

**SMA 6304 / MIT 2.853 / MIT 2.854  
Manufacturing Systems**

**Lecture 11: Forecasting**

Lecturer: Prof. Duane S. Boning

# Agenda

## 1. Regression

- Polynomial regression
- Example (using Excel)

## 2. Time Series Data & Regression

- Autocorrelation – ACF
- Example: white noise sequences
- Example: autoregressive sequences
- Example: moving average
- ARIMA modeling and regression

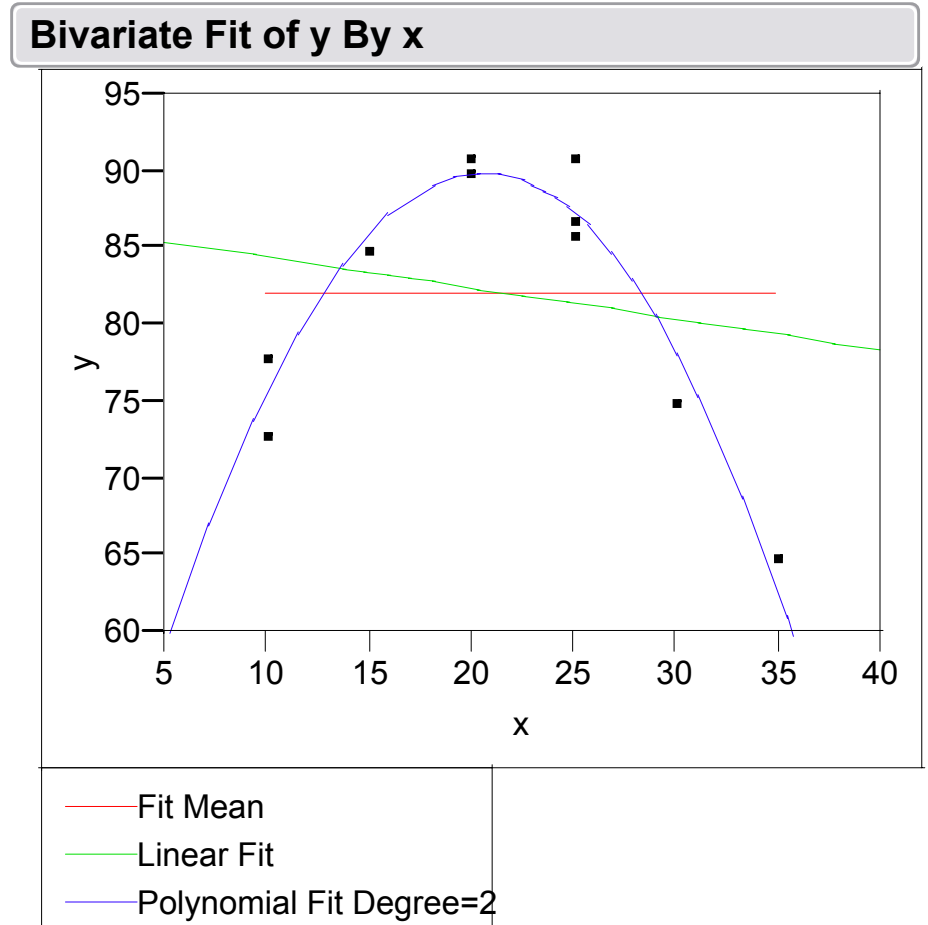
## 3. Forecasting Examples

# Regression – Review & Extensions

- Single Model Coefficient: Linear Dependence  $\eta = \beta x$
- Slope and Intercept (or Offset):  $\eta = \beta_0 + \beta_1 x$
- Polynomial and Higher Order Models:  $\eta = \beta_0 + \beta_1 x + \beta_2 x^2$
- Multiple Parameters  $\eta = \beta_0 + \beta_1 x + \beta_2 w$
- Key point: “linear” regression can be used as long as the model is linear in the coefficients (doesn’t matter the dependence in the independent variable)

# Polynomial Regression Example

Image removed due to copyright considerations.



- Replicate data provides opportunity to check for lack of fit

# Growth Rate – First Order Model

- Mean significant, but linear term not
- Clear evidence of lack of fit

Image removed due to copyright considerations.

# Growth Rate – Second Order Model

- No evidence of lack of fit
- Quadratic term significant

Image removed due to copyright considerations.

# Polynomial Regression In Excel

- Create additional input columns for each input
- Use “Data Analysis” and “Regression” tool

x	x <sup>2</sup>	y
10	100	73
10	100	78
15	225	85
20	400	90
20	400	91
25	625	87
25	625	86
25	625	91
30	900	75
35	1225	65

<i>Regression Statistics</i>	
Multiple R	0.968
R Square	0.936
Adjusted R Square	0.918
Standard Error	2.541
Observations	10

<i>ANOVA</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	2	665.706	332.853	51.555	6.48E-05
Residual	7	45.194	6.456		
Total	9	710.9			

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>
Intercept	35.657	5.618	6.347	0.0004	22.373	48.942
x	5.263	0.558	9.431	3.1E-05	3.943	6.582
x <sup>2</sup>	-0.128	0.013	-9.966	2.2E-05	-0.158	-0.097

# Polynomial Regression

## Analysis of Variance

Source	DF	Sum of Square	Mean Squar	F Ratio
Model	2	665.70617	332.853	51.5551
Error	7	45.19383	6.456	Prob > F
C. Total	9	710.90000		<.0001

• Generated using JMP

## Lack Of Fit

Source	DF	Sum of Square	Mean Squar	F Ratio
Lack Of Fit	3	18.193829	6.0646	0.8985
Pure Error	4	27.000000	6.7500	Prob > F
Total Error	7	45.193829		0.5157
				Max RSq
				0.9620

## Summary of Fit

RSquare	0.936427
RSquare Adj	0.918264
Root Mean Sq Error	2.540917
Mean of Response	82.1
Observations (or Sum Wgts)	10

## Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob> t
Intercept	35.657437	5.617927	6.35	0.0004
x	5.2628956	0.558022	9.43	<.0001
x*x	-0.127674	0.012811	-9.97	<.0001

## Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
x	1	1	574.28553	88.9502	<.0001
x*x	1	1	641.20451	99.3151	<.0001



# Agenda

## 1. Regression

- Polynomial regression
- Example (using Excel)

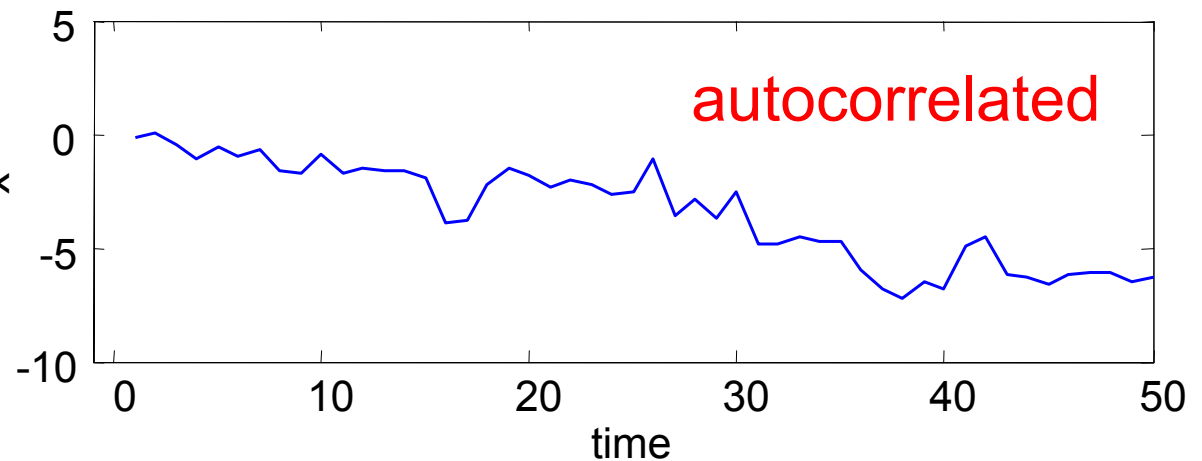
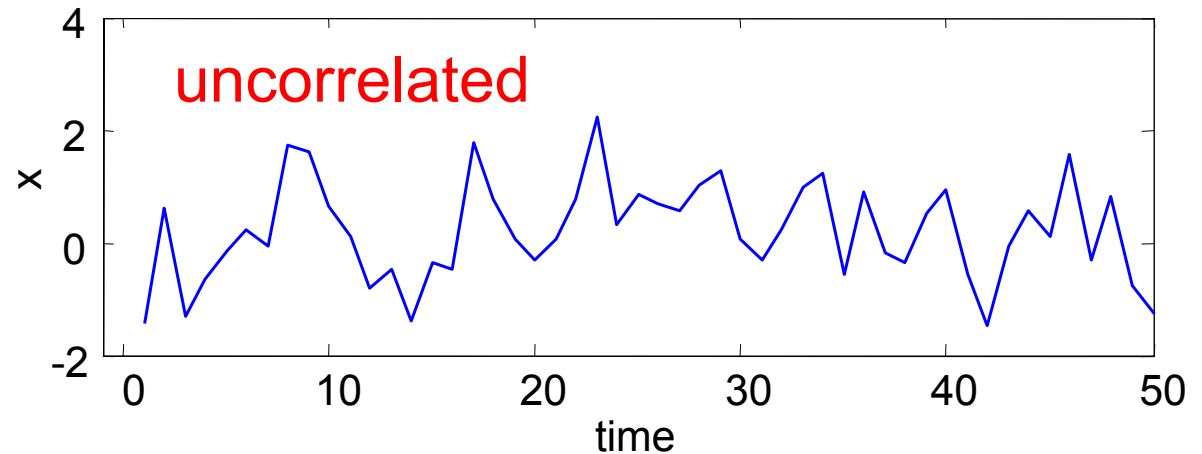
## 2. Time Series Data & Time Series Regression

- Autocorrelation – ACF
- Example: white noise sequences
- Example: autoregressive sequences
- Example: moving average
- ARIMA modeling and regression

## 3. Forecasting Examples

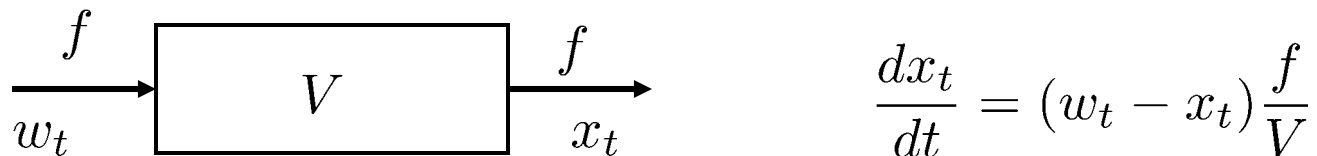
# Time Series – Time as an Implicit Parameter

- Data is often collected with a ***time-order***
- An underlying dynamic process (e.g. due to physics of a manufacturing process) may create ***autocorrelation*** in the data



# Intuition: Where Does Autocorrelation Come From?

- Consider a chamber with volume  $V$ , and with gas flow in and gas flow out at rate  $f$ . We are interested in the concentration  $x$  at the output, in relation to a known input concentration  $w$ .



$$x_t = w_t - \frac{V}{f} \frac{dx_t}{dt} = w_t - \tau \frac{dx_t}{dt}$$

Consider a step change in input of  $w_0$  at  $t = 0$ . Then

$$x_t = w_0(1 - e^{-t/\tau})$$

Discretizing:  $x_t = x_{t-1} + (w_0 - x_{t-1})(1 - e^{-\Delta t/T})$

$$x_t = aw_t + (1 - a)x_{t-1} \quad \text{where } a = 1 - e^{-\Delta t/T}$$

correlation between  $x_t$  &  $x_{t-1}$  is  $\rho = 1 - a = e^{-\Delta t/T}$

# Key Tool: Autocorrelation Function (ACF)

- Time series data: time index  $i$

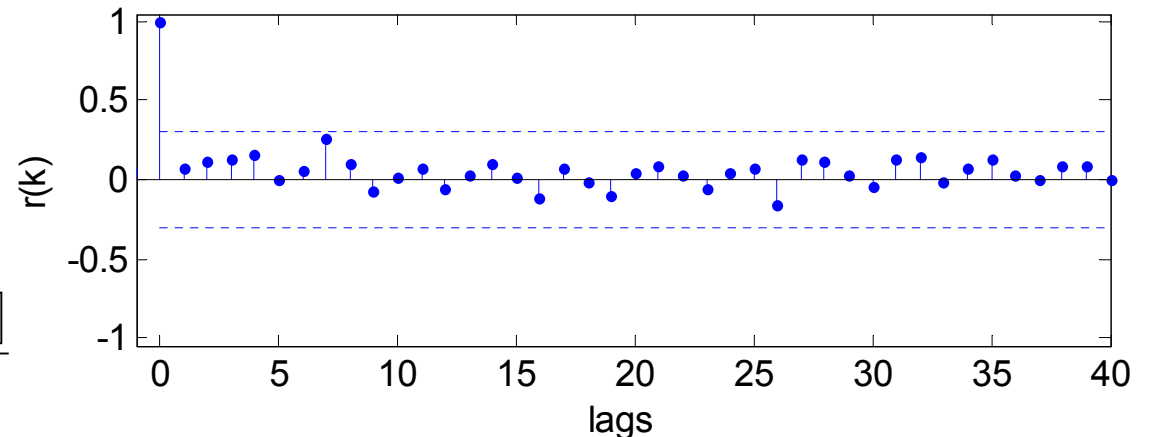
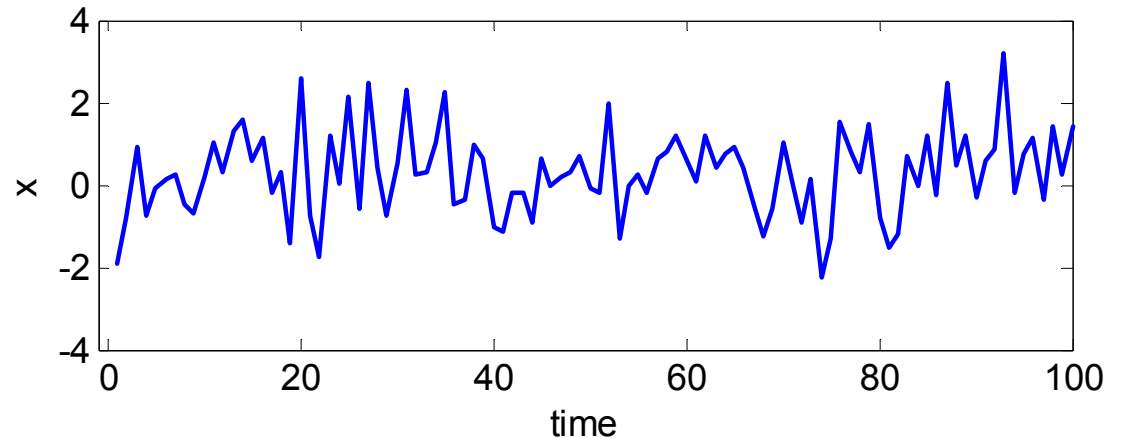
$$x_i \sim N(0, 1)$$

- CCF: cross-correlation function

$$r_{xy}(k) = \frac{1}{N} \sum_{i=1}^{N-1} \frac{[x_i - \bar{x}][y_{i+k} - \bar{y}]}{s_x s_y}$$

- ACF: auto-correlation function

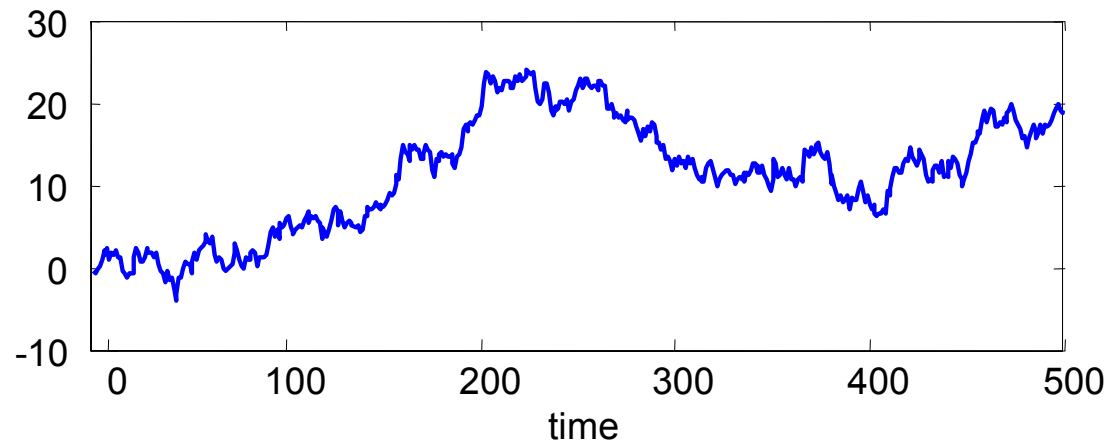
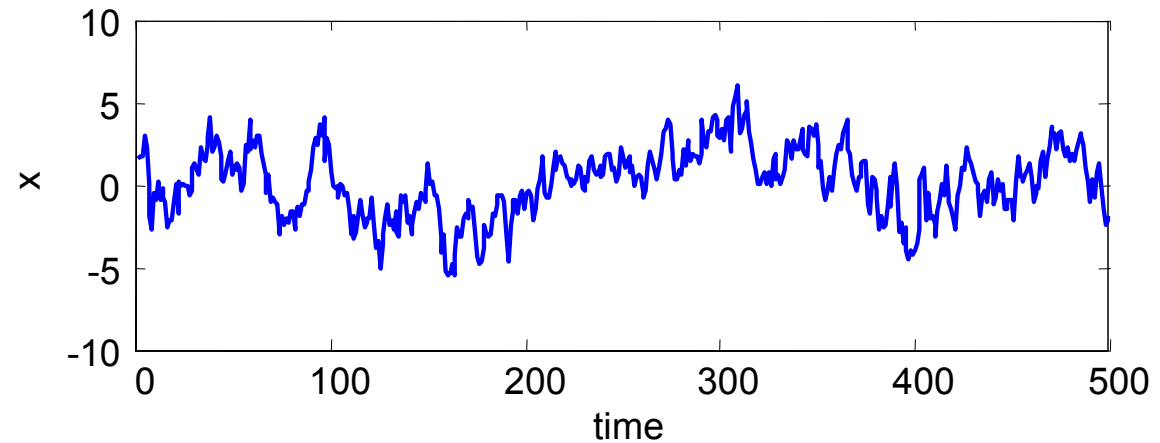
$$r_{xx}(k) = \frac{1}{N} \sum_{i=1}^{N-1} \frac{[x_i - \bar{x}][x_{i+k} - \bar{x}]}{s_x^2}$$



⇒ ACF shows the “similarity” of a signal to a lagged version of same signal

# Stationary vs. Non-Stationary

Stationary series:  
Process has a **fixed** mean



# White Noise – An Uncorrelated Series

- Data drawn from IID gaussian

$$w_i \sim N(0, 1)$$

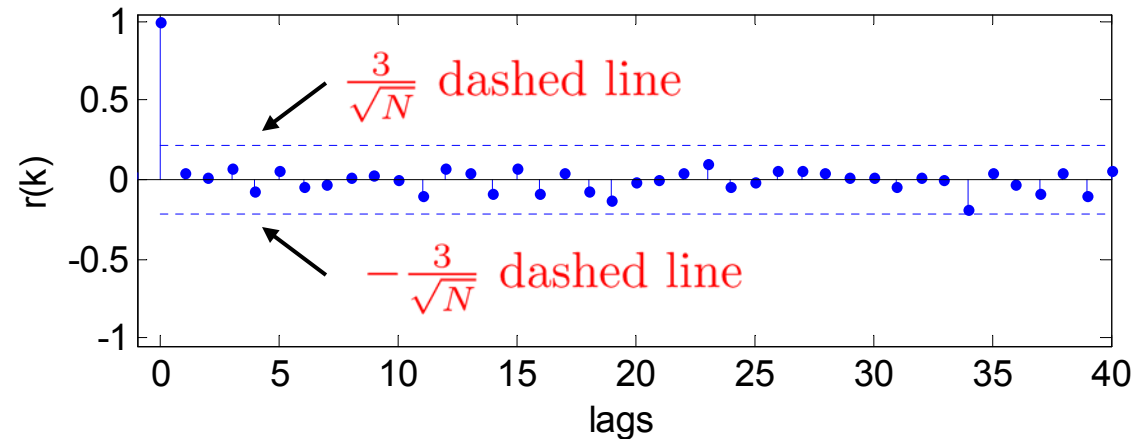
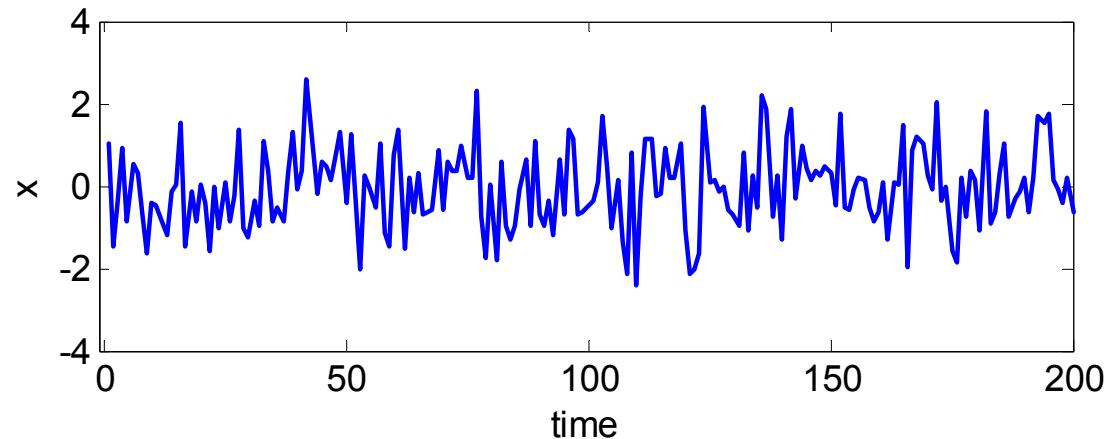
- ACF: We also plot the  $3\sigma$  limits – values within these not significant
- Note that  $r(0) = 1$  always (a signal is always equal to itself with zero lag – perfectly autocorrelated at  $k = 0$ )

- Sample mean

$$\bar{w} = \frac{1}{N} \sum_i^N w_i$$

- Sample variance

$$s_w^2 = \frac{1}{N-1} \sum_i^N (w_i - \bar{w})^2$$



# Autoregressive Disturbances

- Generated by:

$$w_i \sim N(0, 1)$$

$$c_i = \alpha \cdot c_{i-1} + w_i$$

Shown:  $\alpha = 0.9$

- Mean

$$\mu_c = E(c_i) = 0$$

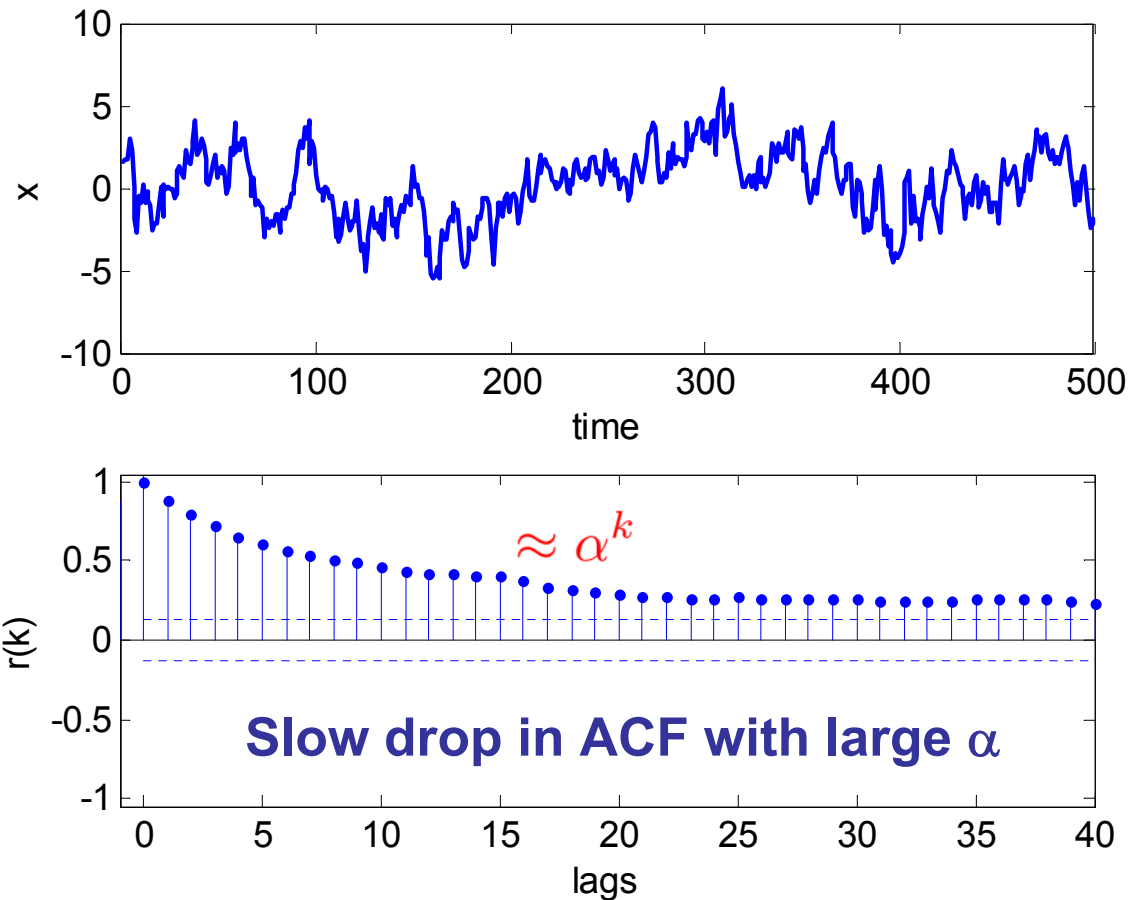
since  $\mu_w = 0$

- Variance

$$\begin{aligned} \sigma_c^2 &= \text{Var}(c_i) = E[(c_i - \bar{c})^2] \\ &= E(\alpha^2 c_{i-1}^2 + 2\alpha c_{i-1} w_i + w_i^2) \\ &= \alpha^2 \text{Var}(c_i) + \text{Var}(w_i) \end{aligned}$$

$$\Rightarrow \sigma_c^2 = \frac{\sigma_w^2}{1 - \alpha^2}$$

So AR (autoregressive) behavior  
increases variance of signal.



# Another Autoregressive Series

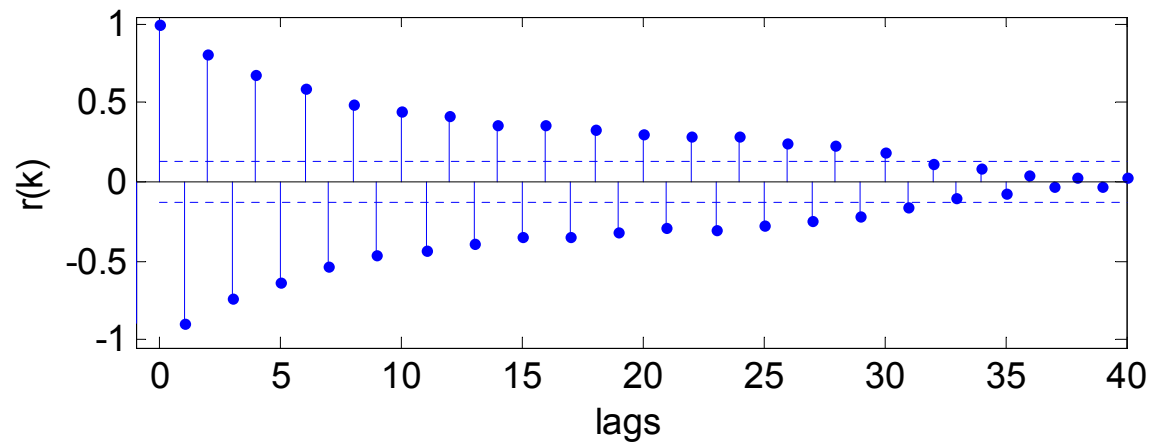
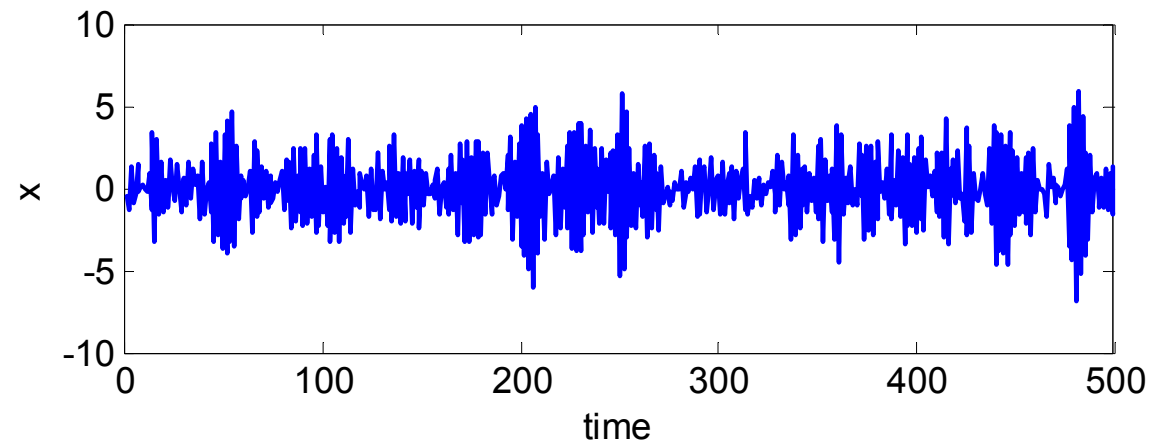
- Generated by:

$$w_i \sim N(0, 1)$$

$$c_i = \alpha \cdot c_{i-1} + w_i$$

Shown:  $\alpha = -0.9$

- High **negative** autocorrelation:



**Slow drop in ACF with large  $\alpha$**

**But now ACF alternates in sign**



# Random Walk Disturbances

- Generated by:

$$w_i \sim N(0, 1)$$
$$c_i = 1 \cdot c_{i-1} + w_i$$

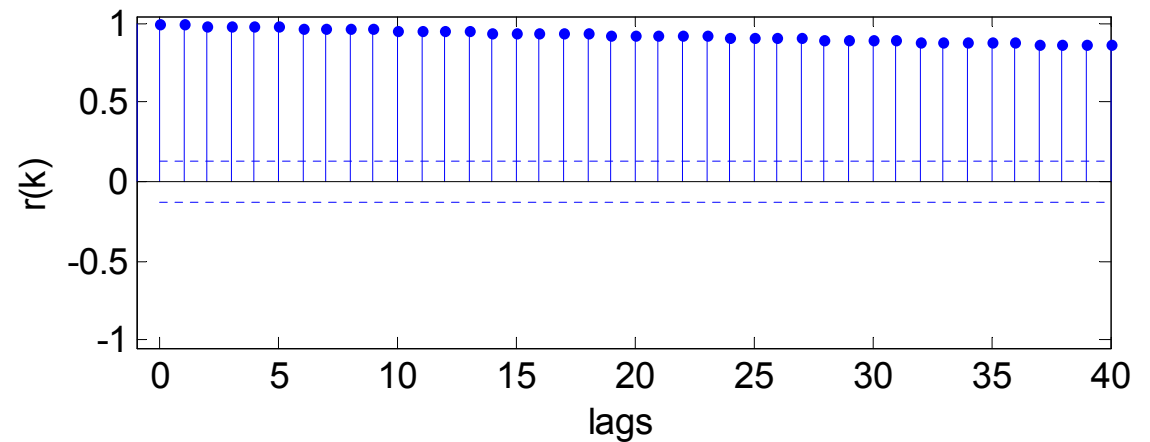
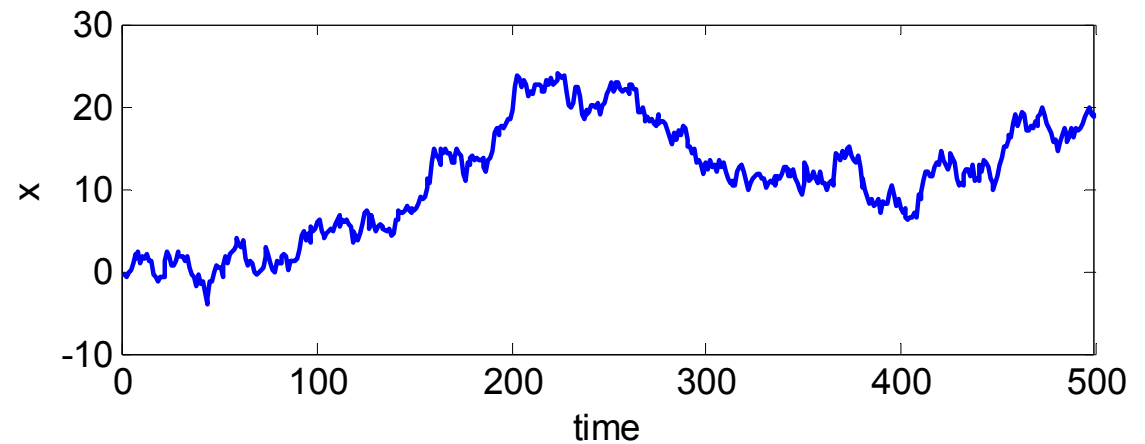
AR with  $\alpha = 1$

- Mean

$\bar{c} \neq 0$  non-stationary

- Variance

Variance increases as  
sequence gets longer



**Very slow drop in ACF for  $\alpha = 1$**

# Moving Average Sequence

- Generated by:

$$w_i \sim N(0, 1)$$

$$c_i = w_i + \beta \cdot w_{i-1}$$

Shown:  $\beta = 0.5$

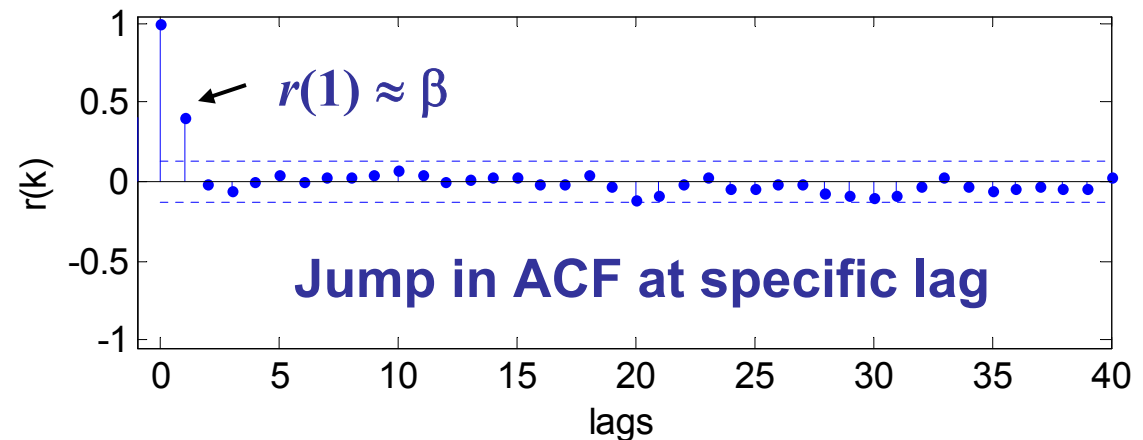
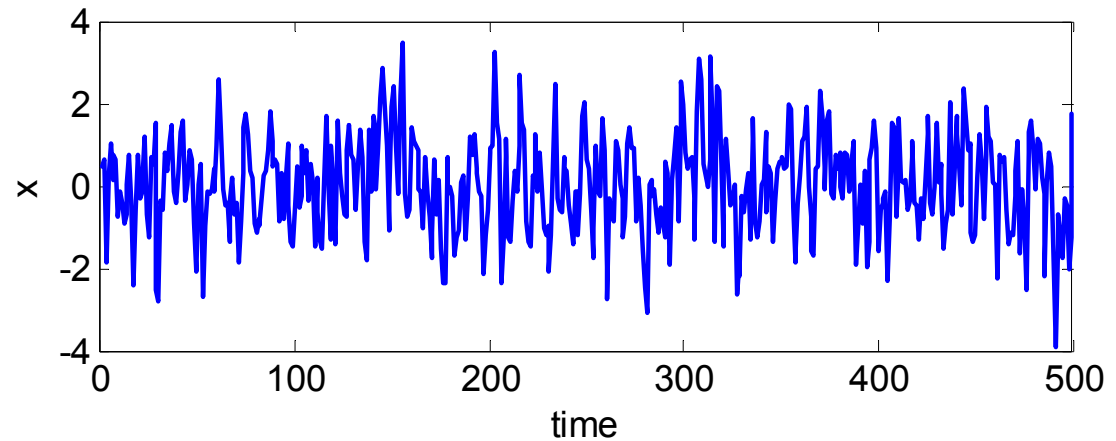
- Mean

$$\mu_c = \mathbb{E}(c_i) = 0$$

since  $\mu_w = 0$

- Variance

$$\begin{aligned} \sigma_c^2 &= \text{Var}(c_i) = \mathbb{E}[(c_i - \bar{c})^2] \\ &= \mathbb{E}(w_i^2 + 2\beta w_i w_{i-1} + \beta^2 w_{i-1}^2) \\ &= (1 + \beta^2) \text{Var}(w_i) \end{aligned}$$



$\Rightarrow \sigma_c^2 = (1 + \beta^2) \sigma_w^2$  So MA (moving average) behavior also *increases* variance of signal.

# ARMA Sequence

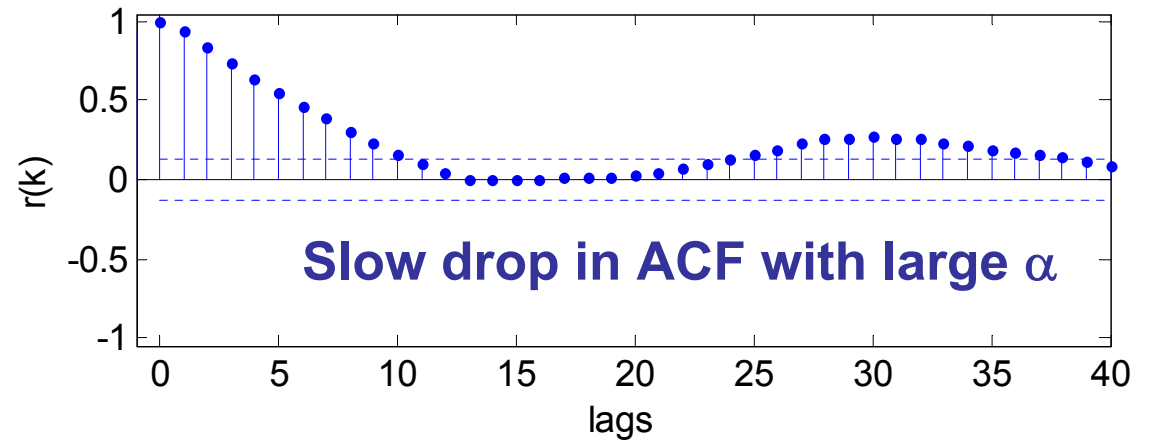
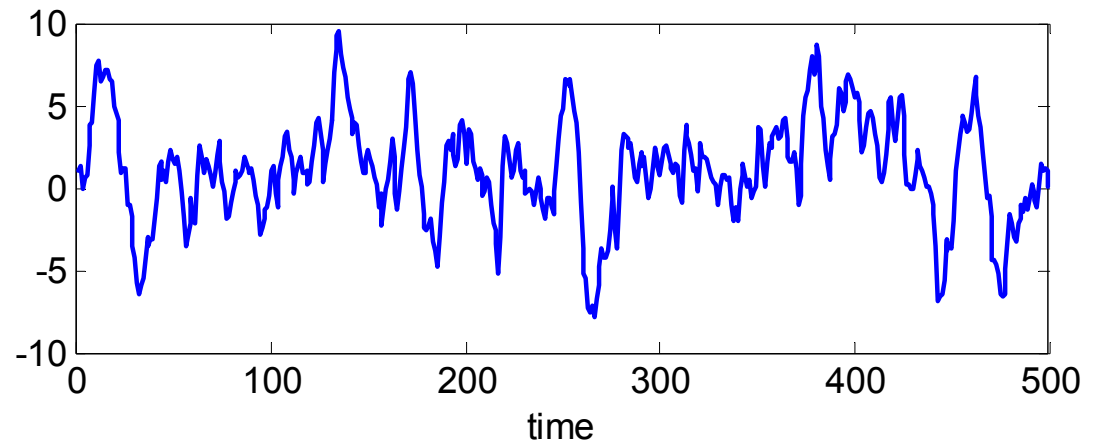
- Generated by:

$$w_i \sim N(0, 1)$$

$$c_i = \alpha \cdot c_{i-1} + w_i + \beta \cdot w_{i-1} \quad \times$$

Shown:  $\alpha = 0.9$ ,  $\beta = 0.5$

- Both AR & MA behavior



# ARIMA Sequence

- Start with ARMA sequence:

$$w_i \sim N(0, 1)$$

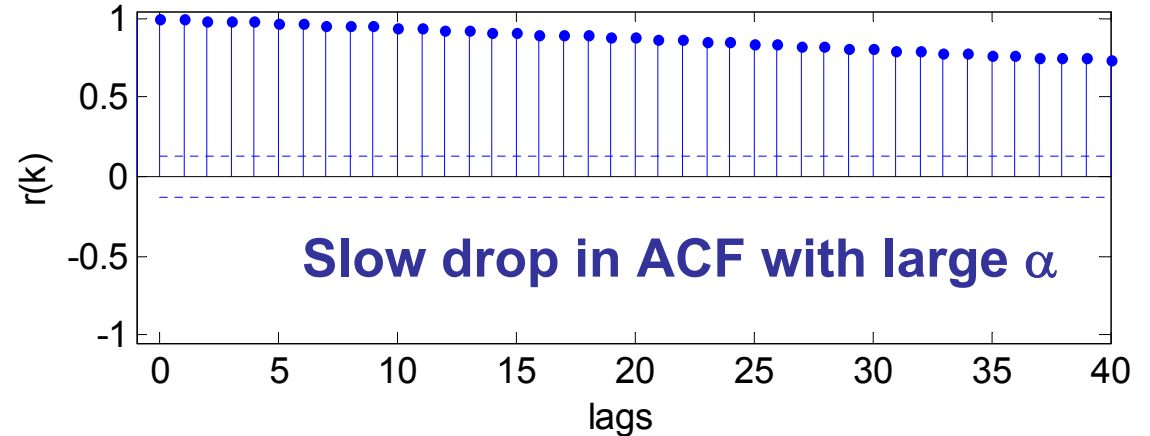
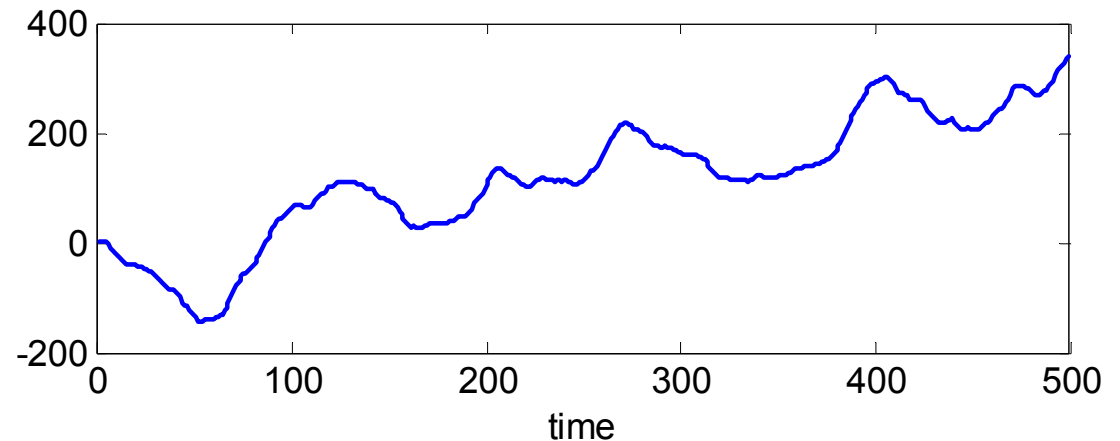
$$c_i = \alpha \cdot c_{i-1} + w_i + \beta \cdot w_{i-1} \quad \times$$

Shown:  $\alpha = 0.9$ ,  $\beta = 0.5$

- Add Integrated (I) behavior

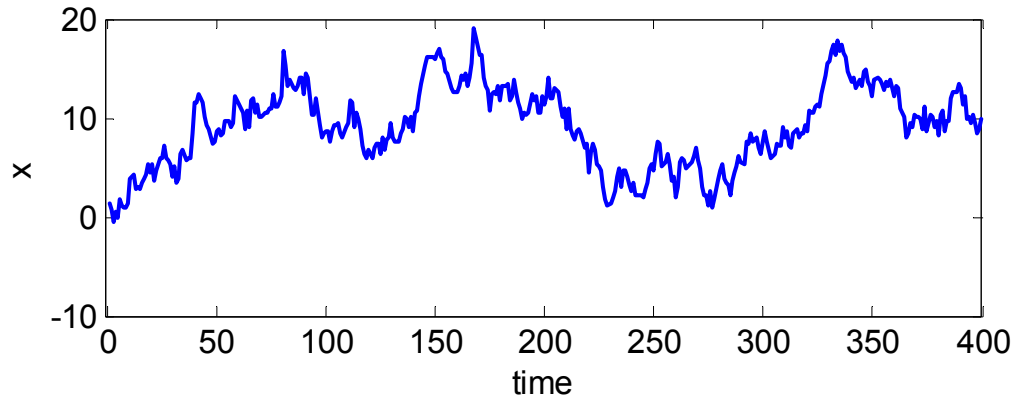
$$x_i = x_{i-1} + c_i$$

random walk (integrative) action

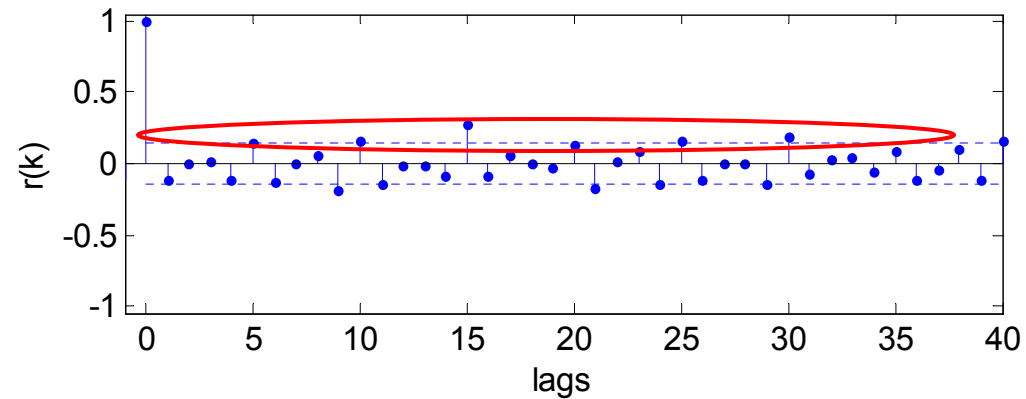
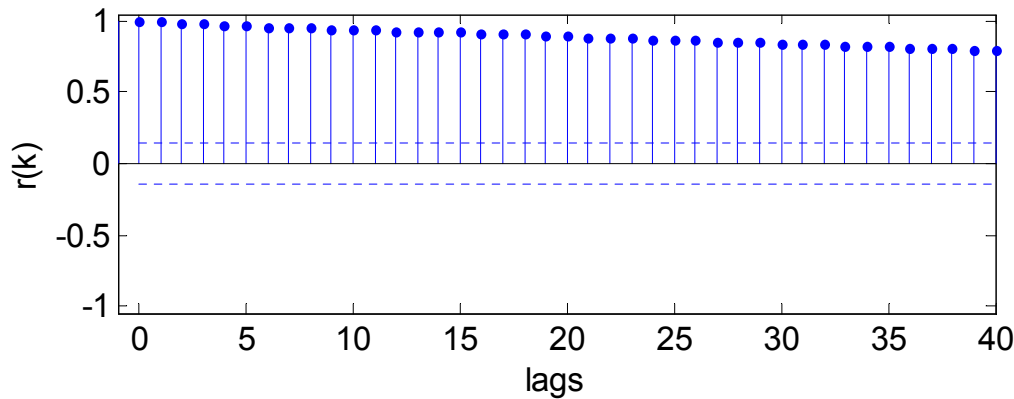
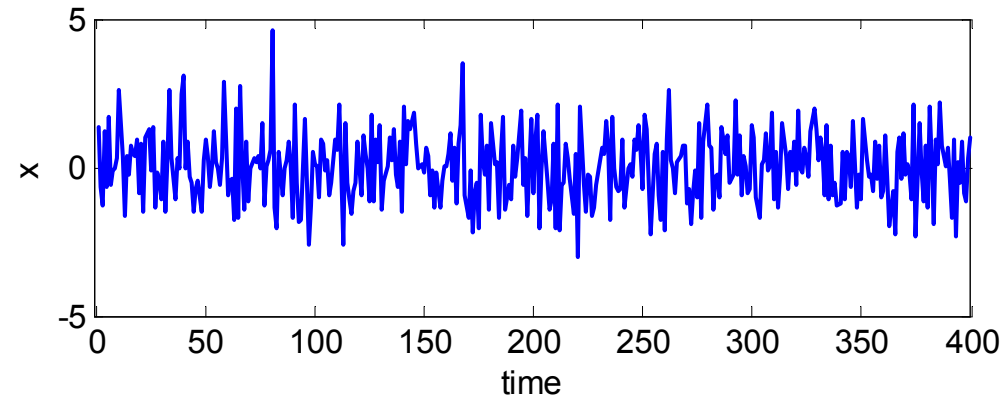


# Periodic Signal with Autoregressive Noise

Original Signal



After Differencing



$$d_i = x_i - x_{i-1}$$

See underlying signal with period = 5

# Agenda

## 1. Regression

- Polynomial regression
- Example (using Excel)

## 2. Time Series Data & Regression

- Autocorrelation – ACF
- Example: white noise sequences
- Example: autoregressive sequences
- Example: moving average
- ARIMA modeling and regression

## 3. Forecasting Examples

# Cross-Correlation: A Leading Indicator

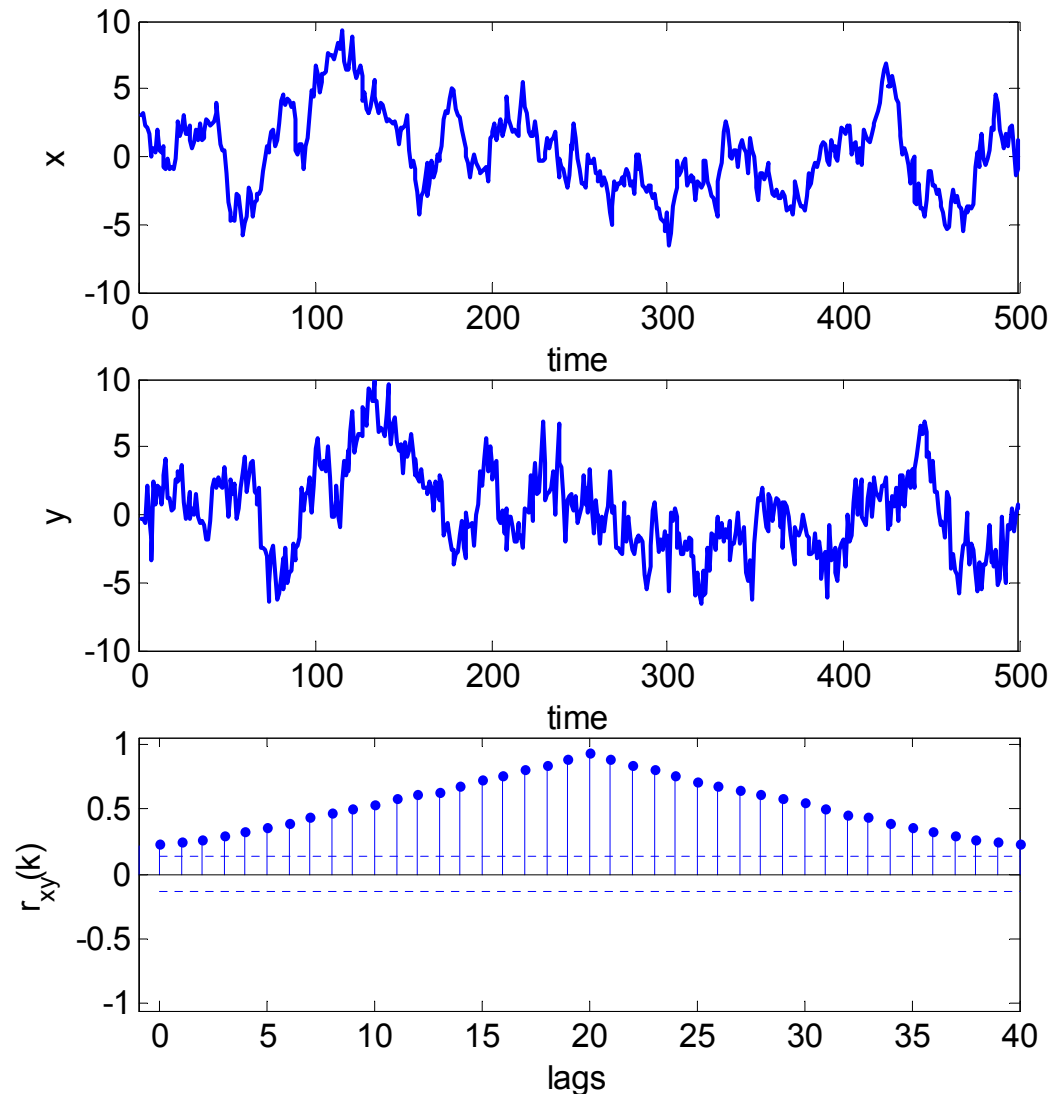
- Now we have two series:
  - An “input” or explanatory variable  $x$
  - An “output” variable  $y$

$$y_i = x_{i-k} + w_i$$

$$w_i \sim N(0, 1)$$

Shown: lag  $k = 20$  and autoregressive  $x$  with  $\alpha = 0.9$

- CCF indicates both AR and lag:



# Regression & Time Series Modeling

- The ACF or CCF are helpful tools in selecting an appropriate model structure
  - Autoregressive terms?
    - $x_i = \alpha x_{i-1}$
  - Lag terms?
    - $y_i = \gamma x_{i-k}$
- One can structure data and perform regressions
  - Estimate *model coefficient* values, significance, and confidence intervals
  - Determine confidence intervals on *output*
  - Check residuals



# Statistical Modeling Summary

## 1. Statistical Fundamentals

- Sampling distributions
- Point and interval estimation
- Hypothesis testing

## 2. Regression

- ANOVA
- Nominal data: modeling of treatment effects (mean differences)
- Continuous data: least square regression  $y = f(\mathbf{x}, \mathbf{b})$

## 3. Time Series Data & Forecasting

- Autoregressive, moving average, and integrative behavior
- Auto- and Cross-correlation functions
- Regression and time-series modeling

$$x_i = f(\mathbf{x}_i, \mathbf{b})$$

$$y_i = f(\mathbf{x}_i, \mathbf{b})$$