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

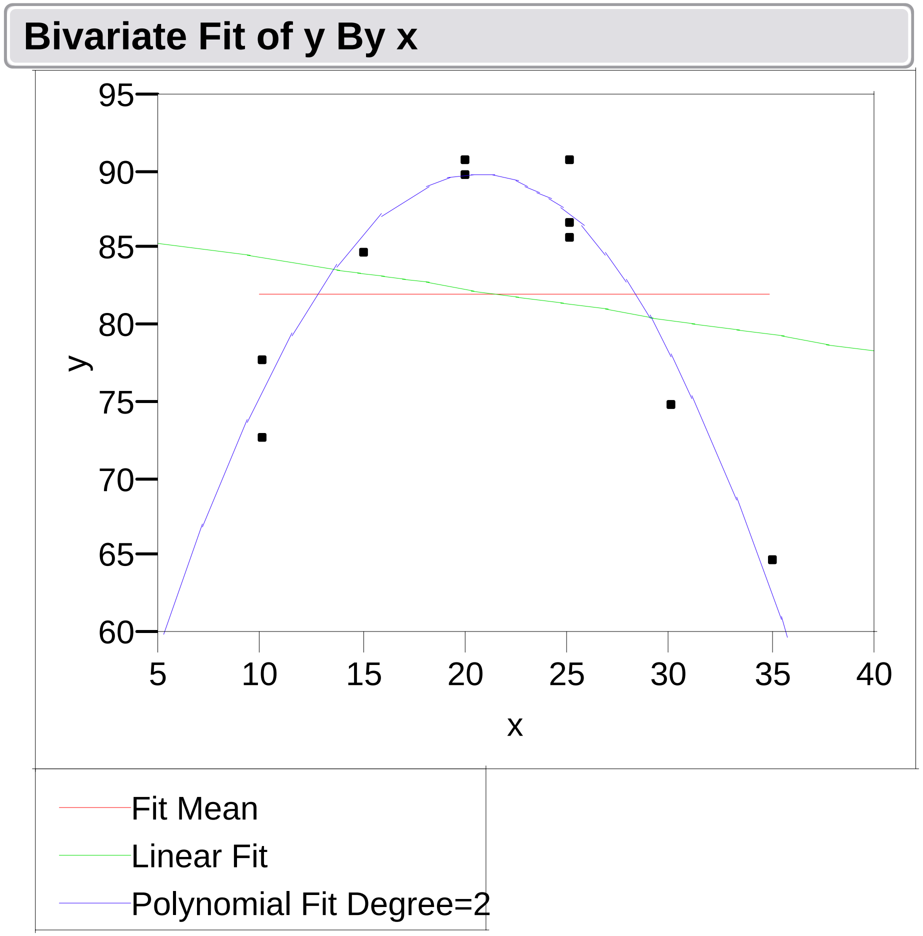
Regression – Review & Extensions

- 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

Growth rate data

observation number	amount of supplement (grams) x	growth rate (coded units) y
1	10	73
2	10	78
3	15	85
4	20	90
5	20	91
6	25	87
7	25	86
8	25	91
9	30	75
10	35	65



- 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

Analysis of variance for growth rate data: straight line model

source		sum of squares		degrees of freedom		mean square	
model	mean	$S_M = 67{,}428.6$	67,404.1	2	1		67,404.1
	extra for linear		24.5		1		24.5
residual	lack of fit	$S_R = 686.4$	$S_L = 659.40$	8	4	85.8	164.85
	pure error		$S_E = 27.0$		4		6.75
total		$S_T = 68{,}115.0$		10			

ratio = 24.42

Growth Rate – Second Order Model

- No evidence of lack of fit
- Quadratic term significant

Analysis of variance for growth rate data: quadratic model

source	sum of squares		degrees of freedom		mean square
model	$S_M = 68{,}071.8$	mean 67,404.1	3	1	67,404.1
		extra for linear 24.5		1	24.5
		extra for quadratic 643.2		1	643.2
residual	$S_R = 43.2$	$S_L = 16.2$	7	3	5.40 ratio = 0.80
		$S_E = 27.0$		4	6.75
total	$S_T = 68{,}115.0$		10		

Polynomial Regression In Excel

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

x	x^2	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

Regression Statistics	
Multiple R	0.968
R Square	0.936
Adjusted R Square	0.918
Standard Error	2.541
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	665.706	332.853	51.555	6.48E-05
Residual	7	45.194	6.456		
Total	9	710.9			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
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^2	-0.128	0.013	-9.966	2.2E-05	-0.158	-0.097

Polynomial Regression

- Generated using JMP

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

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)

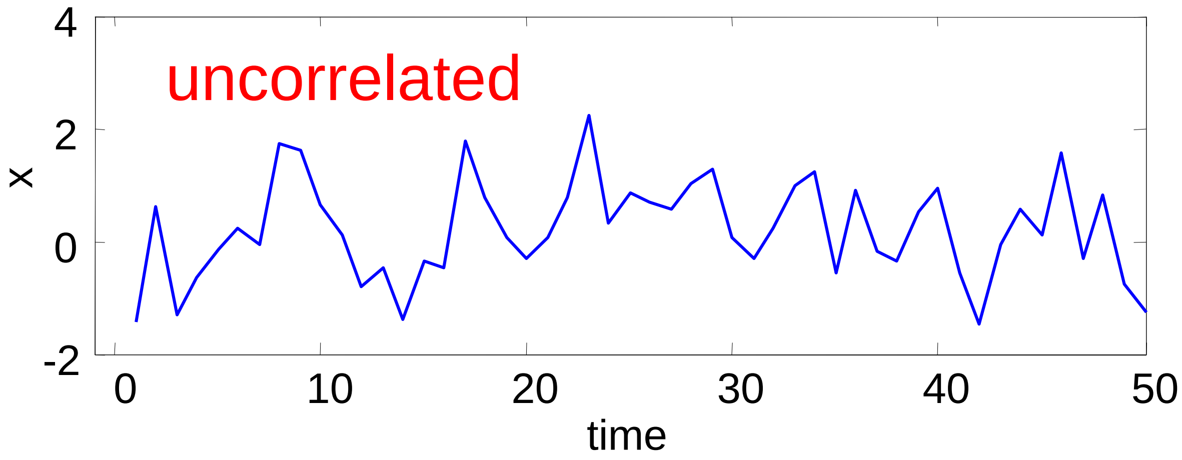
2. **Time Series Data & Time Series Regression**
   - **Autocorrelation – ACF**
   - **Example: white noise sequences**
   - **Example: autoregressive sequences**
   - **Example: moving average**
   - **ARIMA modeling and regression**

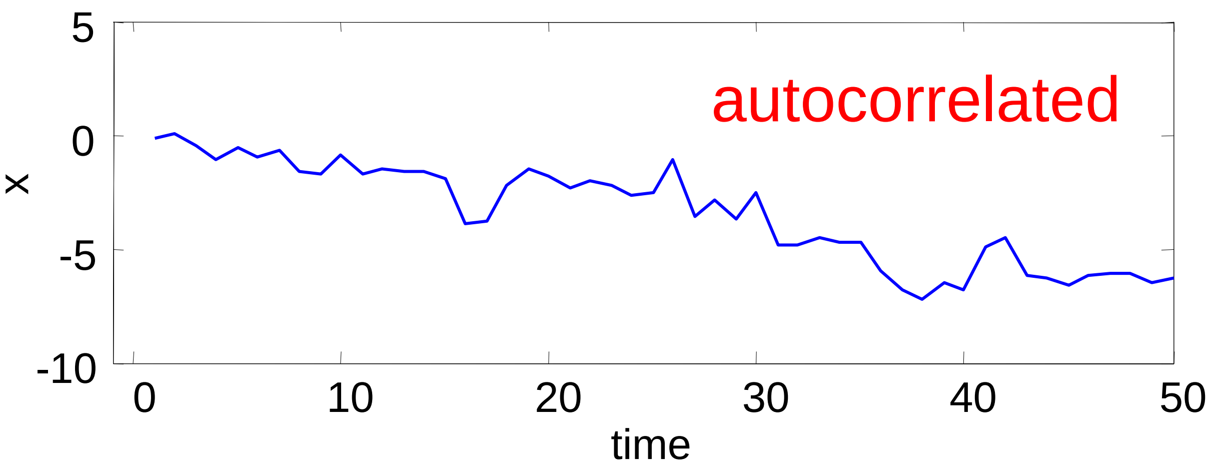
3. Forecasting Examples

Time Series – Time as an Implicit Parameter

- Data is often collected with a ***time-order***

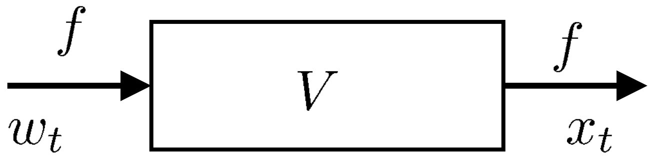
- An underlying dynamic process (e.g. due to physics of a manufacturing process) may create ***autocorrelation*** in the data





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*.



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

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

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

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

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

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

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*

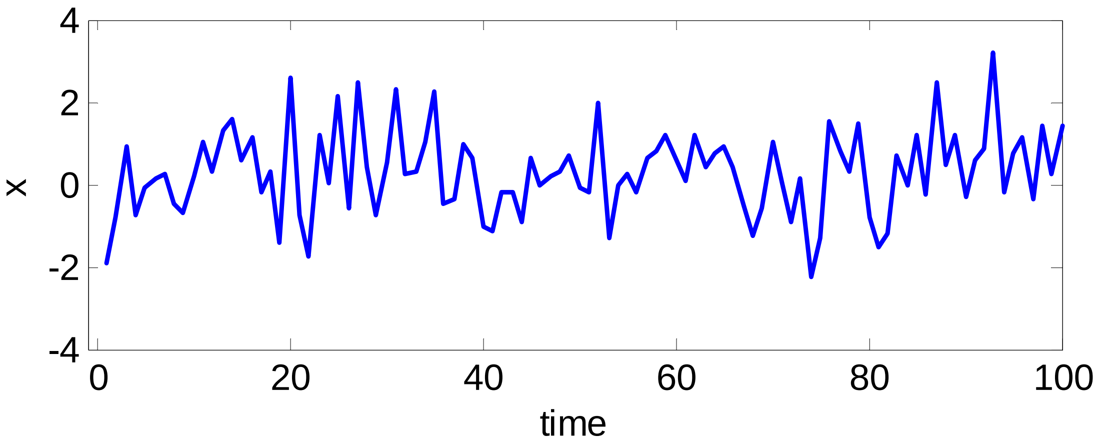
$$x_i \sim N(0,1)$$

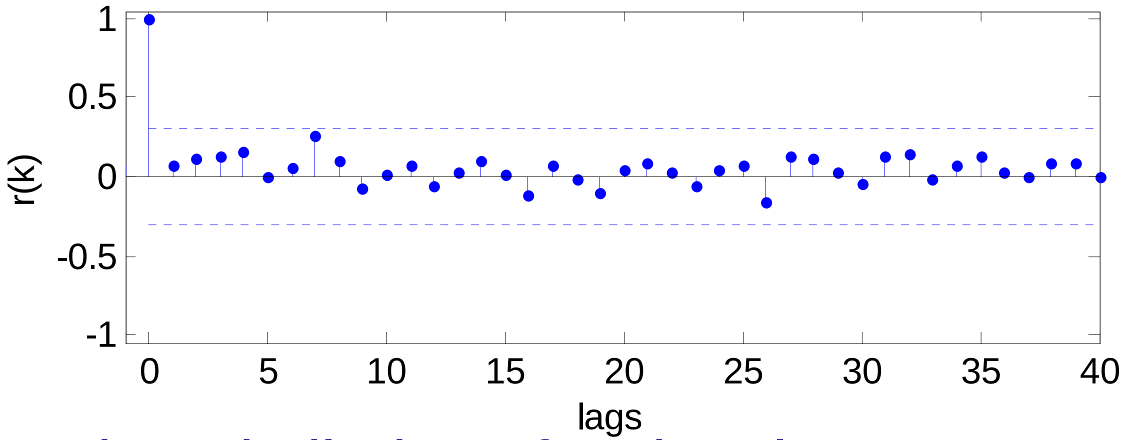
- CCF: cross-correlation function

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

- ACF: auto-correlation function

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

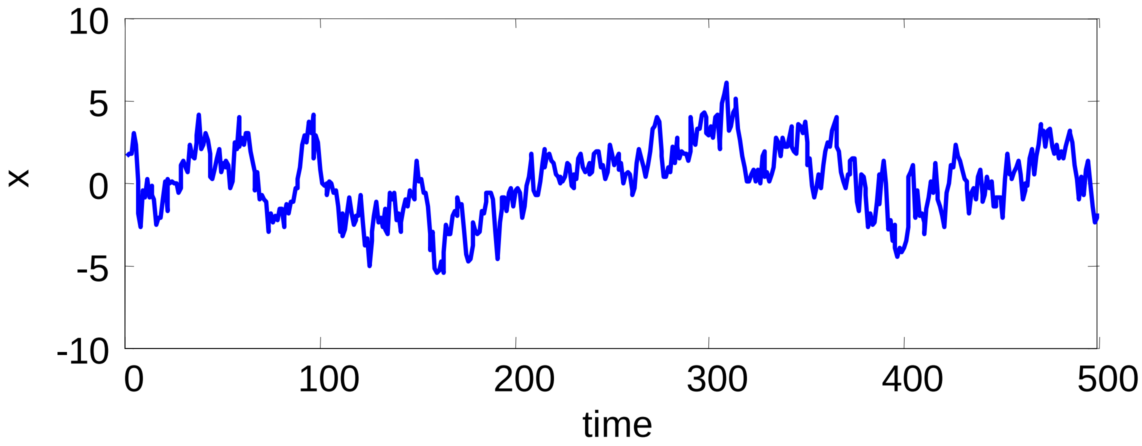


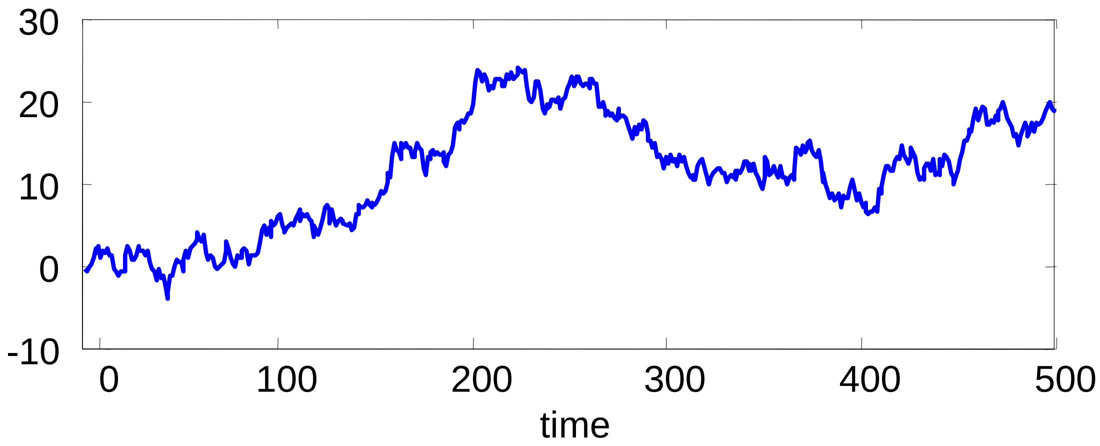


$\Rightarrow$ ACF shows the “similarity” of a signal to a lagged version of same signal

Stationary vs. Non-Stationary

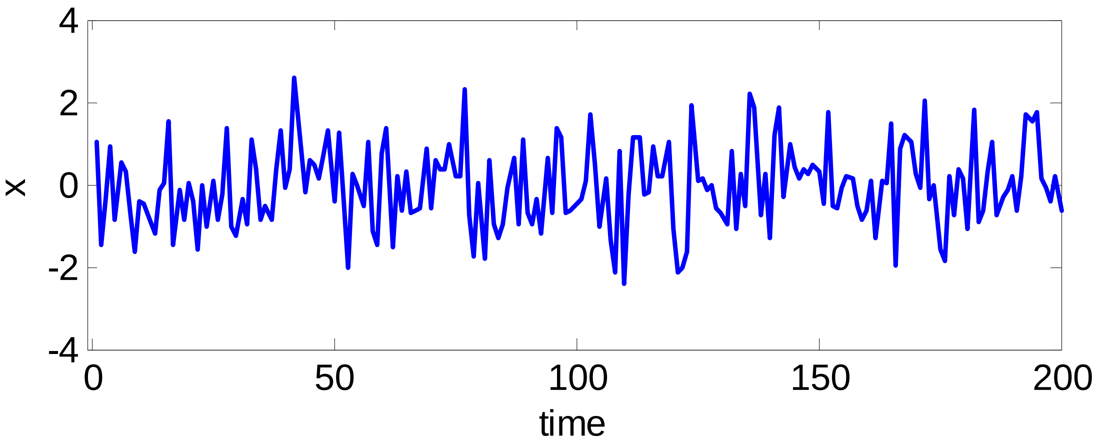
Stationary series:
Process has a **fixed** mean

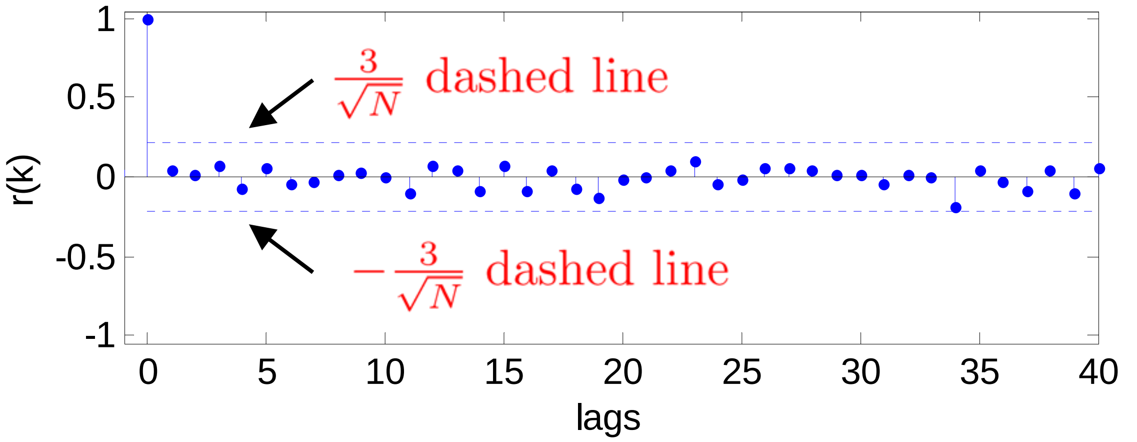




White Noise – An Uncorrelated Series

- Data drawn from IID gaussian
  $$w_i \sim N(0,1)$$
- ACF: We also plot the 3σ 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:

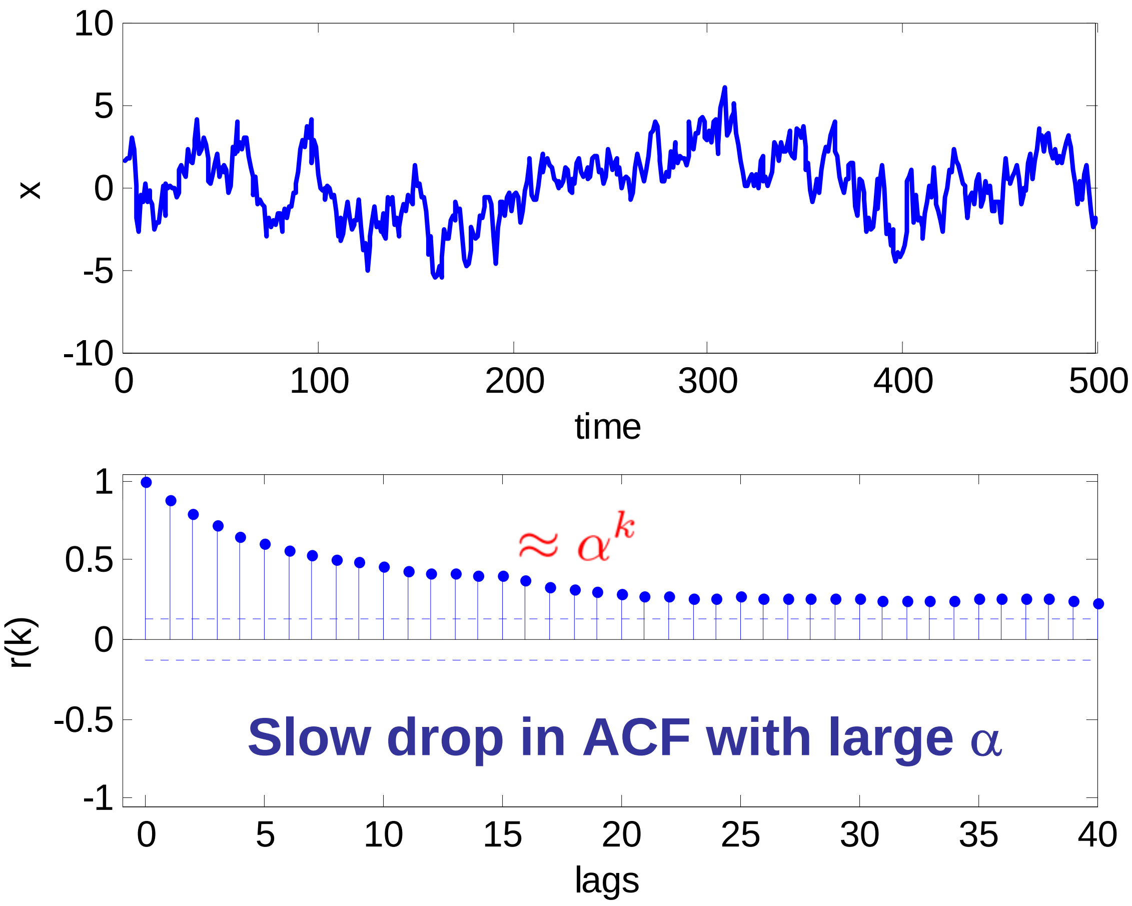
$$w_i \sim N(0,1)$$
$$c_i = \alpha \cdot c_{i-1} + w_i$$
$$\text{Shown: } \alpha = 0.9$$

- Mean

$$\mu_c = \mathrm{E}(c_i) = 0$$
$$\text{since } \mu_w = 0$$

- Variance

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



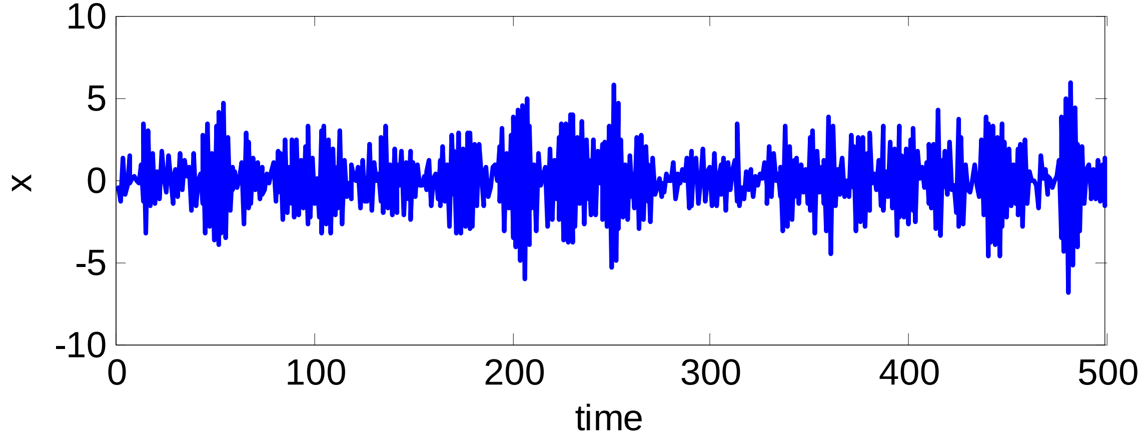
$$\Rightarrow \quad \sigma_c^2 = \frac{\sigma_w^2}{1-\alpha^2}$$

So AR (autoregressive) behavior *increases* variance of signal.

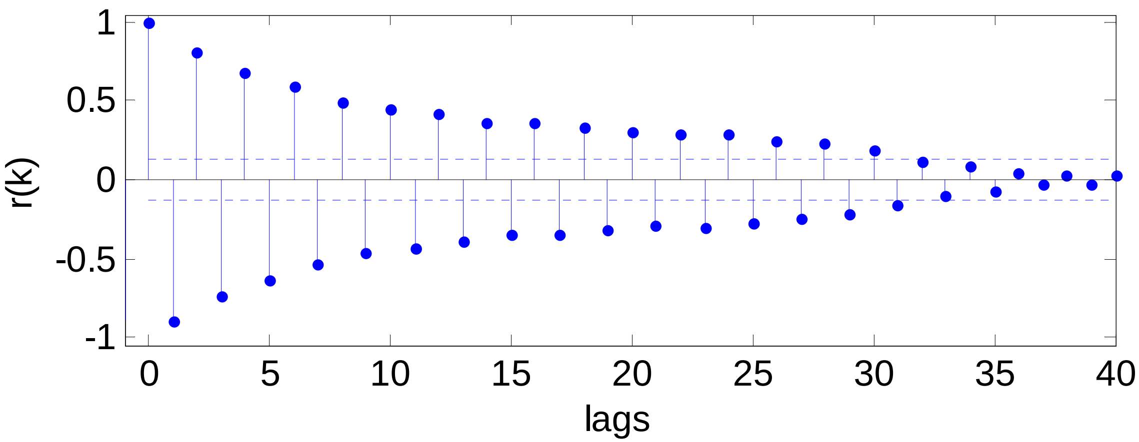
Another Autoregressive Series

- Generated by:

$$w_i \sim N(0,1)$$
$$c_i = \alpha \cdot c_{i-1} + w_i$$
$$\text{Shown: } \alpha = -0.9$$



- High **negative** autocorrelation:



Slow drop in ACF with large α

But now ACF alternates in sign

Random Walk Disturbances

- Generated by:

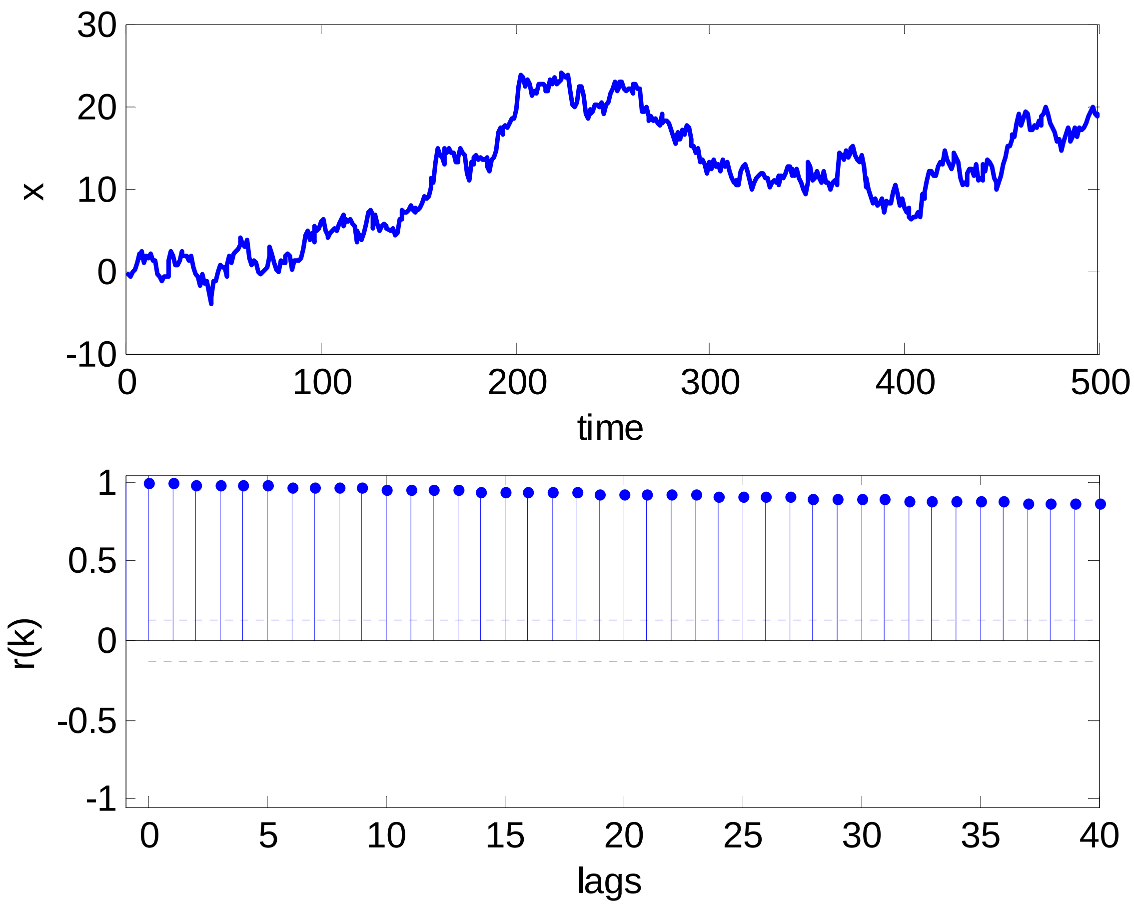
$$w_i \sim N(0,1)$$
$$c_i = 1 \cdot c_{i-1} + w_i$$
$$\text{AR with } \alpha = 1$$

- Mean

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

- Variance

Variance increases as sequence gets longer



Very **slow drop in ACF for α = 1**

Moving Average Sequence

- Generated by:

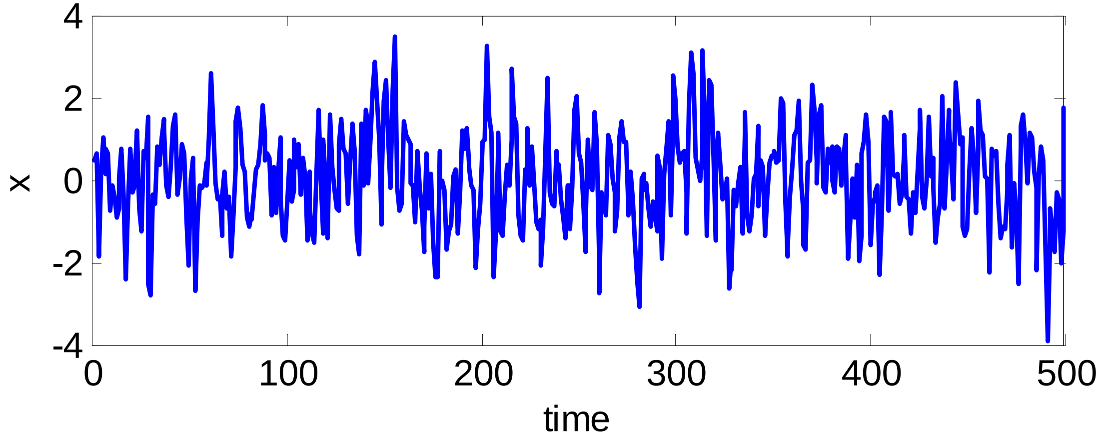
$$w_i \sim N(0,1)$$
$$c_i = w_i + \beta \cdot w_{i-1}$$
$$\text{Shown: } \beta = 0.5$$

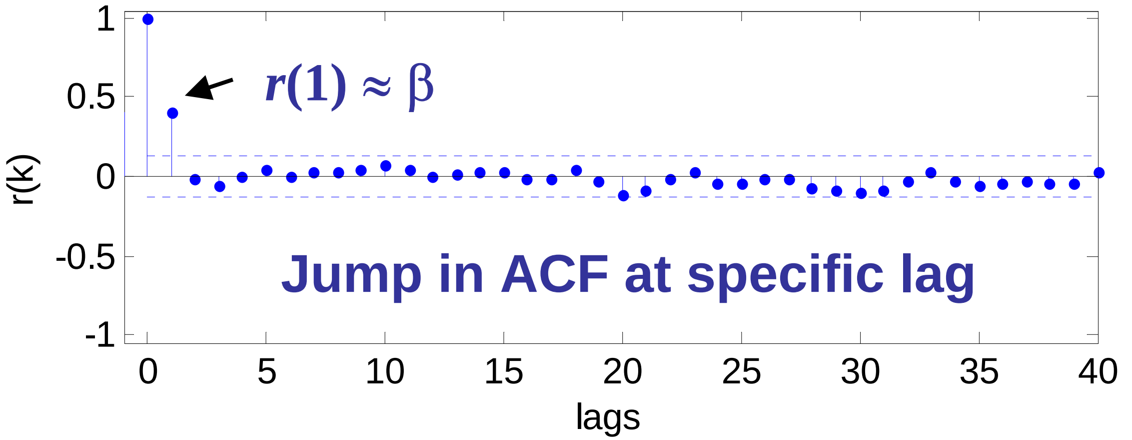
- Mean

$$\mu_c = \mathrm{E}(c_i) = 0$$
$$\text{since } \mu_w = 0$$

- Variance

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





$$\Rightarrow \quad \sigma_c^2 = (1+\beta^2)\sigma_w^2$$

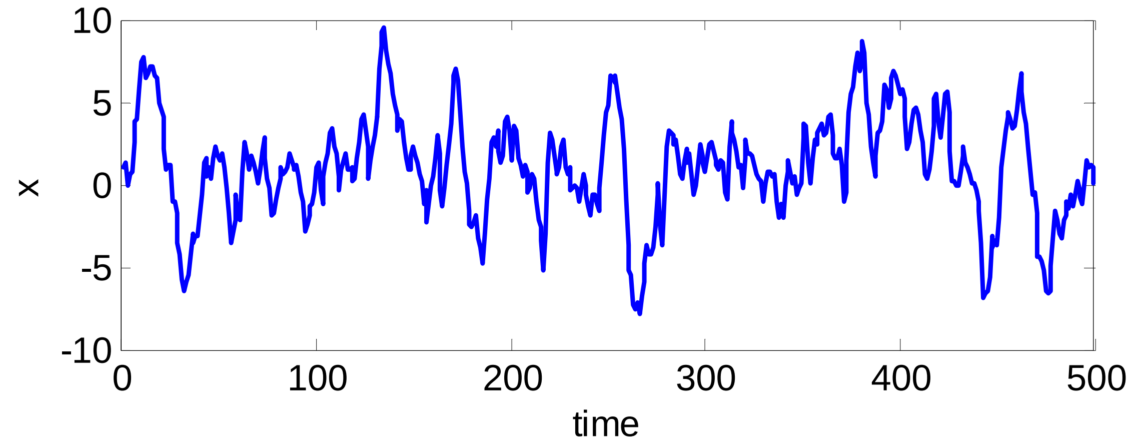
So MA (moving average) behavior also *increases* variance of signal.

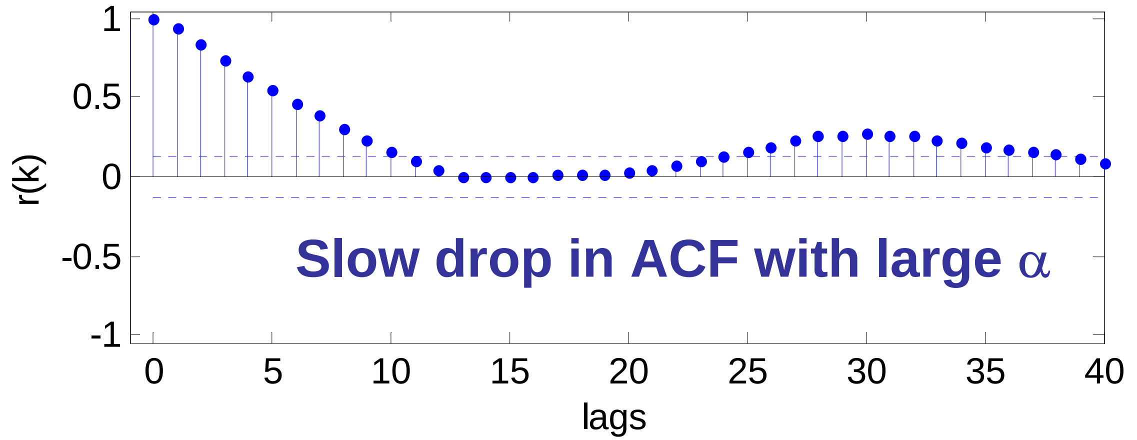
ARMA Sequence

- Generated by:

$$w_i \sim N(0,1)$$
$$c_i = \alpha \cdot c_{i-1} + w_i + \beta \cdot w_{i-1}$$
$$\text{Shown: } \alpha = 0.9, \quad \beta = 0.5$$

- Both AR & MA behavior





ARIMA Sequence

- Start with ARMA sequence:

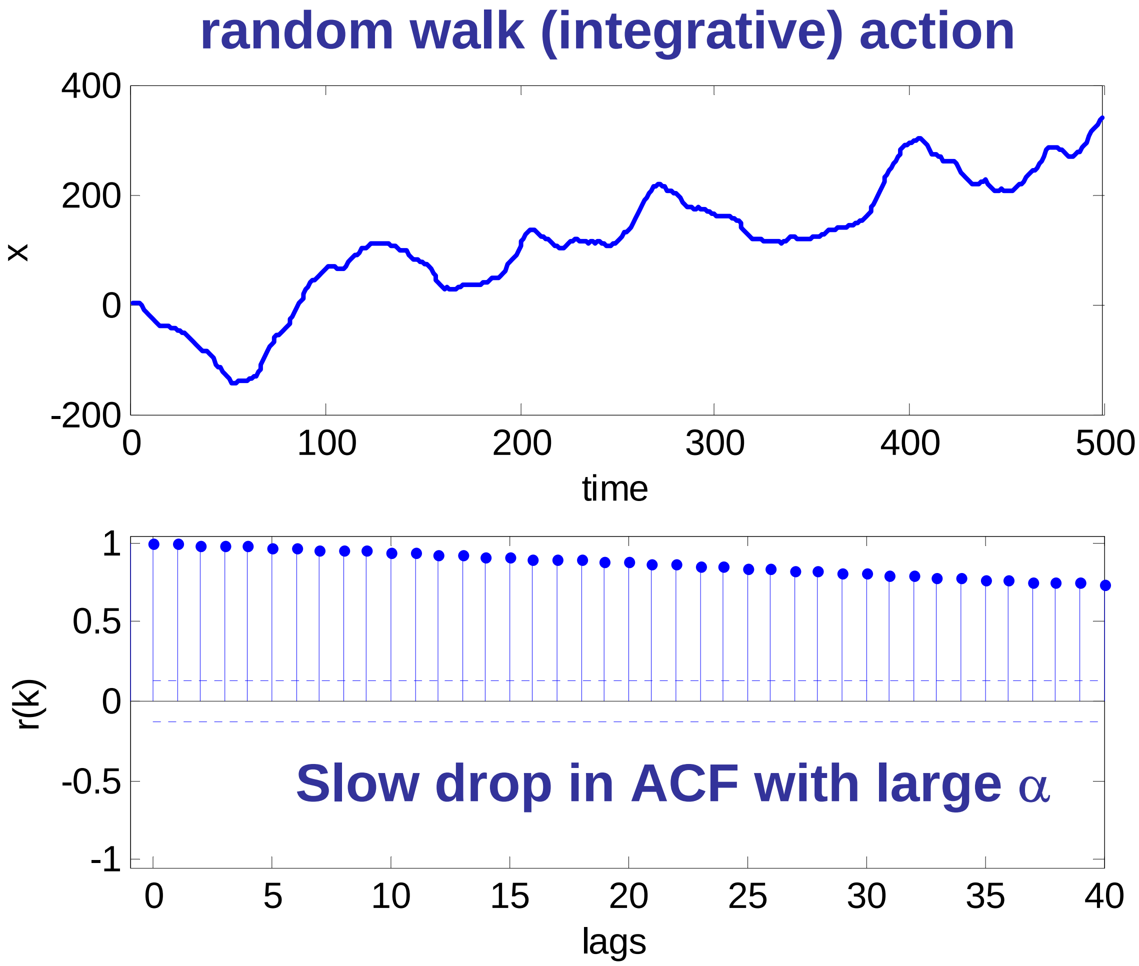
$$w_i \sim N(0,1)$$

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

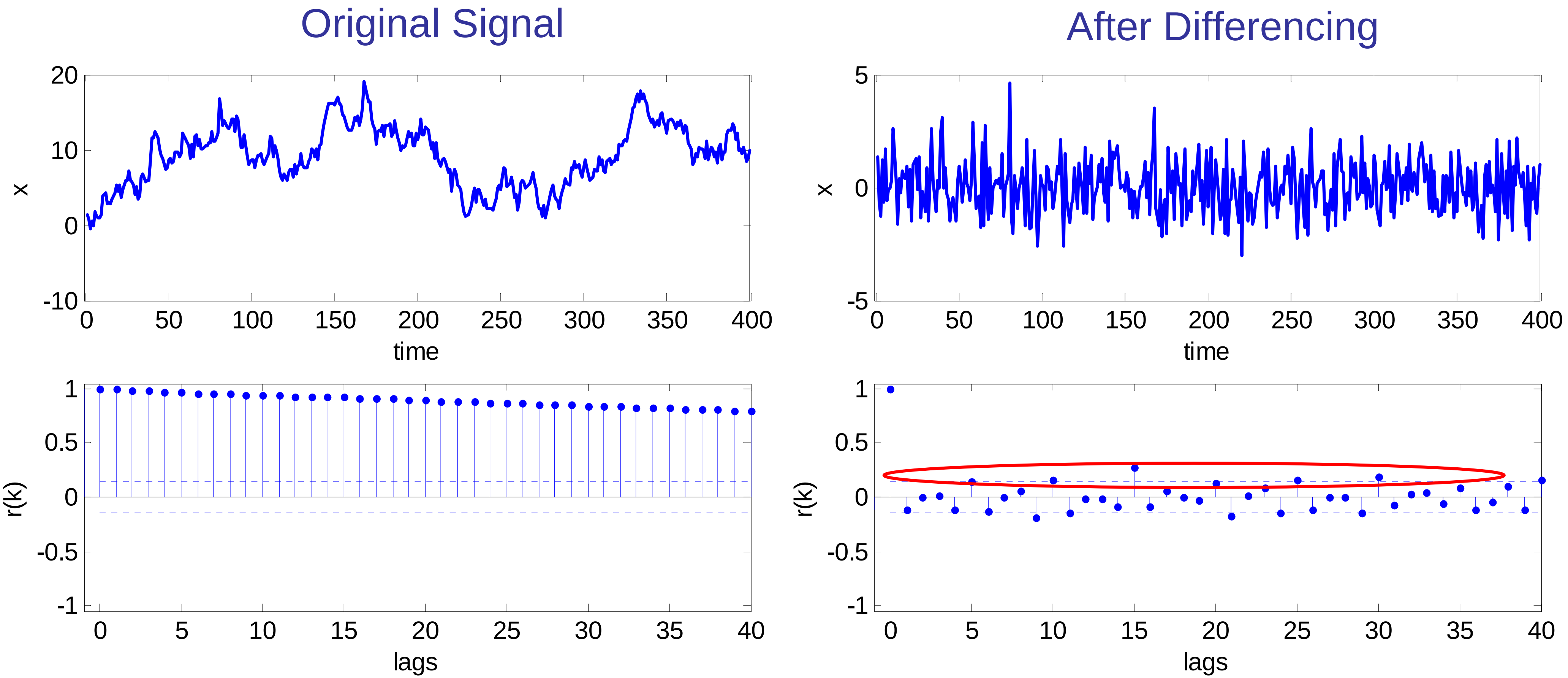
$$\text{Shown: } \alpha = 0.9, \quad \beta = 0.5$$

- Add Integrated (I) behavior

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



Periodic Signal with Autoregressive Noise



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

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

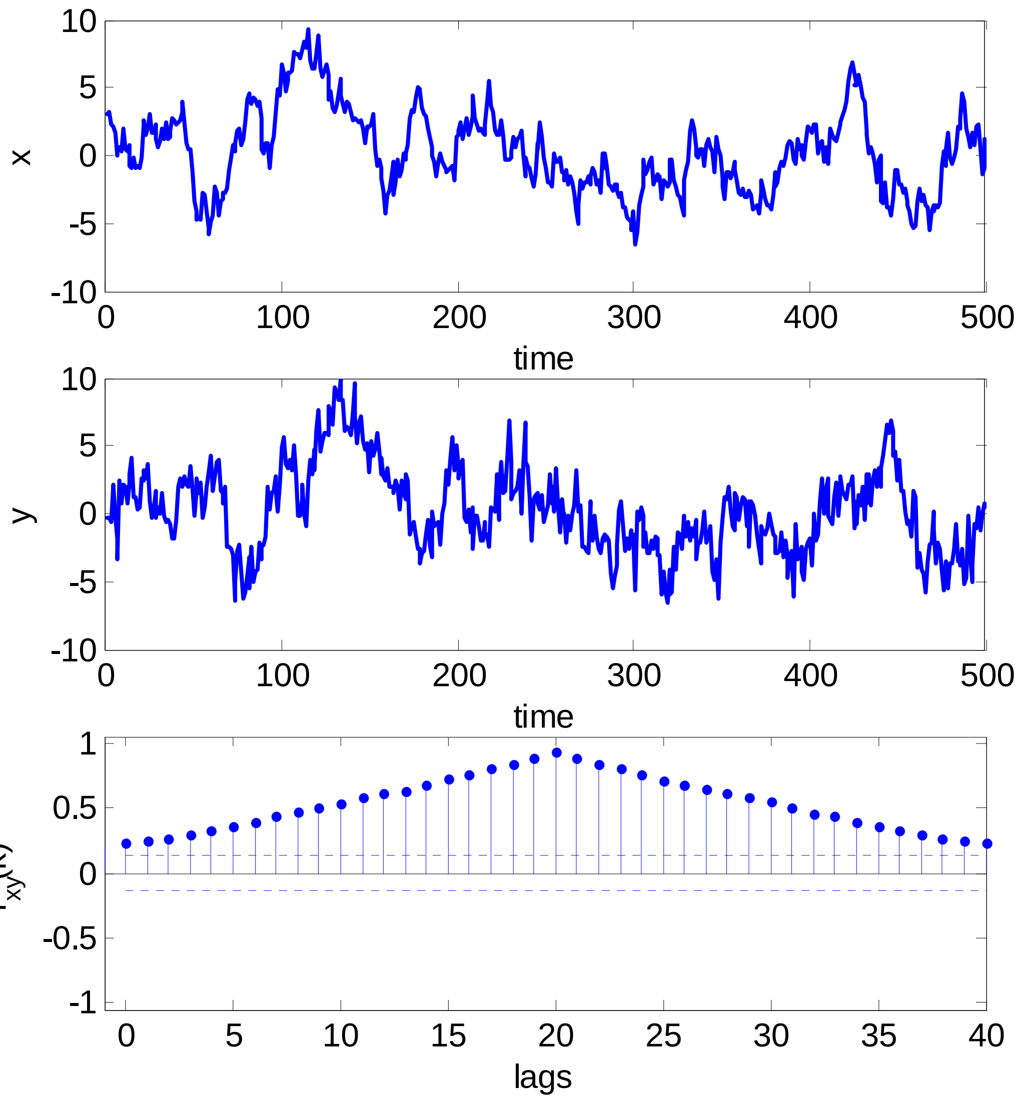
- Now we have two series:
  - An "input" or explanatory variable x
  - An "output" variable y

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

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

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

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