Time Series Analysis: Identifying Underlying Patterns

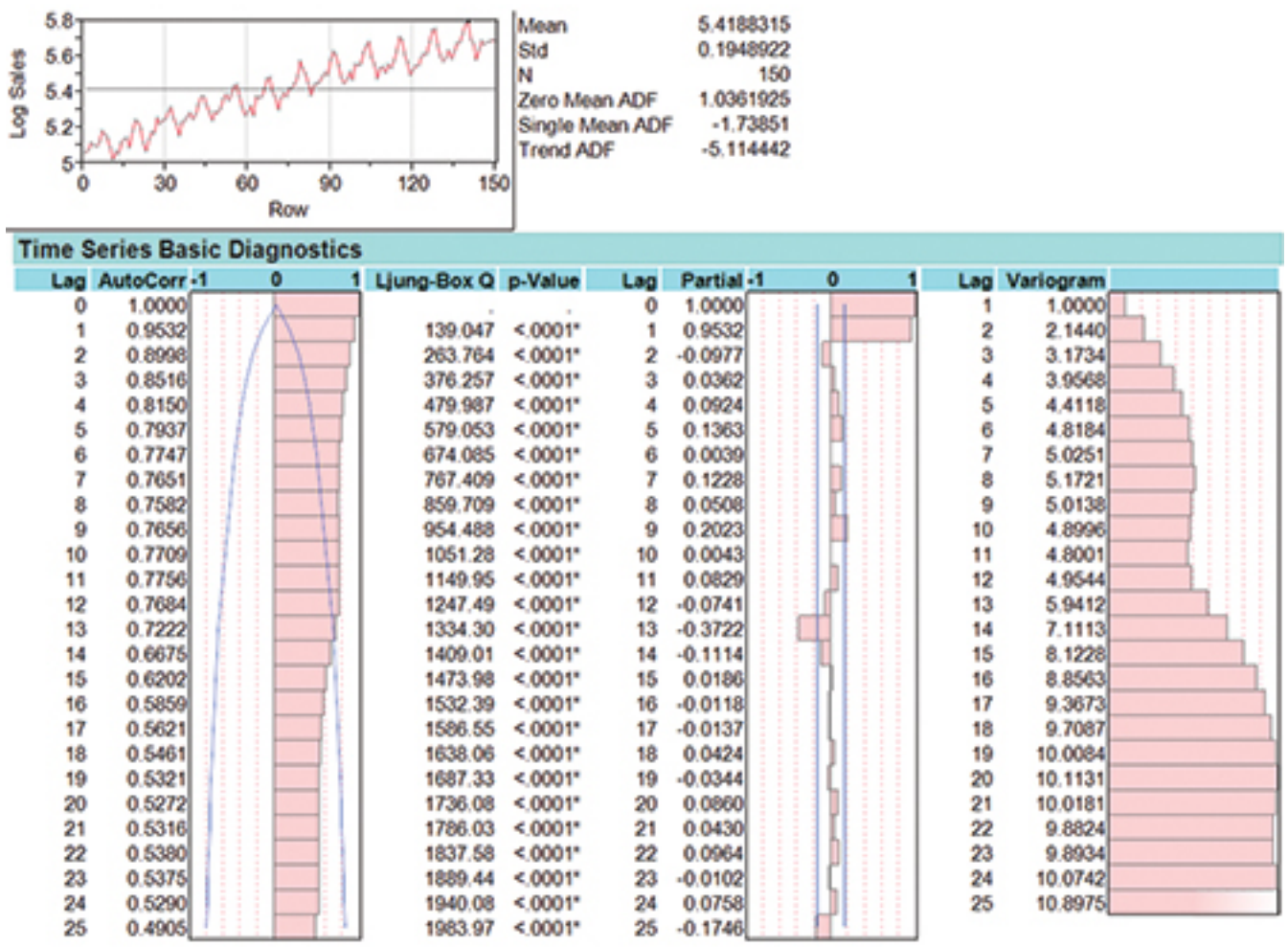
Mark A. Anawis

The goal of this type of analysis can be divided into characterization or prediction

Nelson Mandela said: “We must use time wisely and forever realize that the time is always ripe to do right.” Time series data have a temporal order that makes analysis distinctly different from other data analysis. The goal of time series analysis can be divided into characterization or prediction. There is a consistent pattern contaminated with random noise, which typically requires filtering to aid in identifying the underlying pattern. The pattern itself can be divided into the main trend and a seasonal component.

The main trend can often be described by a linear function, which may need to be transformed to eliminate any non-linearity using an exponential or log function. If there is considerable error masking the trend, smoothing is required, such as a moving average which replaces components of the series with a simple or weighted average.

The seasonal component can be examined via autocorrelation correlograms, which display serial correlation for consecutive lags. A Ljung-Box Q statistic and p-values can be used to determine whether a group of autocorrelation is significant or to distinguish residuals from white noise. A small p-value (< .05) indicates the possibility of non-zero autocorrelation within the first few lags (Figure 1).



By removing dependencies within a lag, partial autocorrelations can provide a better representation of serial dependencies.

Useful in characterization and forecasting, the Autoregressive Integrated Moving Average (ARIMA) model uses a linear combination of prior values and a series of random shock errors (see Figure 2 for the general equation).

A stationarity requirement, defined as a constant mean, variance and autocorrelation through time, constrains parameters to a certain range. Each member of the series can be affected by prior random shock errors. An invertibility requirement, whereby the moving average equation can be recast into an autoregressive form also constrains the moving average parameters. There are three types of parameters:

- autoregressive (p)
- differencing (d)
- moving average (q)

The number of autoregressive and moving average parameters should contain the fewest number to fit the data and rarely needs to be greater than two. Differencing the series aids in meeting the stationarity requirement. Optimizing the number of times the series needs differencing is aided by reviewing the plots and the autocorrelogram. Changes in level or autocorrelations that decline slowly at longer lags require first-order differencing, whereas slope changes require second-order differencing.

ARIMA Model

$$\varphi(B)(\omega_t - \mu) = \theta(B)a_t$$

where $\varphi(B)$ is the autoregressive operator,
$\theta(B)$ is the moving average operator,
B is the backshift operator defined as:

$$B = \frac{y_{t-1}}{y}$$

ω_t is the response series after differencing defined as:
$\omega_t = (1-B)^d y_t$
t is the time index,
μ is the intercept or mean term,
a_t is the sequence of random shocks

Parameter estimation uses a function minimization algorithm, such as the quasi- Newton, to maximize the probability of the observed series. Parameter estimates have t ratios and associated probabilities to determine significance. Seasonal components, such as those seen in the retail industry, add another layer. They also have the three parameters described in the ARIMA model. A seasonal ARIMA can be combined with a basic ARIMA. The seasonality may be either additive or multiplicative in nature.

Transfer function analysis, where an input and output series of data exist, is similar to ARIMA models. Prior to building the model, an exploratory phase using pre-whitening is instructive. It involves finding a model for the input series, applying it to the output series, getting residuals from both series, computing cross-correlations from the residual series and finding the correct transfer function polynomials. Stationarity is tested using Augmented Dickey-Fuller (ADF) tests for random walks. A zero mean, single mean, and trend ADF along with autocorrelation, partial autocorrelation and cross correlation diagnostics can be used.

Model evaluation for forecasts can be done by plotting the observed values and one-step-ahead forecasts. Different models can be compared using criteria such as the Akaike's Information Criterion (AIC) or Schwarz's Bayesian Criterion (SBC) and RSquare, where larger values indicate a better fit and Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) where smaller values indicate a better fit (Figure 3).

Report Graph	Model	DF	Variance	AIC	SBC	RSquare	-2LogLH	Weights .2.4.6.8	MAPE
	Seasonal ARIMA(1, 1, 1)(0, 0, 0)12	146	0.0049895	-363.9352	-354.9233	0.870	-369.9352	0.774018	0.803799
	Seasonal ARIMA(0, 1, 1)(0, 1, 1)12 No Intercept	135	0.0041025	-360.2177	-354.3777	0.871	-364.2177	0.120646	0.459960
	Seasonal ARIMA(1, 1, 1)(1, 1, 1)12	132	0.0040426	-359.9463	-345.3464	0.876	-369.9463	0.105336	0.457908

Spectral analysis is used to decompose a time series into several sinusoidal functions of a certain wavelength in order to identify the seasonal variations of different lengths. This can be extended in cross-spectrum analysis to the simultaneous analysis of two series to uncover their correlations at different frequencies. The cross-spectrum consists of complex numbers that can be smoothed to calculate cross-density and quadrature density values, which are combined into a cross-amplitude, a measure of covariance between the frequency components in the series. Since the sine and cosine functions are orthogonal, their squared coefficients can be added for each frequency to produce a periodogram where this periodogram value can be interpreted in variance terms at a given frequency.

The time requirements associated with performing spectral analysis led to a refinement in the fast Fourier algorithm (FFT) where the time required is proportional to N*log2(N), although the number in the series needs to be padded in order to be a power of 2.

Time series analysis is found in weather forecasting, economics, pattern recognition and statistics. It involves an exploratory phase to characterize the data to understand the underlying trend separate from the random error. External factors which may interrupt the trend need to be taken into account. Analyzing serial dependence, as well as seasonal and cyclic components leads to models which can be compared to identify a best fit model for forecasting.
*Note: Figures 1 and 3 were generated using JMP v.10 software.

Mark Anawis is a Principal Scientist and ASQ Six Sigma Black Belt at Abbott. He may be reached at editor@ScientificComputing.com.

Source URL (retrieved on *07/31/2015 - 7:23am*):
http://www.scientificcomputing.com/blogs/2014/06/time-series-analysis-identifying-underlying-patterns