Time Series Analysis Using Statgraphics Centurion

Neil W. Polhemus, CTO, StatPoint Technologies, Inc.

Copyright 2011 by StatPoint Technologies, Inc.
Web site: www.statgraphics.com

Procedures to be Covered

- Descriptive Methods (time sequence plots, autocorrelation functions, periodograms)
- Smoothing
- Seasonal Decomposition
- Forecasting
  - User-specified models
  - Automatic model selection

Time Series Analysis

Analysis of sequentially ordered data.

Standard notation: \( Y_t, t = 1, 2, \ldots, n \)

Main complication for analysis is that the observations are usually not independent. Instead, they are autocorrelated.

Example #1 – U.S. Unemployment Rate

Defined as the percentage of the civilian labor force that are actively looking for a job but can’t find one. (Note: people who have given up looking for jobs or are not available for work are not considered part of the labor force.)

Source: www.bls.gov

November 2011:
- Unadjusted rate: 8.2%
- Seasonally adjusted rate: 8.6%

Time Sequence Plot

Classic time series model consists of 4 components:
1. Trend (T) – long-term changes in the underlying process mean.
2. Cycle (C) – cyclical variations around the mean with no fixed period.
3. Seasonality (S) – cyclical variations around the mean with a fixed period.
4. Randomness (R) – random fluctuations.
### Basic Models

**Multiplicative model**

\[ Y_t = T_t \times C_t \times S_t \times R_t \]

**Additive model**

\[ Y_t = T_t + C_t + S_t + R_t \]

*Note: taking logarithms turns a multiplicative model into an additive model.*

### Smoothing

Statgraphics has 2 types of smoothers:

- **Linear smoothers** – weighted moving averages that estimate the underlying trend-cycle by averaging out the other components.
- **Nonlinear smoothers** – variations on running medians that reduce the impact of isolated outliers.

### Linear Smoothers

These smoothers are basically weighted moving averages of the form

\[ S_t = \sum_{j=m}^{m} W_j Y_{t-j} \]

Statgraphics offers simple moving averages, exponentially weighted moving averages, and moving averages developed by Spencer and Henderson.

Except for the EWMA, the moving averages are centered around time \( t \).

### Moving Average with Span of 12

![Smoothing Time Series Plot for Unemployment rate (unl) with a moving average of 12 months.](image)

### Nonlinear Smoothers

These smoothers were developed by John Tukey and are based on running medians of either 3 or 5 observations.

A typical smoother is the 3RSSH smoother:
- **3** – smooth the data with a running median of 3.
- **R** – repeat the running median again and again until nothing changes.
- **S** – split the mesas and dales and repeat the 3R operation.
- **S** – split the mesas and dales and repeat the 3R operation.
- **H** – apply a simple 3-term MA as proposed by Hanning.

### Adjustments

Various adjustments may be applied to the time series data before it is smoothed.

![Adjustment Options](image)
Smoothing with 3RSSH

Smoothed Time Series Plot for Unemployment rate (smooth)

Seasonal Adjustment

Removes the seasonal component from a time series, leaving the other components behind.

For a multiplicative model:

\[
Y_{t}^{adj} = \frac{Y_{t}}{S_{t}} = \frac{T_{t}C_{t}S_{t}R_{t}}{S_{t}} = T_{t}C_{t}R_{t}
\]

Step #1 – Smooth data to estimate trend-cycle

Step #2 – Divide data by smooth and then average each season to estimate seasonality

Step #3 – Divide data by seasonal index to remove seasonality
Seasonal Subseries Plot – Shows Changes on a Season by Season Basis

Forecasting

Used to predict future values of the time series.

Types of models:
- Autoprojective models – forecasts the future based solely on the past behavior of the time series itself.
- Input/output models – includes information from other predictor variables, usually leading indicators.

Methods

Statgraphics Procedures

User-Specified Model – fits a particular model selected by the user.

Automatic Model Selection – fits many models and automatically picks the best according to some optimization criterion.
Akaike Information Criterion (AIC)

Minimizes the root mean squared error RMSE plus a penalty based on the number of estimated model parameters $c$.

$$AIC = 2 \ln(RMSE) + \frac{2c}{n}$$

Can be thought of as a tradeoff between model precision and model complexity (which can lead to bias).

Results – Unadjusted Rate

Example #2 – Closing prices of Apple Inc. common stock

Data consist of closing prices on each trading day between Jan. 4, 2010 and Dec. 2, 2011.

Source: [www.yahoo.com](http://www.yahoo.com)

Low during period: $192.05 on Feb. 4, 2010

High during period: $422.24 on Oct. 18, 2011

Latest close: $389.70 on Dec. 2, 2011

Results – Adjusted Rate

Example #2 – Closing prices of Apple Inc. common stock

Data consist of closing prices on each trading day between Jan. 4, 2010 and Dec. 2, 2011.

Source: [www.yahoo.com](http://www.yahoo.com)

Low during period: $192.05 on Feb. 4, 2010

High during period: $422.24 on Oct. 18, 2011

Latest close: $389.70 on Dec. 2, 2011

Time Sequence Plot – AAPL close

Automatic Forecasting – AAPL close
Random Walk Model

The random walk model implies that the differences between consecutive closing prices are random.

\[ \Delta_t = Y_t - Y_{t-1} \]

The past history has no effect on the behavior of the time series in the future. Thus the best forecast for tomorrow's closing price is today's closing price.

Statistical Time Series Models

A time series may be thought of consisting of a deterministic component and a stochastic component.

\[ Y_t = \eta_t + \epsilon_t \]

Deterministic Component

The deterministic component may be:

- a constant mean: \( \mu \)
- a trend over time such as: \( a + bt \)
- a cyclical effect such as: \( a \sin(awt + f) \)
- a function of one or more input variables: \( m \times X_t - k + b \)
- a combination of the above

Stochastic Component

The stochastic component may often be represented as a linear combination of random shocks \( \epsilon_t \) at the present and in the past:

\[ \epsilon_t = \sum_{k=0}^{\infty} \Psi_k a_{t-k} \]

ARIMA Models

Common model forms include:

- Autoregressive AR(p) models:
  \[ Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + \epsilon_t \]
- Moving average MA(q) models:
  \[ Y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q} \]
- Mixed ARMA(p,q) models:
  \[ Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \ldots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \]
- Models involving differencing : ARIMA(p,d,q)
  \[ Y_t - Y_{t-1} = \phi_1(Y_{t-1} - Y_{t-2}) + \epsilon_t - \theta \epsilon_{t-1} \]

Autocorrelation Function

Plots the linear correlation between observations separated by various lags:

\[ r_l = \text{corr}(Y_t, Y_{t-l}) = \frac{\sum_{t=1}^{n-l} (Y_t - \bar{Y})(Y_{t-l} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2} \]

Autoregressive models show exponential decay, while moving average models show spikes at low lags. Failure to decay is a sign of nonstationarity.
Partial Autocorrelation Function

Used to determine the order of autoregressive model needed to represent the time series adequately.

Autoregressive models show spikes at low lags, while moving average models show exponential decay.

Unemployment Rate Model

Automatic forecasting selected an ARIMA(1,1,1) model:

\[ Y_t - Y_{t-1} = 0.925(Y_{t-4} - Y_{t-2}) + a_t - 0.697a_{t-4} \]

\[ \hat{\sigma}_a = 0.116 \]

Periodogram

Performs an analysis of variance across the Fourier frequencies.

\[ Y_t = a_0 + \sum_{i=0}^{n} [a_i \sin(2\pi f_i) + b_i \cos(2\pi f_i)] \quad n = 2q + 1 \quad f_i = \frac{i}{n} \]

\[ I(f_i) = N \frac{1}{2} (a_i^2 + b_i^2) \]
Periodogram for First Differences

Crosscorrelation Function
Plots the linear correlation between two time series separated by various lags:

\[ r_{xy}(k) = \text{corr}(Y_t, X_{t-k}) \]

Plot of Residuals from Random Walk Model for AAPL

Crosscorrelations with for First Differences of BUD

Example #3 – Input/Output Models
Output time series: Anheiser Busch stock price (BUD)
Input time series X: Apple Inc stock price (AAPL)

Showed that there appears to be a small negative correlation between changes in AAPL today and changes in BUD tomorrow.

We will build a model for BUD that includes AAPL as a lagged regressor in an ARIMA model.
Step 1: Forecast Input Time Series and Add Forecasts to Data Sheet

Step 2: Add Regressor to Forecasting Model

Fitted Model

\[ BUD_t = BUD_{t-1} - 0.0289(AAPL_{t-1} - AAPL_{t-2}) \]

Forecasting Results with Input Variable

The effect on the forecasts is very slight since AAPL did not change much on the last day of the observed data.

More Information

Go to www.statgraphics.com

Or send e-mail to info@statgraphics.com