

Introduction to Time Series Analysis

Raj Jain

Washington University in Saint Louis

Saint Louis, MO 63130

Jain@cse.wustl.edu

Audio/Video recordings of this lecture are available at:

<http://www.cse.wustl.edu/~jain/cse567-17/>



- ❑ What is a time series?
- ❑ Autoregressive Models
- ❑ Moving Average Models
- ❑ Integrated Models
- ❑ ARMA, ARIMA, SARIMA, FARIMA models
- ❑ Note: These slides are based on R. Jain, "The Art of Computer Systems Performance Analysis," 2nd Edition (in preparation).

Stochastic Processes

- ❑ Ordered sequence of random observations
- ❑ Example:
 - Number of virtual machines in a server
 - Number of page faults
 - Number of queries over time
- ❑ Analysis Technique: Time Series Analysis
- ❑ Long-range dependence and self-similarity in such processes can invalidate many previous results

Stochastic Processes: Key Questions

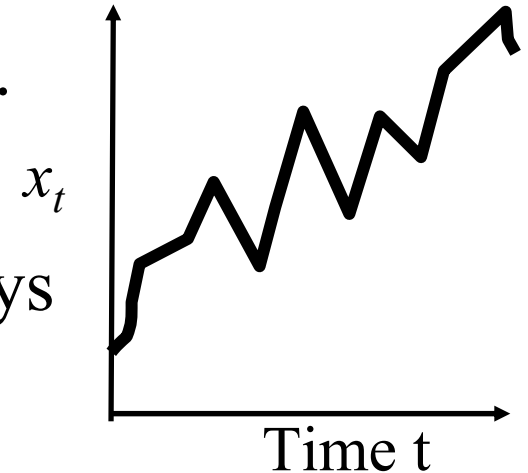
1. What is a time series?
2. What are different types of time series models?
3. How to fit a model to a series of measured data?
4. What is a stationary time series?
5. Is it possible to model a series that is not stationary?
6. How to model a series that has a periodic or seasonal behavior as is common in video streaming?

Stochastic Processes : Key Questions (Cont)

1. What are heavy-tailed distributions and why they are important?
2. How to check if a sample of observations has a heavy tail?
3. What are self-similar processes?
4. What are short-range and long-range dependent processes?
5. Why long-range dependence invalidates many conclusions based on previous statistical methods?
6. How to check if a sample has a long-range dependence?

What is a Time Series

- ❑ Time series = Stochastic Process
- ❑ A sequence of observations over time.
- ❑ Examples:
 - Price of a stock over successive days
 - Sizes of video frames
 - Sizes of packets over network
 - Sizes of queries to a database system
 - Number of active virtual machines in a cloud
- ❑ Goal: Develop models of such series for resource allocation and improving user experience.



Autoregressive Models

- ❑ Predict the variable as a linear regression of the immediate past value: $\hat{x}_t = a_0 + a_1 x_{t-1}$
- ❑ Here, \hat{x}_t is the best estimate of x_t given the past history $\{x_0, x_1, \dots, x_{t-1}\}$
- ❑ Even though we know the complete past history, we assume that x_t can be predicted based on just x_{t-1} .
- ❑ Auto-Regressive = Regression on Self
- ❑ Error: $e_t = x_t - \hat{x}_t = x_t - a_0 - a_1 x_{t-1}$
- ❑ Model: $x_t = a_0 + a_1 x_{t-1} + e_t$
- ❑ Best a_0 and $a_1 \Rightarrow$ minimize the sum of square of errors

Example 37.1

- The number of disk access for 50 database queries were measured to be: 73, 67, 83, 53, 78, 88, 57, 1, 29, 14, 80, 77, 19, 14, 41, 55, 74, 98, 84, 88, 78, 15, 66, 99, 80, 75, 124, 103, 57, 49, 70, 112, 107, 123, 79, 92, 89, 116, 71, 68, 59, 84, 39, 33, 71, 83, 77, 37, 27, 30.

- For this data: $\sum_{t=2}^{50} x_t = 3313$ $\sum_{t=2}^{50} x_{t-1} = 3356$

$$\sum_{t=2}^{50} x_t x_{t-1} = 248147 \quad \sum_{t=2}^{50} x_{t-1}^2 = 272102 \quad n = 49$$

$$\begin{aligned} a_0 &= \frac{\sum x_t \sum x_{t-1}^2 - \sum x_{t-1} \sum x_t x_{t-1}}{n \sum x_{t-1}^2 - (\sum x_{t-1})^2} \\ &= \frac{3313 \times 272102 - 3356 \times 248147}{49 \times 272102 - 3356^2} = 33.181 \end{aligned}$$

Example 37.1 (Cont)

$$\begin{aligned} a_1 &= \frac{n \sum x_t x_{t-1} - \sum x_t \sum x_{t-1}}{n \sum x_{t-1}^2 - (\sum x_{t-1})^2} \\ &= \frac{49 \times 248147 - 3313 \times 3356}{49 \times 272102 - 3356^2} = 0.503 \end{aligned}$$

- The AR(1) model for the series is:

$$x_t = 33.181 + 0.503x_{t-1} + e_t$$

- The predicted value of x_2 given x_1 is:

$$\hat{x}_2 = a_0 + a_1 x_1 = 33.181 + 0.503 \times 73 = 69.880$$

- The actual observed value of is 67. Therefore, the prediction error is:

$$e_2 = x_2 - \hat{x}_2 = 67 - 69.880 = -2.880$$

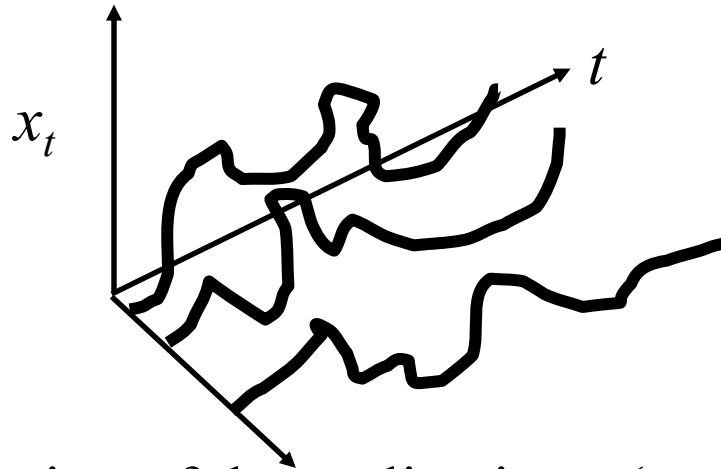
- Sum of squared errors SSE = 32995.57

Exercise 37.1

- Fit an AR(1) model to the following sample of 50 observations: 83, 86, 46, 34, 130, 109, 100, 81, 84, 148, 93, 76, 69, 40, 50, 56, 63, 104, 35, 55, 124, 52, 55, 81, 33, 76, 83, 90, 94, 37, -2, 33, 105, 133, 78, 50, 115, 149, 98, 110, 25, 82, 59, 80, 43, 58, 88, 78, 55, 68. Find a_0 , a_1 and the minimum SSE.

Stationary Process

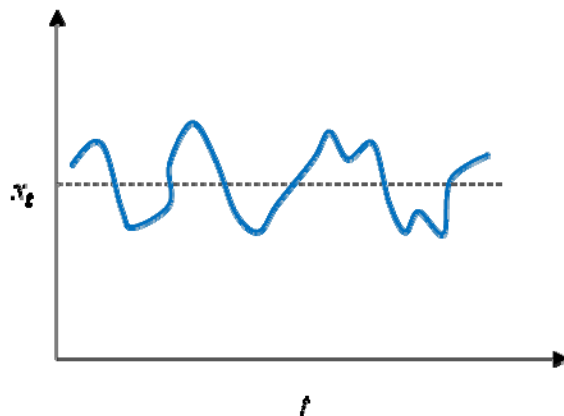
- Each realization of a random process will be different:



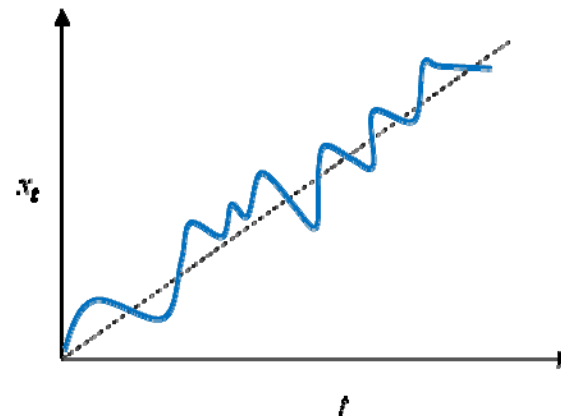
- x is function of the realization i (space) and time t : $x(i, t)$
- We can study the distribution of x_t in space.
- Each x_t has a distribution, e.g., Normal $f(x_t) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_t - \mu)^2}{2\sigma^2}}$
- If this same distribution (normal) with the same parameters μ , σ applies to x_{t+1}, x_{t+2}, \dots , we say x_t is stationary.

Stationary Process (Cont)

- Stationary = Standing in time
⇒ Distribution does not change with time.
- Similarly, the joint distribution of x_t and x_{t-k} depends only on k not on t .
- The joint distribution of $x_t, x_{t-1}, \dots, x_{t-k}$ depends only on k not on t .



(a) Stationary



(b) Non-stationary

Autocorrelation

- ❑ Covariance of x_t and x_{t-k} = Auto-covariance at lag k
Autocovariance of x_t at lag k = $\text{Cov}[x_t, x_{t-k}] = E[(x_t - \mu)(x_{t-k} - \mu)]$
- ❑ For a stationary series:
 - Statistical characteristics do not depend upon time t .
 - Autocovariance depends only on lag k and not on time t

$$\begin{aligned} \text{Autocorrelation of } x_t \text{ at lag } k \quad r_k &= \frac{\text{Autocovariance of } x_t \text{ at lag } k}{\text{Variance of } x_t} \\ &= \frac{\text{Cov}[x_t, x_{t-k}]}{\text{Var}[x_t]} \\ &= \frac{E[(x_t - \mu)(x_{t-k} - \mu)]}{E[(x_t - \mu)^2]} \end{aligned}$$

- ❑ Autocorrelation is dimensionless and is easier to interpret than autocovariance.

Example 37.2

- For the data of Example 37.1, the variance and covariance's at lag 1 and 2 are computed as follows:

$$\text{Sample Mean } \bar{x} = \frac{1}{50} \sum_{t=1}^{50} x_t = \frac{3386}{50} = 67.72$$

$$\text{Var}(x_t) = E[(x_t - \mu)^2] = \frac{1}{49} \sum_{t=1}^{50} (x_t - \bar{x})^2 = \frac{273002 - 50 \times 67.72^2}{49} = 891.879$$

Example 37.2 (Cont)

$$\begin{aligned} \text{Cov}(x_t, x_{t-1}) &= E[(x_t - \mu)(x_{t-1} - \mu)] \\ &= \frac{1}{48} \sum_{t=2}^{50} (x_t - \bar{x}_t)(x_{t-1} - \bar{x}_{t-1}) \\ &= \frac{1}{48} \left[\sum_{t=2}^{50} x_t x_{t-1} - \left(\frac{1}{49} \sum_{t=2}^{50} x_t \right) \sum_{t=2}^{50} x_{t-1} - \sum_{t=2}^{50} x_t \left(\frac{1}{49} \sum_{t=2}^{50} x_{t-1} \right) \right. \\ &\quad \left. + 49 \left(\frac{1}{49} \sum_{t=2}^{50} x_t \right) \left(\frac{1}{49} \sum_{t=2}^{50} x_{t-1} \right) \right] \\ &= \frac{1}{48} \left[\sum_{t=2}^{50} x_t x_{t-1} - \frac{1}{49} \left(\sum_{t=2}^{50} x_t \right) \left(\sum_{t=2}^{50} x_{t-1} \right) \right] \\ &= \frac{1}{48} \left[248147 - \frac{3313 \times 3356}{49} \right] = 442.506 \end{aligned}$$

- ❑ Small Sample $\Rightarrow \bar{x}_t$ and \bar{x}_{t-1} are slightly different.
Not so for large samples.
- ❑ Divisor is 48 since we used sample mean calculated from the same sample

Example 37.2 (Cont)

$$\begin{aligned} \text{Cov}(x_t, x_{t-2}) &= E[(x_t - \mu)(x_{t-2} - \mu)] \\ &= \frac{1}{47} \sum_{t=3}^{50} (x_t - \bar{x}_t)(x_{t-2} - \bar{x}_{t-2}) \\ &= \frac{1}{47} \left[\sum_{t=3}^{50} x_t x_{t-2} - \frac{1}{48} \left(\sum_{t=3}^{50} x_t \right) \left(\sum_{t=3}^{50} x_{t-2} \right) \right] \\ &= \frac{1}{47} \left[229360 - \frac{3246 \times 3329}{48} \right] \\ &= 90.136 \end{aligned}$$

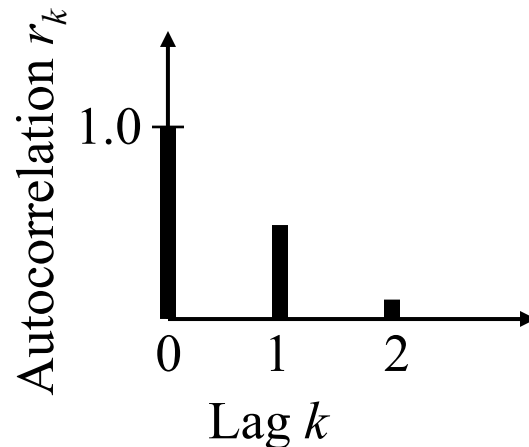
□ Note: Only 48 pairs of $\{x_t, x_{t-1}\} \Rightarrow$ Divisor is 47

Example 37.2 (Cont)

$$\text{Autocorrelation at lag } 0 = r_0 = \frac{\text{Var}(x_t)}{\text{Var}(x_t)} = \frac{891.879}{891.879} = 1$$

$$\text{Autocorrelation at lag } 1 = r_1 = \frac{\text{Cov}(x_t, x_{t-1})}{\text{Var}(x_t)} = \frac{442.506}{891.879} = 0.496$$

$$\text{Autocorrelation at lag } 2 = r_2 = \frac{\text{Cov}(x_t, x_{t-2})}{\text{Var}(x_t)} = \frac{90.136}{891.879} = 0.101$$



White Noise

- Errors e_t are normal independent and identically distributed (IID) with zero mean and variance σ^2
- Such IID sequences are called “**white noise**” sequences.

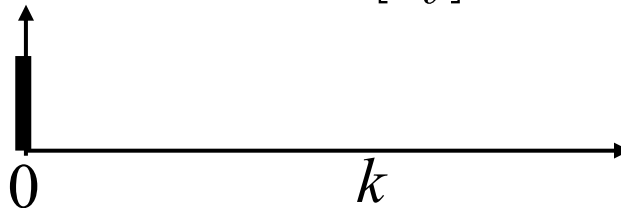
- Properties:

$$E[e_t] = 0 \quad \forall t$$

$$\text{Var}[e_t] = E[e_t^2] = \sigma^2 \quad \forall t$$

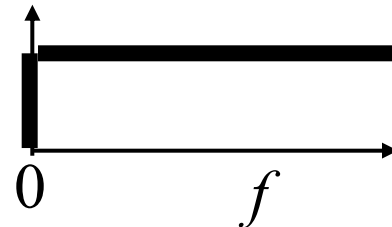
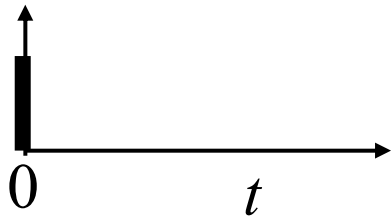
$$\text{Cov}[e_t, e_{t-k}] = E[e_t e_{t-k}] = \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

$$\text{Cor}[e_t, e_{t-k}] = \frac{E[e_t e_{t-k}]}{E[e_t^2]} = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$



White Noise (Cont)

- ❑ The autocorrelation function of a white noise sequence is a spike (δ function) at $k=0$.
- ❑ The Laplace transform of a δ function is a constant. So in frequency domain white noise has a flat frequency spectrum.



- ❑ It was incorrectly assumed that white light has no color and, therefore, has a flat frequency spectrum and so random noise with flat frequency spectrum was called white noise.

White Noise Autocorrelations

- It can be shown that autocorrelations for white noise are normally distributed with mean:

$$E[r_k] \approx \frac{-1}{n}$$

and variance:

$$\text{Var}[r_k] \approx \frac{1}{n}$$

- Therefore, their 95% confidence interval is $-1/n \mp 1.96/\sqrt{n}$
 $z_{0.975} = 1.96$

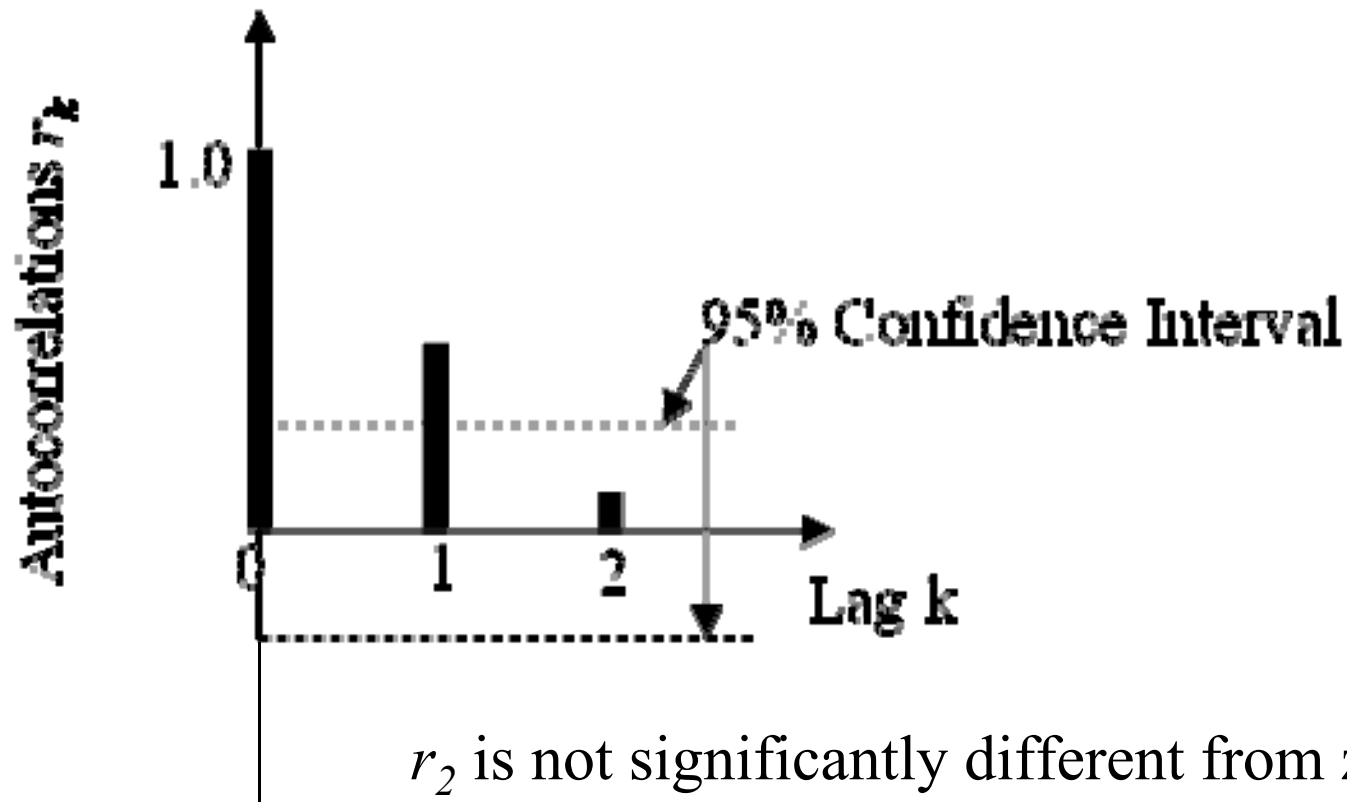
This is generally approximated as $\mp 2/\sqrt{n}$

- This confidence interval can be used to check if a particular autocorrelation is zero.

Example 37.3

- For the data of Example 37.1: $n=50$

$$CI = \mp 2 / \sqrt{(50)} = \mp 0.283$$



Exercise 37.2

- Determine autocorrelations at lag 0 through 2 for the data of Exercise 37.1 and determine which of these autocorrelations are significant at 95% confidence.

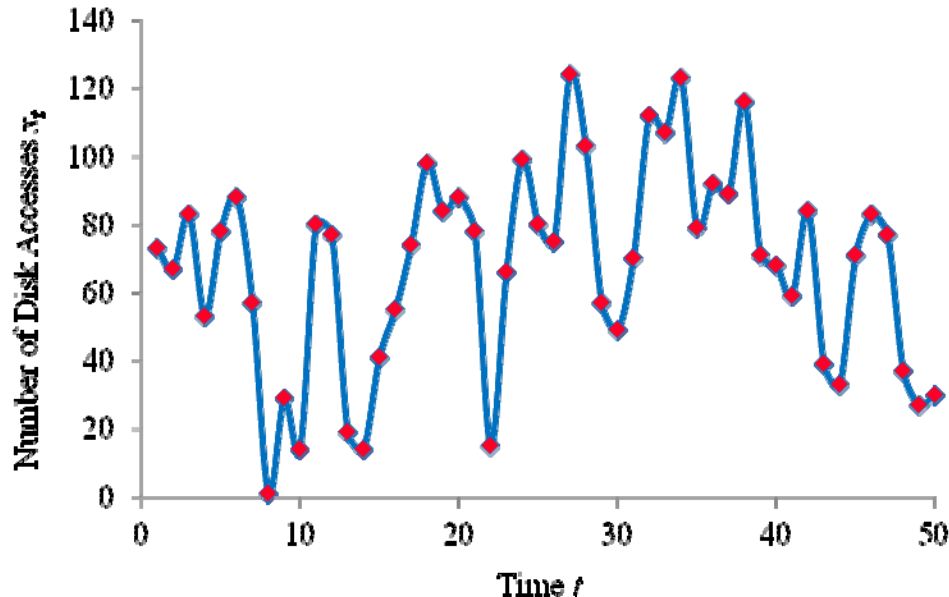
Assumptions for AR(1) Models

1. x_t is a Stationary process
2. Linear relationship between successive values
3. Normal Independent identically distributed errors:
 - a. Normal errors
 - b. Independent errors
4. Additive errors

Visual Tests for AR(1) Models

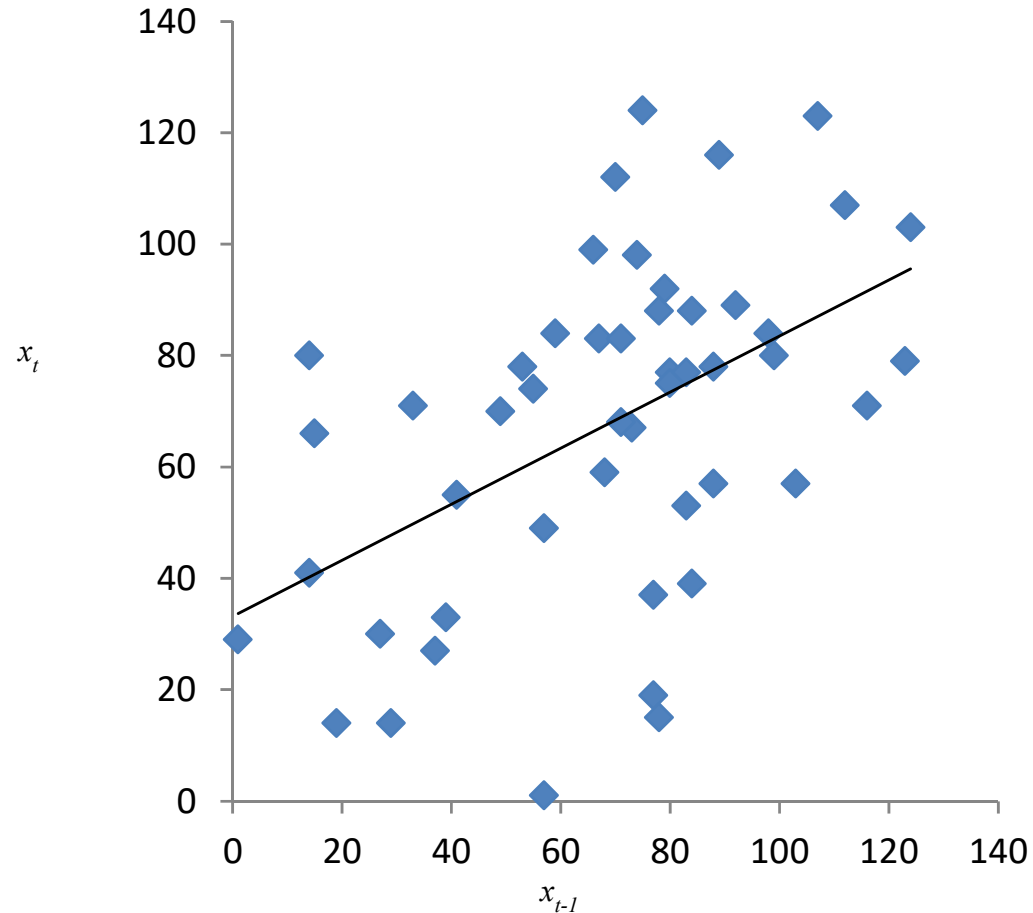
1. Plot x_t as a function of t and look for trends
2. x_t vs. x_{t-1} for linearity
3. Errors e_t vs. predicted values \hat{x}_t for additivity
4. Q-Q Plot of errors for Normality
5. Errors e_t vs. t for IID

1. Plot of x_t



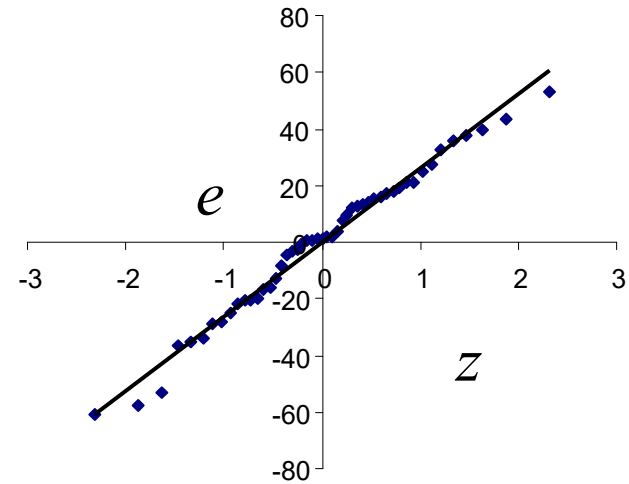
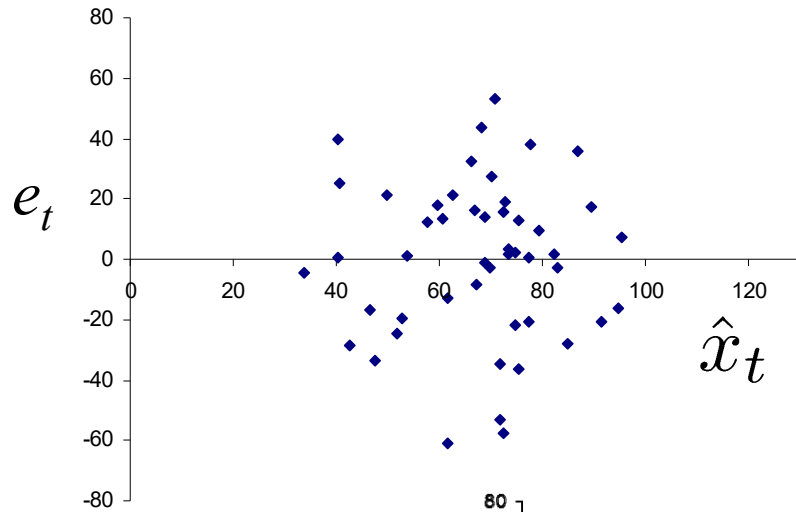
Visual Tests (Cont)

2. Plot of x_t vs. x_{t-1}

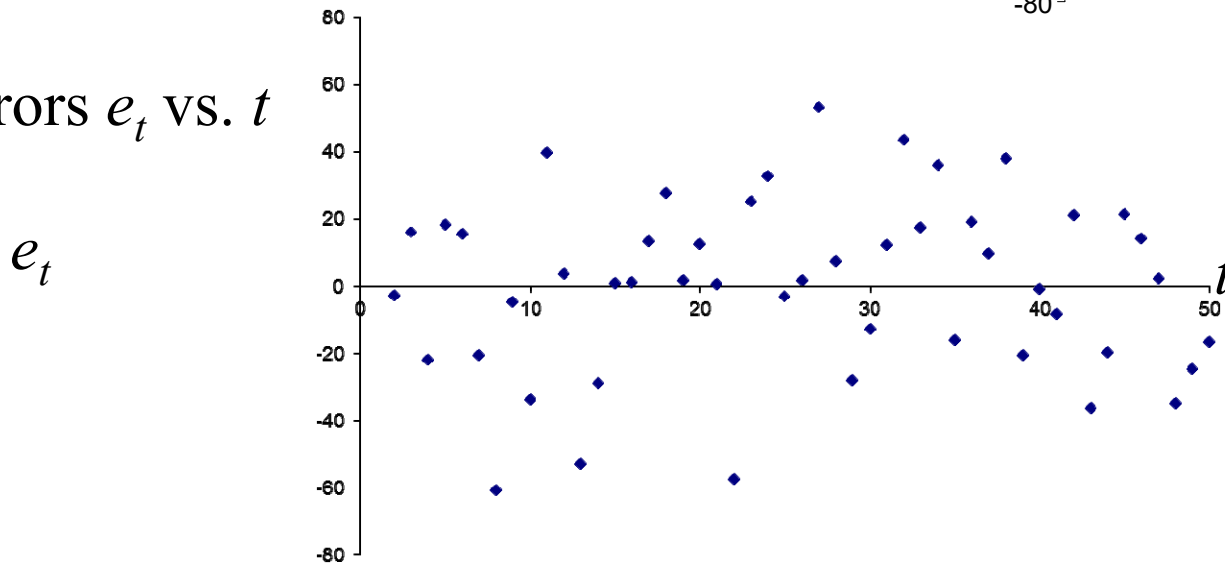


Visual Tests (Cont)

3. Errors e_t vs. predicted values 4. Q-Q Plot of errors



5. Errors e_t vs. t



Exercise 37.3

- Conduct visual tests to verify whether or not the AR(1) model fitted in Exercise 37.1 is appropriate .

AR(p) Model

- x_t is a function of the last p values:

$$x_t = a_0 + a_1x_{t-1} + a_2x_{t-2} + \cdots + a_px_{t-p} + e_t$$

- AR(2): $x_t = a_0 + a_1x_{t-1} + a_2x_{t-2} + e_t$

- AR(3): $x_t = a_0 + a_1x_{t-1} + a_2x_{t-2} + a_3x_{t-3} + e_t$

Backward Shift Operator

$$B(x_t) = x_{t-1}$$

□ Similarly, $B(B(x_t)) = B(x_{t-1}) = x_{t-2}$

□ Or $B^2 x_t = x_{t-2}$

$$B^3 x_t = x_{t-3}$$

$$B^k x_t = x_{t-k}$$

□ Using this notation, AR(p) model is:

$$x_t - a_1 x_{t-1} - a_2 x_{t-2} - \cdots - a_p x_{t-p} = a_0 + e_t$$

$$x_t - a_1 B x_t - a_2 B^2 x_t - \cdots - a_p B^p x_t = a_0 + e_t$$

$$(1 - a_1 B - a_2 B^2 - \cdots - a_p B^p) x_t = a_0 + e_t$$

$$\phi_p(B) x_t = a_0 + e_t$$

□ Here, ϕ_p is a polynomial of degree p .

AR(p) Parameter Estimation

$$x_t = a_0 + a_1x_{t-1} + a_2x_{t-2} + e_t$$

- The coefficients a_i 's can be estimated by minimizing SSE using Multiple Linear Regression.

$$\text{SSE} = \sum e_t^2 = \sum_{t=3}^n (x_t - a_0 - a_1x_{t-1} - a_2x_{t-2})^2$$

- Optimal a_0 , a_1 , and $a_2 \Rightarrow$ Minimize SSE
 \Rightarrow Set the first differential to zero:

$$\frac{d}{da_0} \text{SSE} = \sum_{t=3}^n -2(x_t - a_0 - a_1x_{t-1} - a_2x_{t-2}) = 0$$

$$\frac{d}{da_1} \text{SSE} = \sum_{t=3}^n -2x_{t-1}(x_t - a_0 - a_1x_{t-1} - a_2x_{t-2}) = 0$$

$$\frac{d}{da_2} \text{SSE} = \sum_{t=3}^n -2x_{t-2}(x_t - a_0 - a_1x_{t-1} - a_2x_{t-2}) = 0$$

AR(p) Parameter Estimation (Cont)

- The equations can be written as:

$$\begin{bmatrix} n-2 & \sum x_{t-1} & \sum x_{t-2} \\ \sum x_{t-1} & \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} \\ \sum x_{t-2} & \sum x_{t-1}x_{t-2} & \sum x_{t-2}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum x_t \\ \sum x_t x_{t-1} \\ \sum x_t x_{t-2} \end{bmatrix}$$

Note: All sums are for $t=3$ to n . $n-2$ terms.

- Multiplying by the inverse of the first matrix, we get:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} n-2 & \sum x_{t-1} & \sum x_{t-2} \\ \sum x_{t-1} & \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} \\ \sum x_{t-2} & \sum x_{t-1}x_{t-2} & \sum x_{t-2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_t \\ \sum x_t x_{t-1} \\ \sum x_t x_{t-2} \end{bmatrix}$$

AR(p) Parameter Estimation (Cont)

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} n-p & \sum x_{t-1} & \sum x_{t-2} & \cdots & \sum x_{t-p} \\ \sum x_{t-1} & \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} & \cdots & \sum x_{t-1}x_{t-p} \\ \sum x_{t-2} & \sum x_{t-1}x_{t-2} & \sum x_{t-2}^2 & \cdots & \sum x_{t-2}x_{t-p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum x_{t-p} & \sum x_{t-1}x_{t-p} & \sum x_{t-2}x_{t-p} & \cdots & \sum x_{t-p}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_t \\ \sum x_t x_{t-1} \\ \sum x_t x_{t-2} \\ \vdots \\ \sum x_t x_{t-p} \end{bmatrix}$$

- All sums are from $t=p$ to $t=n$ and have $n-p$ terms.
- For larger data sets: r_k is the autocorrelation at lag k

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} 1 & r_1 & \cdots & r_{p-1} \\ r_1 & 1 & \cdots & r_{p-2} \\ \vdots & \vdots & \vdots & \vdots \\ r_{p-1} & r_{p-2} & \cdots & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{bmatrix}$$

$(i,j)^{\text{th}}$ term = $r_{|i-j|}$ $\nearrow a_0 = (1 - a_1 - a_2 - \cdots - a_p) \bar{x}$

Example 37.5

- Consider the data of Example 37.1 and fit an AR(2) model:

$$\begin{aligned}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} &= \begin{bmatrix} 1 & r_1 \\ r_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0.496 \\ 0.496 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.496 \\ 0.101 \end{bmatrix} \\ &= \begin{bmatrix} 0.592 \\ -0.192 \end{bmatrix}\end{aligned}$$

$$a_0 = (1 - a_1 - a_2)\bar{x} = (1 - 0.592 + 0.192)67.72 = 40.688$$

- SSE= 31979.39
- Small sample \Rightarrow Values of a_0 , a_1 , and a_2 are approximate.
- Exact model by regression:

$$x_t = 39.979 + 0.587x_{t-1} - 0.180x_{t-2} + e_t$$

$$\text{SSE}=31969.99$$

Exercise 37.4

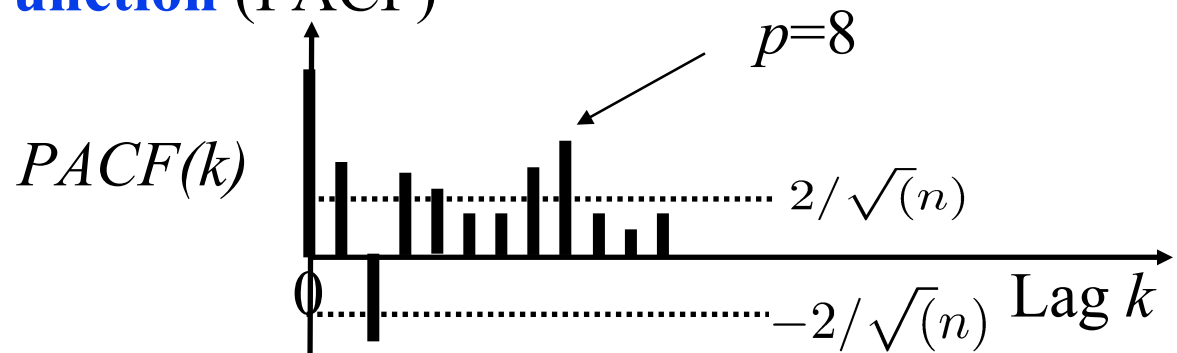
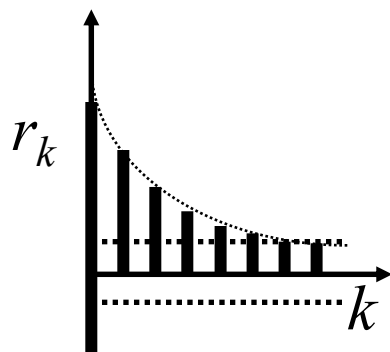
- Fit an AR(2) model to the data of Exercise 37.1. Determine parameters a_0 , a_1 , a_2 and the SSE using multiple regression. Repeat the determination of parameters using autocorrelation function values.

Exercise 37.5

- Fit an AR(3) model to the data of Exercise 37.1. Determine parameters a_0, a_1, a_2, a_3 and the SSE using multiple regression.

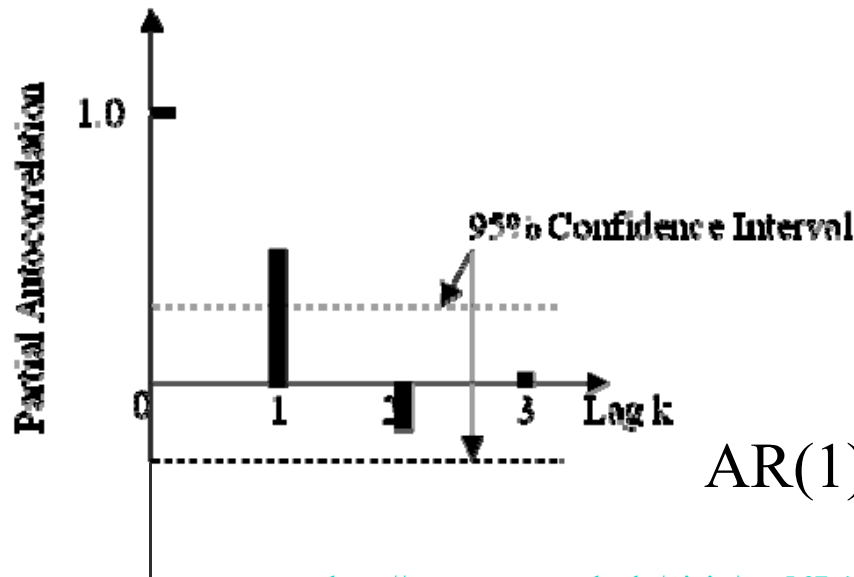
Determining the Order AR(p) $a_p \mp 2/\sqrt{(n)}$

- ❑ ACF of AR(1) is an exponentially decreasing fn of k
- ❑ Fit AR(p) models of order $p=0, 1, 2, \dots$
- ❑ Compute the $(1-\alpha)\%$ confidence intervals of $a_p = a_p \mp \frac{z_{1-\alpha/2}}{\sqrt{n}}$
- ❑ After some p , the last coefficients a_p will not be significant for all higher order models.
- ❑ This highest p is the order of the AR(p) model for the series.
- ❑ This sequence of last coefficients is also called "**Partial Autocorrelation Function (PACF)**"



Example 37.6

- ❑ For the data of Example 37.1, we have:
- ❑ AR(1): $x_t = 33.181 + 0.503x_{t-1} + e_t$
- ❑ AR(2): $x_t = 39.979 + 0.587x_{t-1} - 0.180x_{t-2} + e_t$
- ❑ Similarly, AR(3): $x_t = 37.313 + 0.598x_{t-1} - 0.211x_{t-2} + 0.052x_{t-3} + e_t$
- ❑ PACF at lags 1, 2, and 3 are: 0.503, -0.180, and 0.052



AR(1) is appropriate.

Computing PACF

PACF at lag 1 = $s_1 = a_1$ in AR(1) = r_1 $|M|$ = Determinant of M

PACF at lag 2 = $s_2 = a_2$ in AR(2) = $\frac{\begin{vmatrix} 1 & r_1 \\ r_1 & r_2 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 \\ r_1 & 1 \end{vmatrix}}$

PACF at lag 3 = $s_3 = a_3$ in AR(3) = $\frac{\begin{vmatrix} 1 & r_1 & r_1 \\ r_1 & 1 & r_2 \\ r_2 & r_1 & r_3 \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_1 \\ r_2 & r_1 & 1 \end{vmatrix}}$

Computing PACF (Cont)

$$\text{PACF at lag } k = s_k = a_k \text{ in AR}(k) = \frac{\begin{vmatrix} 1 & r_1 & \cdots & r_{k-2} & r_1 \\ r_1 & 1 & \cdots & r_{k-3} & r_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{k-2} & r_{k-3} & \cdots & 1 & r_{k-1} \\ r_{k-1} & r_{k-2} & \cdots & r_1 & r_k \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & \cdots & r_{k-2} & r_{k-1} \\ r_1 & 1 & \cdots & r_{k-3} & r_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{k-2} & r_{k-3} & \cdots & 1 & r_1 \\ r_{k-1} & r_{k-2} & \cdots & r_1 & 1 \end{vmatrix}}$$

Exercise 37.6

- Using the results of Exercises 37.1, 37.4, and 37.5, determine the partial autocorrelation function at lags 1, 2, 3 for the data of Exercise 37.1. Determine which values are significant. Based on this which $AR(p)$ model will be appropriate for this data?

Moving Average (MA) Models



- Moving Average of order 1: MA(1)

$$x_t - b_0 = e_t + b_1 e_{t-1}$$

b_0 is the mean of the time series.

- The parameters b_0 and b_1 cannot be estimated using standard regression formulas since we do not know errors. The errors depend on the parameters.
- So the only way to find optimal b_0 and b_1 is by iteration.
⇒ Start with some suitable values and change b_0 and b_1 until SSE is minimized and average of errors is zero.

Example 37.4

- Consider the data of Example 37.1.

- For this data: $\bar{x} = \frac{1}{50} \sum_{t=1}^{50} x_t = 67.72$

- We start with $b_0 = 67.72$, $b_1 = 0.4$,
Assuming $e_0 = 0$, compute all the errors and SSE.

$$e_1 = x_1 - b_0 - b_1 e_0 = 73 - 67.72 - 0.4 \times 0 = 5.28$$

$$e_2 = x_2 - b_0 - b_1 e_1 = 67 - 67.72 - 0.4 \times 5.28 = -2.832$$

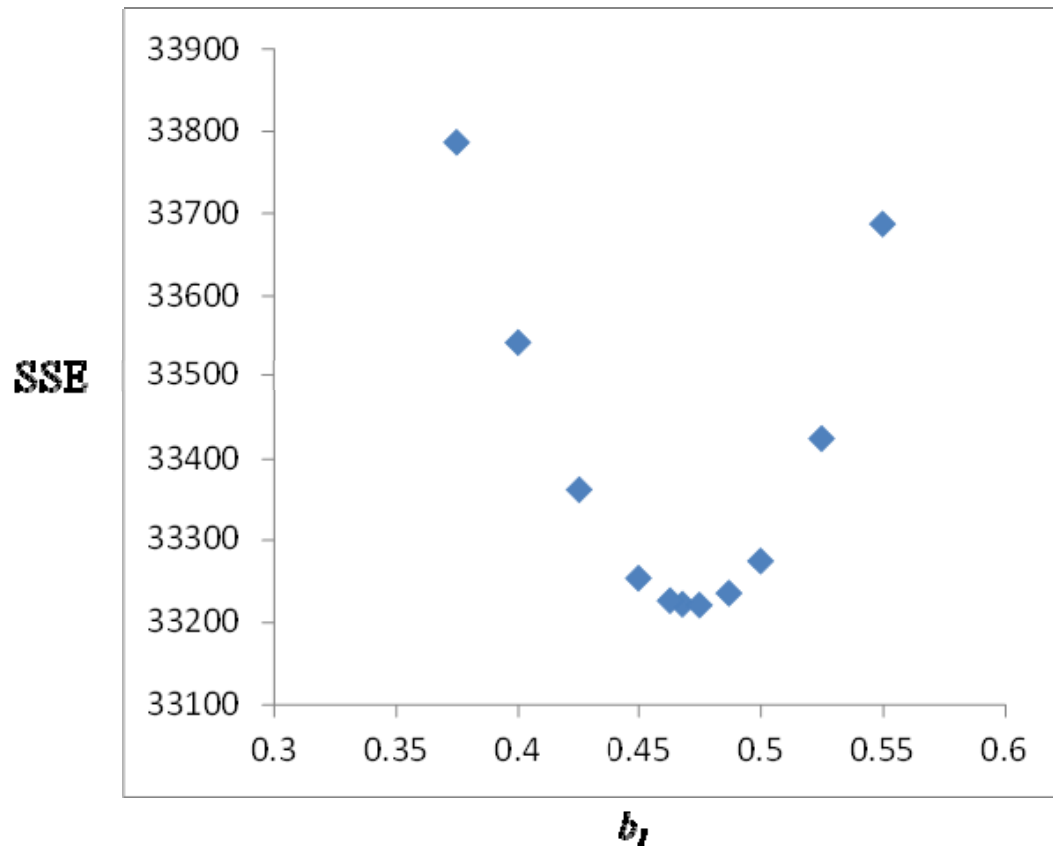
...

$$\bar{e} = \frac{1}{50} \sum_{t=1}^{50} e_t = -0.152 \quad \text{and SSE} = 33542.8$$

- We then adjust a_0 and b_1 until SSE is minimized and mean error is close to zero.

Example 37.4 (Cont)

- The steps are: Starting with $b_0 = \bar{x}$ and trying various values of b_1 . SSE is minimum at $b_1=0.475$. SSE= 33221.06

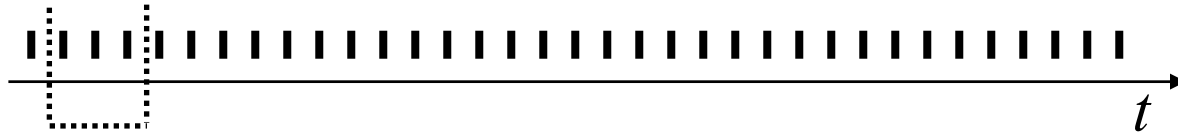


Example 37.4 (Cont)

$$\bar{e} = \frac{1}{50} \sum_{t=1}^{50} e_t = -0.1661$$

- Keeping $b_1=0.475$, try neighboring values of b_0 to get average error as close to zero as possible.
- $b_0= 67.475$ gives $\bar{e}=-0.001$ SSE=33221.93

MA(q) Models



- ❑ Moving Average of order 1: MA(1)

$$x_t - b_0 = e_t + b_1 e_{t-1}$$

- ❑ Moving Average of order 2: MA(2)

$$x_t - b_0 = e_t + b_1 e_{t-1} + b_2 e_{t-2}$$

- ❑ Moving Average of order q : MA(q)

$$x_t - b_0 = e_t + b_1 e_{t-1} + b_2 e_{t-2} + \cdots + b_q e_{t-q}$$

- ❑ Moving Average of order 0: MA(0) (Note: This is also AR(0))

$$x_t - b_0 = e_t$$

$x_t - b_0$ is a white noise. b_0 is the mean of the time series.

Exercise 37.7

- Fit an MA(0) model to the data of Exercise 37.1.
Determine parameter b_0 and SSE

MA(q) Models (Cont)

- Using the backward shift operator B , MA(q):

$$\begin{aligned}x_t - b_0 &= e_t + b_1 B e_t + b_2 B^2 e_t + \cdots + b_q B^q e_t \\ &= (1 + b_1 B + b_2 B^2 + \cdots + b_q B^q) e_t \\ &= \Psi_q(B) e_t\end{aligned}$$

- Here, Ψ_q is a polynomial of order q .

Example 37.8

- Fit MA(2) model to the data of Example 37.1

$$x_t = b_0 + e_t + b_1 e_{t-1} + b_2 e_{t-2}$$

- Round 1: Setting $b_0 = \bar{x}_t = 67.72$, we try 9 combinations of $b_1 = \{0.2, 0.3, 0.4\}$ and $b_2 = \{0.2, 0.3, 0.4\}$.
Minimum SSE is 33490.26 at $b_1 = 0.4$ and $b_2 = 0.2$
- Round 2: Try 4 new points around the current minimum
 $b_1 = \{0.35, 0.45\}$ and $b_2 = \{0.15, 0.25\}$
Minimum SSE is 32551.62 at $b_1 = 0.45$, $b_2 = 0.15$
- Round 3: Try 4 new points around the current minimum.
Try $b_1 = \{0.425, 0.475\}$ and $b_2 = \{0.125, 0.175\}$
Minimum SSE is 32342.61 at $b_1 = 0.475$, $b_2 = 0.125$

Example 37.8 (Cont)

- ❑ Round 4: Try 4 new points around the current minimum.
Try $b_1 = \{0.4625, 0.4875\}$ and $b_2 = \{0.125, 0.175\}$
Minimum SSE is 32201.58 at $b_1 = 0.4875, b_2 = 0.125$
- ❑ Round 5: Try 4 new points around the current minimum.
Try $b_1 = \{0.481, 0.493\}$ and $b_2 = \{0.112, 0.137\}$
Minimum SSE is 32148.21 at $b_1 = 0.493, b_2 = 0.137$
- ❑ Since the decrease in SSN is small (close to 0.1%), we arbitrarily stop here.
- ❑ The model is:

$$x_t = 67.72 + e_t + 0.493e_{t-1} + 0.137e_{t-2}$$

Exercise 38.8

- Fit an MA(1) model to the data of Exercise 37.1. Determine parameters b_0 , b_1 and the minimum SSE.

Autocorrelations for MA(1)

- For this series, the mean is:

$$\mu = E[x_t] = b_0 + E[e_t] + b_1 E[e_{t-1}] = b_0$$

- The variance is:

$$\begin{aligned}\text{Var}[x_t] &= E[(x_t - \mu)^2] = E[(e_t + b_1 e_{t-1})^2] \\ &= E[e_t^2 + 2b_1 e_t e_{t-1} + b_1^2 e_{t-1}^2] \\ &= E[e_t^2] + 2b_1 E[e_t e_{t-1}] + b_1^2 E[e_{t-1}^2] \\ &= \sigma^2 + 2b_1 \times 0 + b_1^2 \sigma^2 = (1 + b_1^2) \sigma^2\end{aligned}$$

- The autocovariance at lag 1 is:

$$\begin{aligned}\text{autocovar at lag 1} &= E[(x_t - \mu)(x_{t-1} - \mu)] \\ &= E[(e_t + b_1 e_{t-1})(e_{t-1} + b_1 e_{t-2})] \\ &= E[e_t e_{t-1} + b_1 e_{t-1} e_{t-1} + b_1 e_t e_{t-2} + b_1^2 e_{t-1} e_{t-2}] \\ &= E[0 + b_1 E[e_{t-1}^2] + 0 + 0] \\ &= b_1 \sigma^2\end{aligned}$$

Autocorrelations for MA(1) (Cont)

- The autocovariance at lag 2 is:

$$\begin{aligned}\text{Covar at lag 2} &= E[(x_t - \mu)(x_{t-2} - \mu)] \\ &= E[(e_t + b_1 e_{t-1})(e_{t-2} + b_1 e_{t-3})] \\ &= E[e_t e_{t-2} + b_1 e_{t-1} e_{t-2} + b_1 e_t e_{t-3} + b_1^2 e_{t-1} e_{t-3}] \\ &= 0 + 0 + 0 + 0 = 0\end{aligned}$$

- For MA(1), the autocovariance at all higher lags ($k > 1$) is 0.

- The autocorrelation is:

$$r_k = \begin{cases} 1 & k = 0 \\ \frac{b_1}{1+b_1^2} & k = 1 \\ 0 & k > 1 \end{cases}$$

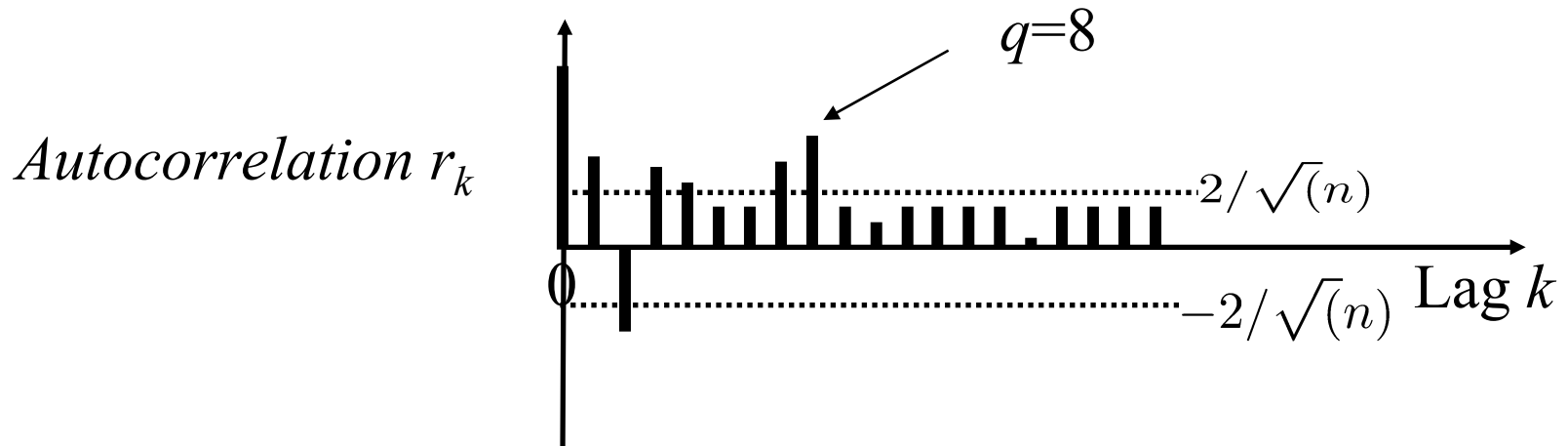
- The autocorrelation of MA(q) series is non-zero only for lags $k \leq q$ and is zero for all higher lags.

Example 37.9

- ❑ For the data of Example 37.1:
- ❑ Autocorrelation is zero for all lags $k > 1$.
- ❑ MA(1) model is appropriate for this data.

Example 37.10

- ❑ The order of the last significant r_k determines the order of the MA(q) model.
- ❑ For the following data, all autocorrelations at lag 9 and higher are zero \Rightarrow MA(8) model would be appropriate



Exercise 37.9

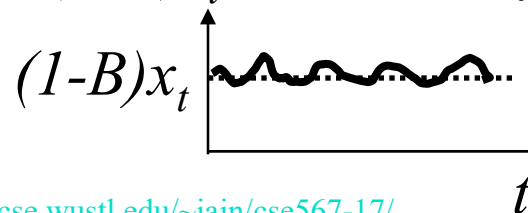
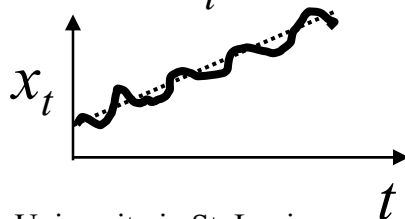
- Fit an MA(2) model to the data of Exercise 37.2. Determine parameters b_0 , b_1 , b_2 and the minimum SSE. For this data, which model would you choose MA(0), MA(1) or MA(2) and why?

Duality of AR(p) vs. MA(q)

- ❑ Determining the coefficients of AR(p) is straight forward but determining the order p requires an iterative procedure
- ❑ Determining the order q of MA(q) is straight forward but determining the coefficients requires an iterative procedure

Non-Stationarity: Integrated Models

- ❑ In the white noise model MA(0): $x_t = b_0 + e_t$
- ❑ The mean b_0 is independent of time.
- ❑ If it appears that the time series is increasing approximately linearly with time, the first difference of the series can be modeled as white noise: $(x_t - x_{t-1}) = b_0 + e_t$
- ❑ Or using the B operator: $(1-B)x_t = x_t - x_{t-1}$
$$(1-B)x_t = b_0 + e_t$$
- ❑ This is called an "integrated" model of order 1 or I(1). Since the errors are integrated to obtain x.
- ❑ Note that x_t is not stationary but $(1-B)x_t$ is stationary.



Integrated Models (Cont)

- If the time series is parabolic, the second difference can be modeled as white noise:

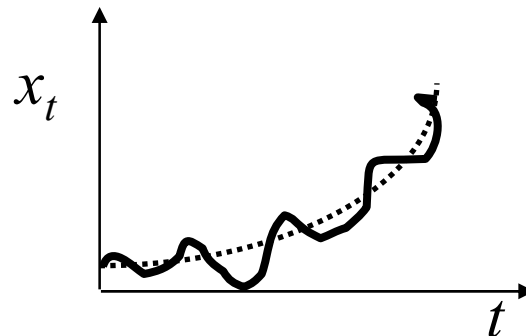
$$(x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = b_0 + e_t$$

- Or $(1 - B)^2 x_t = b_0 + e_t$

This is an I(2) model. Also written as:

$$D^2 x_t = b_0 + e_t$$

Where Operator $D = 1 - B$



ARMA and ARIMA Models

- It is possible to combine AR, MA, and I models

- ARMA(p, q) Model:

$$x_t - a_1x_{t-1} - \dots - a_px_{t-p} = b_0 + e_t + b_1e_{t-1} + \dots + b_qe_{t-q}$$
$$\phi_p(B)x_t = b_0 + \psi_q(B)e_t$$

- ARIMA(p, d, q) Model:

$$\phi_p(B)(1 - B)^d x_t = b_0 + \psi_q(B)e_t$$

- Using algebraic manipulations, it is possible to transform AR models to MA models and vice versa.

Example 37.11

- Consider the MA(1) model: $x_t = b_0 + e_t + b_1 e_{t-1}$
- It can be written as: $(x_t - b_0) = (1 + b_1 B)e_t$

$$(1 + b_1 B)^{-1}(x_t - b_0) = e_t$$

$$(1 - b_1 B + b_1^2 B^2 - b_1^3 B^3 + \dots)(x_t - b_0) = e_t$$

$$(x_t - b_1 x_{t-1} + b_1^2 x_{t-2} - b_1^3 x_{t-3} + \dots) - \frac{b_0}{1 + b_1} = e_t$$

$$x_t = \frac{b_0}{1 + b_1} + b_1 x_{t-1} - b_1^2 x_{t-2} + b_1^3 x_{t-3} - \dots + e_t$$

- If $b_1 < 1$, the coefficients decrease and soon become insignificant. This results in a finite order AR model.

Exercise 39.10

- Convert the following AR(1) model to an equivalent MA model:

$$x_t = a_0 + a_1 x_{t-1} + e_t$$

Non-Stationarity due to Seasonality

- ❑ The mean temperature in December is always lower than that in November and in May it always higher than that in March
⇒ Temperature has a yearly season.
- ❑ One possible model could be I(12):

$$x_t - x_{t-12} = b_0 + e_t$$

- ❑ or

$$(1 - B^{12})x_t = b_0 + e_t$$

Seasonal ARIMA (SARIMA) Models

- SARIMA $(p, d, q) \times (P, R, Q)^s$ Model:

$$\phi_p(B)\Phi_P(B^s)(1 - B^s)^R(1 - B)^d x_t = b_0 + \psi_q(B)\Psi_Q(B^s)e_t$$

- Fractional ARIMA (FARIMA) Models

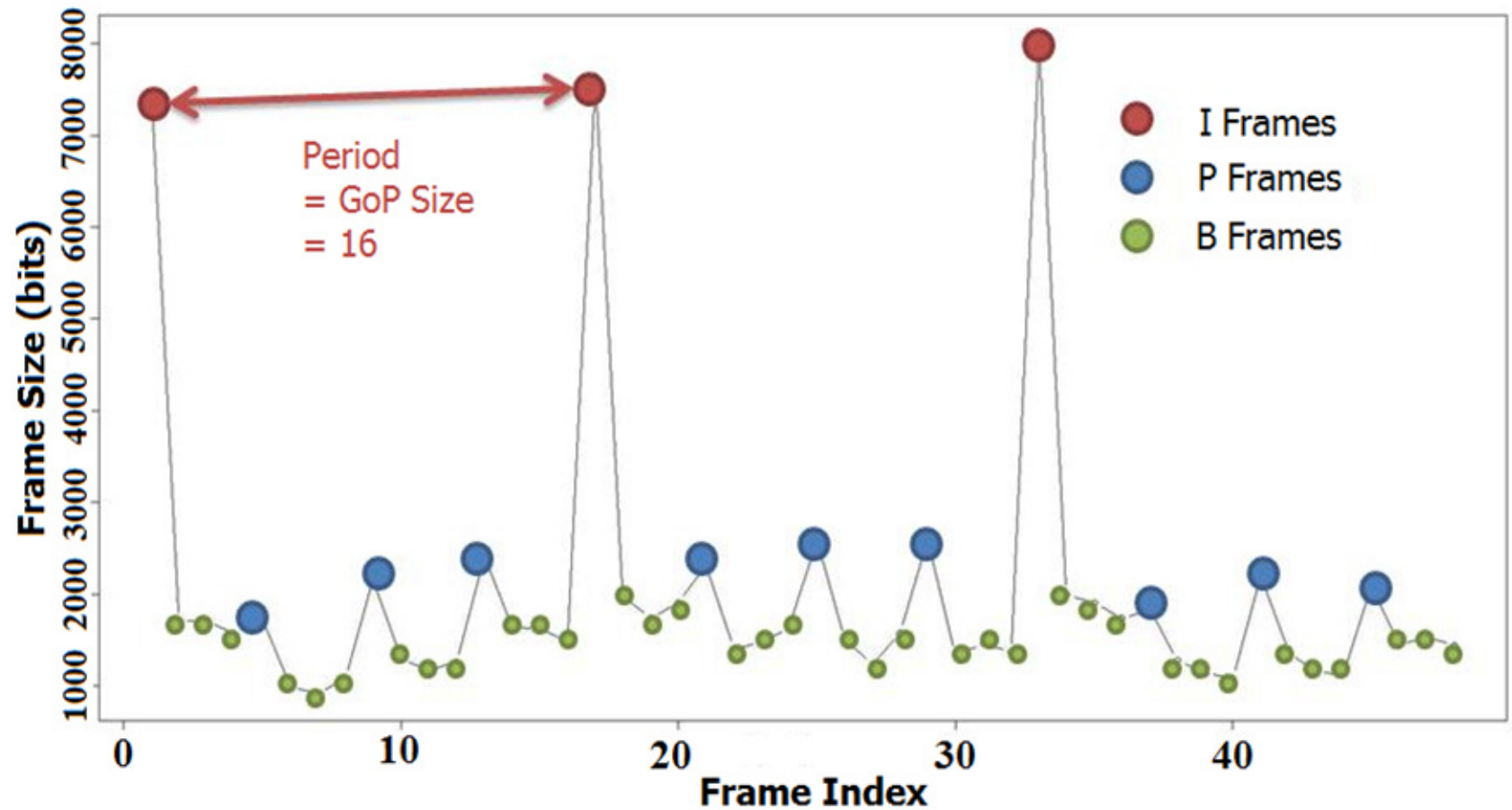
ARIMA(p, d+ δ , q) $-0.5 \leq \delta \leq 0.5$

\Rightarrow Fractional Integration allowed.

Exercise 37.11

- Write the expression for SARIMA(1,0,1)(0,1,0)¹² model in terms of x 's and e 's.

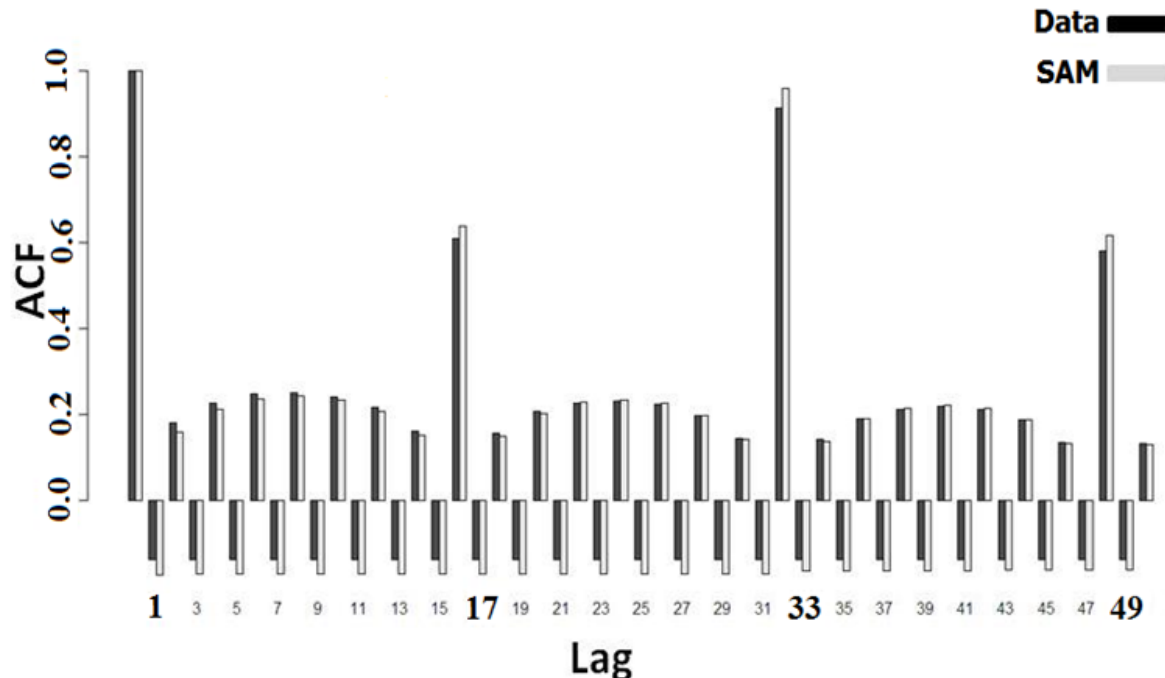
Case Study 37.1: Mobile Video



□ Observation: Every 16th frame is a large (I) frame.

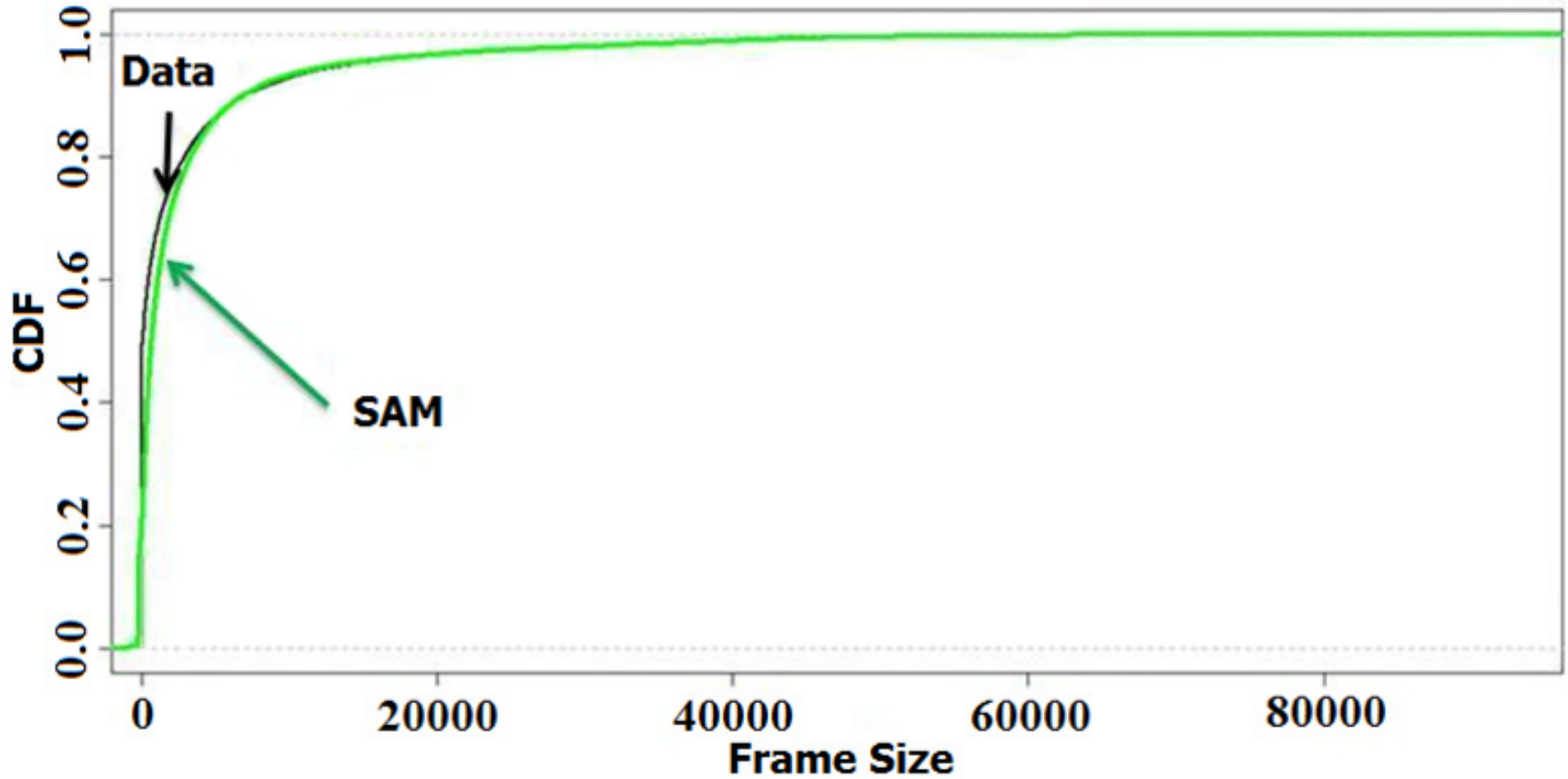
Traffic Modeling – All Frames

- A closer look at the ACF graph shows a strong continual correlation every 16 lag → GOP size

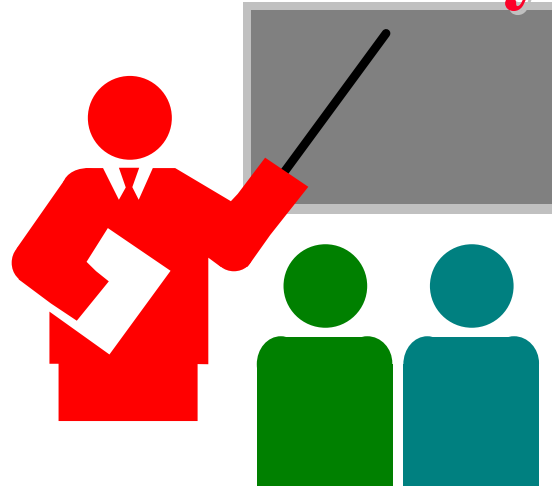


- Result: SARIMA (1, 0, 1) \times (1,1,1)^s Model, s=group size =16

Validation



Summary



- AR(1) Model:

$$x_t = a_0 + a_1 x_{t-1} + e_t$$

- MA(1) Model:

$$x_t - b_0 = e_t + b_1 e_{t-1}$$

- ARIMA(1,1,1) Model:

$$x_t - x_{t-1} = a_0 + a_1 (x_{t-1} - x_{t-2}) + e_t + b_1 e_{t-1}$$

- Seasonal ARIMA (1,0,1)x(0,1,0)¹² model:

$$x_t - x_{t-12} = a_0 + a_1 (x_{t-1} - x_{t-13}) + e_t + b_1 e_{t-1}$$

Scan This to Download These Slides



Raj Jain

<http://rajjain.com>

Related Modules



CSE567M: Computer Systems Analysis (Spring 2013),

https://www.youtube.com/playlist?list=PLjGG94etKypJEKjNAa1n_1X0bWWNyZcof

CSE473S: Introduction to Computer Networks (Fall 2011),

https://www.youtube.com/playlist?list=PLjGG94etKypJWOSPMh8Azcgy5e_10TiDw



Wireless and Mobile Networking (Spring 2016),

https://www.youtube.com/playlist?list=PLjGG94etKypKeb0nzyN9tSs_HCd5c4wXF

CSE571S: Network Security (Fall 2011),

<https://www.youtube.com/playlist?list=PLjGG94etKypKvzfVtutHcPFJXumyyg93u>



Video Podcasts of Prof. Raj Jain's Lectures,

<https://www.youtube.com/channel/UCN4-5wzNP9-ruOzQMs-8NUw>