

# Automatic Time Series Forecasting

■ Sean FitzGerald



Finance, telecommunications, and manufacturing are just a few industries that are constantly looking for ways to build faster and more effective forecast-

ing applications. Analysts working with sequential time series data such as stock market, telemetry and sales performance data, use historical values to predict future values in a particular area of interest. In this process, pressure is high, time is valuable and techniques that can dramatically reduce time to analysis are highly sought after.

The IMSL Auto\_ARIMA (Autoregressive Integrated Moving Average) algorithm from Visual Numerics, includes a number of techniques that reduce the complexity and time to analysis, including estimation of missing values, identification and adjustment for the effects of outliers, seasonal adjustments, selection of the best input parameters for the ARIMA ( $p, d, q$ ) model and output of the forecast.

Developed through Visual Numerics Consulting Services in the C language and designed for use by programmers and researchers, the IMSL Auto\_ARIMA function is easily embedded into applications for real-time analysis of large volumes of time series data. Unlike traditional ARIMA methods, it requires very little pre-processing of time series data and supports both automatic and manual modes of parameter selection. Individual techniques of dealing with missing values, outliers and seasonality can be used independently of the Auto\_ARIMA function providing greater flexibility for the programmer or researcher.

**Model Selection:** In a traditional ARIMA methodology, the user must specify the number of autoregressive parameters ( $p$ ), the level of differencing ( $d$ ) and moving average ( $q$ ) parameters. This process typically requires expert experience and analysis of the series' autocorrelations and partial autocorrelations. In automatic mode, Auto\_ARIMA automatically selects values for  $p$ ,  $d$  and  $q$ . In addition, it can be configured to automatically detect and adjust for the number of seasons ( $s$ ). Control of whether these parameters are specified by the user or automatically selected is accomplished by using one of six models.

**Seasonal Variations:** Engineering and economic time series frequently contain seasonal or cyclical variations from environmental effects, such as normal business cycles like seasonal sales, to effects of thermal deformation in satellite equipment or higher frequency oscillations.

Two input parameters are required for seasonal differencing: the number of differences to use in the model,  $d$ , and the number of seasons,  $s$ . Both can be user-specified or entered as an array of possible values, with the best-fit values determined automatically. An AR( $p$ ) model (identified using a minimum AIC method) is used to evaluate the best differenced output time series for the combinations of input parameters  $d$  and  $s$ . To remove variances within the series, Auto\_ARIMA uses seasonal differencing and requires as input parameters the number of seasons in the time series. Missing values in the time series are estimated before determining values for  $d$  and  $s$ .

**Missing Values:** Missing values are common in time series for many reasons, including recording errors and equipment failures. Missing values in a

series must be replaced with estimates before model and parameter estimation. Auto\_ARIMA *estimate\_missing* function estimates missing values automatically using one of four methods.

If the missing value occurs early in the series, there may be insufficient data to compute an AR( $p$ ) model. If this happens, *estimate\_missing* falls back to Method 1, replacing a missing value with the median value of previously observed data.

**Outliers:** Outliers are spurious observations that are not consistent with the rest of the time series and are transferred to the resulting forecast.

Auto\_ARIMA employs the Chen-Liu algorithm (Chen & Liu, 1993), a joint estimation method, to identify outliers. Outputs of the function include the number of outliers identified, times at which they occurred, and the outlier classifications. The *ts\_outlier\_identification* function is employed for identification and adjustment of the outliers. Outliers are classified into one of five categories based on the Chen-Liu standardized statistic for each outlier type.

**Forecasting:** Forecasting is done after identifying the best-fit model and outliers. The effects of outliers near the beginning of a long series usually have little effect on future forecasts. However, outliers, other than additive outliers, can have a dramatic impact on future forecasts. Their effects must be identified and incorporated into future forecasts to improve the accuracy of those forecasts. Auto\_ARIMA not only automatically identifies these outliers, but also incorporates their effects into future forecasts to improve forecast accuracy.

*Sean FitzGerald is VP of Technology and Consulting with Visual Numerics, Inc. He may be contacted at [sceditor@scimag.com](mailto:sceditor@scimag.com).*