SPC for Autocorrelated Data Using Automated Time Series Forecasting

Centre for Marketing Analytics and Forecasting at Lancaster University Friday Forecasting Talk January 29, 2021

John Noguera, P. Eng. - CTO & Co-founder, SigmaXL, Inc.

jnoguera@SigmaXL.com

CMAF

SigmaXL
Simply Smart Software

Agenda

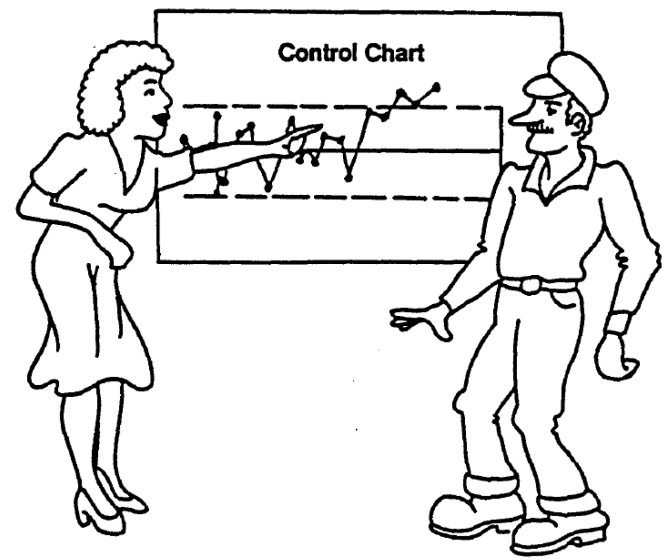
- Introduction to Statistical Process Control (SPC)
- SPC for Autocorrelated Data
- Software
- Autocorrelation
- Example 1: Chemical Process Concentration
- Simple Exponential Smoothing (EWMA)
- Example 2: Ln(Monthly Airline Passengers-Modified)

Agenda

- Error, Trend, Seasonal (ETS) Exponential Smoothing models
- Autoregressive Integrated Moving Average (ARIMA) models
- ARIMA with Predictors
  - Analyze control chart outlier versus shift
- Questions
- References
- Appendix

Introduction to Statistical Process Control (SPC)

- Statistical process control (SPC) is a method of quality control which employs statistical methods to monitor and control a process.
- SPC was pioneered by Walter A. Shewhart at Bell Laboratories in the early 1920s. Shewhart developed the control chart in 1924 and the concept of a state of statistical control.



Ref: https://en.wikipedia.org/wiki/Statistical_process_control

Introduction to Statistical Process Control (SPC)

- Shewhart concluded that while every process displays variation, some processes display variation that is natural to the process ("common" causes of variation); these processes he described as being in statistical control.
- Other processes additionally display variation that is not present in the causal system of the process at all times ("special" causes of variation), which Shewhart described as not in control.

Ref: https://en.wikipedia.org/wiki/Statistical_process_control

Introduction to Statistical Process Control (SPC)

- Special cause variation is also known as "assignable" cause variation and can appear as a shift in the process mean, process variance or as an outlier.
- Examples of special causes include standard operating procedure not being followed, deficient batch of raw material, or equipment malfunction.
- If an out-of-control signal occurs, the operator must find the root cause and take corrective action to restore the process to a state of statistical control. Out-of-Control Action Plans (OCAPs) are used to streamline the troubleshooting effort.

Introduction to Statistical Process Control (SPC)

The Shewhart Individuals & Moving Range chart is a type of control chart used to monitor variables data from a business or industrial process for which it is impractical to use rational subgroups. In practice this is the most commonly used type of control chart.

Center line $\bar{x} = \frac{\sum_{i=1}^{m} x_i}{m}$

Control limits $\bar{x} \pm 3\frac{\overline{MR}}{d_2}$

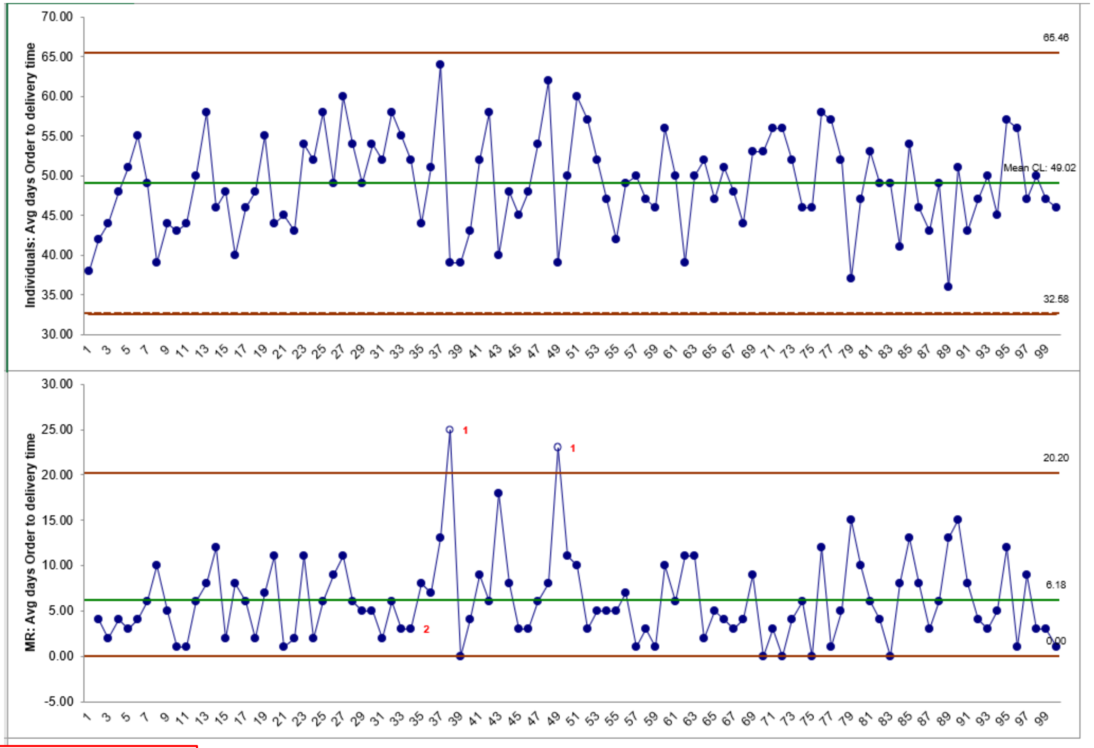
Plotted statistic x_i

Center line $\overline{MR} = \frac{\sum_{i=2}^{m} |x_i - x_{i-1}|}{m-1}$

Upper control limit $D_4 \overline{MR}$

Lower control limit $D_3 \overline{MR}$

Plotted statistic $MR_i = |x_i - x_{i-1}|$



Ref: https://en.wikipedia.org/wiki/Shewhart_individuals_control_chart

Introduction to Statistical Process Control (SPC)

- For an in-control process, observations are assumed to be normally distributed, independent, with fixed mean and constant variance.
- Lack of normality can be addressed using rational subgroups (central limit theorem), a Box-Cox power transformation or use of an Exponentially Weighted Moving Average (EWMA) chart.
- The presence of serial autocorrelation in data will adversely affect the performance of a Shewhart chart. In the process industries, autocorrelation is typically positive due to inertial elements, resulting in false alarms. If, however, the autocorrelation is negative, the Shewhart limits will be too wide, and significant process shifts will not be detected.

SPC for Autocorrelated Data

- Statistical process control for autocorrelated processes typically use the EWMA (Exponentially Weighted Moving Average) one-step-ahead forecast model.
- The time series model forecasts the motion in the mean and an Individuals control chart is plotted of the residuals to detect assignable causes.

SPC for Autocorrelated Data

- The challenge with this approach is that if there is seasonality or negative autocorrelation in the data, the user needs an advanced level of knowledge in forecasting methods to pick the correct model, e.g., Seasonal Exponential Smoothing models or Seasonal Autoregressive Integrated Moving Average (ARIMA) models are required.

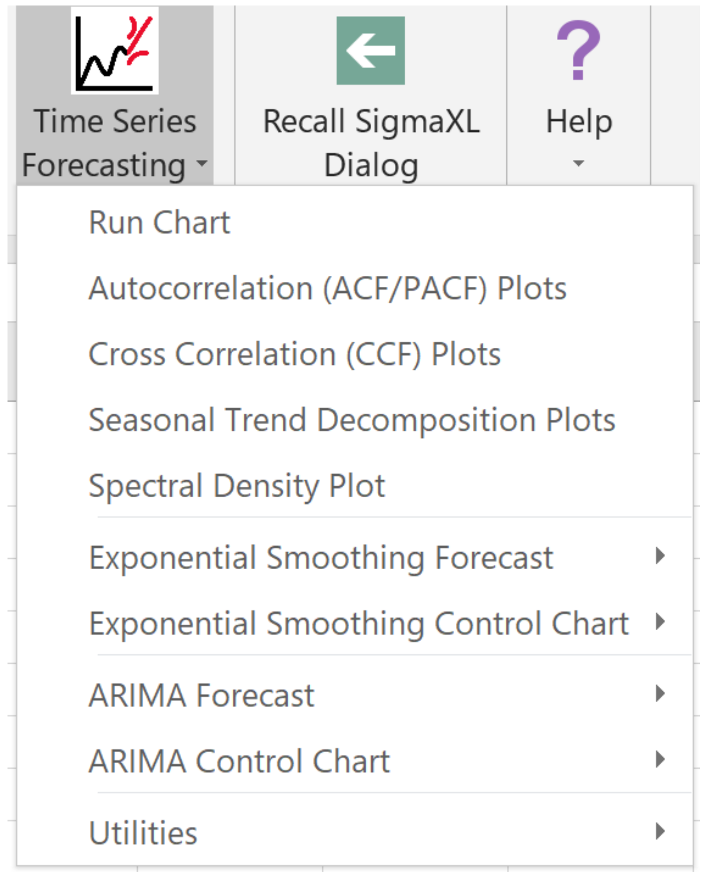
SPC for Autocorrelated Data

- We will review simple exponential smoothing/EWMA, then introduce recent developments in time series forecasting that use automatic model selection to accurately pick the time series model that produces a minimum forecast error.
- An accurate forecast for your time series means the residuals will most often have the right properties to correctly apply a control chart, thus leading to an improved control chart with reduced false alarms and misses.

Software

- Control charts for autocorrelated data using automated time series forecasting will be demonstrated using SigmaXL, a commercial Excel Add-in for statistical and graphical analysis, that is menu based and easy to use.
- The methods demonstrated here can also be implemented using R time series packages such as forecast (auto.arima, ets, mstl) or adam, store the residuals and then create an Individuals control chart with qcc (Quality Control Charts).

SigmaXL Version 9
Time Series Forecasting Menu



Autocorrelation

- Just as correlation measures the extent of a linear relationship between two variables, autocorrelation (AC) measures the linear relationship between lagged values of data.
- A plot of the data vs. the same data at lag k will show a positive or negative trend. If the slope is positive, the AC is positive; if there is a negative slope, the AC is negative.
- The Autocorrelation Function (ACF) formula is:

$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \overline{y})(y_{t-k} - \overline{y})}{\sum_{t=1}^{T} (y_t - \overline{y})^2}$$

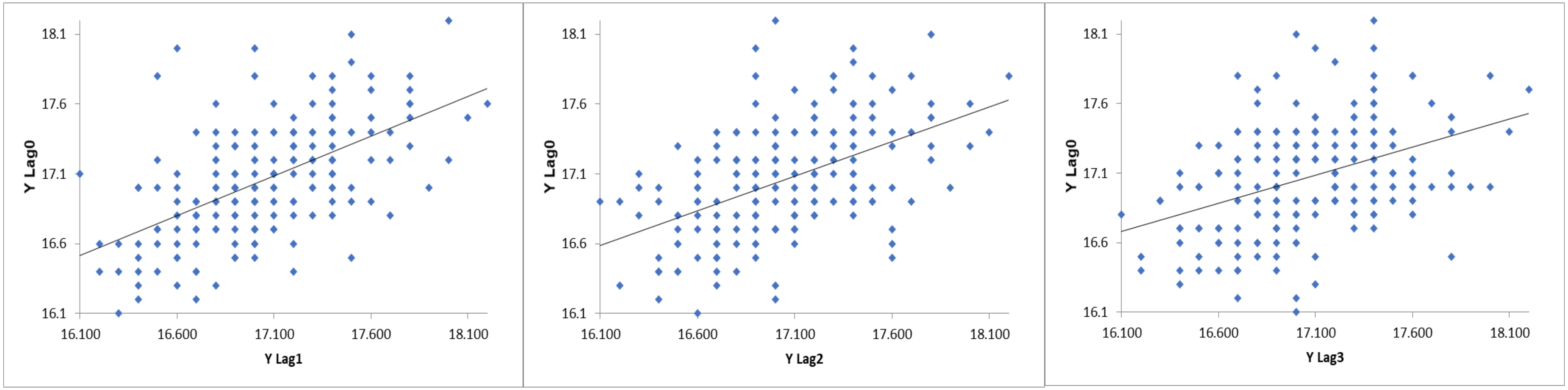
where T is length of the time series [4].

Autocorrelation

Y Lag0	Y Lag1	Y Lag2	Y Lag3
17			
16.6	17		
16.3	16.6	17	
16.1	16.3	16.6	17
17.1	16.1	16.3	16.6
16.9	17.1	16.1	16.3
16.8	16.9	17.1	16.1
17.4	16.8	16.9	17.1
17.1	17.4	16.8	16.9

Pearson Correlations	Y Lag1	Y Lag2	Y Lag3
Y Lag0	0.571	0.498	0.407

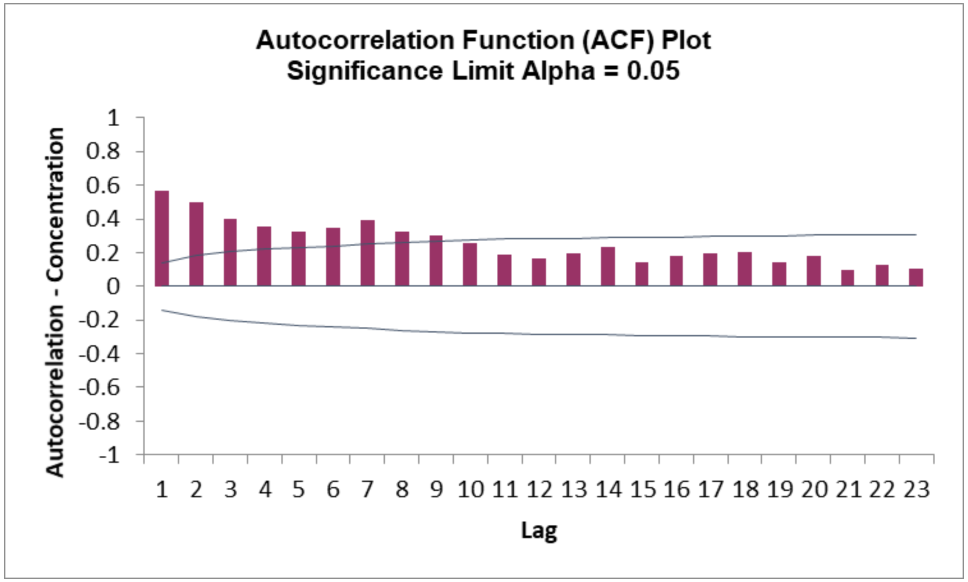
Pearson correlations are used here for demonstration purposes. They are approximately equal to the ACF correlation values.



Any statistically significant correlation ($r_k > 2/\sqrt{N}$) will adversely affect the performance of a Shewhart control chart.

The Ljung-Box test is used to determine if a group of autocorrelations are significant (see formula in Appendix).

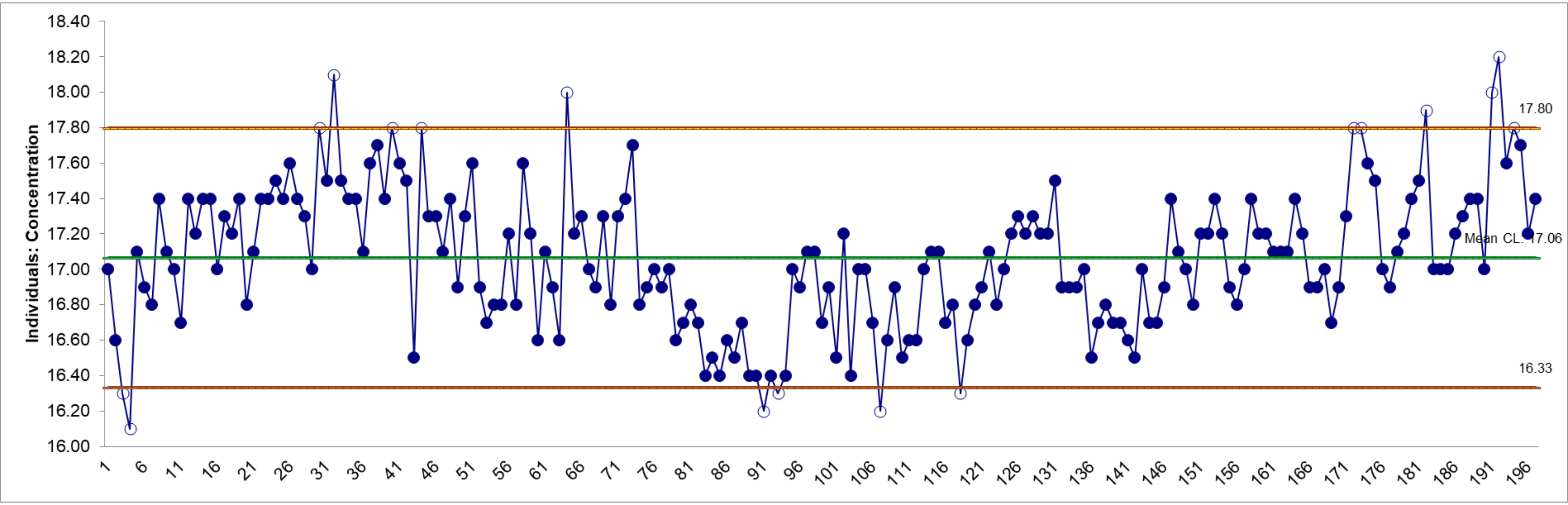
Example 1: Box-Jenkins Series A - Chemical Process Concentration - Autocorrelation Function (ACF) Plot



SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots

Example 1: Chemical Process Concentration - Series A.xlsx - Concentration

Example 1: Box-Jenkins Series A - Chemical Process Concentration - Individuals Control Chart



17 out-of-control data points

SigmaXL > Control Charts > Individuals

Autocorrelation

Guidelines from Woodall & Faltin [10]:

- If possible, one should first attempt to remove the source of the autocorrelation.
- If the source of autocorrelation cannot be removed directly, then it may be possible to model the autocorrelation and use a feedback control scheme to reduce variability about a specified target value.
- If the source of the autocorrelation cannot be removed directly, and feedback control is not a viable option, then it is important to monitor the process with control charts which do not repeatedly give signals due to presence of the autocorrelation.

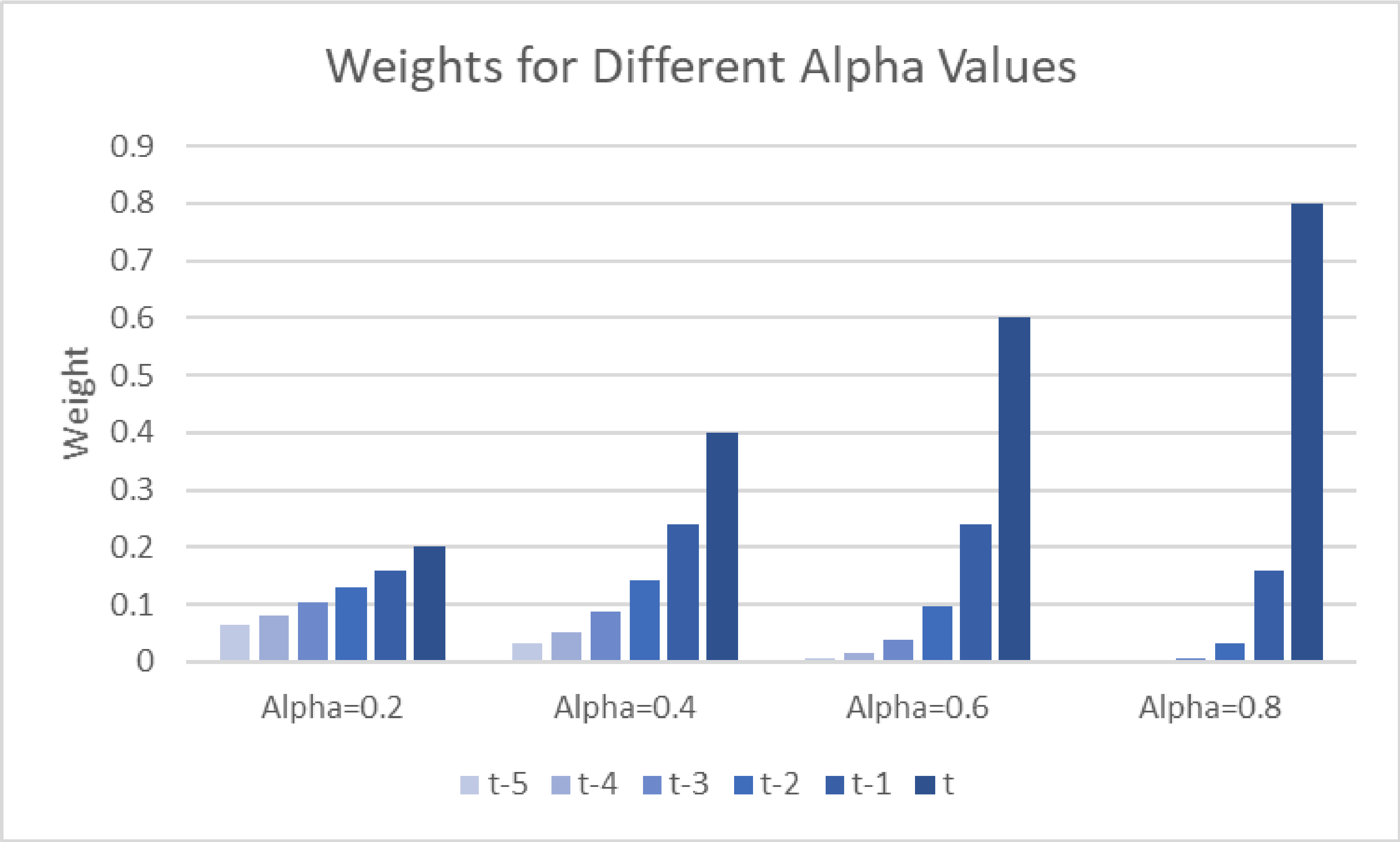
Simple (Single) Exponential Smoothing
Exponentially Weighted Moving Average (EWMA)

Forecasts are calculated using weighted averages, where the weights decrease exponentially as observations come from further in the past with the smallest weights associated with the oldest observations:

$$\hat{y}_{t+1} = \alpha\ y_t + \alpha(1-\alpha)\ y_{t-1} + \alpha(1-\alpha)^2\ y_{t-2} + \cdots$$

where $0 \leq \alpha \leq 1$ is the level smoothing parameter [4].

Simple (Single) Exponential Smoothing
Exponentially Weighted Moving Average (EWMA)



Simple (Single) Exponential Smoothing
Exponentially Weighted Moving Average (EWMA)

- An equivalent formulation for simple exponential smoothing is:

$$\hat{y}_{t+1} = \alpha y_t + (1 - \alpha)\hat{y}_t$$

  with the starting forecast value (initial level) $\hat{y}_1$ typically estimated as y_1.

- The formula used for EWMA is the same, but the smoothing parameter λ is typically used instead of α and X_t instead of y_t:

$$\mathrm{EWMA}_{t+1} = \lambda X_t + (1 - \lambda)\mathrm{EWMA}_t$$

  with the starting forecast value EWMA_1 estimated as the data mean or target value.

Simple (Single) Exponential Smoothing
Exponentially Weighted Moving Average (EWMA)

- In the case of an EWMA control chart, the smoothing parameter λ is determined by desired average run length characteristics and is typically 0.2.
- For forecasting or SPC for autocorrelated data, the smoothing parameter and initial level are determined by minimizing the sum-of-square forecast errors (residuals):

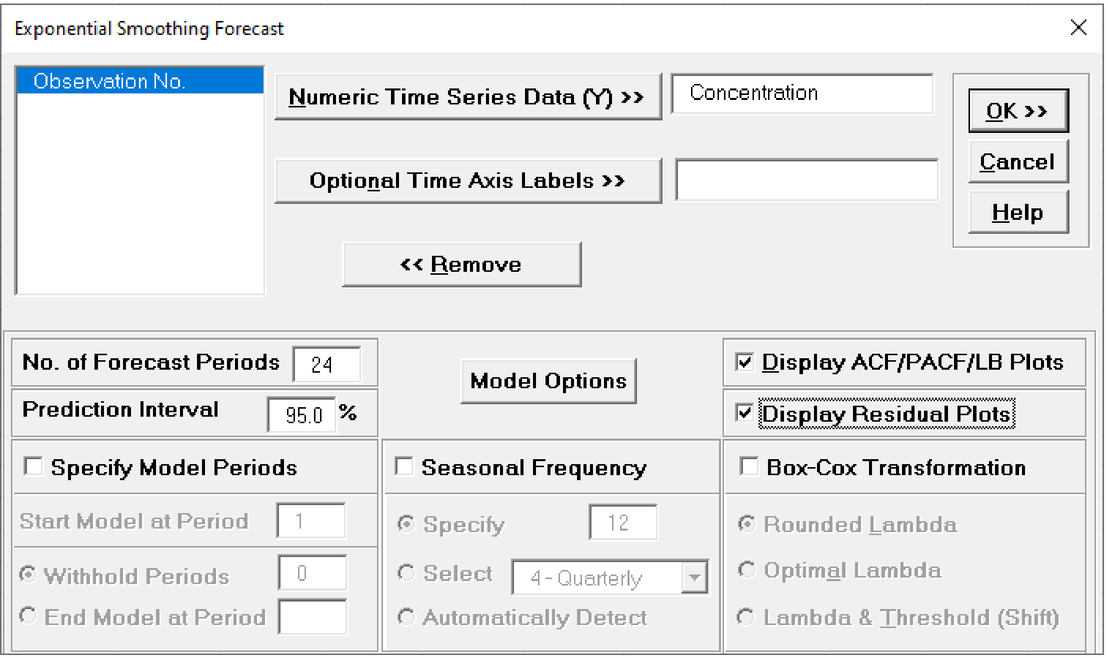
$$\text{SSE} = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 = \sum_{t=1}^{T} e_t^2 .$$

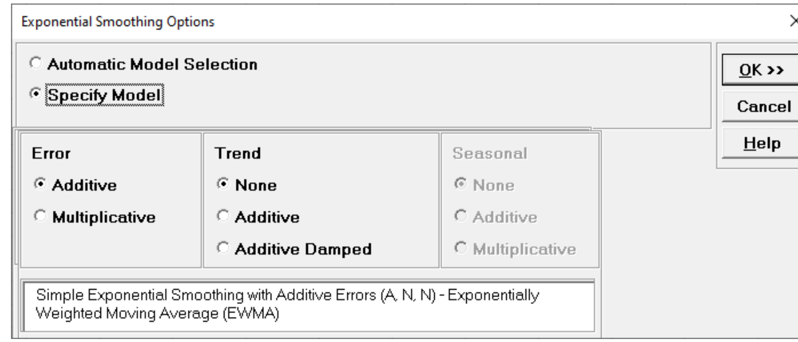
- This involves non-linear minimization methods like Newton-Raphson or Nelder-Mead Simplex.

Simple (Single) Exponential Smoothing
Exponentially Weighted Moving Average (EWMA)

- As usual for any statistical model, the residuals should be normal, independent and identically distributed.
- If this is achieved, this also means that the assumptions for a Shewhart control chart are satisfied.

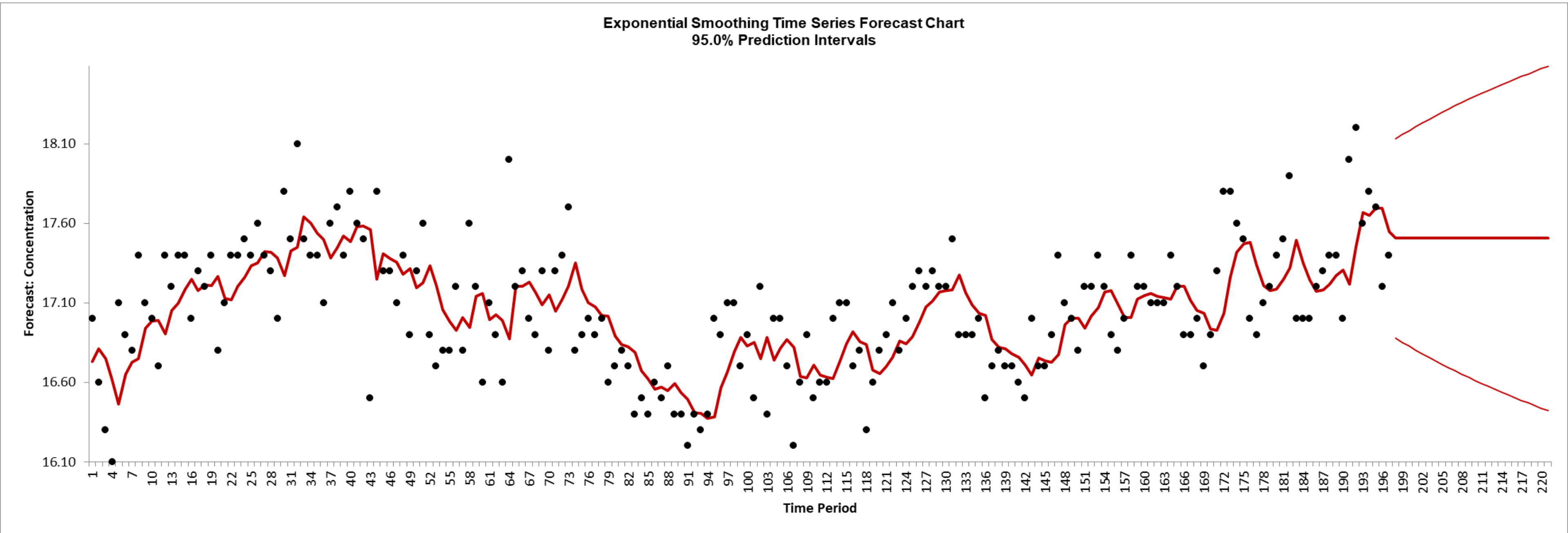
Example 1: Box-Jenkins Series A - Chemical Process Concentration - Simple Exponential Smoothing (EWMA) Time Series Forecast





SigmaXL > Time Series Forecasting > Exponential Smoothing Forecast > Forecast

Example 1: Box-Jenkins Series A - Chemical Process Concentration - Simple Exponential Smoothing (EWMA) Time Series Forecast



Exponential Smoothing Model: Concentration

Model Type: Simple Exponential Smoothing with Additive Errors (A, N, N) - Exponentially Weighted Moving Average (EWMA)

Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

Exponential Smoothing Model Information	
Seasonal Frequency	1
Model selection criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

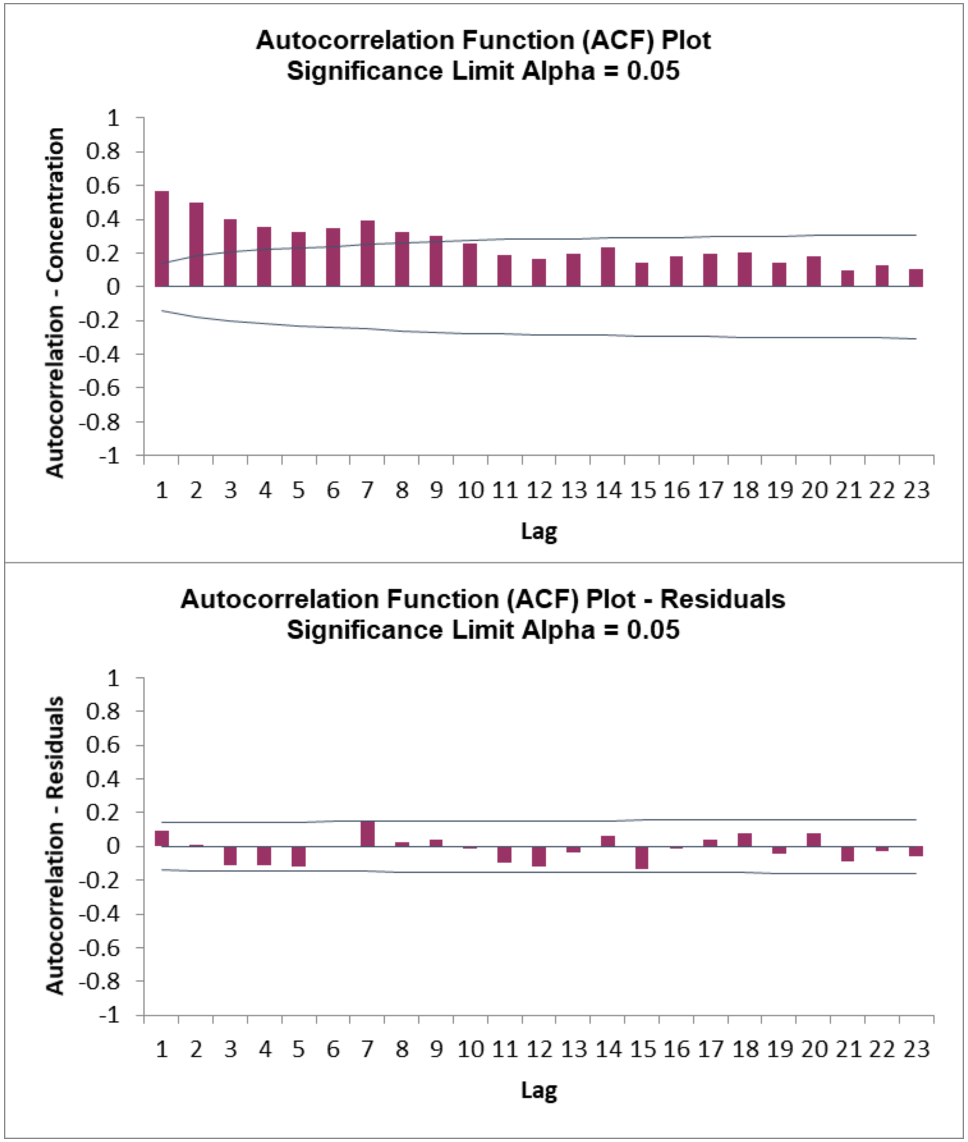
Parameter Estimates	
Term	Coefficient
alpha (level smoothing)	0.294785988
l (level initial state)	16.73121246

Exponential Smoothing Model Statistics	
No. Observations	197
DF	194
StDev	0.319007644
Variance	0.101765877
Log-Likelihood	-293.8036067
AICc	593.7315658
AIC	593.6072135
BIC	603.4568246

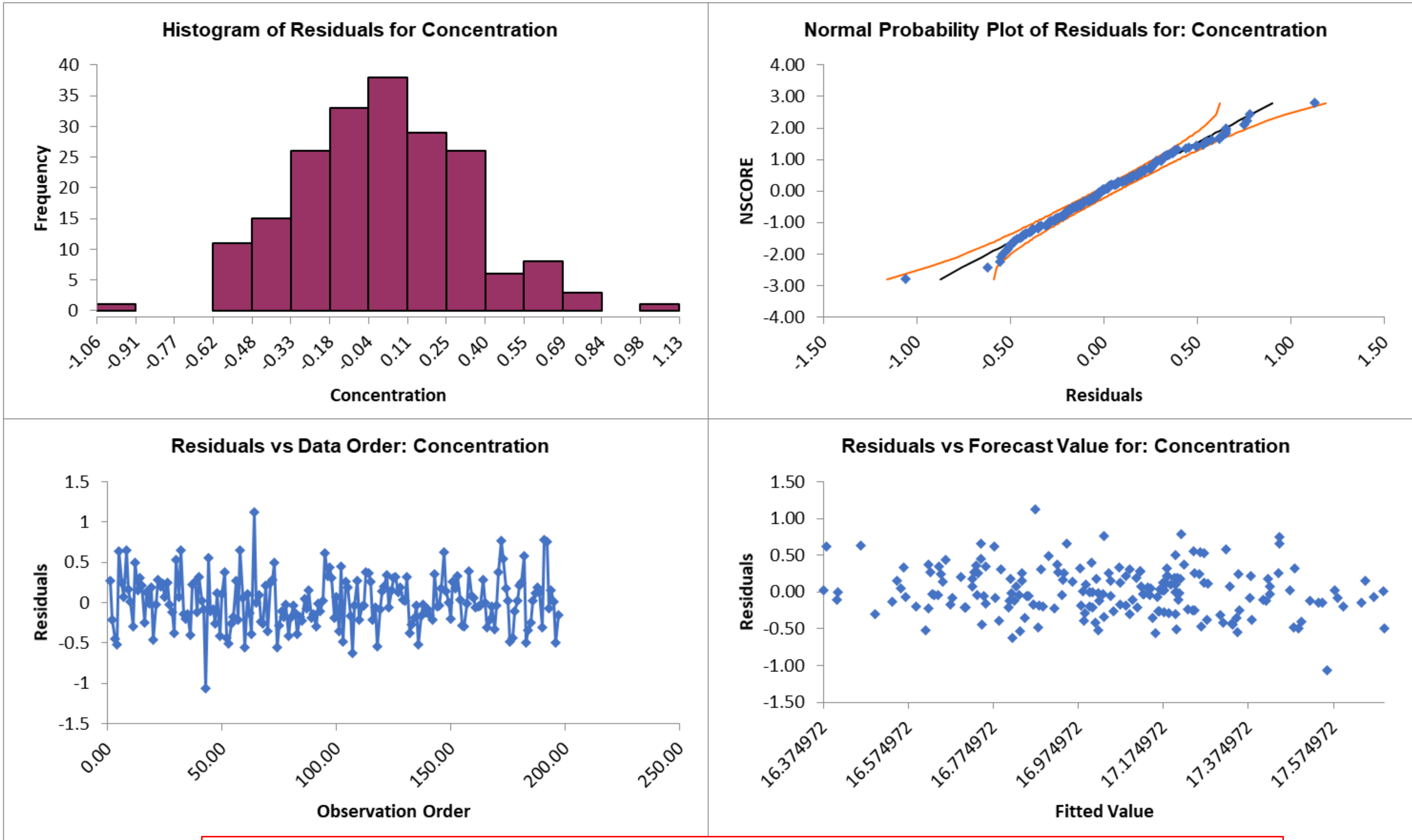
Forecast Accuracy			
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast	Out-of-Sample (Withhold) Full Period Forecast
N	197		
RMSE	0.316569334		
MAE	0.247329038		
MAPE	1.446520183		
MASE	0.897712804		

Simple Exponential Smoothing (EWMA) specified. 95% Prediction Intervals for forecast.

Example 1: Box-Jenkins Series A - Chemical Process Concentration - ACF Plots (Raw Data versus Residuals)

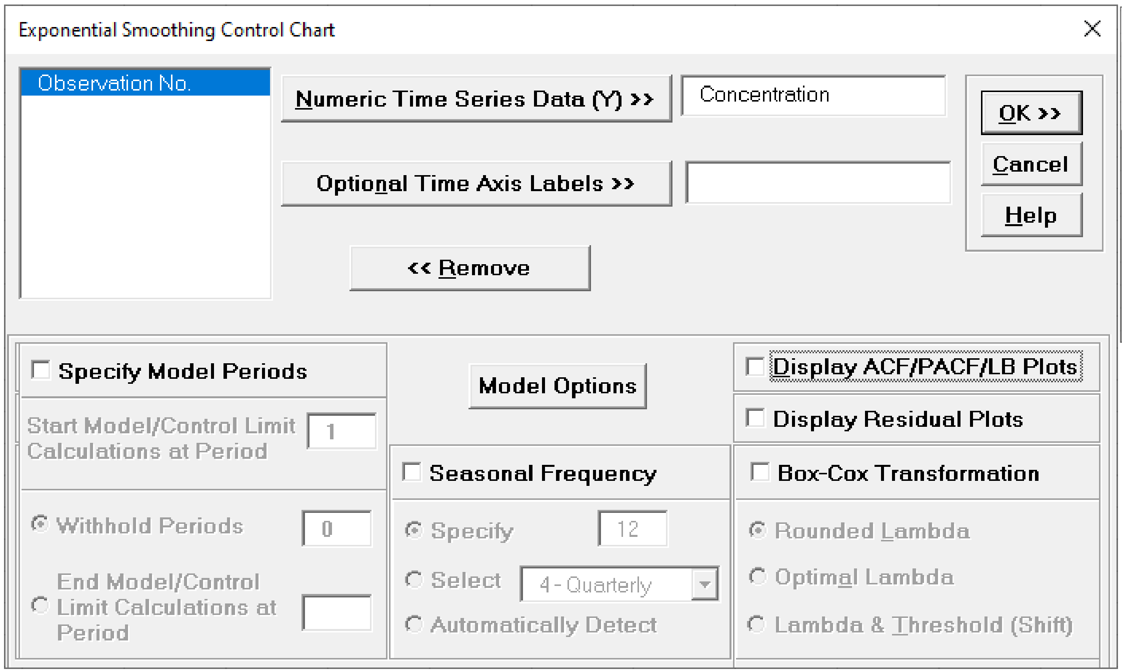


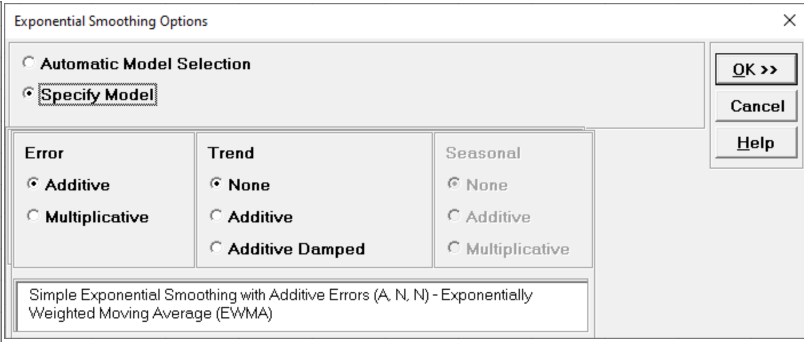
Example 1: Box-Jenkins Series A - Chemical Process Concentration - Residuals



Residuals look good – approximately normal with equal variance.

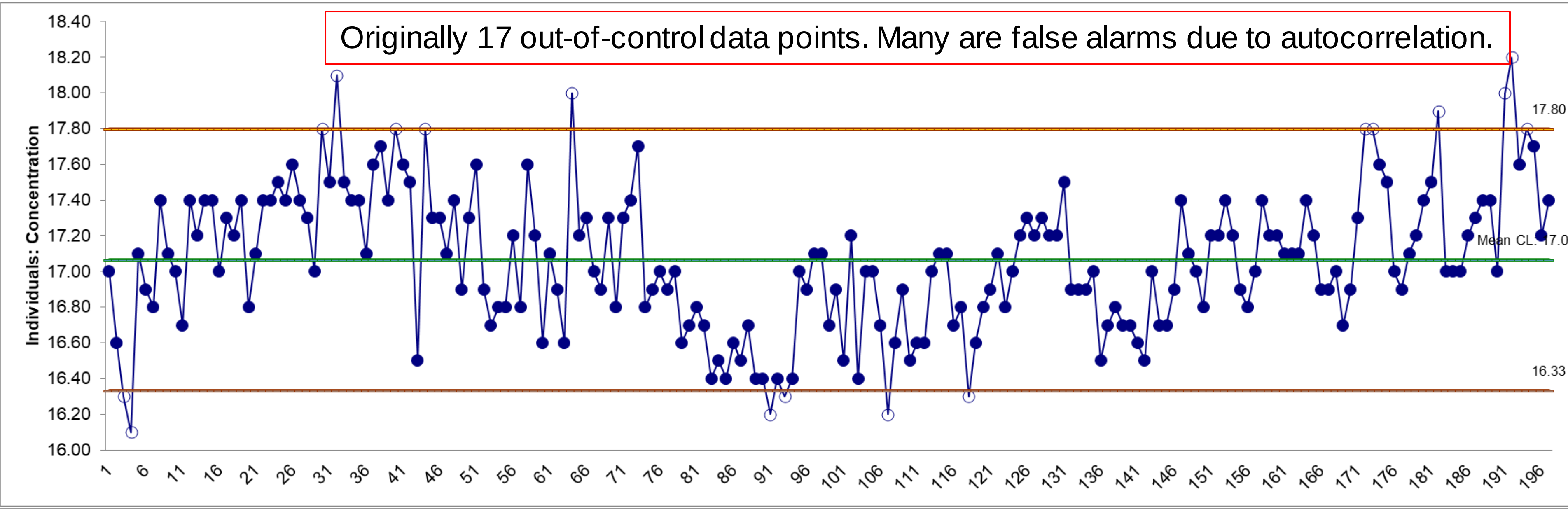
Example 1: Box-Jenkins Series A - Chemical Process Concentration - Simple Exponential Smoothing (EWMA) Control Chart

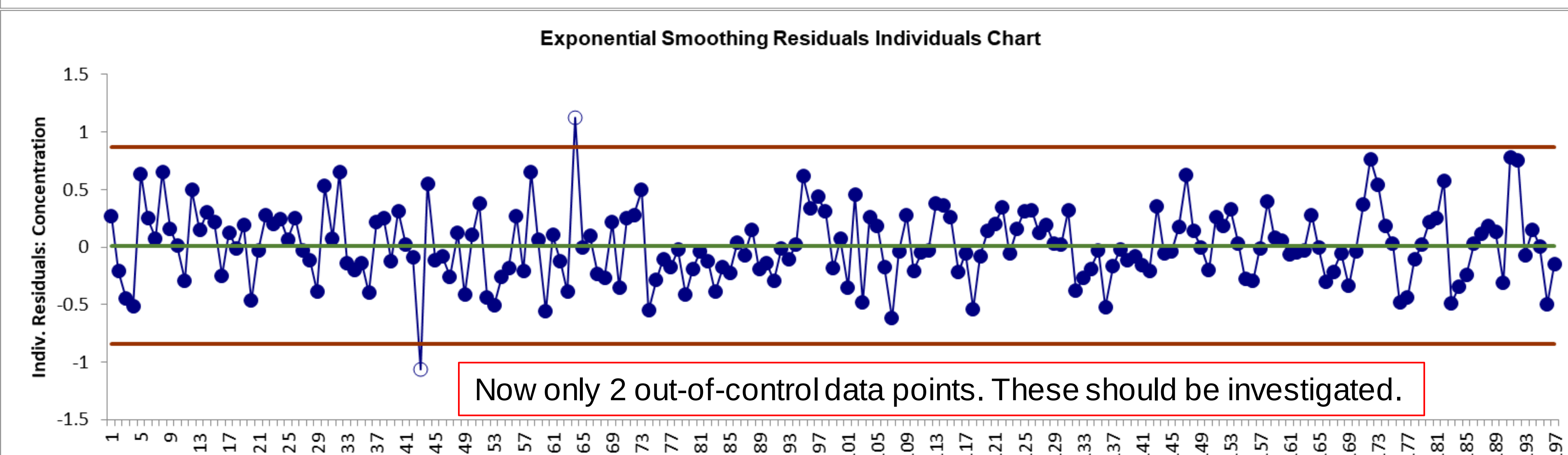




SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Control Chart

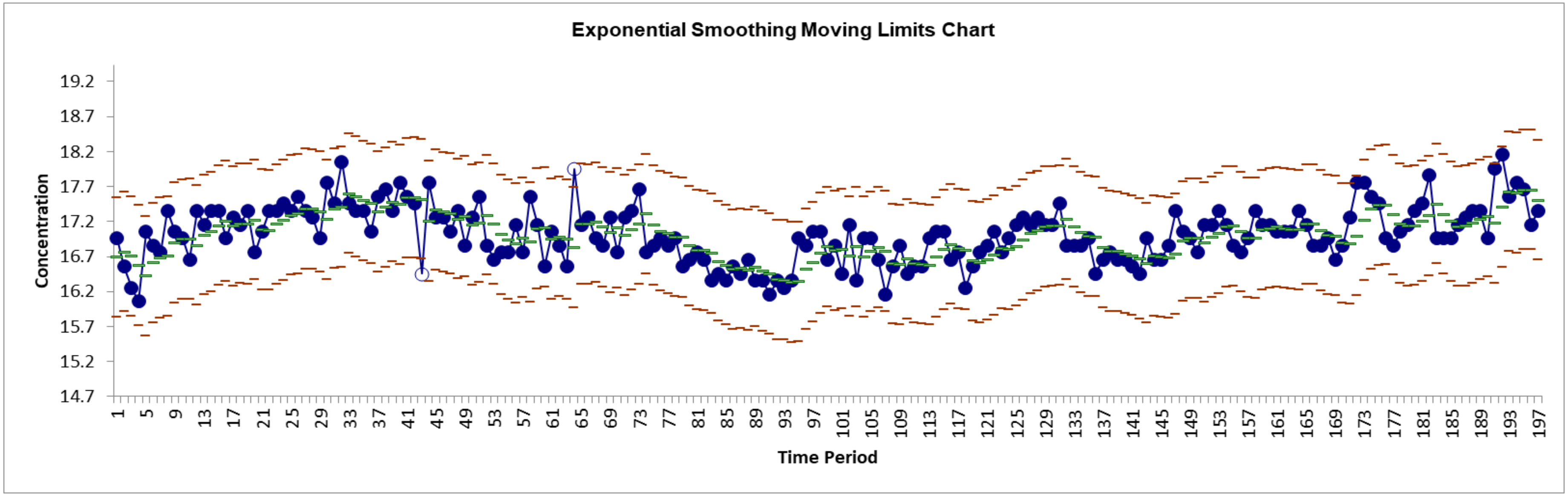
Example 1: Box-Jenkins Series A - Chemical Process Conc. - Individuals Control Chart (Raw Data versus Residuals)





SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Control Chart

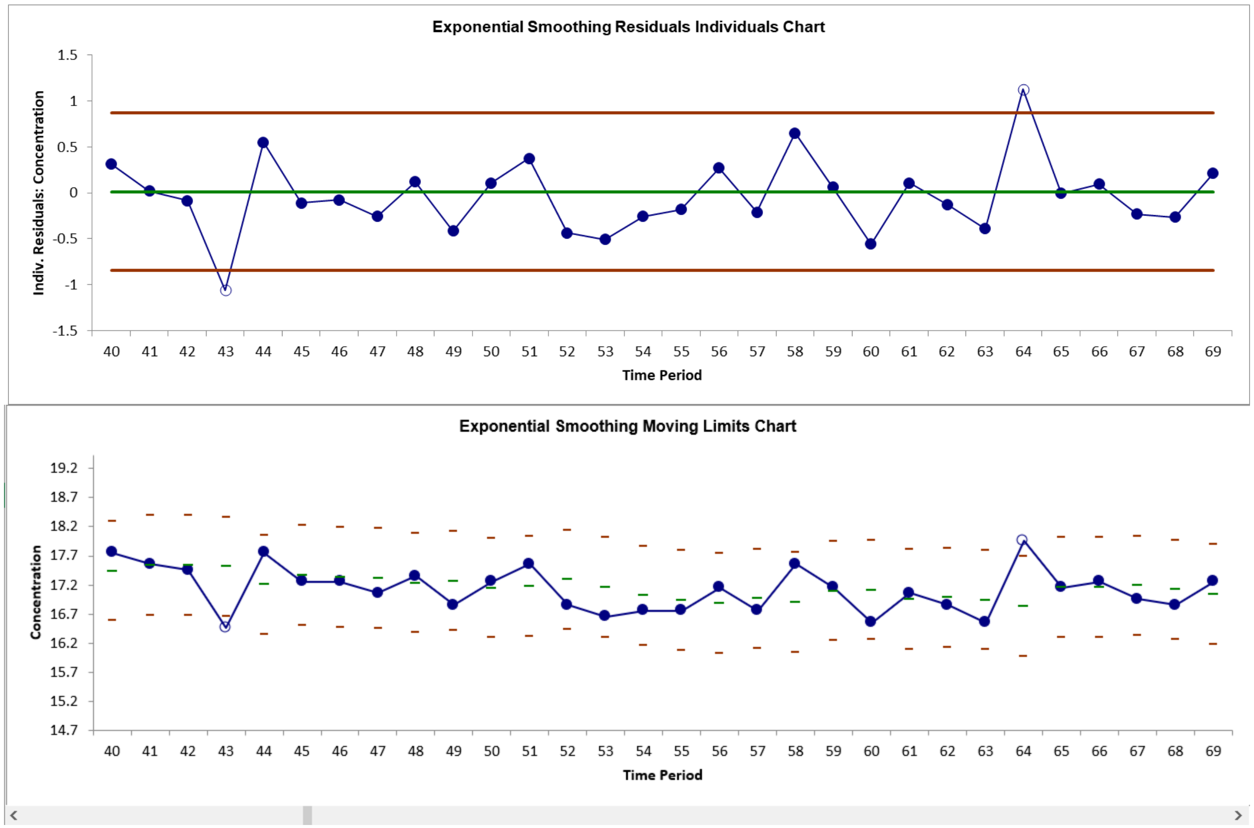
Example 1: Box-Jenkins Series A - Chemical Process Concentration – Moving Limits Control Chart

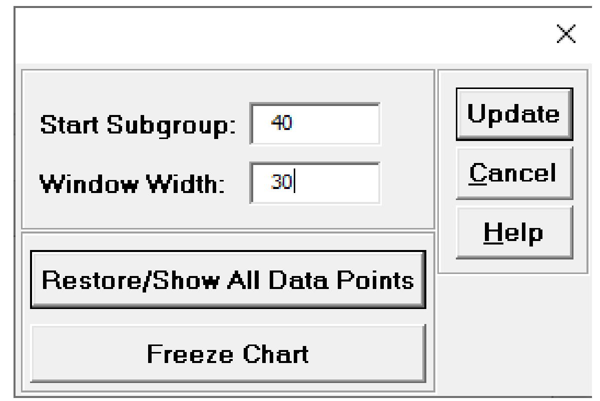


The Moving Limits chart uses the one step prediction as the center line, so the control limits will move with the center line.

SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Control Chart

Example 1: Box-Jenkins Series A - Chemical Process Concentration - Individuals Control Chart: Enable Scrolling





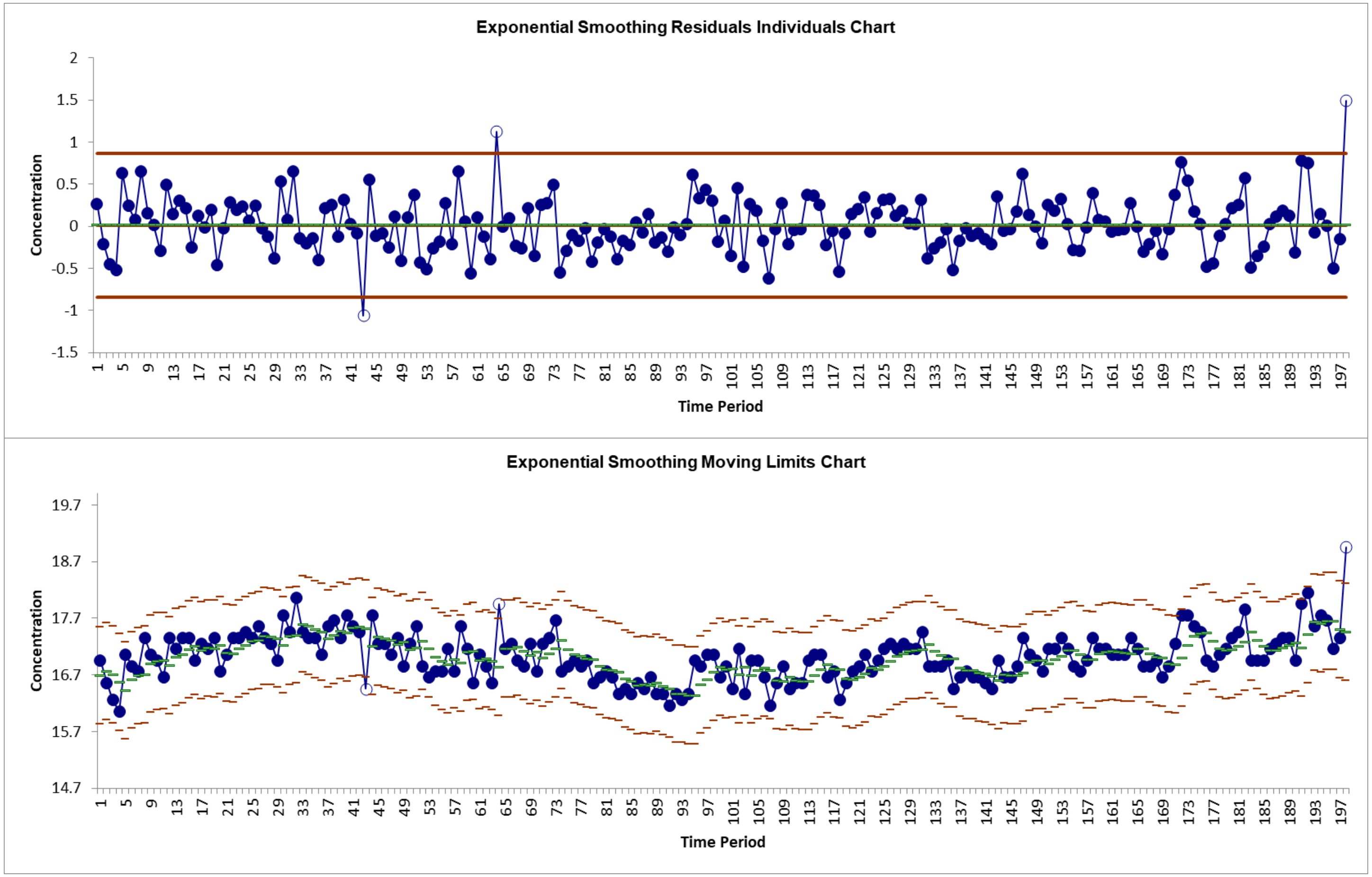
SigmaXL Chart Tools > Enable Scrolling

Example 1: Box-Jenkins Series A - Chemical Process Concentration - Individuals Control Chart: Add Data

Now we will add a new data point to the Series A Concentration Data. The residuals will be computed using the same model as above without re-estimation of the model parameters or recalculation of the control limits. This is also known as the "Phase II" application of a Control Chart, where an out-of-control signal should lead to an investigation into the assignable cause and corrective action or process adjustment applied.

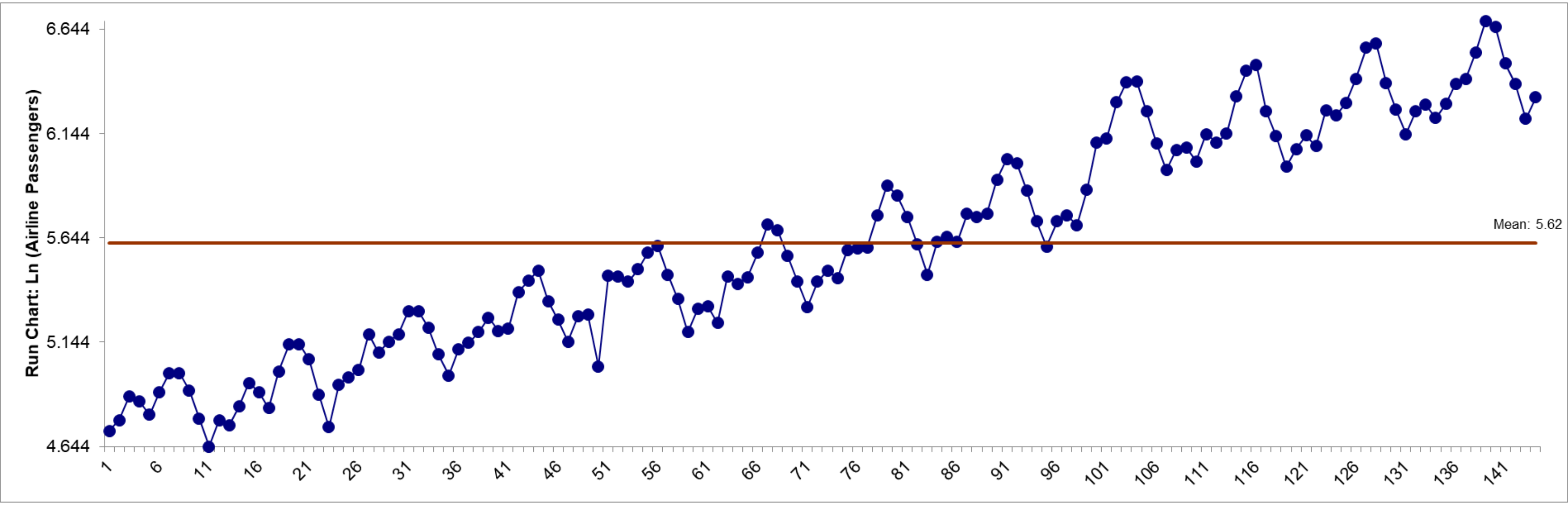
196	195	17.7
197	196	17.2
198	197	17.4
199	198	19

Example 1: Box-Jenkins Series A - Chemical Process Concentration - Individuals Control Chart: Add Data



SigmaXL Chart Tools > Add Data to this Control Chart

Example 2a: Box-Jenkins Series G – Ln(Monthly Airline Passengers) - Run Chart



Data modified with negative outlier at 50 (-.25) and level shift (+.25) starting at 100.

Data shows strong positive trend and strong seasonality (monthly data).

SigmaXL > Time Series Forecasting > Run Chart

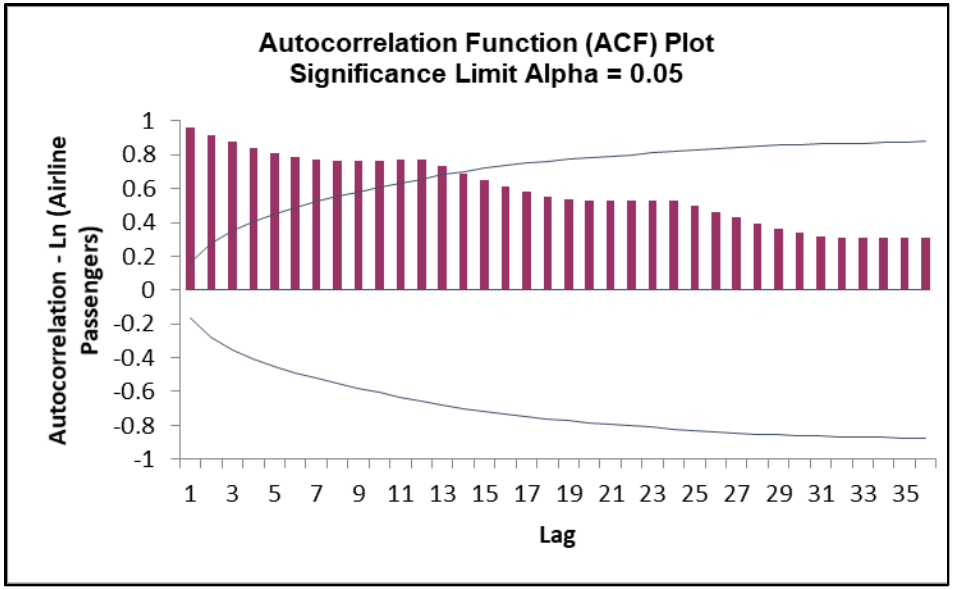
Example 2: Airline Passengers Modified.xlsx – Ln(Airline Passengers)

Example 2a: Box-Jenkins Series G – Ln(Monthly Airline Passengers) - Individuals Control Chart

The control chart signals here are meaningless.

SigmaXL > Control Charts > Individuals

Example 2a: Box-Jenkins Series G – Ln(Monthly Airline Passengers) - Autocorrelation (ACF) Plot



SigmaXL > Time Series Forecasting > Autocorrelation (ACF/PACF) Plots

Error, Trend, Seasonal (ETS) Exponential Smoothing Models

- Error, Trend, Seasonal (ETS) models expand on simple exponential smoothing to accommodate trend and seasonal components as well as additive or multiplicative errors.
- Simple Exponential Smoothing is an Error Model.
- Error, Trend model is Holt’s Linear, also known as double exponential smoothing.

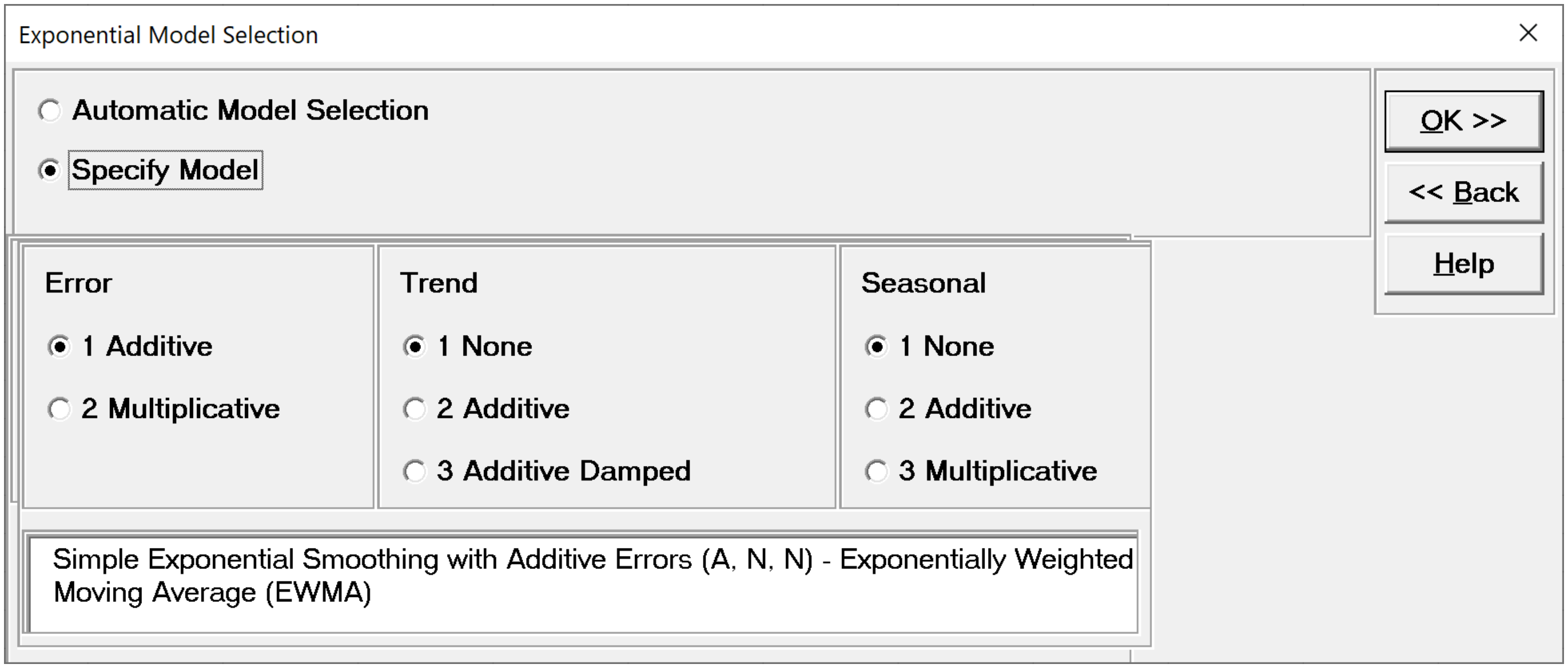
Error, Trend, Seasonal (ETS) Exponential Smoothing Models

- Error, Trend, Seasonal model is Holt-Winters, also known as triple exponential smoothing.
  - Seasonal frequency must be specified:
    - Quarterly data = 4 (observations per year)
    - Monthly data = 12 (observations per year)
    - Daily data = 7 (observations per week)
    - Hourly data = 24 (observations per day)
  - Frequency is the number of observations per “cycle”. This is the opposite of the definition of frequency in physics, or in engineering Fourier analysis, where “period” is the length of the cycle, and “frequency” is the inverse of period.

Reference: https://robjhyndman.com/hyndsight/seasonal-periods/

See Appendix for more information on ETS Models

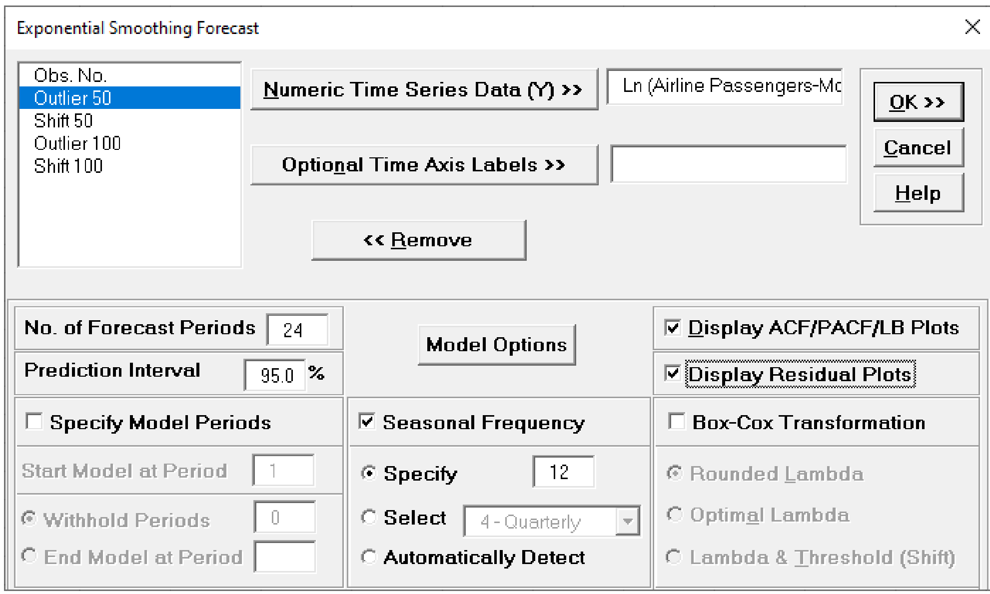
Error, Trend, Seasonal (ETS) models
Hyndman's Taxonomy

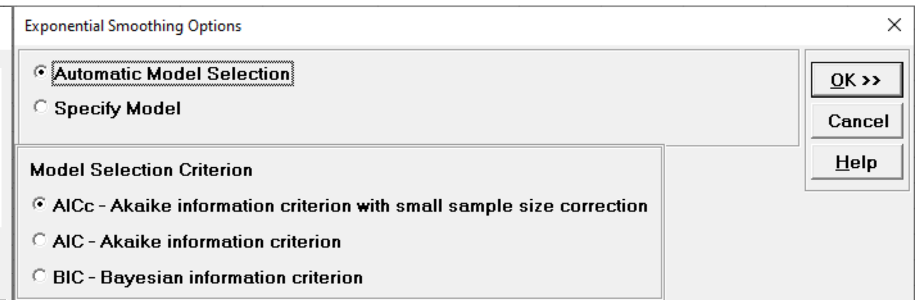


SigmaXL > Time Series Forecasting > Exponential Smoothing Forecast > Forecast

Shorthand notation (Error, Trend, Seasonal): (A, N, N) is Additive Error, No Trend, No Seasonal.

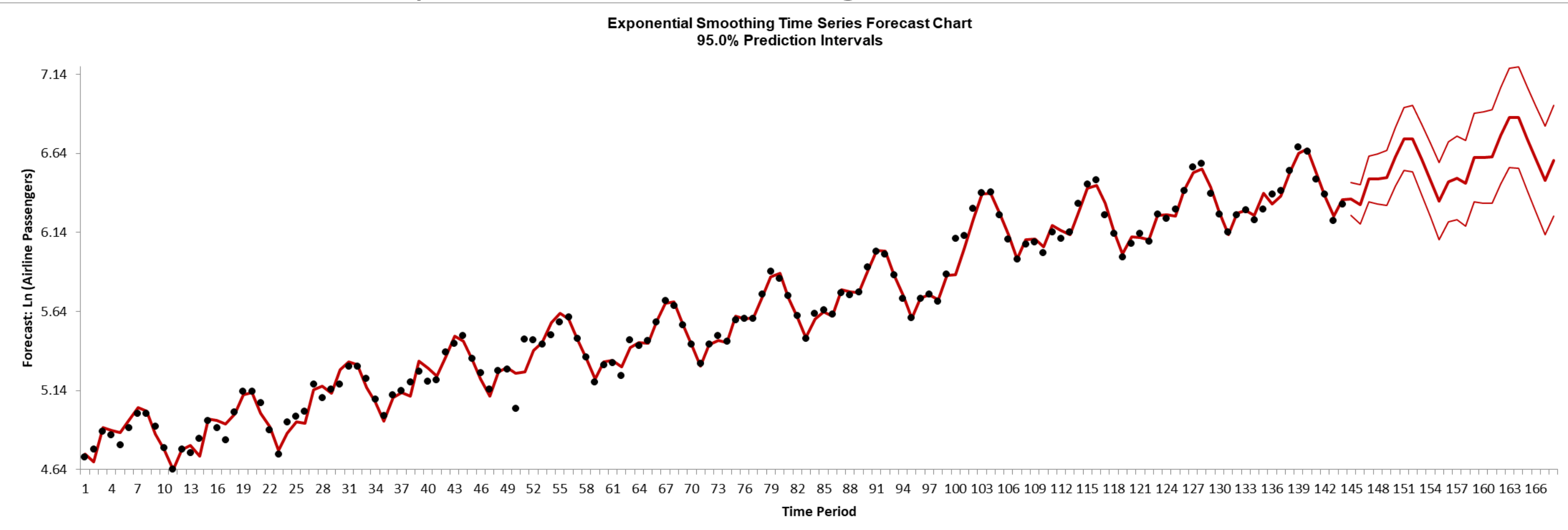
Example 2b: Box-Jenkins Series G – Ln(Monthly Airline Passengers) – Exponential Smoothing Forecast with Automatic Model Selection





SigmaXL > Time Series Forecasting > Exponential Smoothing Forecast > Forecast

Example 2b: Box-Jenkins Series G – Ln(Monthly Airline Passengers) Seasonal Exponential Smoothing with Trend (Holt-Winters)



Exponential Smoothing Model: Ln (Airline Passengers)

Model Type: Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters) (A, A, A)

Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

Exponential Smoothing Model Information	
Seasonal Frequency	12
Model selection criterion	AICc
Box-Cox Transformation	N/A
Lambda	
Threshold	

Parameter Estimates	
Term	**Coefficient**
alpha (level smoothing)	0.674949361
beta (trend smoothing)	0.0001
gamma (seasonal smoothing)	0.0001
l (level initial state)	4.821176207
b (trend initial state)	0.01120374
s1 (seasonal initial state)	-0.106060076
s2 (seasonal initial state)	-0.218770169
s3 (seasonal initial state)	-0.073902798
s4 (seasonal initial state)	0.065376172
s5 (seasonal initial state)	0.211146551
s6 (seasonal initial state)	0.221577218
s7 (seasonal initial state)	0.118793494
s8 (seasonal initial state)	-0.00223835
s9 (seasonal initial state)	0.001331568

Exponential Smoothing Model Statistics	
No. Observations	144
DF	127
StDev	0.052717869
Variance	0.002779174
Log-Likelihood	74.98184762
AICc	-111.1065524
AIC	-115.9636952
BIC	-65.47686914

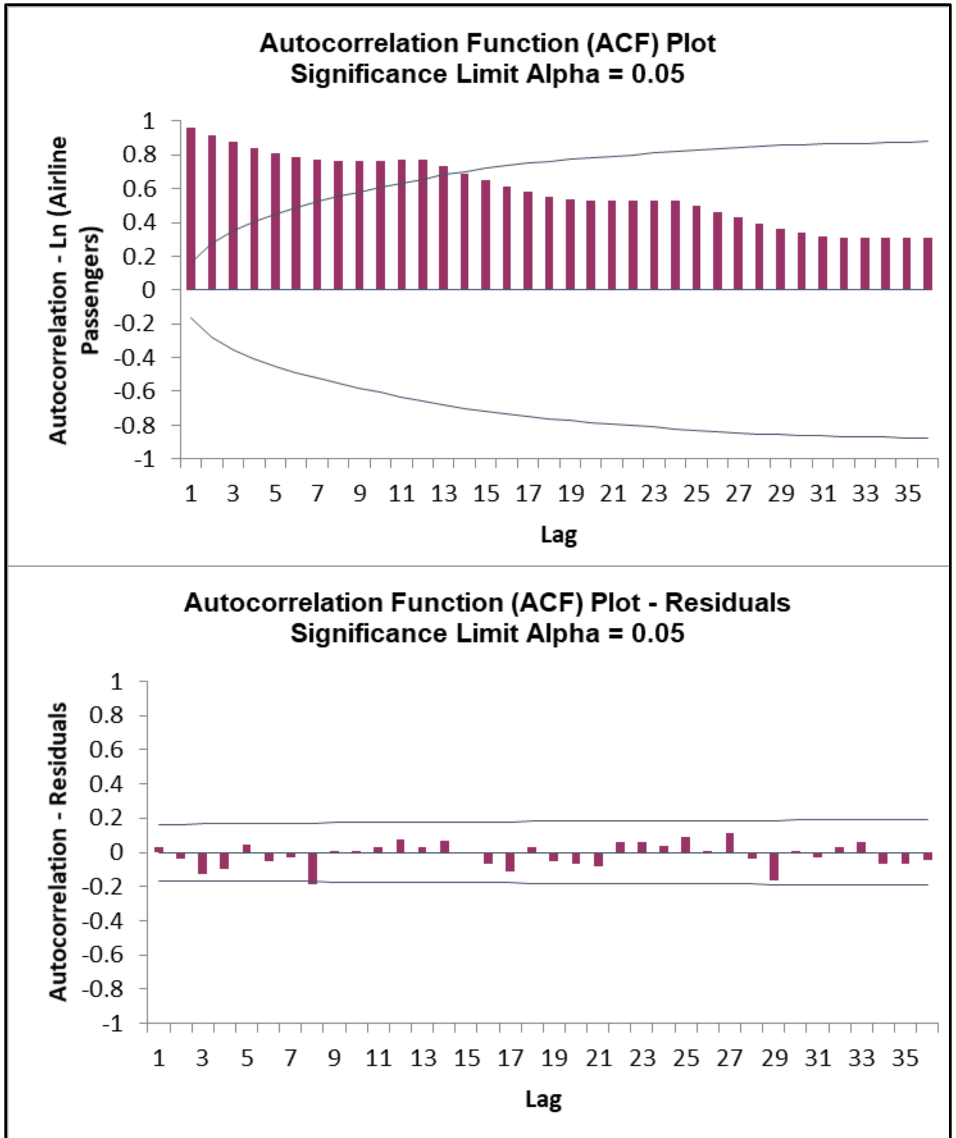
Forecast Accuracy			
Metric	**In-Sample (Estimation) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) One-Step-Ahead Forecast**	**Out-of-Sample (Withhold) Full Period Forecast**
N	144		
RMSE	0.049508351		
MAE	0.033506668		
MAPE	0.609538124		
MASE	0.230072112		

Chart Area

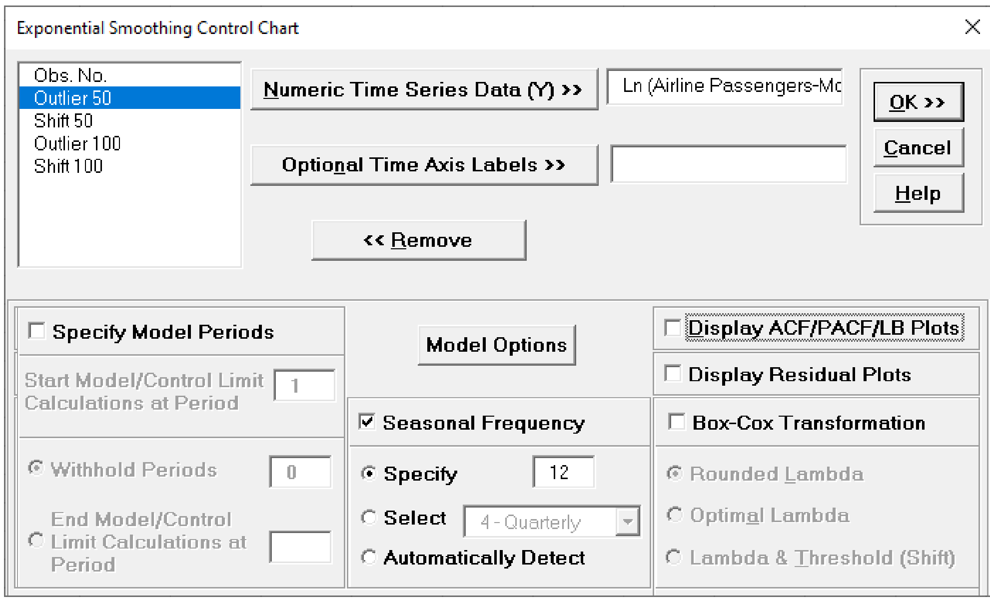
ETS Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters) (A, A, A) **automatically selected**. Seasonal Frequency = 12 (Monthly data).

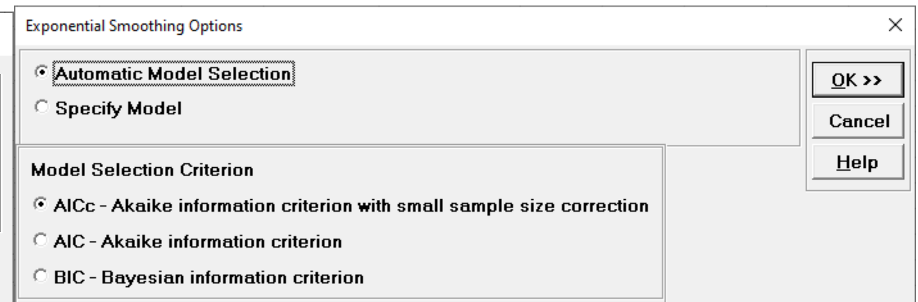
SigmaXL > Time Series Forecasting > Exponential Smoothing Forecast > Forecast

Example 2b: Box-Jenkins Series G – Ln(Monthly Airline Passengers) - ACF Plots (Raw Data versus Residuals)



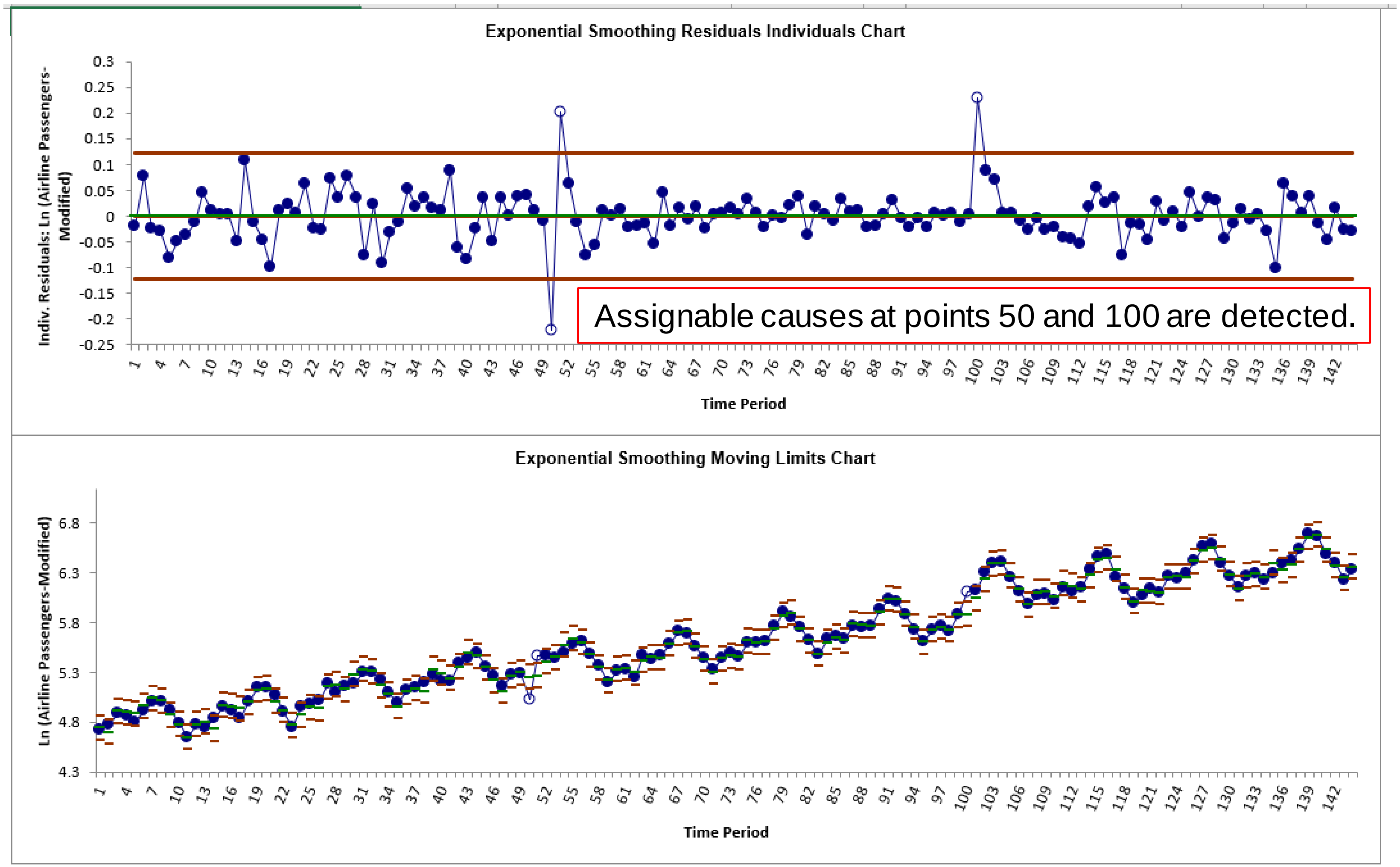
Example 2b: Box-Jenkins Series G – Ln(Monthly Airline Passengers) – Exponential Smoothing Control Chart with Automatic Model Selection





SigmaXL > Time Series Forecasting > Exponential Smoothing Control Chart > Control Chart

Example 2b: Box-Jenkins Series G – Ln(Monthly Airline Passengers) Exponential Smoothing Control Charts



Exponential Smoothing Model: Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters) (A, A, A) - Model Automatically Selected
Model Periods: All observations are used in the Exponential Smoothing model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

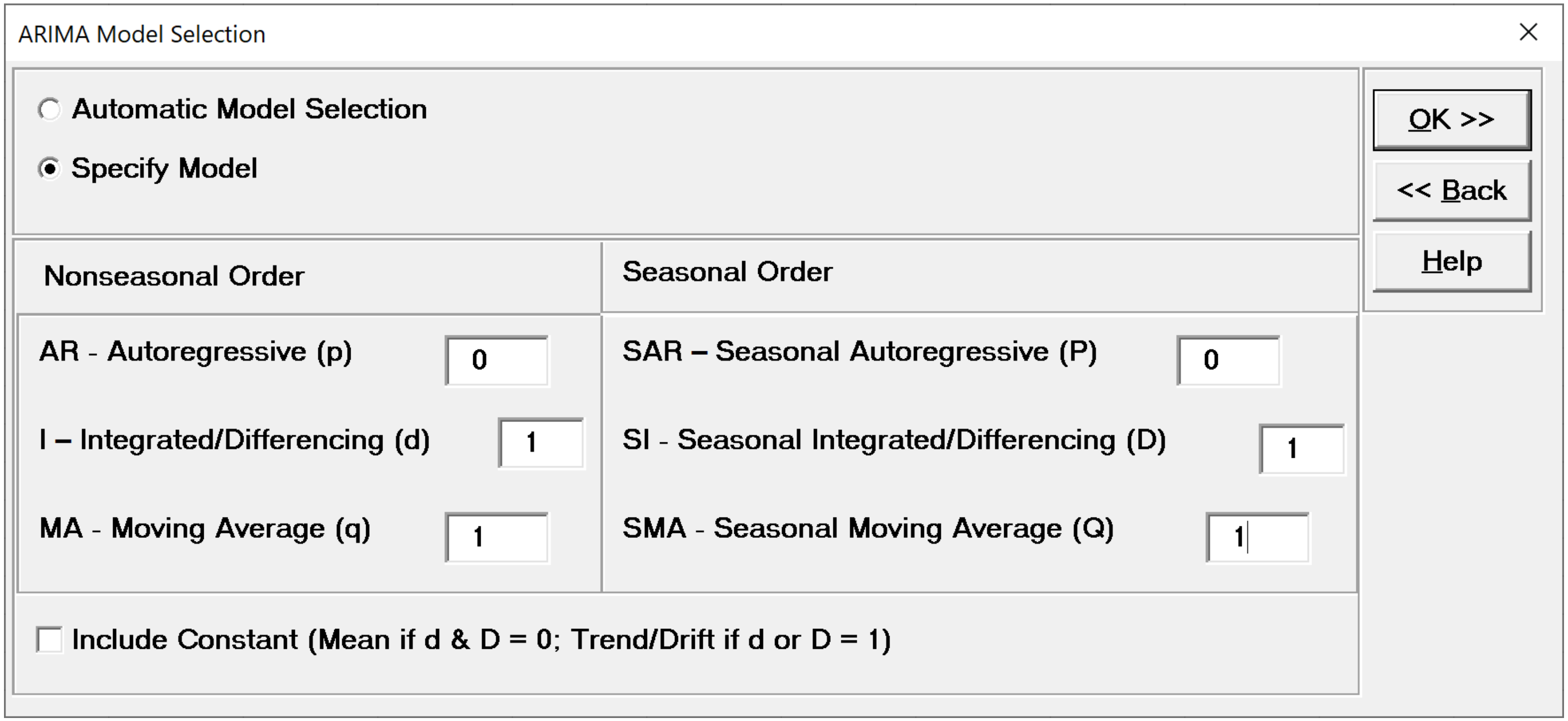
Box-Jenkins AutoRegressive Integrated Moving Average (ARIMA) Models

- An ARIMA model includes an Autoregressive (AR) component of order p, an Integrated/Differencing component of order d and a Moving Average component of order q and an optional constant.

- An ARIMA Seasonal model includes a Seasonal Autoregressive (SAR) component of order P, a Seasonal Integrated/Differencing component of order D and a Seasonal Moving Average component of order Q.

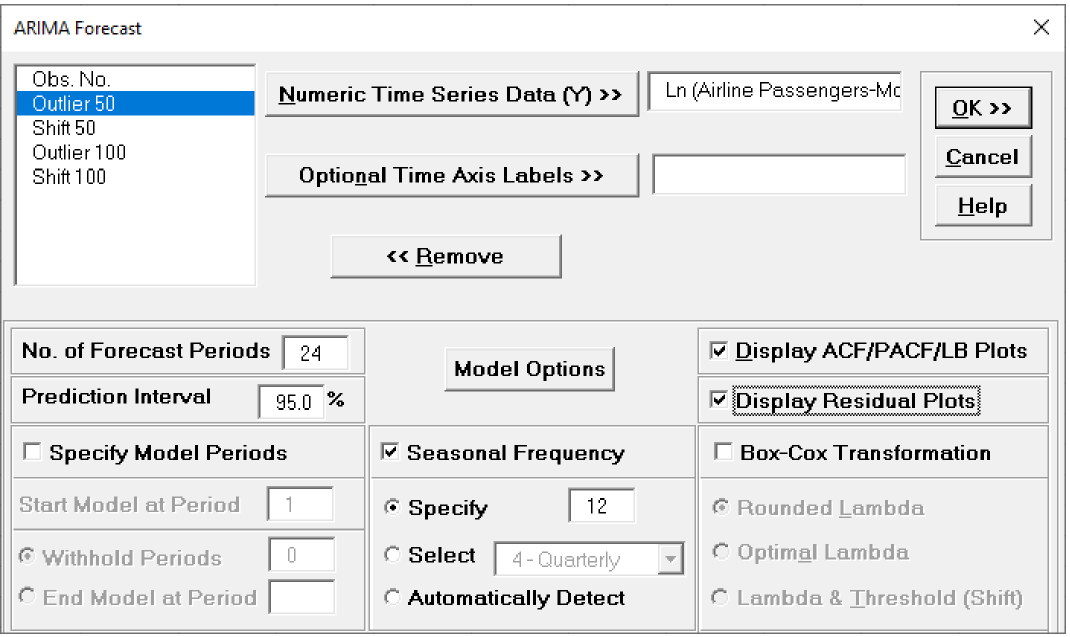
See Appendix for more information on ARIMA Models

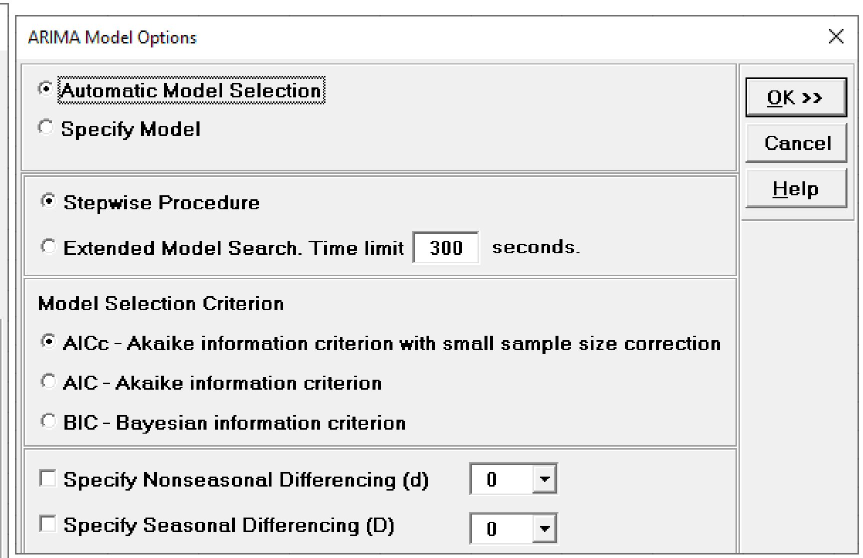
Box-Jenkins AutoRegressive Integrated Moving Average (ARIMA) Models



SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast

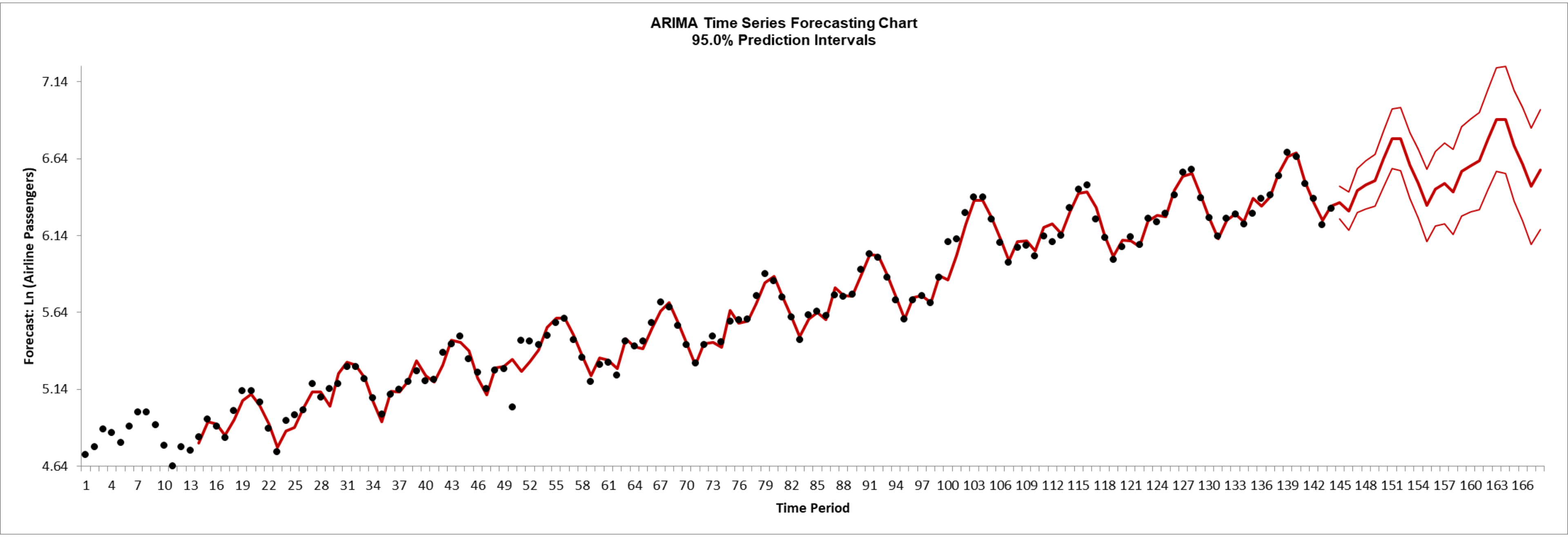
Example 2c: Box-Jenkins Series G – Ln(Monthly Airline Passengers) – ARIMA Forecast with Automatic Model Selection





SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast

Example 2c: Box-Jenkins Series G – Ln(Monthly Airline Passengers) Seasonal ARIMA Time Series Forecast



ARIMA Model: Ln (Airline Passengers)

Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

ARIMA Model Summary	
AR Order (p)	0
I Order (d)	1
MA Order (q)	1
SAR Order (P)	0
SI Order (D)	1
SMA Order (Q)	1
Seasonal Frequency	12
Include Constant	0
No. of Predictors	0
Model selection criterion	AICc
Box-Cox Transformation	N/A
Lambda	

Parameter Estimates				
Term	Coefficient	SE Coefficient	T	P
MA_1	0.387990136	0.091662663	4.2328046	0.0000
SMA_1	0.686391161	0.074905258	9.1634577	0.0000

ARIMA Model Statistics	
No. Observations	144
DF	129
StDev	0.054403004
Variance	0.002959687
Log-Likelihood	191.5831814
AICc	-376.9773864
AIC	-377.1663627
BIC	-368.5407708

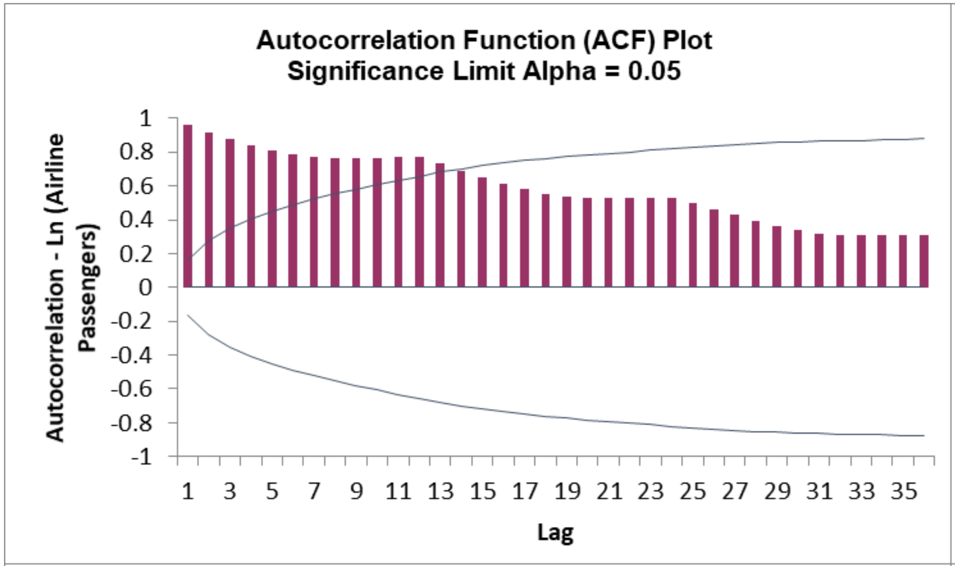
Residuals Randomness Runs Test	
P-Value	0.2925

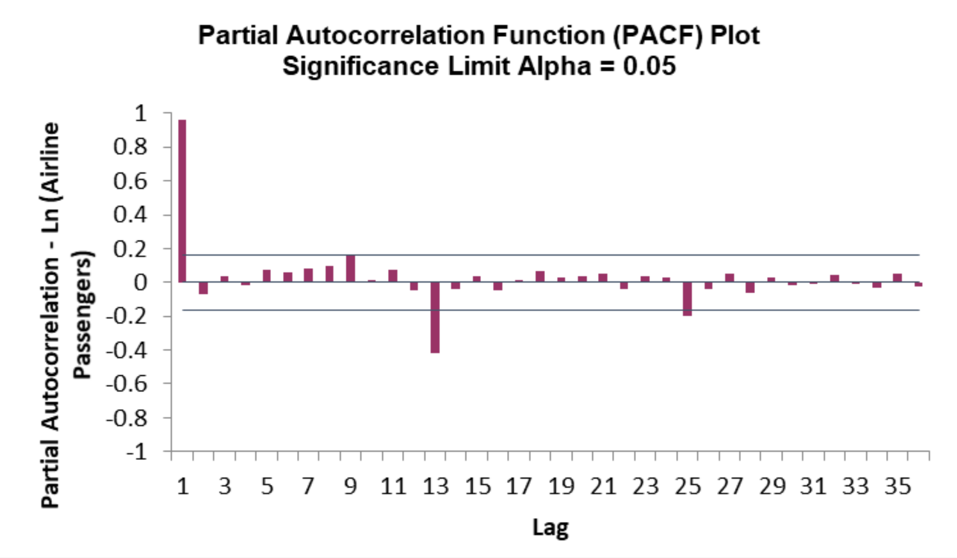
Forecast Accuracy		
Metric	In-Sample (Estimation) One-Step-Ahead Forecast	Out-of-Sample (Withhold) One-Step-Ahead Forecast
N	144	
RMSE	0.055615558	
MAE	0.036285476	
MAPE	0.648152199	
MASE	0.249152676	

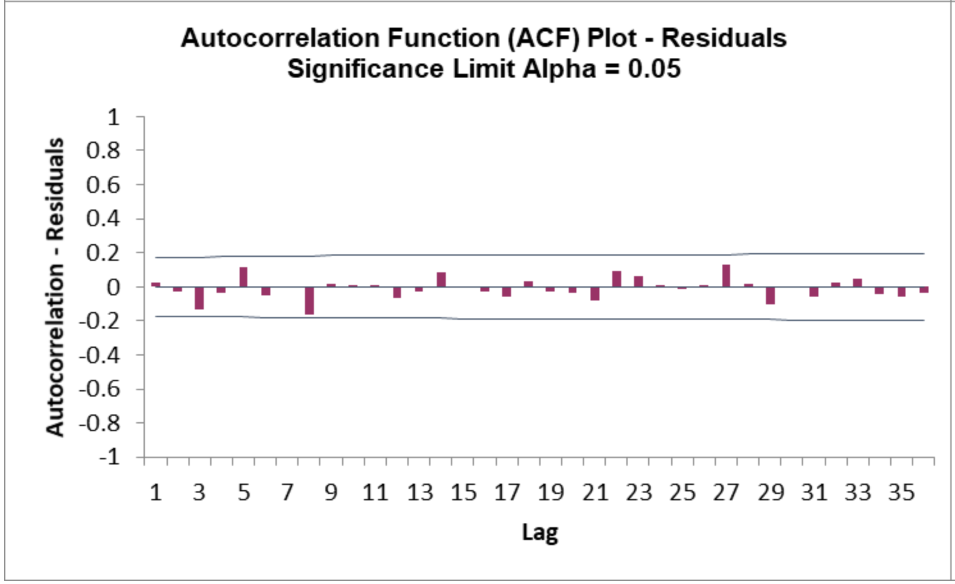
ARIMA (0,1,1) (0,1,1) **automatically selected**. Seasonal Frequency = 12 (Monthly data).

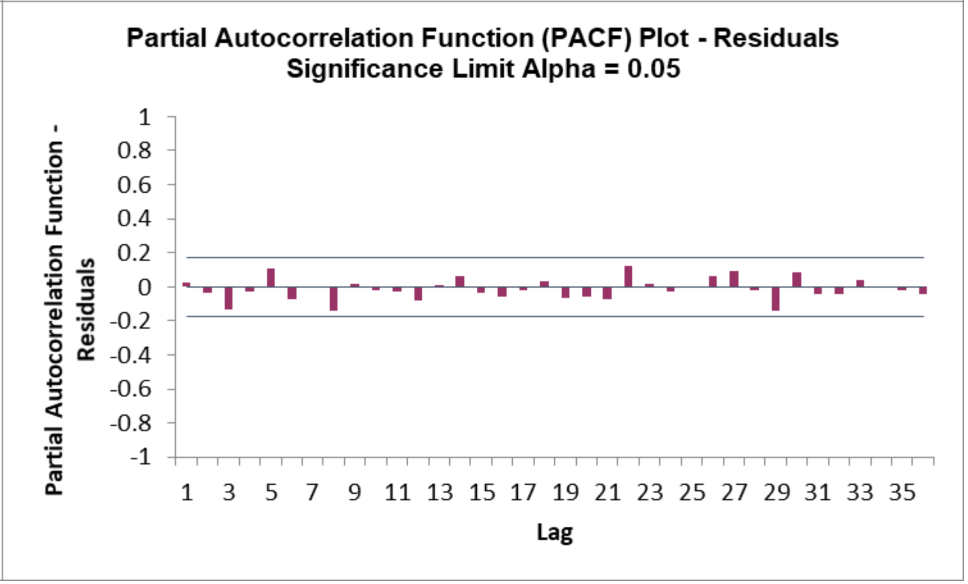
SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast

Example 2c: Box-Jenkins Series G – Ln(Monthly Airline Passengers) ACF/PACF Plots – Raw Data versus Residuals

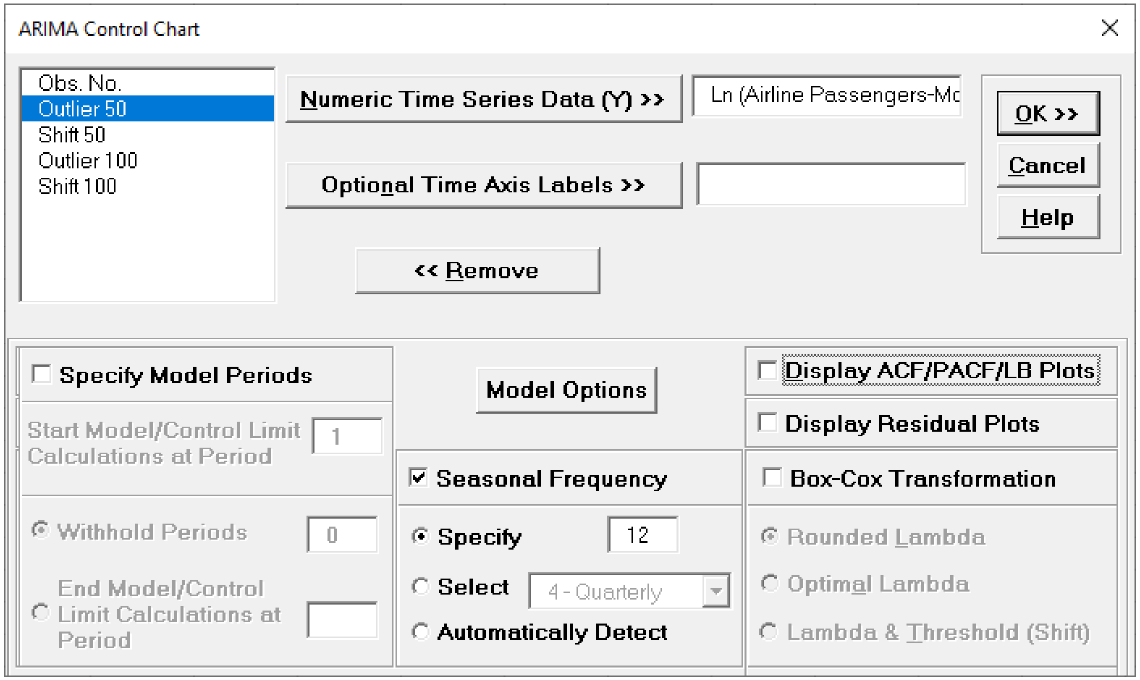


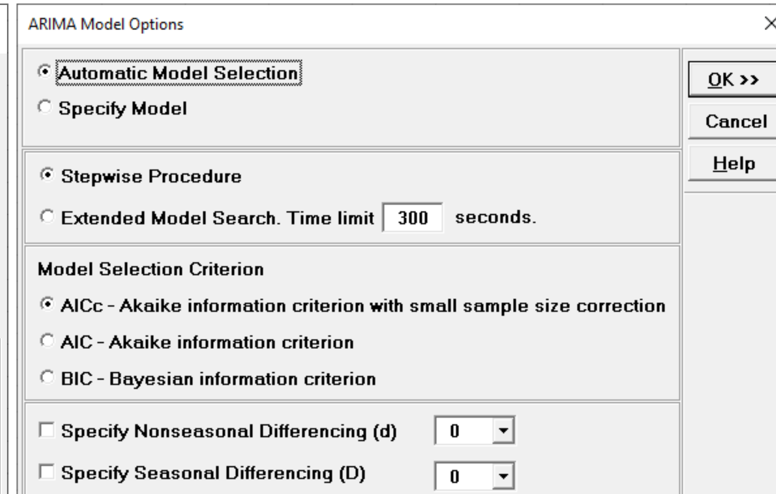






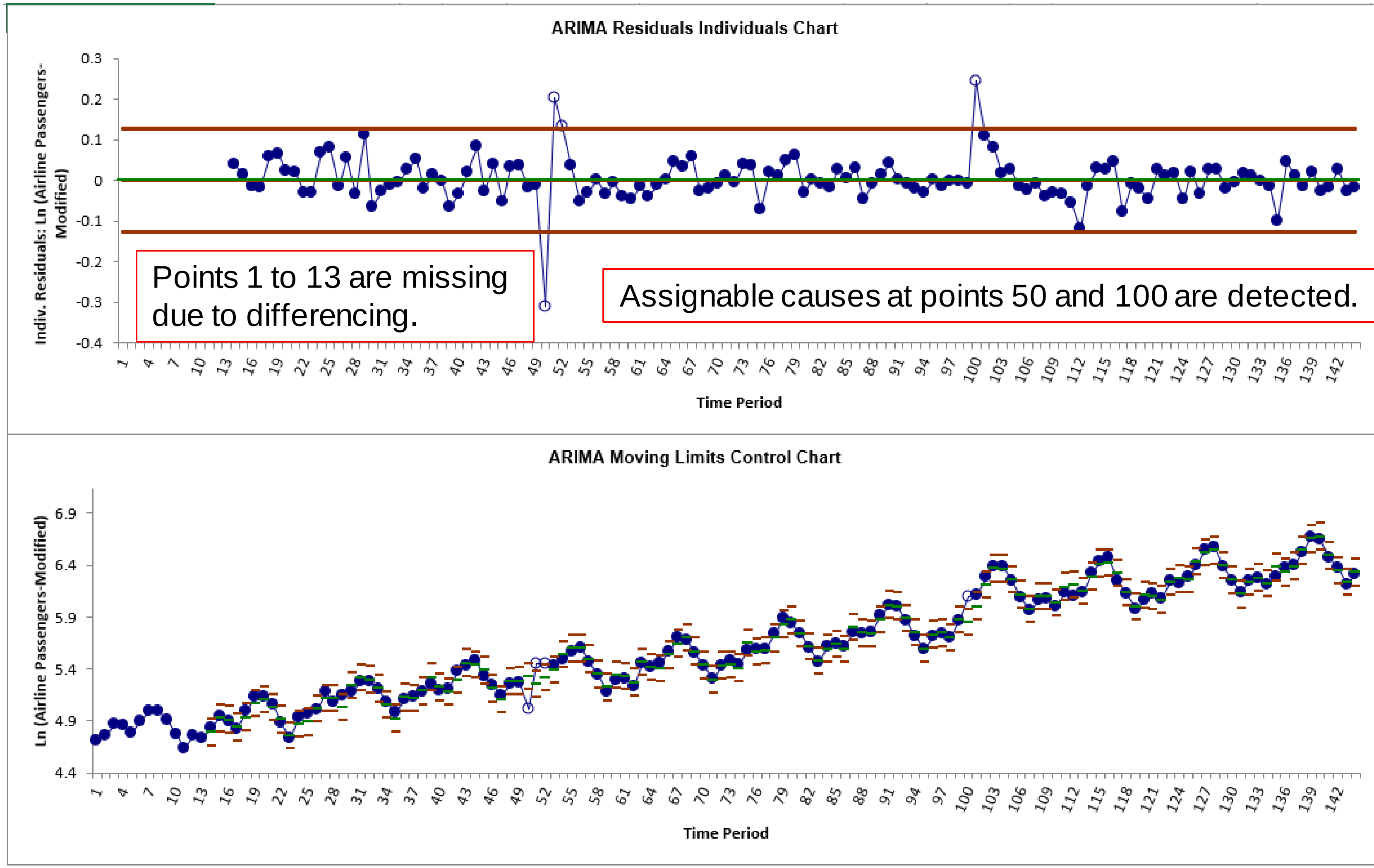
Example 2c: Box-Jenkins Series G – Ln(Monthly Airline Passengers) – ARIMA Control Chart with Automatic Model Selection





SigmaXL > Time Series Forecasting > ARIMA Control Chart > Control Chart

Example 2c: Box-Jenkins Series G – Ln(Monthly Airline Passengers) ARIMA Control Charts



ARIMA Model: Ln (Airline Passengers-Modified) - Model Automatically Selected

Model Periods: All observations are used in the ARIMA model estimation. No withhold periods available for out-of-sample forecast accuracy evaluation.

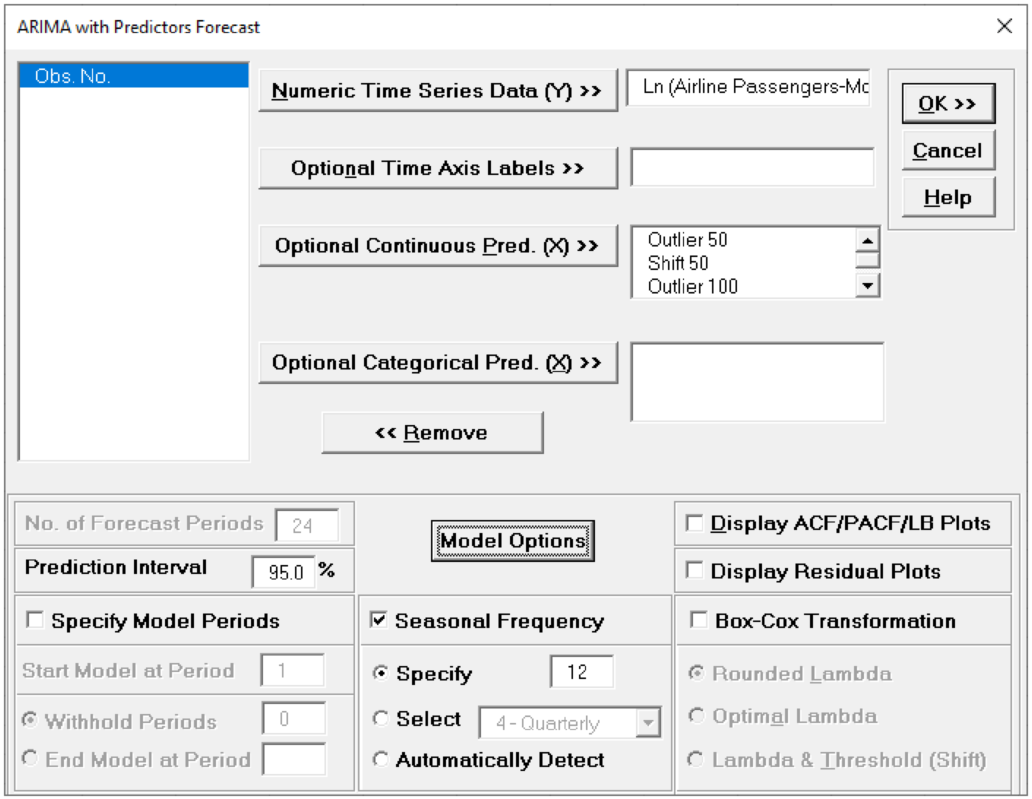
ARIMA with Predictors

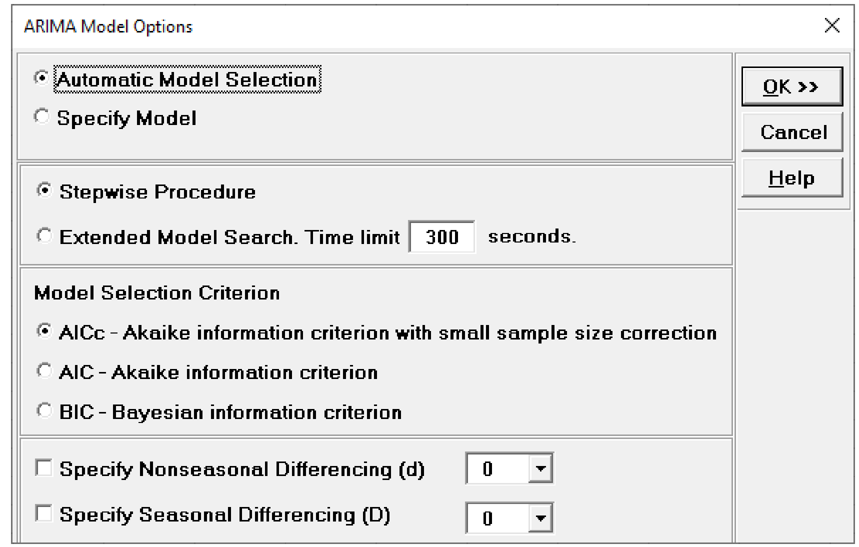
- The ARIMA model supports continuous or categorical predictors, similar to multiple regression.
- As with multiple linear regression, predictors should not be strongly correlated.

Example 2d: Box-Jenkins Series G – Ln(Monthly Airline Passengers) – Outlier versus Shift Coded Predictors

Obs. No.	Ln (Airline Passengers-Modified)	Outlier 50	Shift 50	Outlier 100	Shift 100
49	5.278114659	0	0	0	0
50	5.028114659	1	1	0	0
51	5.463831805	0	1	0	0
52	5.459585514	0	1	0	0
53	5.433722004	0	1	0	0
54	5.493061443	0	1	0	0
55	5.575949103	0	1	0	0
99	5.874930731	0	1	0	0
100	6.10220248	0	1	1	1
101	6.122117789	0	1	0	1
102	6.295005314	0	1	0	1
103	6.392037406	0	1	0	1
104	6.396329258	0	1	0	1
105	6.251414878	0	1	0	1

Example 2d: Box-Jenkins Series G – Ln(Monthly Airline Passengers) – ARIMA Forecast with Predictors: Outlier versus Shift Coded Predictors





SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast with Predictors

Example 2d: Box-Jenkins Series G – Ln(Monthly Airline Passengers) – Outlier versus Shift Coded Predictors

Using ARIMA Forecast with Predictors, we can see that *Outlier50* and *Shift100* are significant denoting Obs. No. 50 as an outlier and 100 as a shift. This is, of course, what we expected since that's how the Ln Airline Passenger data was modified.

This method to identify outlier versus shift is intended as a complement to process knowledge and the search for assignable causes used in classical SPC.

Parameter Estimates				
Term	**Coefficient**	**SE Coefficient**	**T**	**P**
MA_1	0.395854258	0.086277876	4.588132	**0.0000**
SMA_1	0.556926317	0.073521579	7.575005	**0.0000**
Outlier 50	-0.296267312	0.032292885	9.174384	**0.0000**
Shift 50	0.044130277	0.035505653	1.242908	0.2162
Outlier 100	-0.000597974	0.031822469	0.018791	0.9850
Shift 100	0.249993132	0.035384072	7.065132	**0.0000**

SigmaXL > Time Series Forecasting > ARIMA Forecast > Forecast with Predictors

SPC for Autocorrelated Data Using Automated Time Series Forecasting

Questions?

John Noguera, P. Eng. - CTO & Co-founder, SigmaXL, Inc.

jnoguera@SigmaXL.com

References

1. Alwan, L.C., and Roberts, H.V. (1988), "Time Series Modeling for Statistical Process Control," **Journal of Business and Economic Statistics**, 6, 87-95.
2. Box, G. E. P. , Jenkins, G. M. , Reinsel, G. C. and Ljung, G.M. (2016). *Time Series Analysis, Forecasting and Control*, 5th edition, Wiley.
3. Hunter, J.S. (1986), "The Exponentially Weighted Moving Average," **Journal of Quality Technology**, 18, 203-210.
4. Hyndman, R.J., & Athanasopoulos,G. (2018). *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2.
5. Hyndman, R. J. and Y. Khandakar, (2008). "Automatic Time Series Forecasting: The forecast Package for R. " **Journal of Statistical Software**, 27(3), 1-22.

References

6. Montgomery, D. C., and Mastrengelo, C.M. (1991), “Some Statistical Process Control Methods for Autocorrelated Data,” **Journal of Quality Technology**, 23, 179-204.
7. Montgomery, D. C. (2013). *Introduction to Statistical Quality Control*, 7th edition, Wiley.
8. Montgomery, D. C., Jennings, C.L., and Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting*, 2nd edition, Wiley.
9. *NIST/SEMATECH e-Handbook of Statistical Methods*, https://www.itl.nist.gov/div898/handbook.
10. Woodall, W.H. and Faltin, F.W. “Autocorrelated Data and SPC” **ASQC Statistics Division Newsletter**, 13(4).

Appendix

- Autocorrelation: Ljung-Box Test
- Error, Trend, Seasonal (ETS) models
  Hyndman's Taxonomy
- Box-Jenkins AutoRegressive Integrated Moving Average (ARIMA) Models
- Assess Forecast Accuracy

Autocorrelation: Ljung-Box Test

- In addition to looking at the ACF plot, we can also do a more formal test for autocorrelation by considering a whole set of r_k values as a group, rather than treating each one separately.

$$Q = T(T+2)\sum_{k=1}^{h} (T-k)^{-1} r_k^2,$$

  where h is the maximum lag being considered and T is the number of observations.

- If the autocorrelations did come from a white noise series, then Q would have a χ^2 distribution with $(h - k)$ degrees of freedom, where k is the number of parameters in the model [4].

Error, Trend, Seasonal (ETS) models Hyndman’s Taxonomy

- Rob Hyndman has developed a complete taxonomy that describes all of the combinations of exponential smooth models in a consistent manner. [4]
- Error:
  - Additive or Multiplicative
  - The point forecasts produced by the models are identical if they use the same smoothing parameter values. Multiplicative will, however, generate different prediction intervals to accommodate change in variance.
  - An alternative to multiplicative is to use the Ln transformation (Box-Cox transformation with Lambda = 0).
  - Error models include the smoothing parameter α and initial level value.

Error, Trend, Seasonal (ETS) models Hyndman's Taxonomy

- Trend:
  - None, Additive, Additive Damped
  - Multiplicative Trend is not recommended as they tend to produce poor forecasts
  - Trend models add a smoothing parameter β and initial trend value.
  - Damped trend models add a smoothing parameter ϕ that "dampens" the trend to a flat line some time in the future.

Error, Trend, Seasonal (ETS) models
Hyndman's Taxonomy

- Seasonal:
  - None, Additive, Multiplicative
  - Seasonal models add a smoothing parameter γ and initial seasonal values.
    - # of initial values = seasonal frequency – 1
    - constrained to sum to 0 for additive or 12 for multiplicative
- Some of the model combinations lead to numerical instability and are not considered in the selection process: (A,N,M) (A,A,M) (A,Ad,M)

Error, Trend, Seasonal (ETS) models Hyndman's Taxonomy - Examples

Short hand (Error, Trend, Seasonal)	Method
(A, N, N)	Simple Exponential Smoothing with Additive Errors – Exponentially Weighted Moving Average (EWMA)
(M, N, N)	Simple Exponential Smoothing with Multiplicative Errors
(A, A, N)	Additive Trend Method with Additive Errors (Holt's Linear)
(M, A, N)	Additive Trend Method with Multiplicative Errors (Holt's Linear)
(A, A, A)	Additive Trend, Additive Seasonal Method with Additive Errors (Holt-Winters)
(M, A, A)	Additive Trend, Additive Seasonal Method with Multiplicative Errors (Holt-Winters)
(A, Ad, A)	Additive Damped Trend, Additive Seasonal Method with Additive Errors
(M, Ad, A)	Additive Damped Trend, Additive Seasonal Method with Multiplicative Errors

Error, Trend, Seasonal (ETS) models
Hyndman's Taxonomy (State Space Models)

ADDITIVE ERROR MODELS

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$	$y_t = \ell_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/\ell_{t-1}$
A	$y_t = \ell_{t-1} + b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t$ $b_t = b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + b_{t-1})$
A$_d$	$y_t = \ell_{t-1} + \phi b_{t-1} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$	$y_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t$ $b_t = \phi b_{t-1} + \beta\varepsilon_t$ $s_t = s_{t-m} + \gamma\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m} + \varepsilon_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha\varepsilon_t/s_{t-m}$ $b_t = \phi b_{t-1} + \beta\varepsilon_t/s_{t-m}$ $s_t = s_{t-m} + \gamma\varepsilon_t/(\ell_{t-1} + \phi b_{t-1})$

MULTIPLICATIVE ERROR MODELS

Trend	Seasonal N	Seasonal A	Seasonal M
N	$y_t = \ell_{t-1}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$	$y_t = (\ell_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$y_t = \ell_{t-1}s_{t-m}(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1}(1 + \alpha\varepsilon_t)$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A	$y_t = (\ell_{t-1} + b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$
A$_d$	$y_t = (\ell_{t-1} + \phi b_{t-1})(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + \varepsilon_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$y_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}(1 + \varepsilon_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha\varepsilon_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$ $s_t = s_{t-m}(1 + \gamma\varepsilon_t)$

https://otexts.com/fpp2/ets.html

Error, Trend, Seasonal (ETS) Automatic Model Selection

- Parameters are estimated by maximizing the Log-Likelihood function (which is similar to minimizing the residual sum-of-squares).
- Given a set of candidate models for the data, the preferred model is the one with the minimum Information Criteria value:
  - The Information Criteria rewards goodness of fit (as assessed by the likelihood function), but it also includes a penalty that is an increasing function of the number of estimated parameters.
  - The penalty discourages overfitting, because increasing the number of parameters in the model almost always improves the goodness of the fit.

  Reference:
  https://en.wikipedia.org/wiki/Akaike_information_criterion

Error, Trend, Seasonal (ETS) Automatic Model Selection

- Akaike's Information Criterion

$$\text{AIC} = -2\log(L) + 2k,$$

  where L is the likelihood of the model and k is the total number of parameters and initial states that have been estimated.

- The AIC corrected for small sample bias (AICc) is defined as:

$$\text{AIC}_\text{C} = \text{AIC} + \frac{k(k+1)}{T-k-1},$$

- The Bayesian Information Criterion (BIC) is:

$$\text{BIC} = \text{AIC} + k[\log(T) - 2]$$

Error, Trend, Seasonal (ETS) Automatic Model Selection

- AICc is recommended as the default Information Criteria, based on forecast error performance with M3 competition data [4].
- If a Box-Cox transformation is used, Multiplicative models are not considered.

Box-Jenkins AutoRegressive Integrated Moving Average (ARIMA) Models - Stationarity

- ARIMA assumes that the time series is stationary, i.e., it has the property that the mean, variance and autocorrelation structure do not change over time.
- If a time series mean is not stationary (e.g. trending), this can be corrected by differencing, computing the differences between consecutive observations for non-seasonal and between consecutive periods (e.g. months) for seasonal data (Jan 2019 – Jan 2018, etc.).
- For non-seasonal, this may involve 1 or 2 orders of differencing. This order is the Integrated term *d.*
- For seasonal, this may involve 1 order of differencing. This order is the Seasonal Integrated term *D*.

Box-Jenkins AutoRegressive Integrated Moving Average (ARIMA) Models - Stationarity

- If d+D = 0, a constant term in the model is the mean.
- If d+D = 1, a constant term in the model is a trend/drift.
- If d+D > 1, a constant term would be a quadratic trend, so constant should not be included.
- It is recommended that d+D should not be > 3.
- If the variance is not stationary, use a Box-Cox transformation.
- In the Ln(Monthly Airline Passenger) data we are starting with Ln data to deal with non-stationary variance in the raw data.

Box-Jenkins AutoRegressive Integrated Moving Average (ARIMA) Models – AR

- In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. The term autoregression indicates that it is a regression of the variable against itself.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

where ε_t is white noise [4].

Box-Jenkins AutoRegressive Integrated Moving Average (ARIMA) Models – MA

- Rather than using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model [4].

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q}$$

- Model parameters are solved using Kalman Filters and nonlinear minimization. This permits exact calculations (backcasting is not required) and can handle missing values.

Box-Jenkins AutoRegressive Integrated Moving Average (ARIMA) Models

If we combine differencing with autoregression and a moving average model, we obtain a non-seasonal ARIMA model.

$$y'_t = c + \phi_1 y'_{t-1} + \cdots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where y'_t is the differenced series.

- For seasonal, the model consists of terms that are similar to the non-seasonal components of the model. The seasonal model is ARIMA (*P,D,Q*) and combined we have ARIMA (*p,d,q*) (*P,D,Q*).

Partial Autocorrelation (PACF)

- Partial Autocorrelation plots are similar to Autocorrelation plots but adjust for correlation inherent in lags, e.g., y_t and y_{t-2} might be correlated, simply because they are both connected to y_{t-1}, rather than because of any new information contained in y_{t-2} [4].
- Each partial autocorrelation can be estimated as the last coefficient in an autoregressive model. Specifically, α_k, the kth partial autocorrelation coefficient, is equal to the estimate of ϕ_k in an AR(k) model.
- They are typically used in ARIMA to help determine the order of terms in the model, but are also useful as a general diagnostic tool.

Box-Jenkins AutoRegressive Integrated Moving Average (ARIMA) Models – Model Selection

- ACF and PACF plots may be used to assist in determining what order values to use, but this requires a high level of expertise.
- Hyndman and Khandakar [5] give a stepwise procedure to determine optimal order values:
  - Use a Seasonal strength test to determine if D=0 or 1
  - Use a test for stationarity (KPSS) to determine if d=0, 1 or 2
  - With the differenced data, apply a stepwise procedure to solve for p, q, P, Q selecting models with minimum AICc.

Assess Forecast Accuracy

- Information Criteria AIC/AICc cannot be used to compare ARIMA and ETS models or ARIMA models with different d, D values.
- A forecast error e_t is the difference between an observed value and its forecast.
  - It is not the same as a model residual because residuals may be in different units due to Box-Cox transformation or a Multiplicative model.
- Types of forecast error:
  - In-Sample One-Step-Ahead Forecast. This is less useful because the model may be over-fitted.
  - Out-of-Sample (Withhold) One-Step-Ahead. Model parameter estimates do not use any withhold data, but the forecast updates with every new withhold observation.
  - Out-of-Sample (Withhold) Full Period Forecast. This is important if one is assessing forecast accuracy over a horizon. This is used in forecast competitions.

Assess Forecast Accuracy

- Common forecast accuracy measures include:

$$\text{Root mean squared error: RMSE} = \sqrt{\text{mean}(e_t^2)}$$

$$\text{Mean absolute error: MAE} = \text{mean}(|e_t|)$$

$$\text{Mean absolute percentage error: MAPE} = \text{mean}\left(\left|\frac{100\, e_t}{y_t}\right|\right)$$

$$\text{Mean absolute scaled error: MASE} = \text{mean}(|e_t|)/scale$$

  - Scale is the MAE of the in-sample naïve or seasonal naïve forecast (set all forecasts to be the value of the last observation/period)
  - A scaled error is less than one if it arises from a better forecast than the average naïve/seasonal naïve forecast. Conversely, it is greater than one if the forecast is worse than the average naïve forecast [4].