Time Series Analysis Using Statgraphics Centurion

Neil W. Polhemus, CTO, 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

2

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*.

3

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

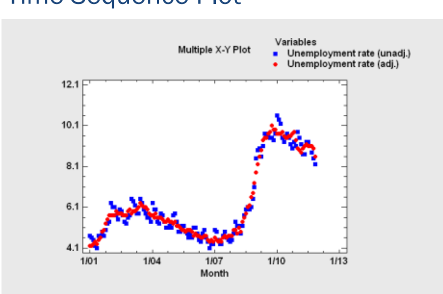
November 2011:

Unadjusted rate: 8.2%

Seasonally adjusted rate: 8.6%

4

Time Sequence Plot



5

Time Series Components

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.

6

Basic Models

Multiplicative model

$$Y_t = T_t * C_t * S_t * R_t$$

Additive model

$$Y_t = T_t + C_t + S_t + R_t$$

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

STATPOINT TECHNOLOGIES, INC.

7

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.

STATPOINT TECHNOLOGIES, INC.

8

Linear Smoothers

These smoothers are basically weighted moving averages of the form

$$S_t = \sum_{j=-m}^{m} W_j Y_{t+j}$$

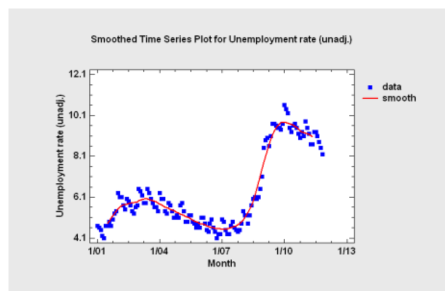
Statgraphics offers simple moving averages, exponentially weighted moving averages, and moving averages developed by Spencer and Henderson.

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

STATPOINT TECHNOLOGIES, INC.

9

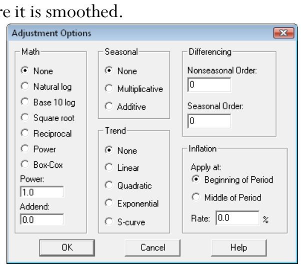
Moving Average with Span of 12



10

Adjustments

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



STATPOINT TECHNOLOGIES, INC.

11

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.

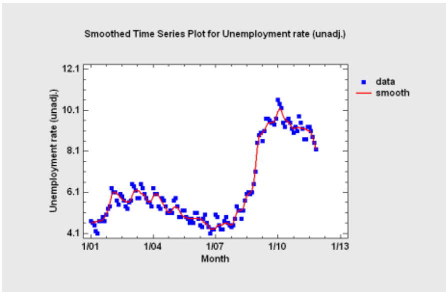
S – split the mesas and dales and repeat the 3R operation.

H – apply a simple 3-term MA as proposed by Hanning.

STATPOINT TECHNOLOGIES, INC.

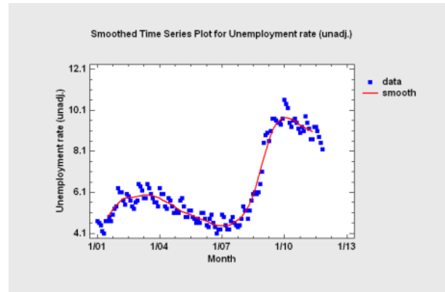
12

Smoothing with 3RSSH



13

Smoothing with 3RSSH and MA(12)



14

Seasonal Adjustment

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

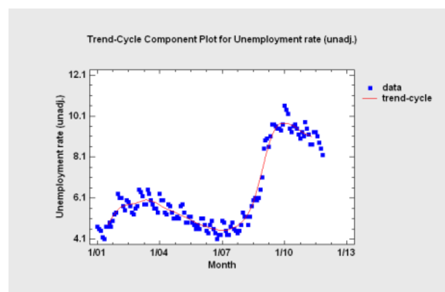
For a multiplicative model:

$$Y_t^{adj} = \frac{Y_t}{S_t} = \frac{T_t C_t S_t R_t}{S_t} = T_t C_t R_t$$

STATPOINT TECHNOLOGIES, INC.

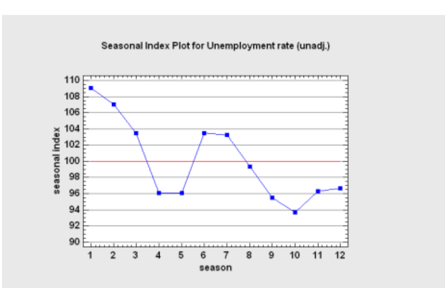
15

Step #1 – Smooth data to estimate trend-cycle



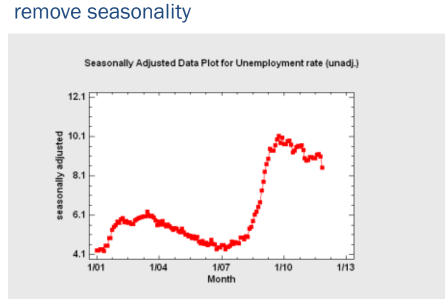
16

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



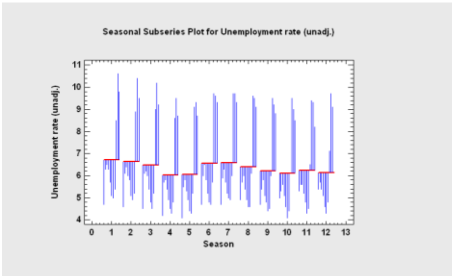
17

Step #3 – Divide data by seasonal index to remove seasonality



18

Seasonal Subseries Plot – Shows Changes on a Season by Season Basis



19

Forecasting

Used to predict future values of the time series.

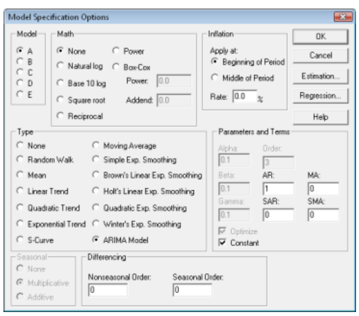
Types of models:

- 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.

STATPOINT TECHNOLOGIES, INC.

20

Methods



21

Statgraphics Procedures

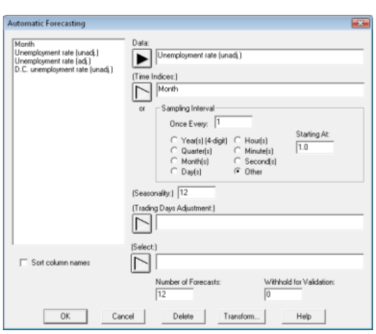
User-Specified Model – fits a particular model selected by the user.

Automatic Model Selection – fits many models and automatically picks the best according to some optimization criterion.

STATPOINT TECHNOLOGIES, INC.

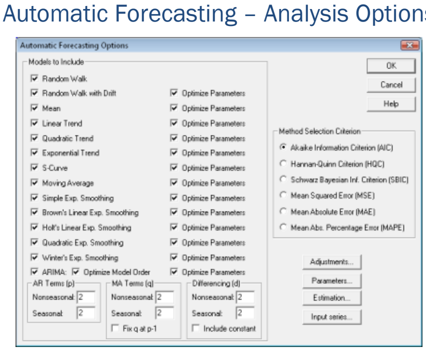
22

Automatic Forecasting – Data Input



23

Automatic Forecasting – Analysis Options



24

Akaike Information Criterion (AIC)

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

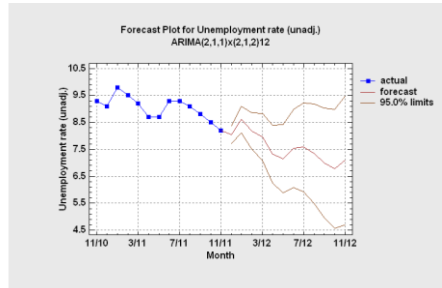
$$AIC = 2\ln(RMSE) + \frac{2c}{n}$$

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

STATPOINT TECHNOLOGIES, INC.

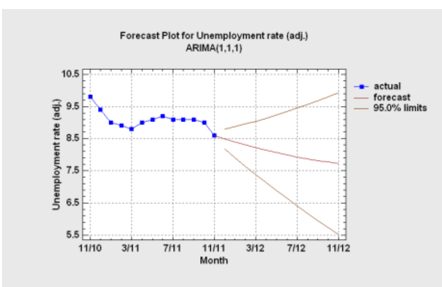
25

Results – Unadjusted Rate



26

Results – Adjusted Rate



27

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.yahoofinance.com

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

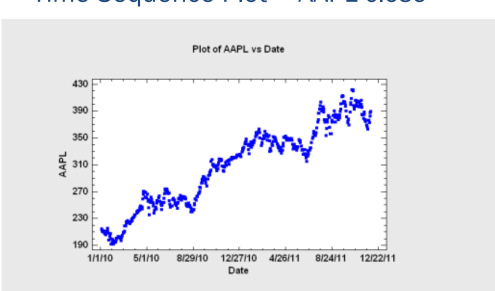
High during period: \$422.24 on Oct. 18, 2011

Latest close: \$389.70 on Dec. 2, 2011

STATPOINT TECHNOLOGIES, INC.

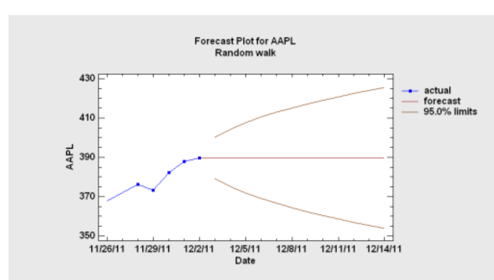
28

Time Sequence Plot – AAPL close



29

Automatic Forecasting– AAPL close



30

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.

31

Statistical Time Series Models

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

$$Y_t = \eta_t + \varepsilon_t$$

32

Deterministic Component

The deterministic component may be:

- a constant mean: μ
- a trend over time such as: $a + b*t$
- a cyclical effect such as: $\alpha*sin(\omega*t+\phi)$
- a function of one or more input variables: $m*X_{t-k} +b$
- a combination of the above

33

Stochastic Component

The stochastic component may often be represented as a linear combination of random shocks a_{t-k} at the present and in the past:

$$\varepsilon_t = \sum_{k=0}^{\infty} \Psi_k a_{t-k}$$

34

ARIMA Models

Common model forms include:

- Autoregressive AR(p) models:

$$Y_t = \mu + \phi_1 (Y_{t-1} - \mu) + \phi_2 (Y_{t-2} - \mu) + a_t$$

- Moving average MA(q) models:

$$Y_t = \mu + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

- Mixed ARMA(p,q) models:

$$Y_t = \mu + \phi_1 (Y_{t-1} - \mu) + a_t - \theta_1 a_{t-1}$$

- Models involving differencing : ARIMA(p,d,q)

$$Y_t - Y_{t-1} = \phi_1 (Y_{t-1} - Y_{t-2}) + a_t - \theta_1 a_{t-1}$$

35

Autocorrelation Function

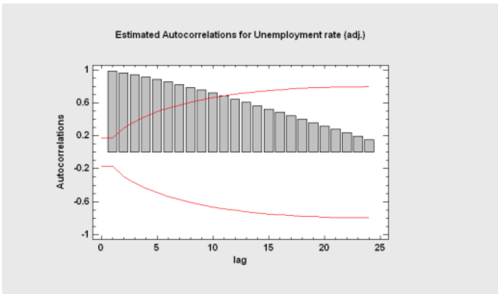
Plots the linear correlation between observations separated by various *lags*:

$$r_k = corr(Y_t, Y_{t-k}) = \frac{\sum_{t=1}^{n-k} (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_{t=1}^{n} (Y_t - \bar{Y})^2}$$

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

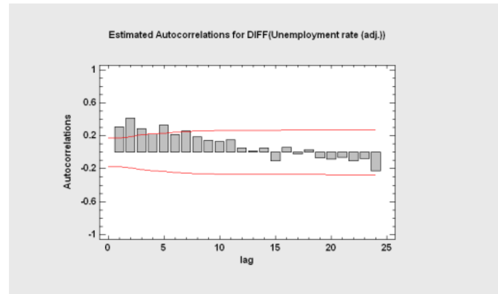
36

ACF for Adjusted Unemployment Rate



37

ACF for First Differences



38

Partial Autocorrelation Function

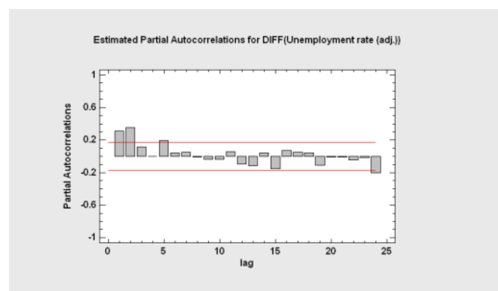
Used to determine the order of autoregressive model needed to represent the time series adequately.

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

STATPOINT TECHNOLOGIES, INC.

39

PACF for First Differences



40

Unemployment Rate Model

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

$$Y_t - Y_{t-1} = 0.925\left(Y_{t-1} - Y_{t-2}\right) + a_t - 0.697 a_{t-1}$$

$$\hat{\sigma}_a = 0.116$$

STATPOINT TECHNOLOGIES, INC.

41

Periodogram

Performs an analysis of variance across the Fourier frequencies.

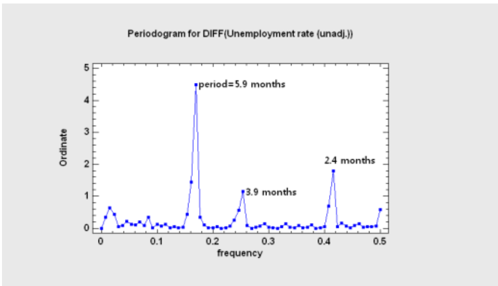
$$Y_t = a_0 + \sum_{i=1}^{q}\left[a_i \sin(2\pi f_i) + b_i \cos(2\pi f_i)\right] \qquad n = 2q+1 \qquad f_i = \frac{i}{n}$$

$$I(f_i) = \frac{N}{2}\left(a_i^2 + b_i^2\right)$$

STATPOINT TECHNOLOGIES, INC.

42

Periodogram for First Differences



43

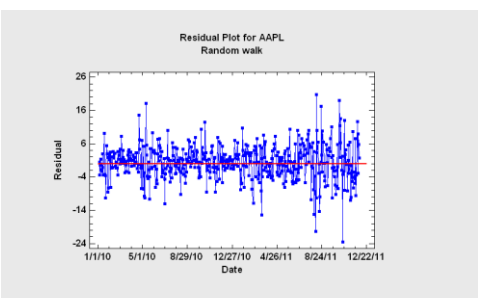
Crosscorrelation Function

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

$$r_{xy}(k) = corr(Y_t, X_{t-k})$$

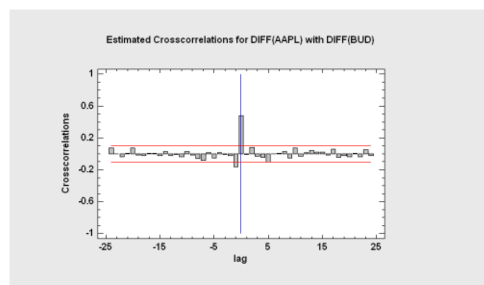
44

Plot of Residuals from Random Walk Model for AAPL



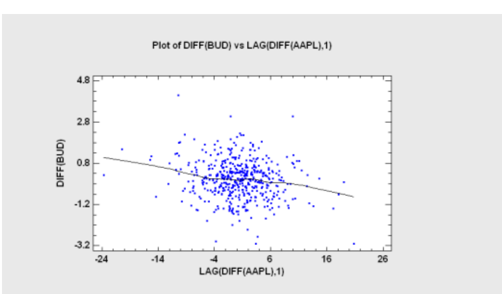
45

Crosscorrelations with for First Differences of BUD



46

X-Y Plot with LOWESS Smooth



47

Example #3 – Input/Output Models

Output time series: Anheiser Busch stock price (BUD)

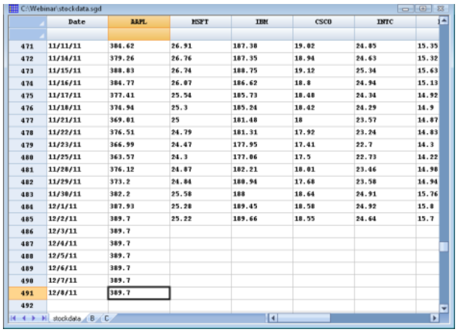
Input time series X: Apple Inc stock price (AAPL)

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.

48

Step 1: Forecast Input Time Series and Add Forecasts to Data Sheet

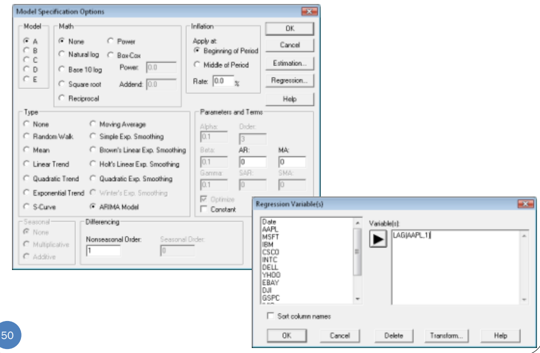

C:\Webinar\stockdata.sgd

	Date	AAPL	MSFT	IBM	CSCO	INTC	
471	11/11/11	384.62	26.91	187.38	19.02	24.85	15.35
472	11/14/11	379.26	26.76	187.35	18.94	24.63	15.32
473	11/15/11	388.83	26.74	188.75	19.12	25.34	15.63
474	11/16/11	384.77	26.07	186.62	18.8	24.94	15.13
475	11/17/11	377.41	25.54	185.73	18.48	24.34	14.92
476	11/18/11	374.94	25.3	185.24	18.42	24.29	14.9
477	11/21/11	369.01	25	181.48	18	23.57	14.87
478	11/22/11	376.51	24.79	181.31	17.92	23.24	14.83
479	11/23/11	366.99	24.47	177.95	17.41	22.7	14.3
480	11/25/11	363.57	24.3	177.06	17.5	22.73	14.22
481	11/28/11	376.12	24.87	182.21	18.01	23.46	14.98
482	11/29/11	373.2	24.84	180.94	17.68	23.58	14.94
483	11/30/11	382.2	25.58	188	18.64	24.91	15.76
484	12/1/11	387.93	25.28	189.45	18.58	24.92	15.8
485	12/2/11	389.7	25.22	189.66	18.55	24.64	15.7
486	12/3/11	389.7					
487	12/4/11	389.7					
488	12/5/11	389.7					
489	12/6/11	389.7					
490	12/7/11	389.7					
491	12/8/11	389.7					
492							

stockdata B C

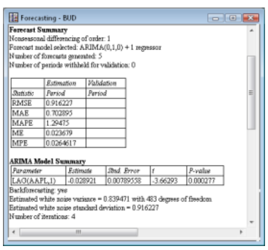
49

Step 2: Add Regressor to Forecasting Model



50

Fitted Model


Forecasting - BUD

Forecast Summary
Nonseasonal differencing of order: 1
Forecast model selected: ARIMA(0,1,0) + 1 regressor
Number of forecasts generated: 5
Number of periods withheld for validation: 0

	Estimation	*Validation*
Statistic	*Period*	*Period*
RMSE	0.916227	
MAE	0.702895	
MAPE	1.29475	
ME	0.023679	
MPE	0.0264617	

ARIMA Model Summary

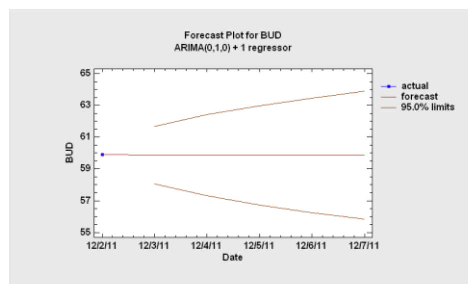
Parameter	*Estimate*	*Stnd. Error*	*t*	*P-value*
LAG(AAPL,1)	-0.028921	0.00789558	-3.66293	0.000277

Backforecasting: yes
Estimated white noise variance = 0.839471 with 483 degrees of freedom
Estimated white noise standard deviation = 0.916227
Number of iterations: 4

$$BUD_t = BUD_{t-1} - 0.0289(AAPL_{t-1} - AAPL_{t-2})$$

51

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.

52

More Information

Go to www.statgraphics.com

Or send e-mail to info@statgraphics.com

STATPOINT TECHNOLOGIES, INC.

53