# The Use of ARIMA Model Approach in Marketing – Application to Sales Forecasting

**Juznik Rotar, L.**
Higher Education Centre Novo mesto, School of Business and Management, Na Loko 2, 8000 Novo mesto, Slovenia

Corresponding author E-mail: laura_juznik@yahoo.com

**Abstract:** *Sales forecasting is becoming more important in everyday businesses. Good forecasting models can increase efficiency of businesses, they are saving money on excess inventory, increase profit and serve its customers better. Sales forecasting is a crucial part of the financial planning of a business. This paper presents a forecasting technique to model quarterly sales of a departmental store using well known ARIMA model to analyze and forecast time series.*
***Keywords**: forecasting, ARIMA model, sales.*

## 1. INTRODUCTION

Sales forecasting is a difficult area of management. Sales forecasting uses past figures to predict short-term or long-term performance. It is a very demanding job, because so many different factors can affect future sales: economic downturns, employee turnover, changing trends and fashions, increased competition, manufacturer recalls and other factors. But there are several methods that can produce consistently accurate sales forecasts. One of them is ARIMA forecasting technique, which we apply in empirical research.

## 2. EMPIRICAL APPLICATION

The problem of this research is to model the quarterly sales of a departmental store to be able to use it for forecasting. Quarterly sales data for the period 2006–2010 is used. For the modeling purposes we use ARIMA method according to the Box and Jenkins (Yaffee, 1999). The general ARIMA method is formulated as following:

$$\phi(B)s_t = \Phi(B)\varepsilon_t \qquad (1)$$

Where $s_t$ is the sale at time t, φ(B) and Φ(B)ε are functions of the backshift operator B:$B_s^1$ =$s_{t-1}$ and $\varepsilon_t$ is the error term. ARIMA model types are listed using standard notation of ARIMA (p,d,q) and (P,D,Q) are their seasonal counterparts (IBM SPSS, 2010; Jakasa, Androcec and Sprcic, 2011; Choong, 2010).

- Autoregressive (p): the number of autoregressive orders in the model. Autoregressive orders specify which previous values from the series are used to predict current values.
- Difference (d): specifies the order of differencing applied to time series before estimating models. Differencing is necessary when trends are present (series with trends are typically nonstationary and ARIMA modeling assumes stationarity) and is used to remove their effect. The order of differencing corresponds to the degree of series trend-first-order differencing accounts for linear trends, second-order differencing accounts for quadratic trends, and so on.
- Moving average (q): the number of moving average orders in the model. Moving average orders specify how deviations from the series mean for previous values are used to predict current values.
- The Expert Modeler within Software tool SPSS is used to automatically finds hte best-fitting model for each dependent series. If independent (predictor) variables are specified the Expert Modeler selects for inclusion in ARIMA models those that have a statistically significant relationship with the dependent series. Expert Modeler considers seasonal models. This option is only enabled if a periodicity has been defined for the active dataset. The Expert Modeler also includes a constant in model.

This time series data, namely quarterly sales data, as many macroeconomic time series are integrated or non stationary. To prepare data for statistical modeling, series are transformed to stationarity either by taking the natural log, by taking a difference, or by taking residuals from a regression. Applying sequence chart (Figure 1) shows the presence of a periodic date component in the active dataset – quarterly periodicity of 4 (quarters).

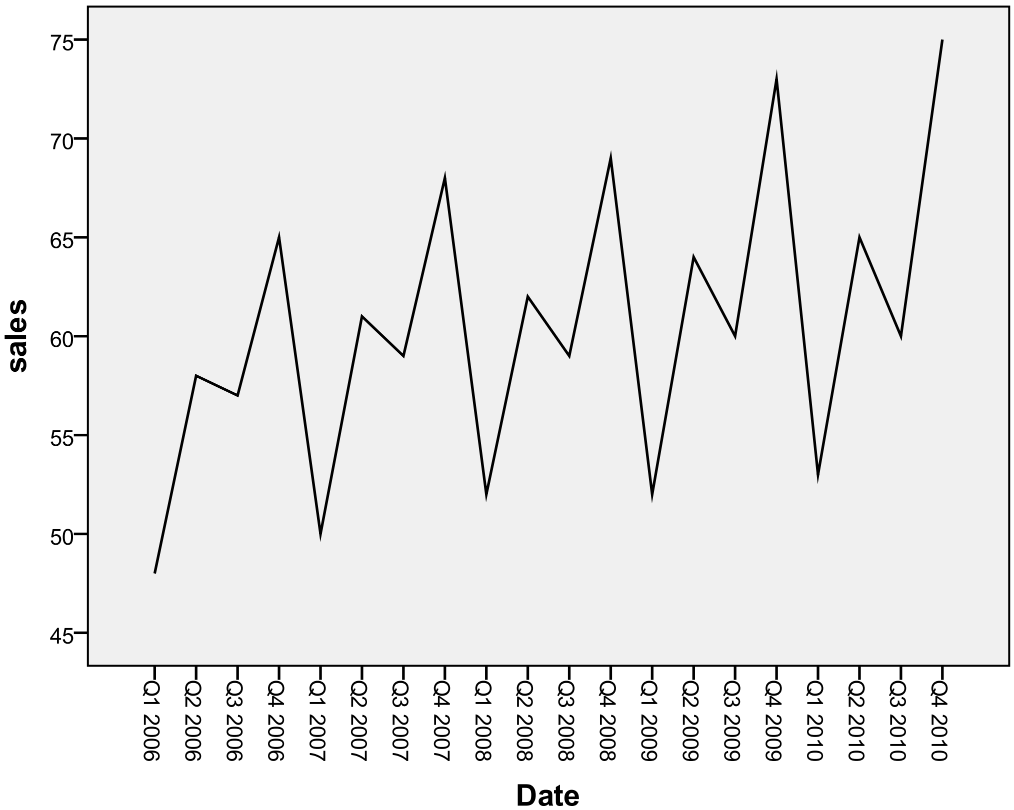


**Figure 1:** Quarterly sales of a departmental store
Source: own calculations.

The series exhibit a number of peaks, which appear to be equally spaced. Given the seasonal nature of sales, with typical highs during the December holiday season, the time series probably has an quarterly periodicity. The seasonal variations also appear to grow with the upward series trend, suggesting that the seasonal variations may be proportional to the level of the series, which implies a multiplicative model rather than an additive model.

After preparing dataset for software tool SPSS, we analyze the main characteristics of the time series (trend, cycle, season) and find out that this time series is a non stationary so we perform differencing transformation. The autocorrelation and partial autocorrelation functions are used, as basic instruments to identify stationarity of time series. Figures 2 and 3 show the ACF and PACF functions of the first-order differenced transformed sales data for the initial dataset (20 observations). The partial autocorrelation function in figure 3 shows one coefficient as significantly non-zero, suggesting that this could be AR(1) model for the data.

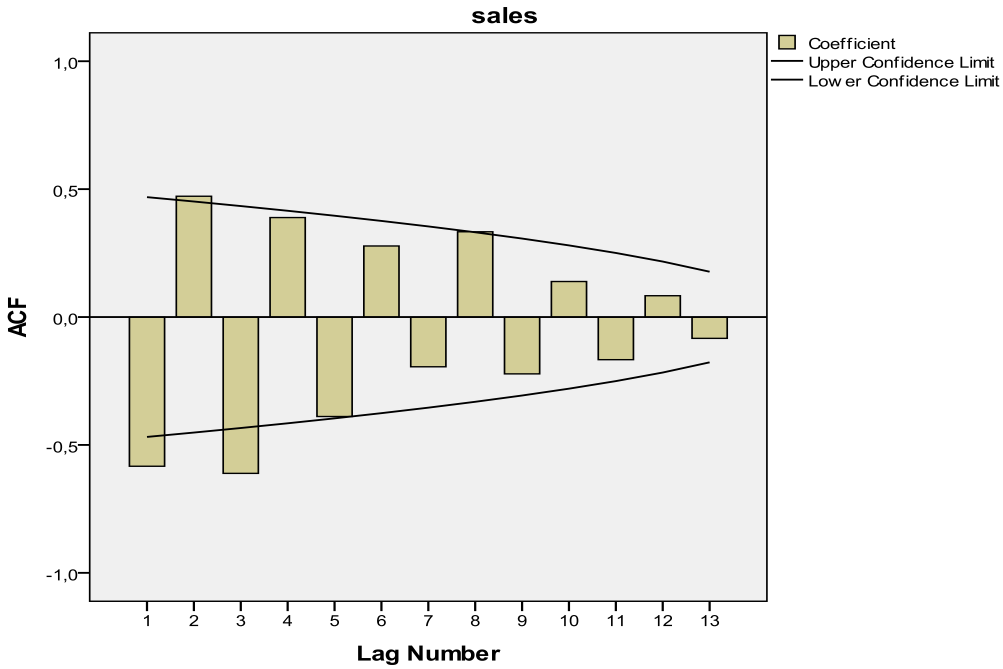


**Figure 2:** Autocorrelation plot for a first-order differenced sales data
Source: own calculations.

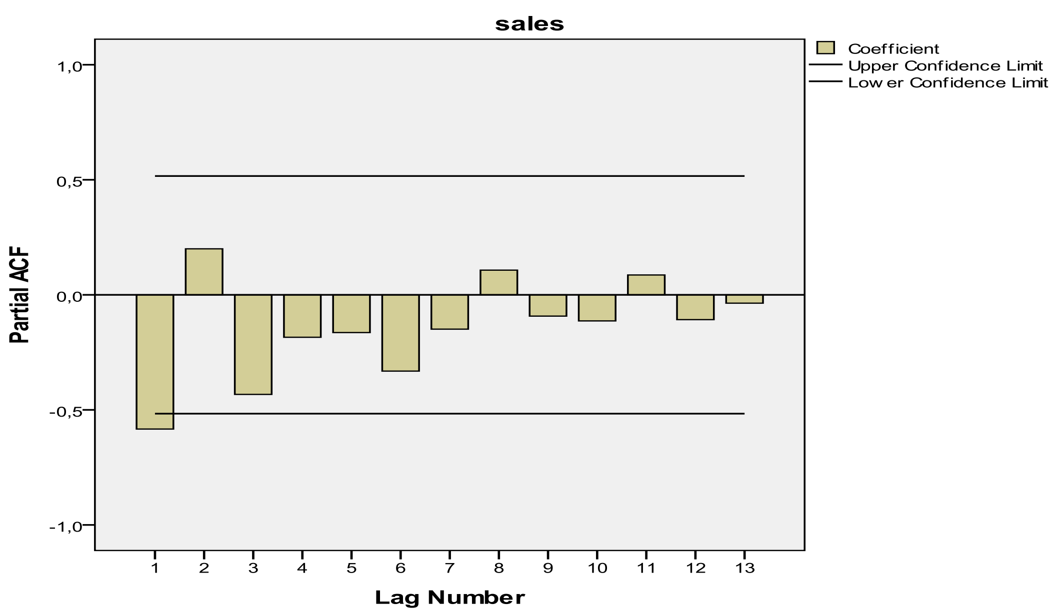


**Figure 3:** Partial autocorrelation plot for a first-order differenced sales data
Source: own calculations.

The Expert Modeler has been applied to predict quarterly sales of a departmental store. Therefore, we have quarterly sales for the period 2006 – 2010. The dataset (20 observations) has been divided in two parts: first part (12 observation) for building a model and second part (8 observations) for testing the model. Using the Expert Modeler we found the best ARIMA model for the dataset, it is ARIMA (1,1,0) (0,1,0), consisting of non seasonal and seasonal parts. Consequently, we apllied the same ARIMA model to the second dataset (8 observations) to see the goodness of fit. The Expert Modeler is also applied to the second set and statistics of goodness of fit have been compared. An autoregressive order of 1 specifies that the value of the series one time period in the past be used to predict the current value. The order of differencing corresponds to the degree of series trend – first-order differencing accounts for linear trends.

Table 1 shows goodness of fit statistics for the first data set. R-squared represents an estimate of the proportion of the total variation in the series that is explained by the model. Largest values (up to a maximum value of 1) indicate better fit. A value of 0,974 means that the model does an excellent job of explaining the observed variations in the series. Mean percentage error (MAPE) for the model is 1,23%. A measure of how much a dependent series varies from its model-predicted level. Maximum absolute percentage error (MaxAPE) represents the largest forecasted error, expressed as a percentage. This measure is useful for imagining a worst-case scenario for your forecasts.

**Table 1:** Model statistics – building the model

**Model Fit**

| | | | | | Percentile | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Fit Statistic | Mean | SE | Minimum | Maximum | 5 | 10 | 25 | 50 | 75 | 90 | 95 |
| Stationary R-squared | ,120 | . | ,120 | ,120 | ,120 | ,120 | ,120 | ,120 | ,120 | ,120 | ,120 |
| R-squared | ,974 | . | ,974 | ,974 | ,974 | ,974 | ,974 | ,974 | ,974 | ,974 | ,974 |
| RMSE | 1,028 | . | 1,028 | 1,028 | 1,028 | 1,028 | 1,028 | 1,028 | 1,028 | 1,028 | 1,028 |
| MAPE | 1,225 | . | 1,225 | 1,225 | 1,225 | 1,225 | 1,225 | 1,225 | 1,225 | 1,225 | 1,225 |
| MaxAPE | 1,881 | . | 1,881 | 1,881 | 1,881 | 1,881 | 1,881 | 1,881 | 1,881 | 1,881 | 1,881 |
| MAE | ,769 | . | ,769 | ,769 | ,769 | ,769 | ,769 | ,769 | ,769 | ,769 | ,769 |
| MaxAE | 1,110 | . | 1,110 | 1,110 | 1,110 | 1,110 | 1,110 | 1,110 | 1,110 | 1,110 | 1,110 |
| Normalized BIC | ,611 | . | ,611 | ,611 | ,611 | ,611 | ,611 | ,611 | ,611 | ,611 | ,611 |

Source: own calculations.

After applying the same model to second dataset we obtain following results (Table 2).

**Table 2:** Model statistics – testing the model

**Model Fit**

| | | | | | Percentile | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Fit Statistic | Mean | SE | Minimum | Maximum | 5 | 10 | 25 | 50 | 75 | 90 | 95 |
| Stationary R-squared | ,427 | . | ,427 | ,427 | ,427 | ,427 | ,427 | ,427 | ,427 | ,427 | ,427 |
| R-squared | ,977 | . | ,977 | ,977 | ,977 | ,977 | ,977 | ,977 | ,977 | ,977 | ,977 |
| RMSE | 1,635 | . | 1,635 | 1,635 | 1,635 | 1,635 | 1,635 | 1,635 | 1,635 | 1,635 | 1,635 |
| MAPE | 1,200 | . | 1,200 | 1,200 | 1,200 | 1,200 | 1,200 | 1,200 | 1,200 | 1,200 | 1,200 |
| MaxAPE | 1,920 | . | 1,920 | 1,920 | 1,920 | 1,920 | 1,920 | 1,920 | 1,920 | 1,920 | 1,920 |
| MAE | ,799 | . | ,799 | ,799 | ,799 | ,799 | ,799 | ,799 | ,799 | ,799 | ,799 |
| MaxAE | 1,156 | . | 1,156 | 1,156 | 1,156 | 1,156 | 1,156 | 1,156 | 1,156 | 1,156 | 1,156 |
| Normalized BIC | 1,715 | . | 1,715 | 1,715 | 1,715 | 1,715 | 1,715 | 1,715 | 1,715 | 1,715 | 1,715 |

Source: own calculations.

The results in table 2 shows a good performance of the ARIMA model (1,1,0) (0,1,0). The Ljung-Box statistics indicate that the model is specified correctly, a significance value is greater than 0,05. MAPE value is 1,2%, lower than in table 1.

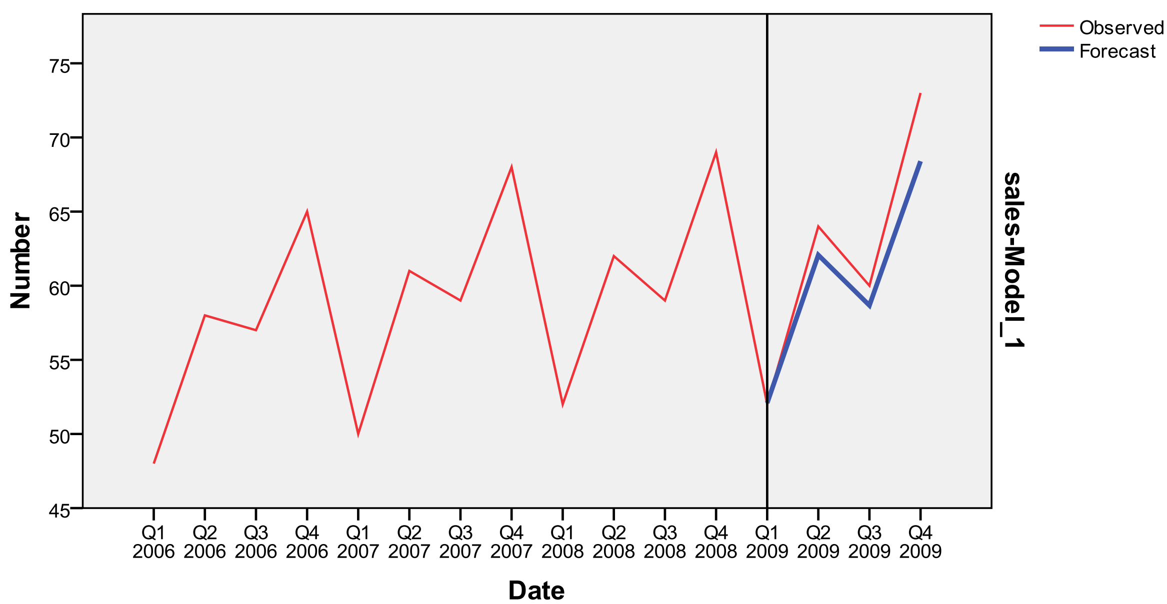


**Figure 4:** The observed and forecast values for the quarterly sales for the period 2006-2010
Source: own calculations.

Figure 4 shows a good agreement between observed and forecast values, indicating that the model has satisfactory predictive ability.

## 3. BUSINESS IMPLICATIONS

Businesses are forced to look well ahead in order to plan their investments, launch new products, decide when to close or withdraw products and so on (Groncutt, 2004; Helensen, 2004; Tvede and Ohnemus, 2001). The sales forecasting process is a critical one for most businesses and is a crucial part of the financial planning of a business. Sales helps you pay employees, cover operating expenses, buy more inventory, market new products and attract more investors. Sales forecasts are also an important part of starting a new business. Almost all new businesses need loans or start-up capital to purchase everything necessary (office space, equipment, inventory, employee salaries, marketing). What an entrepreneur would need is a business plan to prove the business is viable. And a central part of that business plan will be the sales forecast. The overall effect of accurate sales forecasting is a business that runs more efficiently, saving money on excess inventory, increasing profit and serving its customers better.

## 4. CONCLUSION

This paper focuses on sales forecasting using ARIMA model approach and Expert Modeler. Expert Modeler is based on Box and Jenkins method to find the best fitted ARIMA model. The best fitted ARIMA model is (1,1,0) (0,1,0). One quarter is needed to predict the next quarter sales. Seasonal component is recognized and adequately modeled. Mean percentage error (MAPE) for the model is 1,23%.

## References


Choong, J. (2010), *Powerful Forecasting With MS Excel*, XLPert Enterprise.

Groncutt, J. et al. (2004), *Marketing*, Kofen Page, London.

Helensen, S. (2004), *Global Marketing. A Decision-Oriented Approach*, Prentice Hall, London.

IBM SPSS (2010), *IBM SPSS Forecasting 19*, IBM SPSS Company.

Jakasa, T., Androcec, I. and Sprcic, P. (2011), »Electricity Price Forecasting – ARIMA Model Approach«, available at: http://bib.irb.hr/datoteka/526506.EEM11_Session06_06_1351.pdf (accessed 5 March 2012).

Tvede, L. and Ohnemus, P. (2001), *Marketing Strategies for the New Economy*, Wiley, Chichester.

Yaffee, R. (1999), *Introduction to Time Series Analysis and Forecasting,* Academic Press, NY.