Chapter 9

Autocorrelation in Quality Control

What is Autocorrelation?

- *Autocorrelation measures the correlation between a time series and its lagged values (previous time points).*
- *It helps identify patterns or dependencies in the data over time.*
- *Autocorrelation Function (ACF): Shows the correlation of the time series with its own lags.*

$$\mathrm{ACF}(h) = \frac{\mathrm{Cov}(Y_t, Y_{t+h})}{\sqrt{\mathrm{Var}(Y_t) \cdot \mathrm{Var}(Y_{t+h})}}$$

- *Where h is the lag at which we measure the correlation.*

Components of ARIMA Models

- ***AR (Autoregressive):*** *Relates the current value of the series to its previous values.*

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + \epsilon_t$$

- ***I (Integrated):*** *Difference the data to make it stationary.*

$$Y_t' = Y_t - Y_{t-1}$$

- ***MA (Moving Average):*** *Models the error term as a linear combination of past error terms.*

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

ARIMA Model Notation

- ***ARIMA(p, d, q):*** *A time series model where:*
  - ***p*** *is the order of the AR part (number of lagged values).*
  - ***d*** *is the degree of differencing required to achieve stationarity.*
  - ***q*** *is the order of the MA part (number of lagged forecast errors).*

Stationary Time Series

- *A time series is stationary if its statistical properties do not change over time (mean, variance, autocovariance).*
- *Stationarity condition: Autocovariance function γ(h) should depend only on the lag h and **not** on the time t.*
  - ***Mean (μ)** is constant over time:*

$$E(Y_t) = \mu \quad \forall t$$

  - ***Variance (σ^2)** is constant over time:*

$$\text{Var}(Y_t) = \sigma^2 \quad \forall t$$

  - ***Autocovariance (γ)** depends only on the lag h, not time t:*

$$\text{Cov}(Y_t, Y_{t+h}) = \gamma(h) \quad \forall t, h$$

Finite Order AR and MA Processes

- *AR(1) Process:*

$$Y_t = \alpha_0 + \phi_1 Y_{t-1} + \epsilon_t$$

- *MA(1) Process:*

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}$$

- *ARMA(p, q) combines both AR and MA components.*

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

Nonstationary Time Series and Differencing

- *Nonstationary Time Series: A series where mean and variance change over time.*
- *Differencing: Makes nonstationary data stationary by subtracting previous values.*
- *ARIMA(p, d, q) models nonstationary data where differencing (d) is applied.*
  - *For d=1:*

$$Y_t' = Y_t - Y_{t-1}$$

  - *For d=2:*

$$\Delta^2 Y_t = \Delta Y_t - \Delta Y_{t-1}$$

  - *In general:*

$$\Delta^d Y_t = \sum_{k=0}^{d} \binom{d}{k} (-1)^k Y_{t-k}$$

Statistical Tests for Stationarity

- *Augmented Dickey-Fuller (ADF) Test*
  - *H_0: The series is nonstationary.*
  - *H_1: The series is stationary.*
  - *Test Statistic:*

$$\Delta Y_t = \alpha + \beta t + \gamma Y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta Y_{t-i} + \epsilon_t$$

  - $\Delta Y_t = Y_t - Y_{t-1}$
  - *γ is the coefficient of interest*
- *Interpretation:*
  - *If γ is significantly negative (p-value < 0.05), reject H_0; the series is stationary.*

ARIMA Model Identification

- *Partial Autocorrelation Function (PACF): Helps identify the AR(p) part.*
  - *PACF shows the correlation after removing the effect of shorter lags.*

- *Augmented Dickey-Fuller (ADF) Test: Helps to identify the appropriate differencing d to achieve a stationary time series.*
  - *ADF shows the appropriate stationary value of the series.*

- *Autocorrelation Function (ACF): Helps identify the MA(q) part.*
  - *ACF shows the correlation of the series with its lags.*

ARIMA Model Identification

- *Partial Autocorrelation Function (PACF)*
  - *Remove the influence of intermediate lags, providing a clearer picture of the direct relationship between a variable and its past values.*

$$\phi_k = \frac{\operatorname{cov}(X_t, X_{t-k} | X_{t-1}, X_{t-2}, \ldots, X_{t-k+1})}{\sqrt{\operatorname{var}(X_t | X_{t-1}, X_{t-2}, \ldots, X_{t-k+1}) \cdot \operatorname{var}(X_{t-k} | X_{t-1}, X_{t-2}, \ldots, X_{t-k+1})}}$$

- *Autocorrelation Function (ACF):*
  - *Measures the linear relationship between a time series and its lagged values.*

$$\rho_k = \frac{\operatorname{Cov}(X_t, X_{t-k})}{\sqrt{\operatorname{Var}(X_t) \cdot \operatorname{Var}(X_{t-k})}}$$

Building an ARIMA Model

1. *Plot the Time Series:*
   - *A plot reveals a trend, indicating the series might not be stationary.*
2. *Check for Stationarity (ADF Test):*
   - *If $p>0.05$, fail to reject H_0, and conclude the series is nonstationary.*
3. *Differencing to Achieve Stationarity:*
   - *Compute differenced value (starting from one), to reach stationarity (d).*
4. *Plot ACF and PACF of the Stationary Series:*
   - *PACF helps identify p (AR terms).*
   - *ACF helps identify q (MA terms).*

Building an ARIMA Model

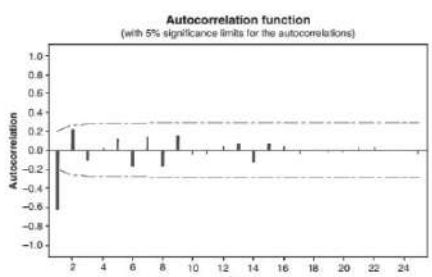
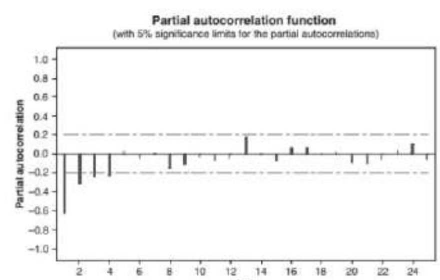
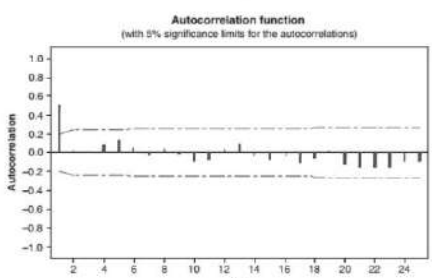
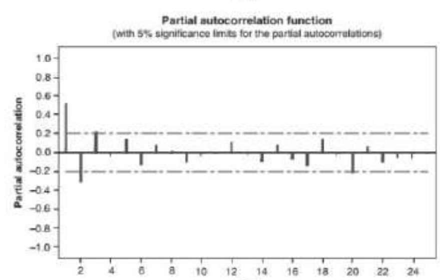
5. *Estimate Parameters:*
   - *Using statistical software (e.g., R or Python)*
6. *Residual Analysis:*
   - *Check if residuals are white noise (uncorrelated and normally distributed).*
   - *Box-Ljung Test (If p-value>0.05, fail to reject H_0, and conclude residuals are white noise.)*
   - *H_0: The data is not correlated.*
   - *H_1: The data exhibit serial correlation.*

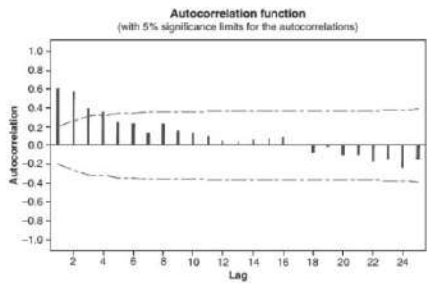
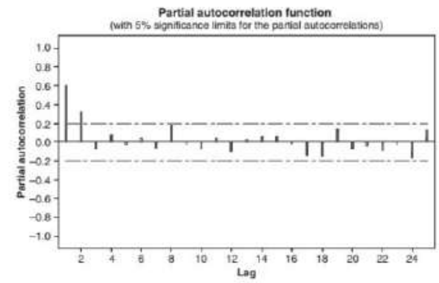
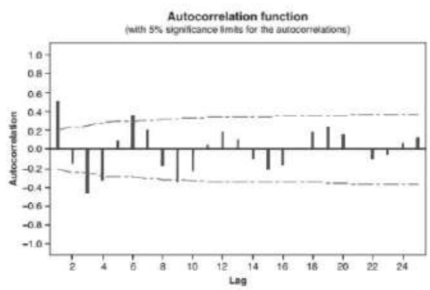
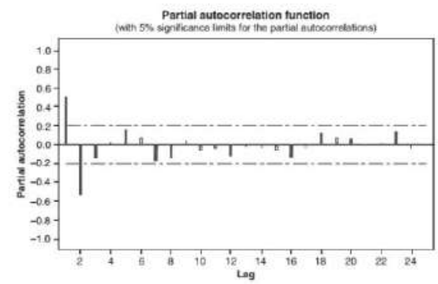
   $$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$$

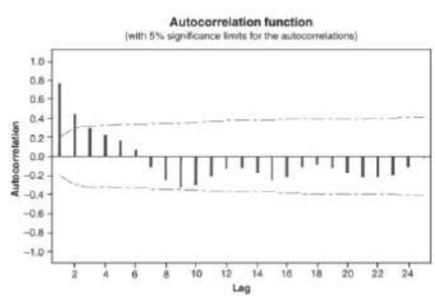
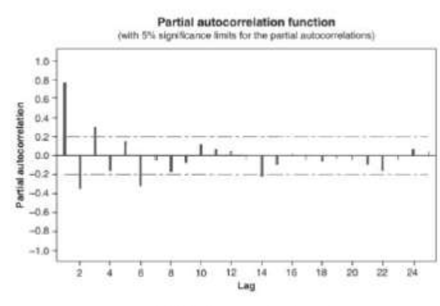
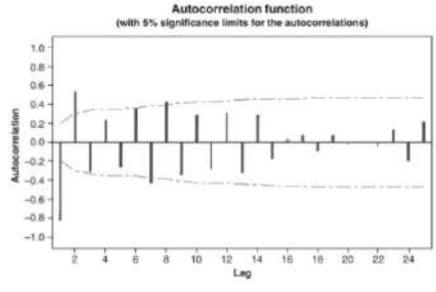
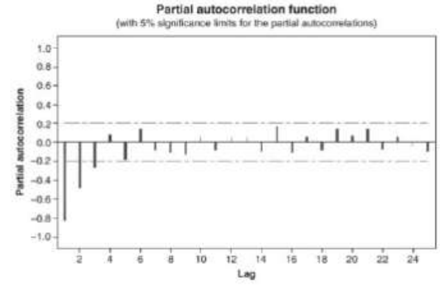
   - *For significance level α, the critical region for rejection of the hypothesis of randomness is: $Q > \chi^2_{1-\alpha,h}$*
7. *Enjoy your model ☺*

ARIMA Model Identification

Model	ACF	PACF
MA(q)	Cuts off after lag q	Exponential decay and/or damped sinusoid
AR(p)	Exponential decay and/or damped sinusoid	Cuts off after lag p
ARMA(p,q)	Exponential decay and/or damped sinusoid	Exponential decay and/or damped sinusoid

Model	Sample ACF	Sample PACF
MA(1) $y_t = 40 + \varepsilon_t - 0.8\varepsilon_{t-1}$		
$y_t = 40 + \varepsilon_t + 0.8\varepsilon_{t-1}$		

Model	Sample ACF	Sample PACF
AR(2) $y_t = 4 + 0.4y_{t-1} + 0.5y_{t-2} + \varepsilon_t$		
$y_t = 4 + 0.8y_{t-1} - 0.5y_{t-2} + \varepsilon_t$		

Model	Sample ACF	Sample PACF
ARMA(1.1) $y_t = 16 + 0.6y_{t-1} + \varepsilon_t + 0.8\varepsilon_{t-1}$		
$y_t = 16 - 0.7y_{t-1} + \varepsilon_t - 0.6\varepsilon_{t-1}$		

Autocorrelation in Quality Control

- *SPC traditionally assumes independent observations.*
- *Autocorrelation in data, common in sequential measurements, requires adjusted SPC methods.*
- *Types of Autocorrelation:*
  - ***Mild Autocorrelation:*** *Typically occurs in processes with low-frequency measurements.*
  - ***Moderate Autocorrelation:*** *Observed in more closely spaced measurements, with significant correlation at smaller lags.*
  - ***Complex (High) Autocorrelation:*** *Seen in continuous or high-frequency data, with persistence across multiple time lags.*

Measuring Autocorrelation

- *Overview of autocorrelation, focusing on how it is measured and interpreted.*
- *Equations:*
  - *Autocorrelation at lag **k**:*

$$\rho_k = \frac{\mathrm{Cov}(x_t, x_{t-k})}{V(x_t)} \quad k = 0, 1, \ldots$$

  - *Sample Autocorrelation Function:*

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2} \quad k = 0, 1, \ldots, K$$

Measuring Autocorrelation

- *Interpretation by Degree:*
  - ***Mild:*** $\mathbf{r_k}$ *quickly declines as* $\boldsymbol{k}$ *increases (e.g., weekly data with minimal persistence).*
  - ***Moderate:*** $\mathbf{r_k}$ *shows slow decay, indicating moderate persistence across multiple lags.*
  - ***Complex:*** *Strong and sustained* $\mathbf{r_k}$ *across several lags, such as in data sampled every second in automated systems*

Autoregressive Models for Autocorrelated Processes

- *Autoregressive (AR) models are useful for handling mild and moderate autocorrelation.*
- *AR models are best for relatively simple, predictable autocorrelation structures*
- *AR(1) model:*

$$x_t = \phi x_{t-1} + \epsilon_t, \quad |\phi| < 1$$

- *Applications by Degree:*
  - *Mild Autocorrelation: AR(1) or AR(2) models can effectively capture patterns.*
  - *Moderate Autocorrelation: AR models (AR(3) or AR(4)) may be required.*
  - *Complex Autocorrelation: Higher-order AR or ARIMA models are required.*

Process Monitoring with Residual Control Charts

- *Using residuals from AR models to create de-autocorrelated data for SPC.*
- *Residual calculation:*

$$e_t = x_t - \phi x_{t-1}$$

- *Key Points:*
  - *Mild Autocorrelation: Residuals from a simple AR(1) model often suffice.*
  - *Moderate Autocorrelation: Higher-order AR models help in producing accurate residuals.*
  - *Complex Autocorrelation: For complex structures, ARIMA may be required before residual plotting to stabilize control charts.*

EWMA Chart for Mild to Moderate Autocorrelation

- *EWMA control charts, suitable for monitoring processes with mild to moderate autocorrelation.*
  - *EWMA statistic:*

$$z_t = \lambda x_t + (1 - \lambda) z_{t-1}$$

- *Interpretation by Degree:*
  - ***Mild Autocorrelation:*** *Small values of λ provide appropriate smoothing.*
  - ***Moderate Autocorrelation:*** *Adjust λ to balance sensitivity while accounting for persistence in data.*
- *EWMA is less effective for complex autocorrelation, where ARIMA or batch means may be needed for greater accuracy*

ARIMA Models for Complex Autocorrelation

- *ARIMA (Autoregressive Integrated Moving Average) models are useful for complex autocorrelation.*
  - *ARIMA(1,0,1) model example:*

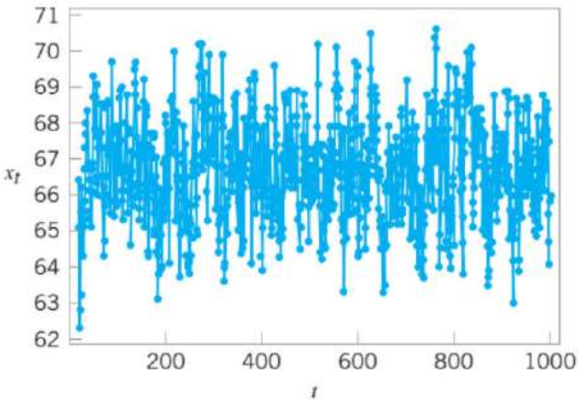
$$x_t = \phi x_{t-1} + \theta \epsilon_{t-1} + \epsilon_t$$

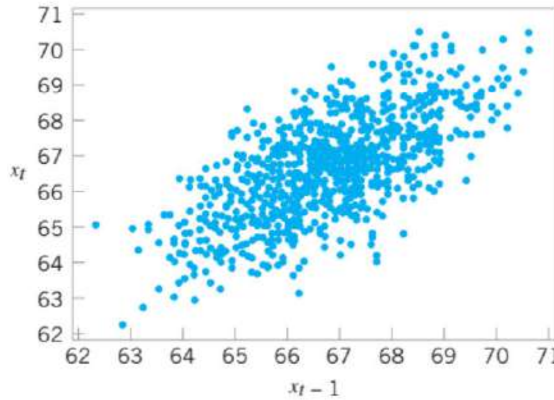
- *Application*
  - ***Complex (High) Autocorrelation:*** *ARIMA models handle mixed autoregressive and moving average structures, capturing long-range dependencies.*
- *Continuous data with periodic patterns, like chemical processes with both short- and long-term dependencies*

Selecting Charts Based on Autocorrelation Degree

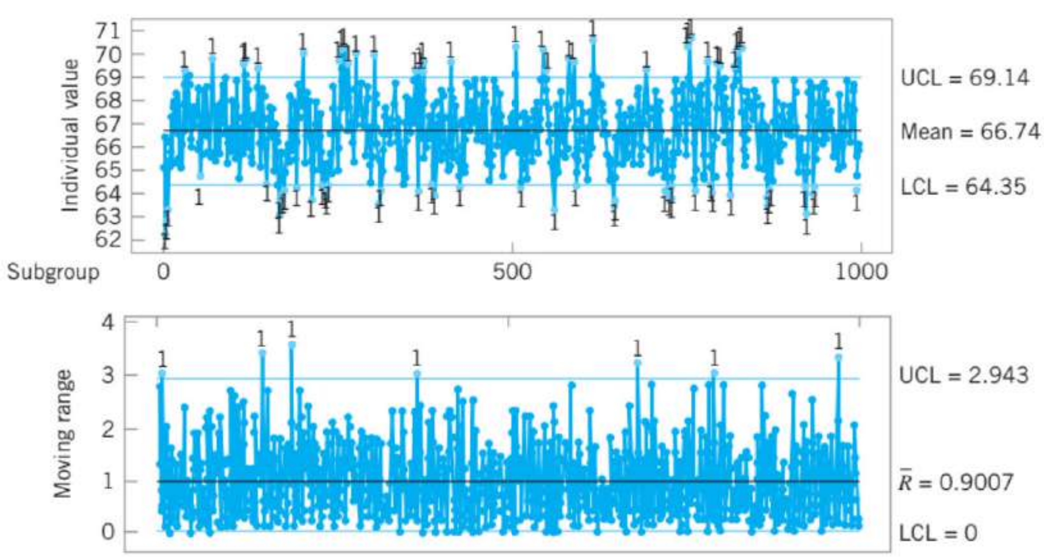
- *Decision-making process for selecting SPC methods based on autocorrelation degree.*
  - ***Mild Autocorrelation:*** *Use residual charts from AR(1) or EWMA charts.*
  - ***Moderate Autocorrelation:*** *Use residual charts from AR models or apply EWMA with adjusted smoothing.*
  - ***Complex Autocorrelation:*** *Apply ARIMA or batch means control charts, depending on data collection frequency and persistence*

A Process Variable with Autocorrelation



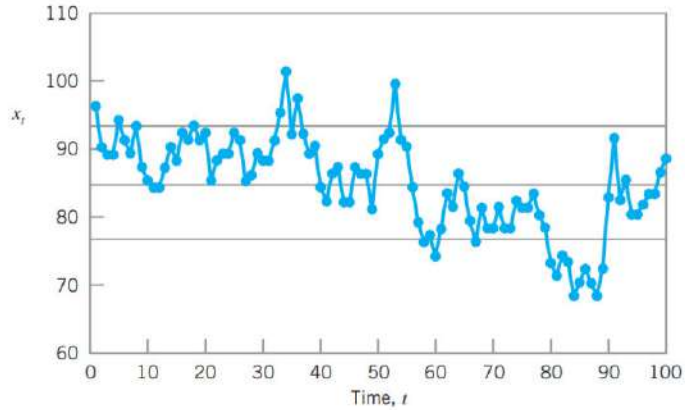


Related Shewhart Charts



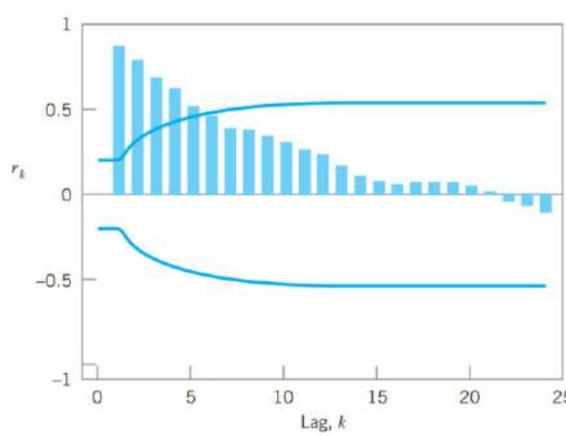
Control Charts of the Residuals (ARIMA)

- *Example 10.2. Figure presents a control chart for individual measurements applied to viscosity measurements from a chemical process taken every hour.*



Control Charts of the Residuals

- *Solution 10.2.*
- *The sample autocorrelation function for the viscosity data is:*
- *There is a strong positive correlation.*
- *Autocorrelation is sufficiently high to distort greatly the performance of a Shewhart control chart.*
- *Correlation greatly increases the frequency of false alarms, so we should be very suspicious about the out-of-control signals on the control chart.*



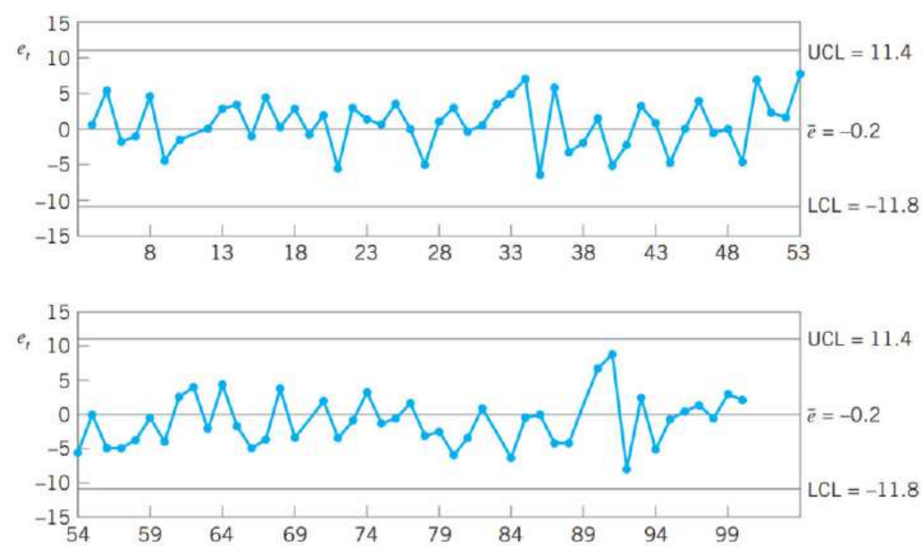
Control Charts of the Residuals (ARIMA)

- *The parameters in the autoregressive model may be estimated by the method of least squares.*
- *The fitted value of this model for the viscosity data is*

$$x_t = 13.04 + 0.847x_{t-1}$$

- *Figure is an individual's control chart of the residuals from the fitted first-order autoregressive model.*

Control Charts of the Residuals (ARIMA)



Control Charts of the Residuals (EWMA)

- *Consider that:*

$$\hat{x}_{t+1}(t) = z_t$$

$$z_t = \lambda x_t + (1 - \lambda) z_{t-1}$$

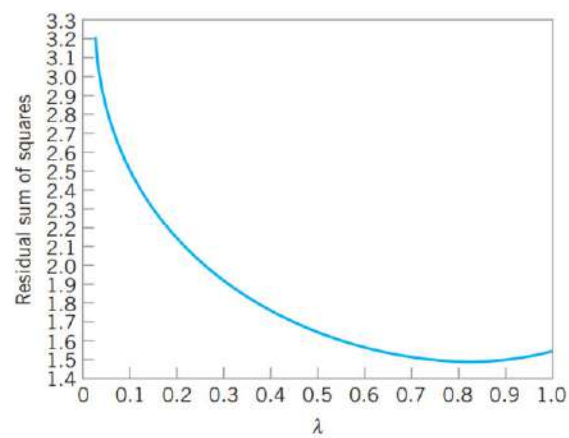
- *The sequence of one-step-ahead prediction errors:*

$$e_t = x_t - \hat{x}_t(t-1)$$

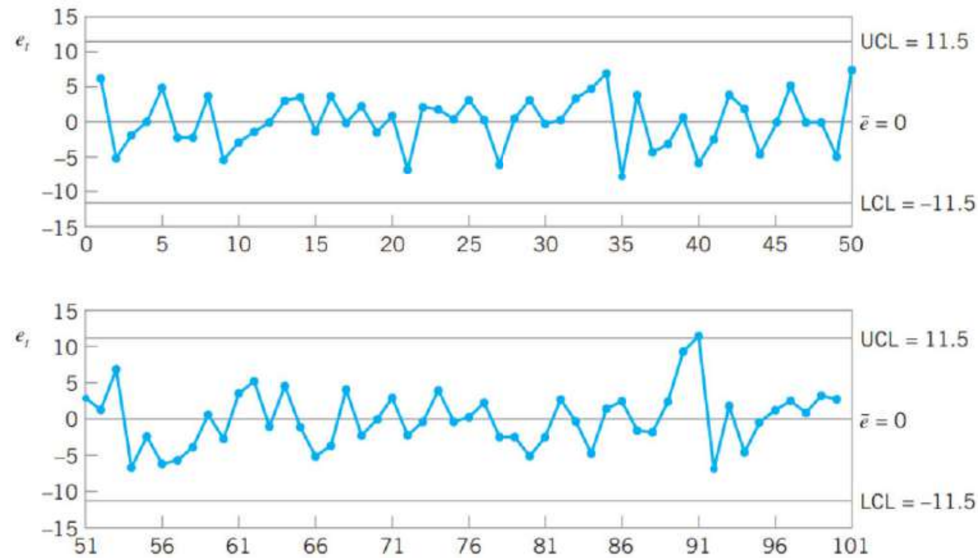
It is independently and identically distributed with mean zero.

Control Charts of the Residuals (EWMA)

- *The parameter λ (or equivalently, θ) would be found by minimizing the sum of squares of the errors e_t*
- *The least squared prediction error occurs at λ = 0.825.*



Control Charts of the Residuals (EWMA)



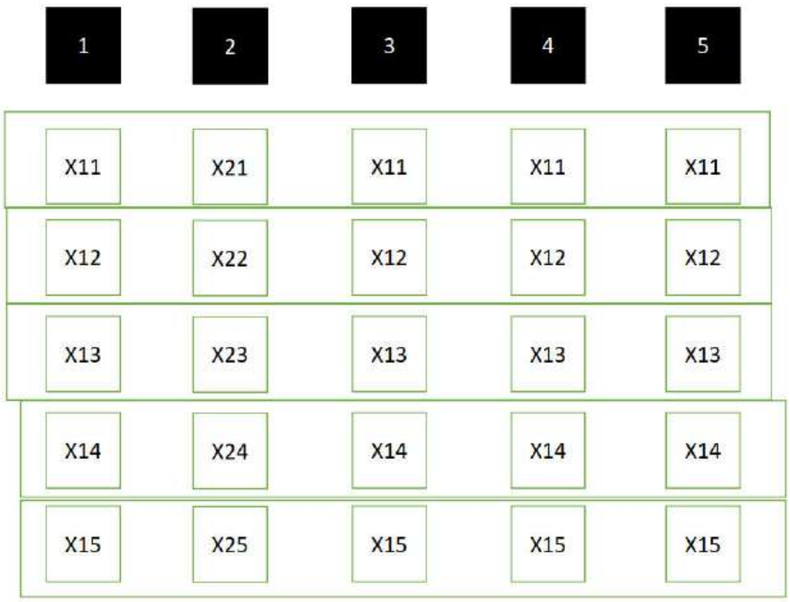
Batch Means Control Chart

- *Runger and Willemain (1996) proposed a control chart based on unweighted batch means for monitoring autocorrelated process data.*
- *The* ***Unweighted Batch Means (UBM)*** *control chart breaks successive groups of sequential observations into batches, with equal weights assigned to every point in the batch.*
- *Let the j^{th} unweighted batch mean be*

$$\bar{x}_j = \frac{1}{b}\sum_{i=1}^{b} x_{(j-1)b+i} \quad j = 1, 2, \ldots$$

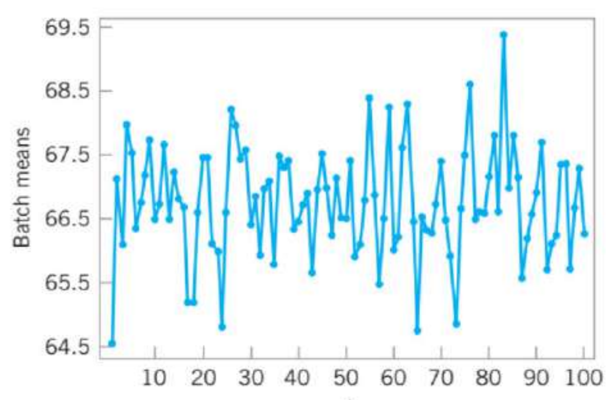
- *Suitable for high-frequency or continuous sampling, where autocorrelation persists over multiple lags*

Batch Means Control Chart

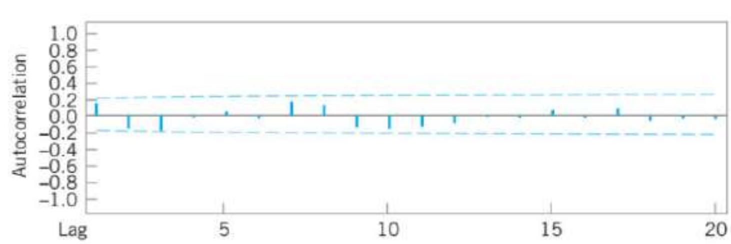


UBM Control Chart

- *In Figure a plot of batch means computed using b = 10.*



- *Related sample autocorrelation function:*



UBM Control Chart

- *The general indication is that the process is stable.*