004) and forecast values for one year ahead (2005).

Figure 3 shows that the company sells more electricity in the period at the end of the year. The best sales of electricity were obtained at the end of 1999. Figure 6 shows predicted values obtained with the use of the Inflation rate with the above model,

Figure 6: Plot of values

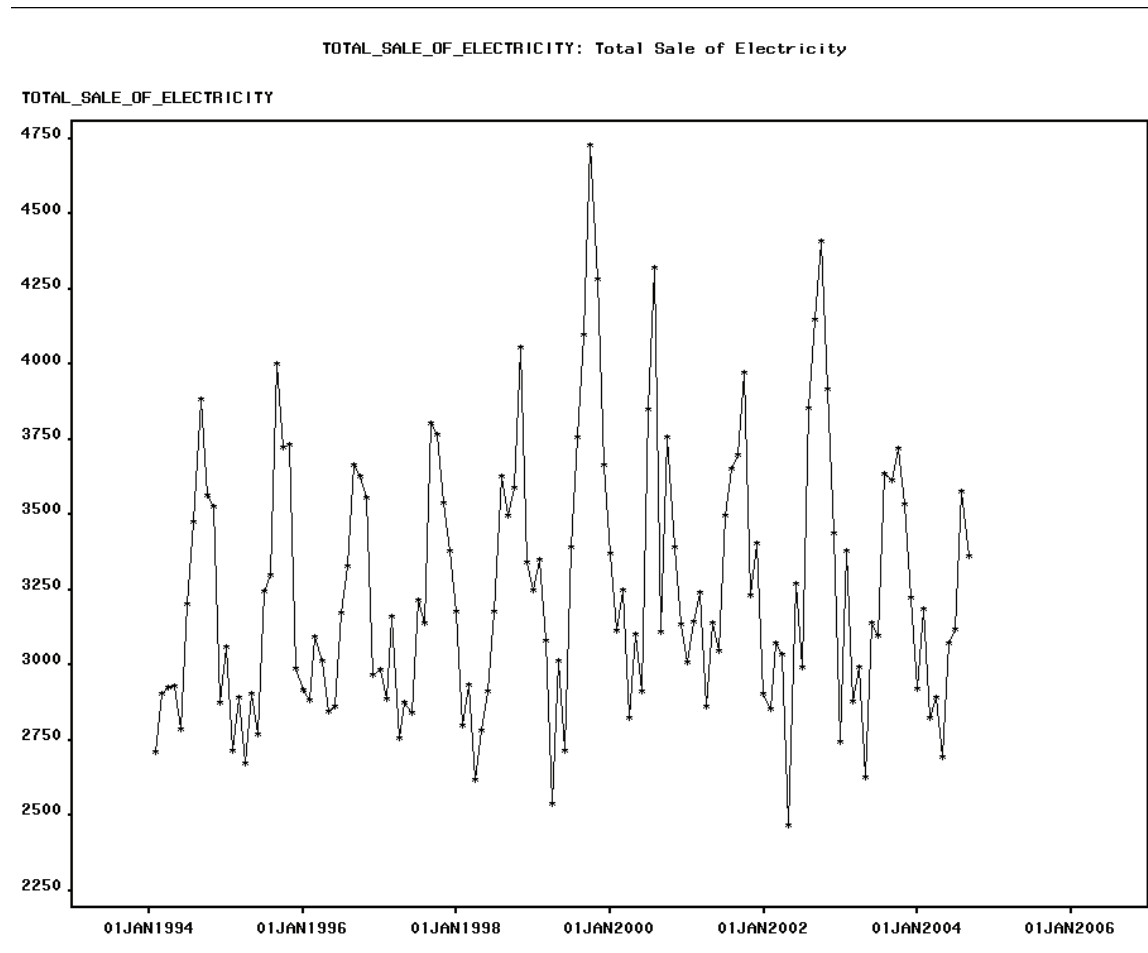


Figure 6 shows better sales in the period at the end of the year with the best sale at the end of 1999.

Figure 7 shows that most of the differences between actual values and the predicted values are between -600 and 600 (these values are in thousands of dollars). Except for two values, the absolute value of the errors between the actual and the predicted values is at most 600. The actual values gave a mean of \$3250.62; so, this error, which is 18.46 percent of the mean value, is very small. This proves that the Inflation rate $[N(1)]$ + Total electric usage + ARIMA (1, 0, 0)₁₂ model predicted well. Therefore, the forecasts of this model are reliable.

Figure 7: Plot of predicted values

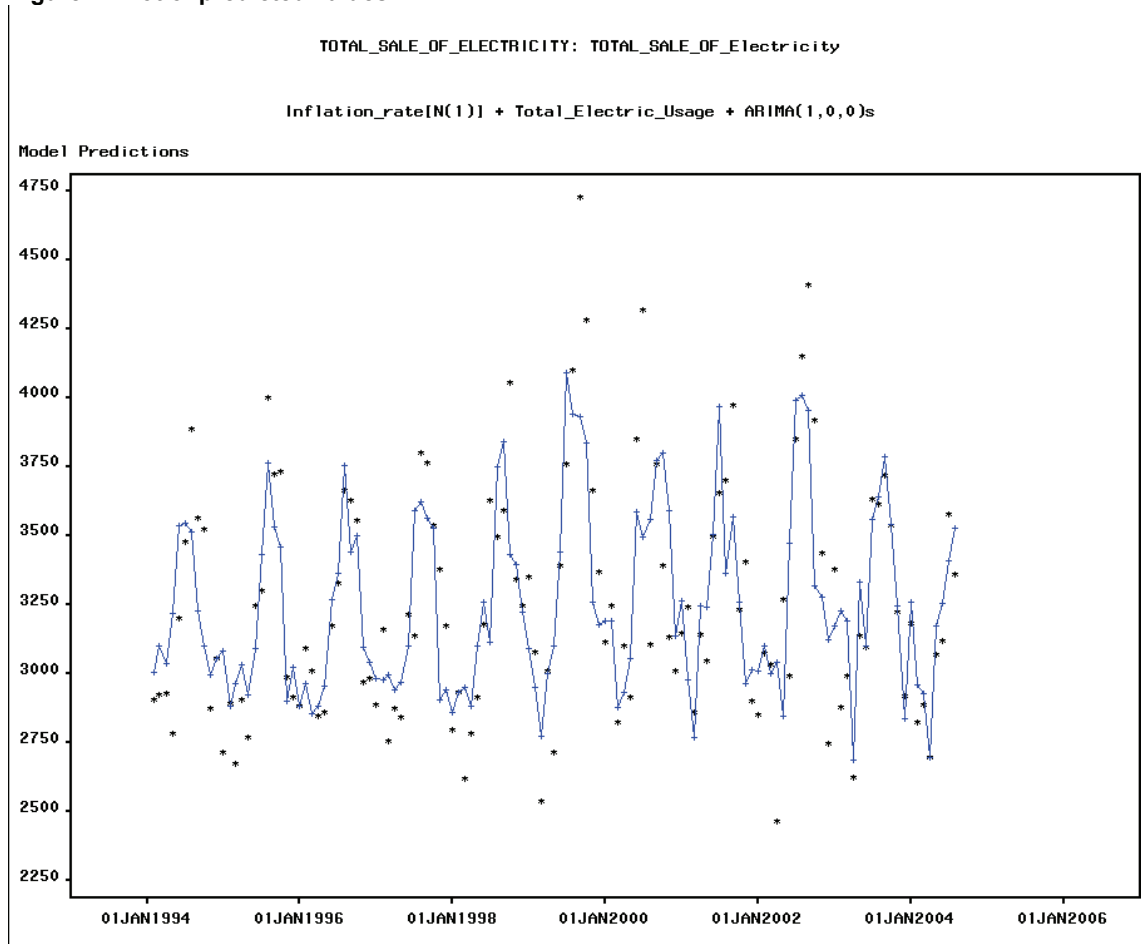
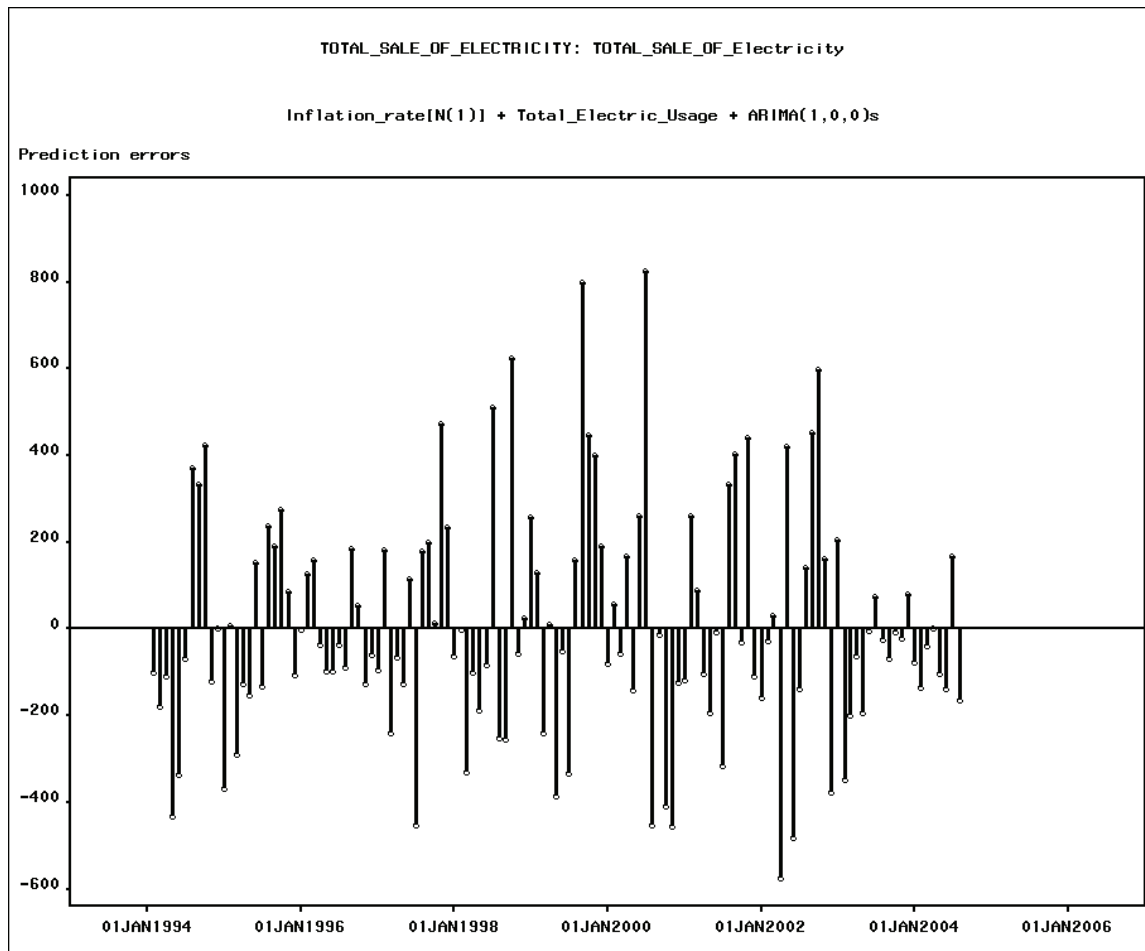


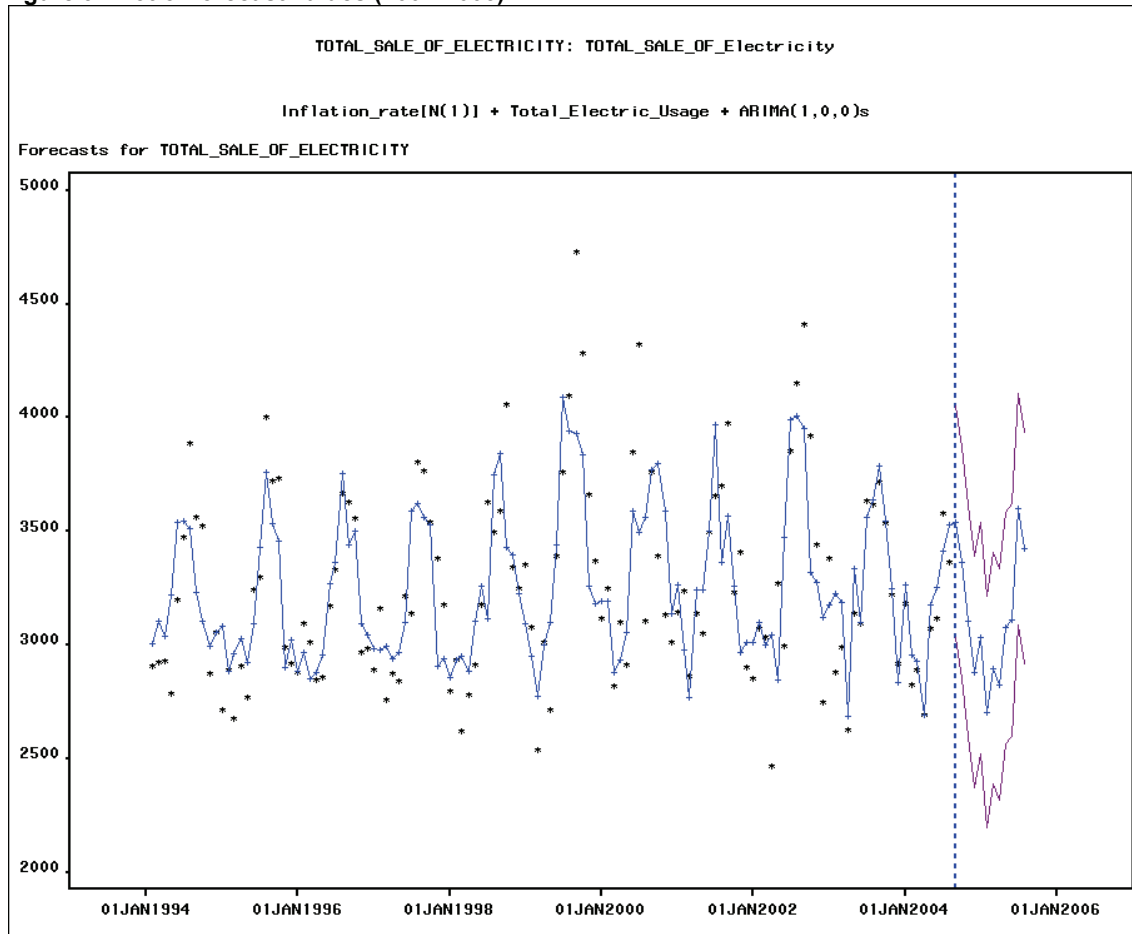
Figure 8 shows that the forecasts are between 2000 and 4250 (with a 95% confidence interval) with predicted high sales at the end of 2004 and improved sales at the end of 2005. The inflation rate [N (1))] +Total electric usage +ARIMA (1, 0, 0)₁₂ predicts a seasonal sale for 2005 and gives the likelihood interval of values.

Figure 8: Plot of prediction errors



The prediction is given in Figure 9.

Figure 9: Plot of forecast values (2004-2005)



The plot above gives the forecast values for the year 2005. The first part of the graph represents the predicted values of the years 1994 to 2004 (these compare with the actual values) and the second part gives the values and the behavior of the forecasts with a 95% confidence interval.

Table 3 gives the forecast values of the total sale of electricity for one whole year ahead. These are eleven values on which the electric company can rely on to plan the business for 2005. (U95: upper limit of the 95%confidence interval, L95: lower limit of the 95% confidence interval)

Table3: Forecast data set from 01 October 2004 to 01 August 2005

Forecast Data Set							
TOTAL_SALE_OF_ELECTRICITY: TOTAL_SALE_OF_Electricity							
Inflation_rate[N(1)] + Total_Electric_Usage + ARIMA(1,0,0)s							
DATE	PREDICT	U95	L95	ERROR	NEERROR	Inflation_rate	Total_Electric_Usage
01OCT2004	3364	3873	2855	.	.	0.0216	3662
01NOV2004	3107	3616	2598	.	.	0.0216	3350
01DEC2004	2880	3389	2371	.	.	0.0216	3327
01JAN2005	3033	3542	2524	.	.	0.0216	3469
01FEB2005	2706	3215	2197	.	.	0.0216	3103
01MAR2005	2899	3408	2390	.	.	0.0216	3362
01APR2005	2827	3336	2318	.	.	0.0216	3305
01MAY2005	3077	3585	2568	.	.	0.0216	3709
01JUN2005	3111	3620	2602	.	.	0.0216	3905
01JUL2005	3598	4107	3089	.	.	0.0216	4324
01AUG2005	3426	3935	2917	.	.	0.0216	4321

*/

CONCLUSIONS

This study demonstrates how ARIMA time series and Regression models are useful to study and forecast sales for a particular company. This paper demonstrates also how the Time Series Forecasting System can be used to construct a model of forecasting. The Inflation rate $[N(1)]$ + Total electric usage + ARIMA $(1, 0, 0)_{12}$ predicted the data considerably well and gave reliable forecasts. According to the data presented, this model was best in forecasting the sales, but could not tell why the sales will contain outliers. The SAS Time Series forecasting system helped construct a model, the ARIMA time series and the Regression, which is effective for forecasting and can be applied to other businesses in order to plan their sales. However, it would be interesting to do further research on the factors that influence the sales, such as the growth of the population of consumers, the industrial growth in the region, the immigration, and so on; this would consolidate better this company's planning.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Beatrice Ugiliweneza

Department of Mathematics

University of Louisville

Louisville, KY 40292

E-mail: b0ugil01@louisville.edu

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.