MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

IDSS

Applied Data Science Program

TIME SERIES

Munther A. Dahleh

Learning Time-Series

- Time Series are everywhere
  - Finance, weather, control systems

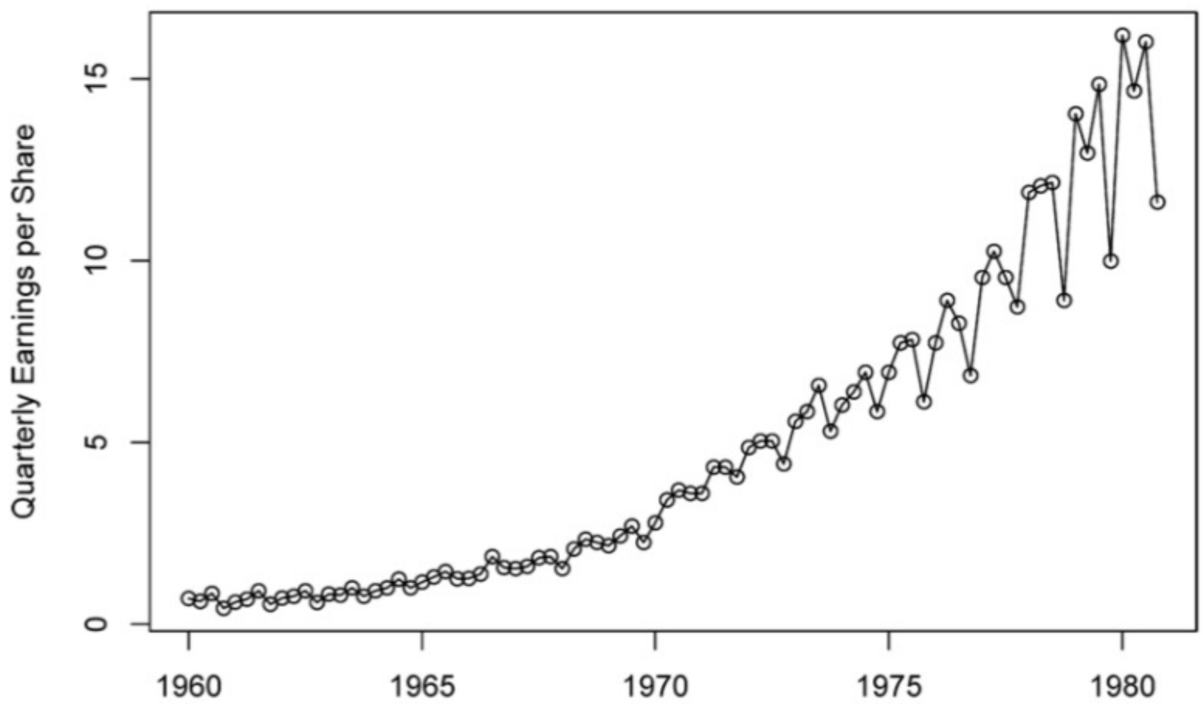
- Questions:
  - Modeling
  - Forecasting
  - Decisions and policy

Learning Time-Series: Outline

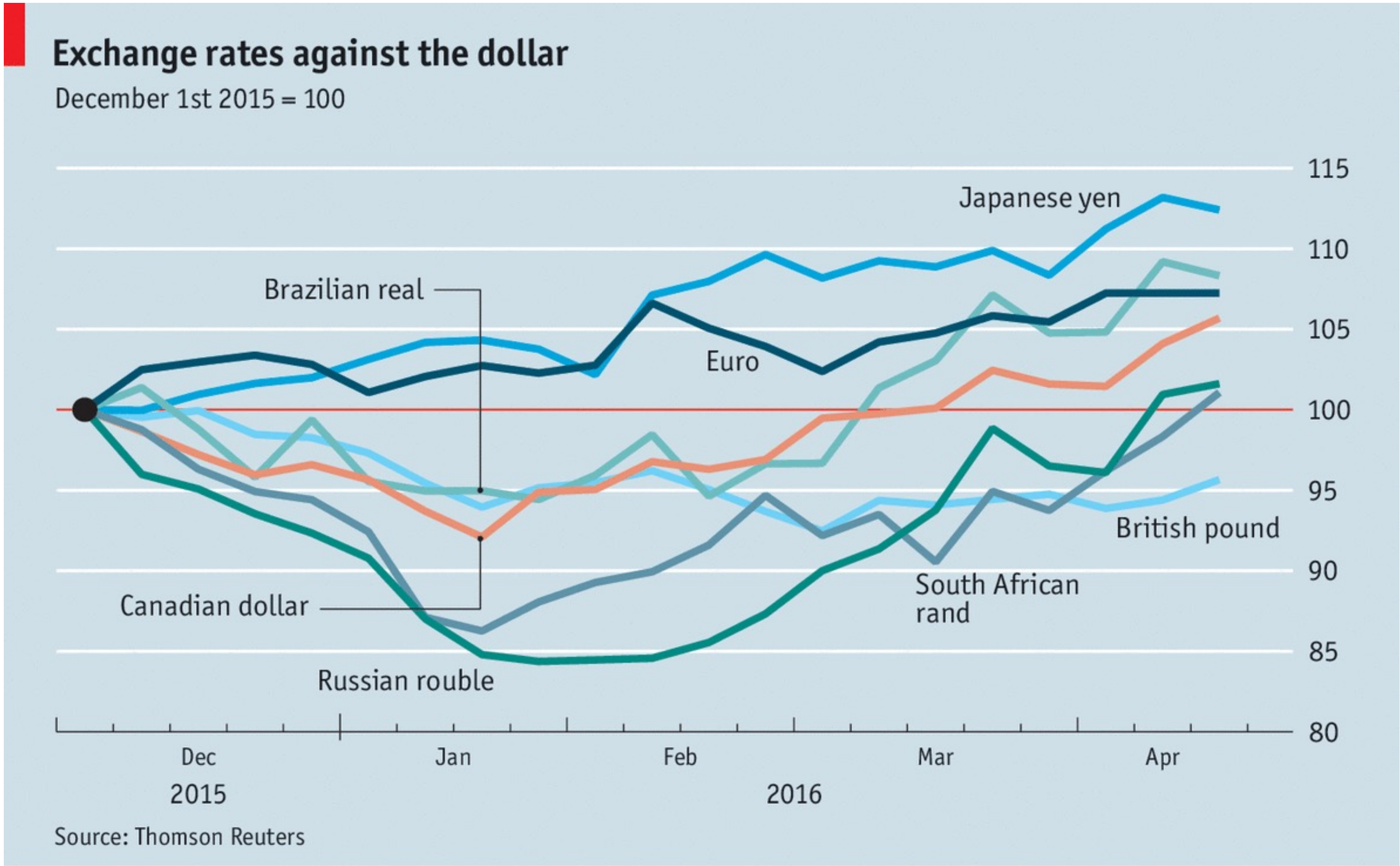
- Part I: Introduction to Time Series
  - Stationarity
  - Trends
- Part II: Models of Time Series
- Part 3: Learning Time Series

Part I: Introduction

Johnson & Johnson quarterly earnings per share

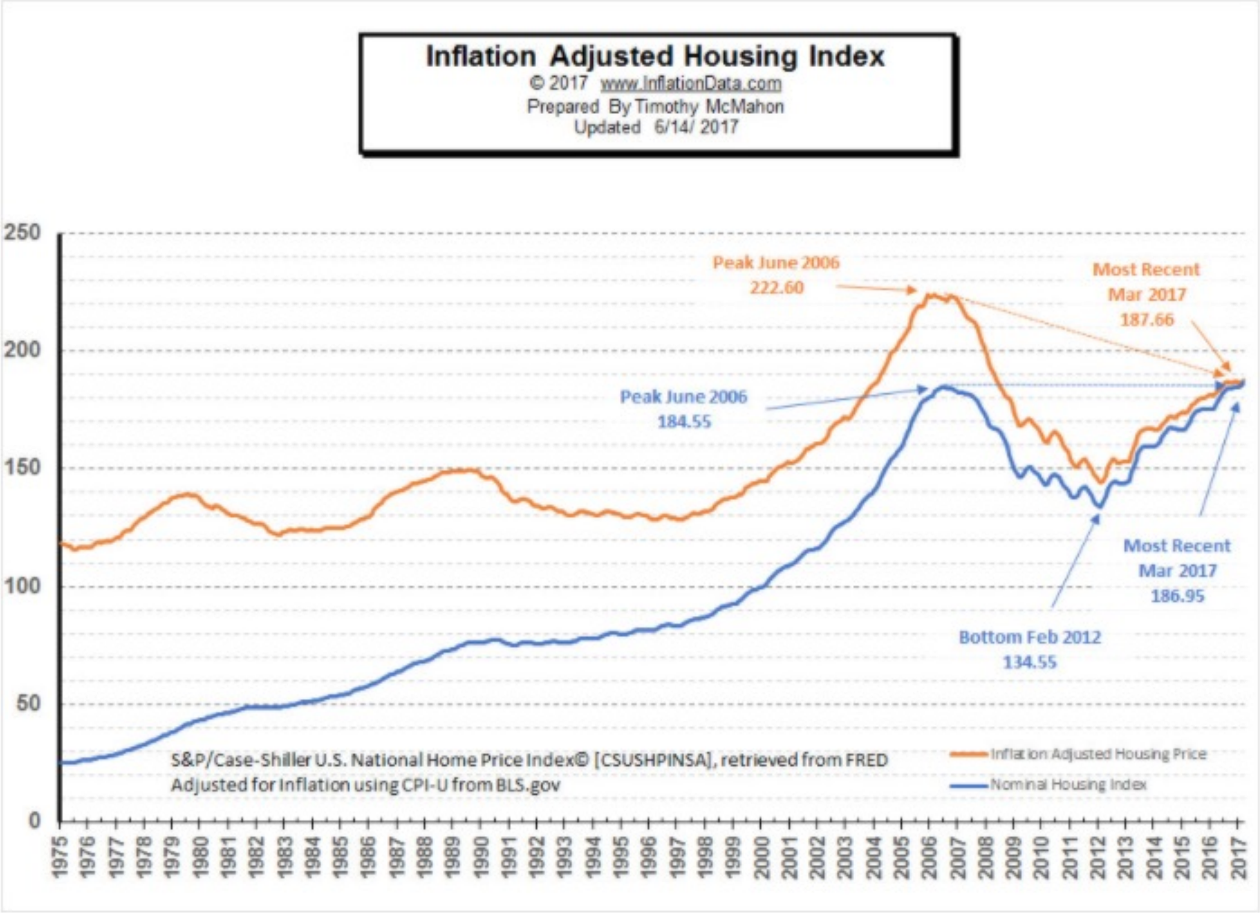


Exchange Rates Against the Dollar

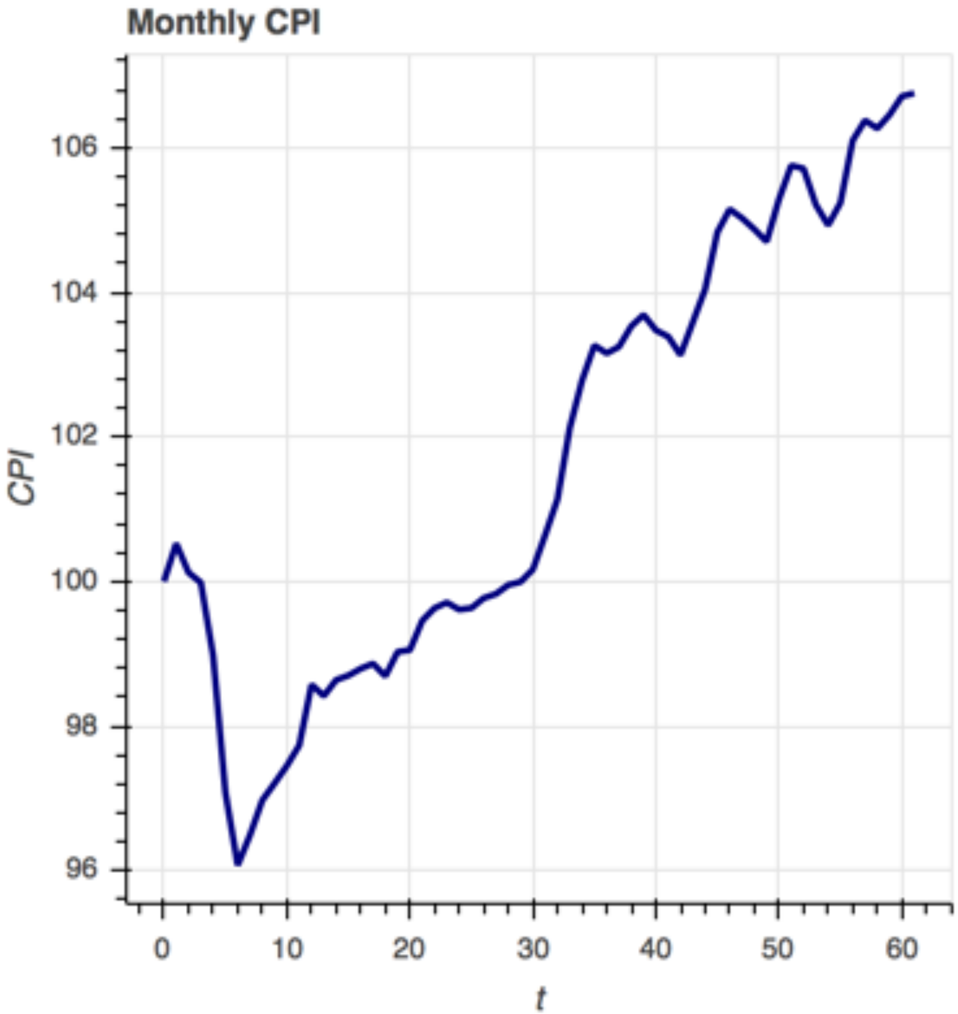


Economist.com

House Prices



Consumer Price Index (billion Prices)



…What is new here?

- What you measure today depends on yesterday
  - Could be simple or complicated dependence
  - Memory can be high
  - Memory is typically unknow
- The variation in the data can be due to a time-varying average: trends
  - Linear, quadratic (deterministic)
  - Periodic (seasonal)
- Transformation of the data may help

Stochastic process

- A sequence of Random Variables:

$$X_0, X_1, X_2, \ldots$$

- Generally described in terms of joint probability distribution over any sequence.

- Stationary stochastic process
  - Strict-sense stationary
  - Wide-sense

Strong-Sense Stationary

- Joint probability distribution is the same if you shift the data

- Distribution of $X_t, X_{t+1}, X_{t+2}, \ldots$
  is the same as $X_{t+h}, X_{t+h+1}, X_{t+h+2}, \ldots \qquad \forall h$

- Generally hard to verify!

- Settle with a weaker notion.

Autocovariance

- Mean of process

$$\mu_t = \mathbf{E}(X_t)$$

- Autocovariance

$$R_X(t_1, t_2) = \mathbf{E}\left((X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})\right)$$

Wide-Sense Stationary

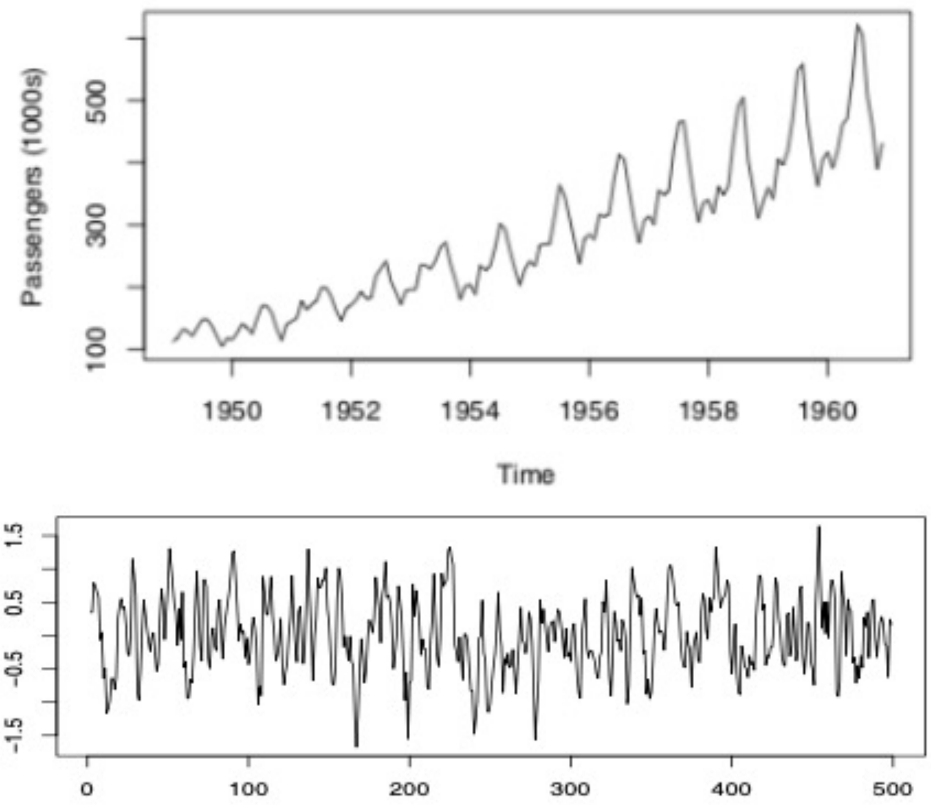
- Mean of process is constant over time

$$\mathbf{E}(X_t) = \mu$$

- Autocovariance is a function of the time difference

$$R_X(t_1 - t_2) = R_X(t_2 - t_1)$$

Testing Stationarity



Testing Stationarity

- Compute Sample mean for each λ

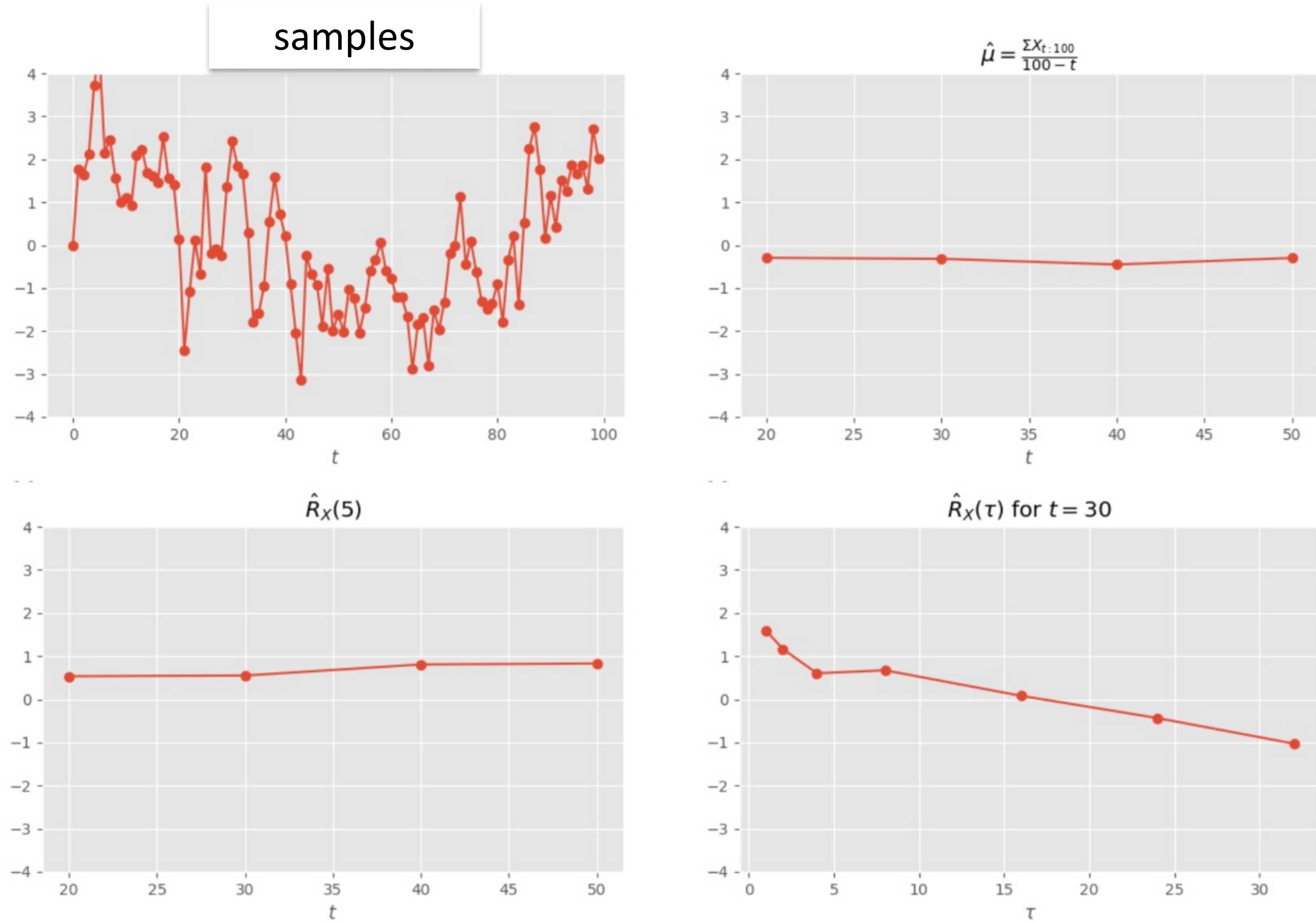
$$\hat{\mu} = \frac{1}{N - \lambda} \sum_{i=\lambda}^{N-1} X_i$$

- Sample Autocovariance for each λ

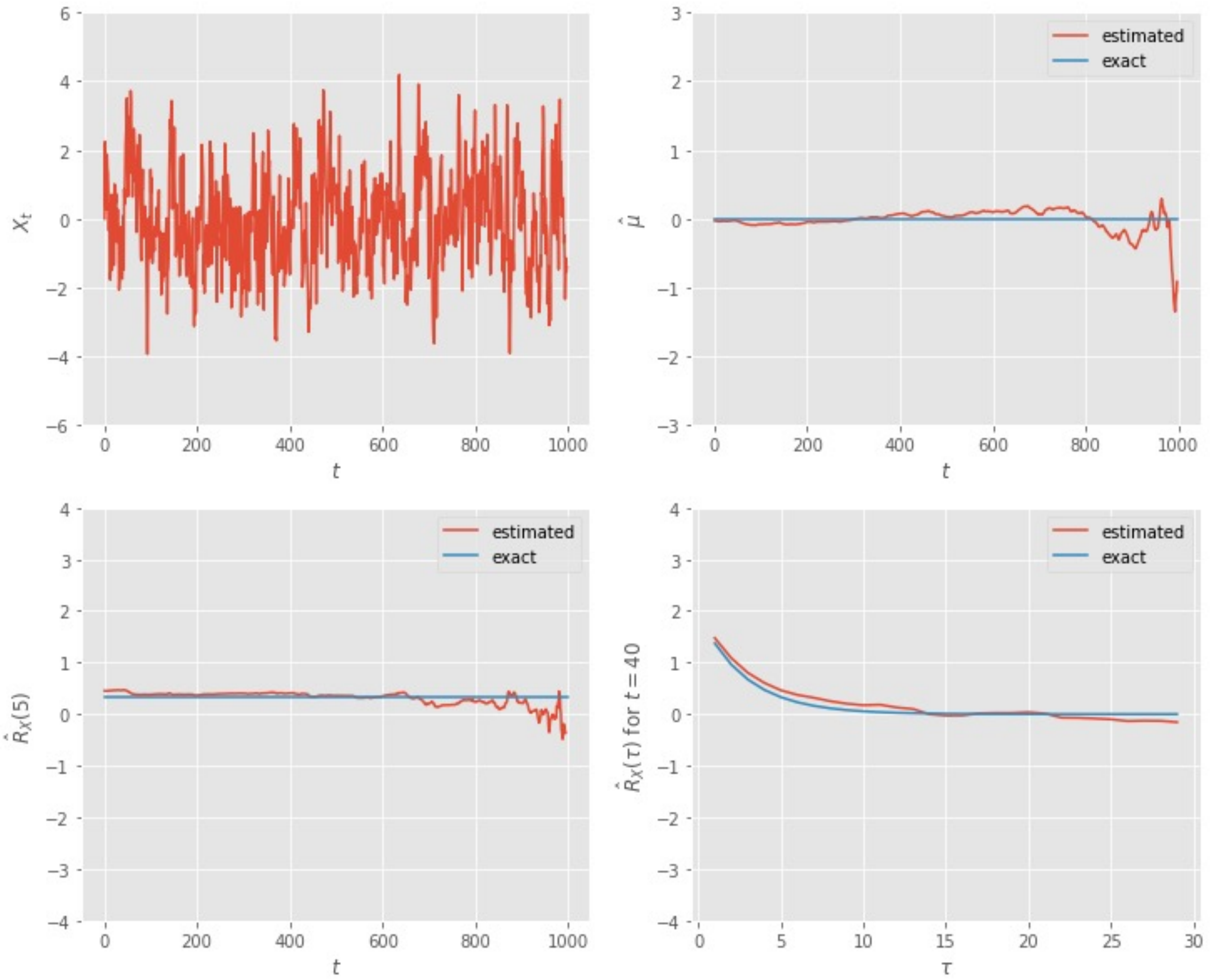
$$\hat{R}_X(\tau) = \frac{1}{N - \lambda} \sum_{i=\lambda}^{N-1} (X_i - \hat{\mu})(X_{i+\tau} - \hat{\mu})$$

- Should not see large variations over λ

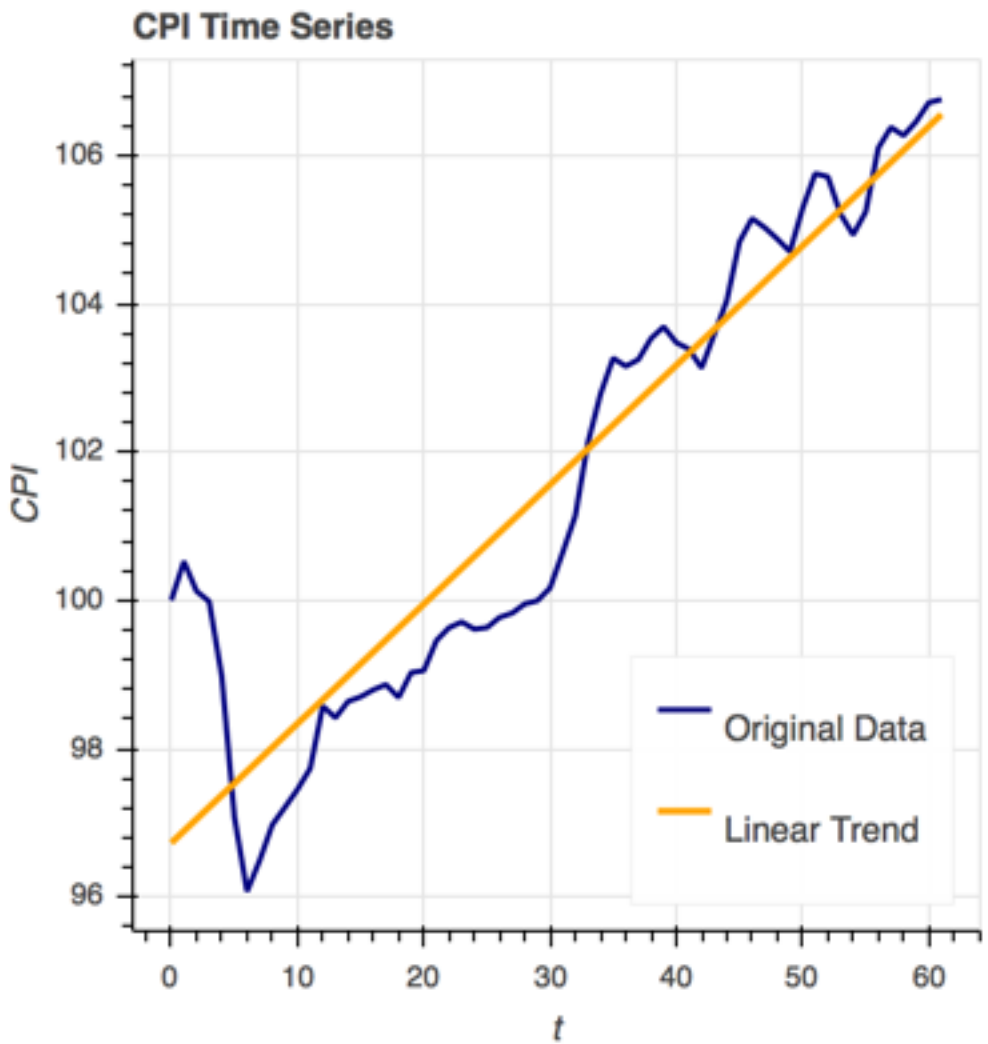
Demonstrate: 100 samples



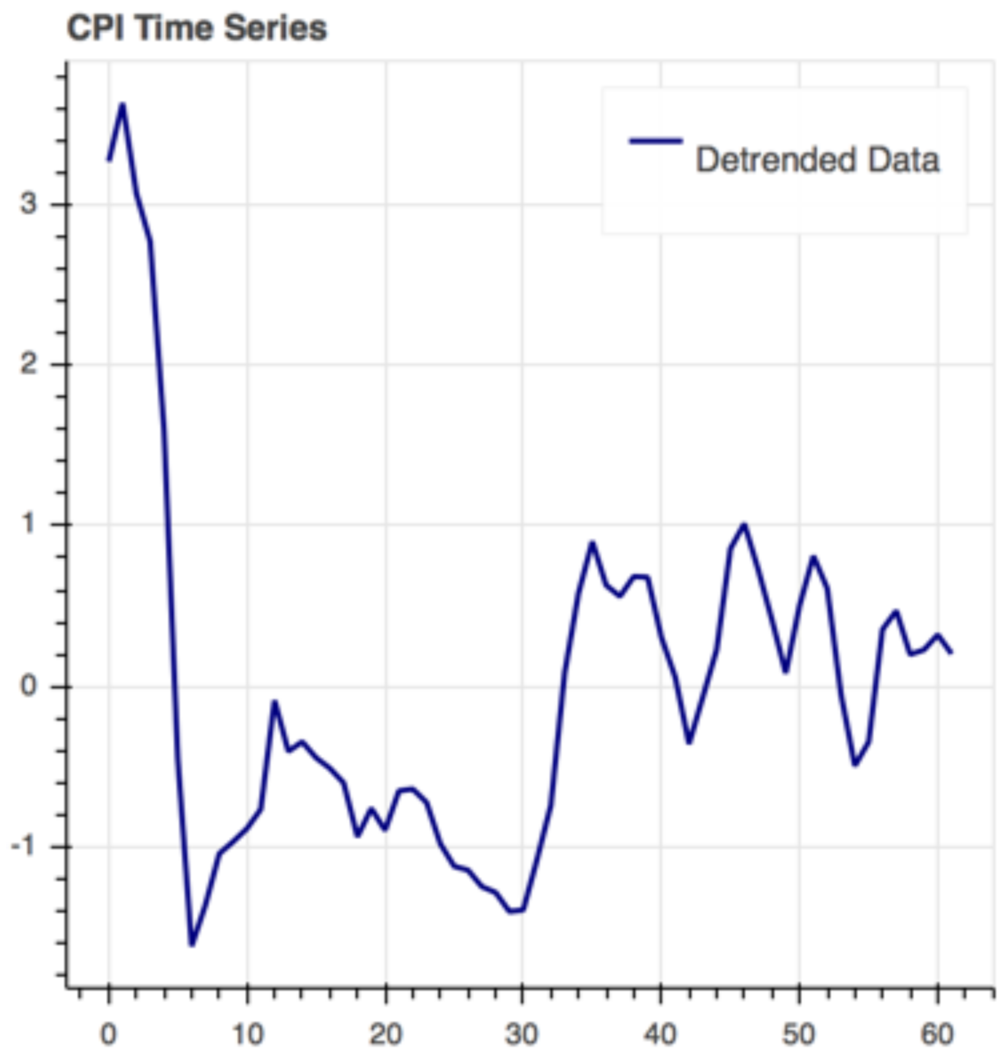
Demonstrate: 1000 samples



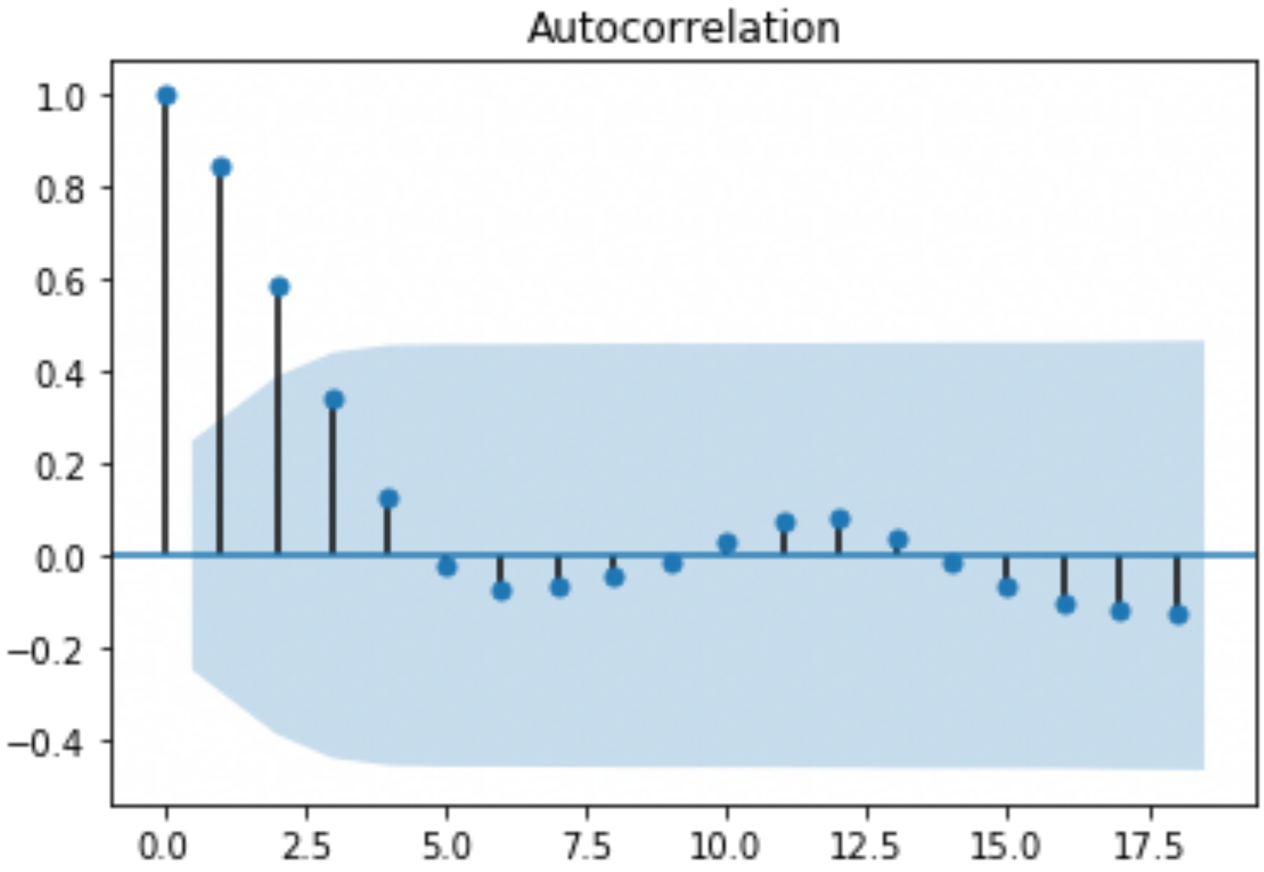
Trends in the Consumer Price Index (CPI) data



De-trended

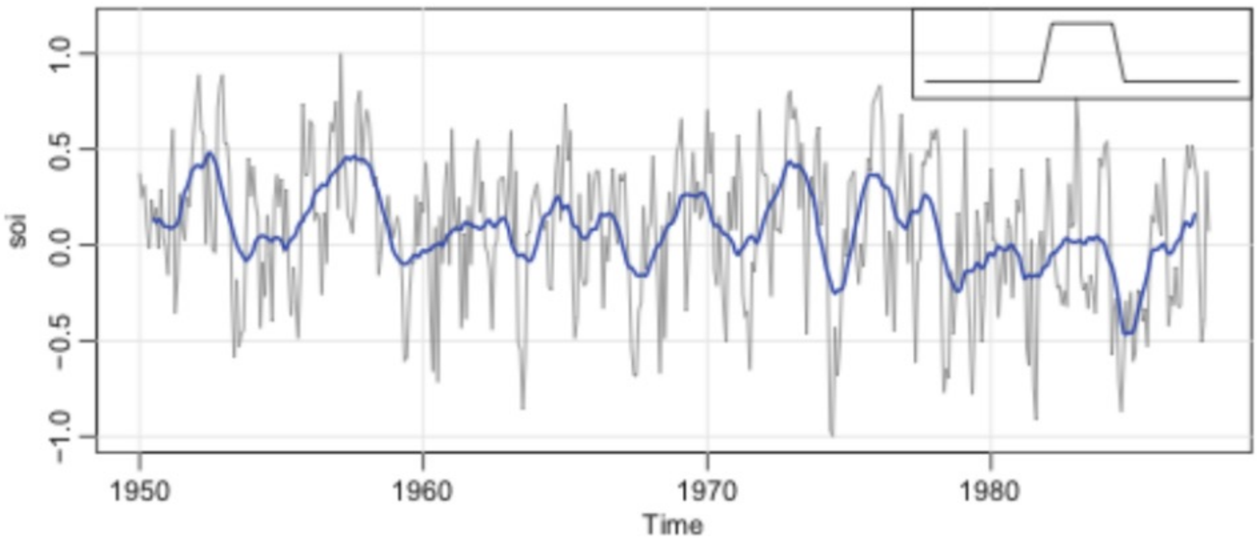


Empirical Autocovariance



Removing Seasonality

- $X_t = S_t + Y_t$
- Estimate S_t using a periodic regression
- Smoothing: $\hat{Y}_t = \sum_{h=-k}^{k} \gamma_h X_{t+h}.$



MIT Institute of Technology

IDSS MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Part II: Models of time series

Massachusetts
Institute of
Technology

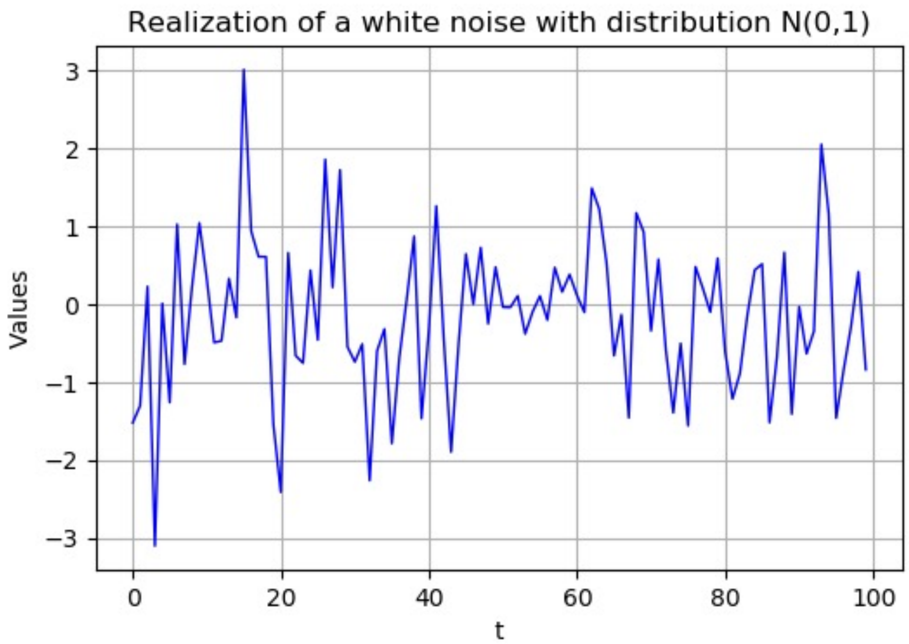
IDSS
MIT INSTITUTE FOR DATA,
SYSTEMS, AND SOCIETY

Models of Time-Series

- White noise
- Auto-regressive (AR)
- Moving Average (MA)
- Auto-regressive moving average (ARMA)
- Autocorrelation of these examples

White noise

- Process $X_t = w$ $\mathbf{E}(w_t) = 0$



- Autocovariance =

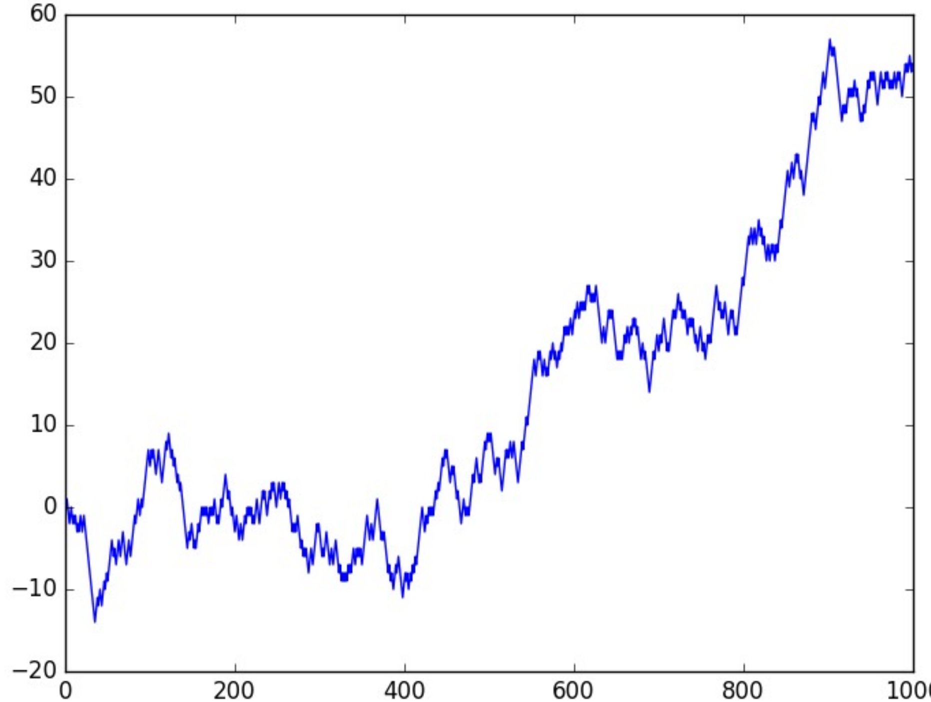
$$\mathbf{E}(w_{t_1} w_{t_2}) = \sigma^2 \, \delta(t_1 - t_2)$$

Massachusetts Institute of Technology

MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Random Walk

$$X_t = X_{t-1} + w_t$$



Autoregressive Process AR(p)

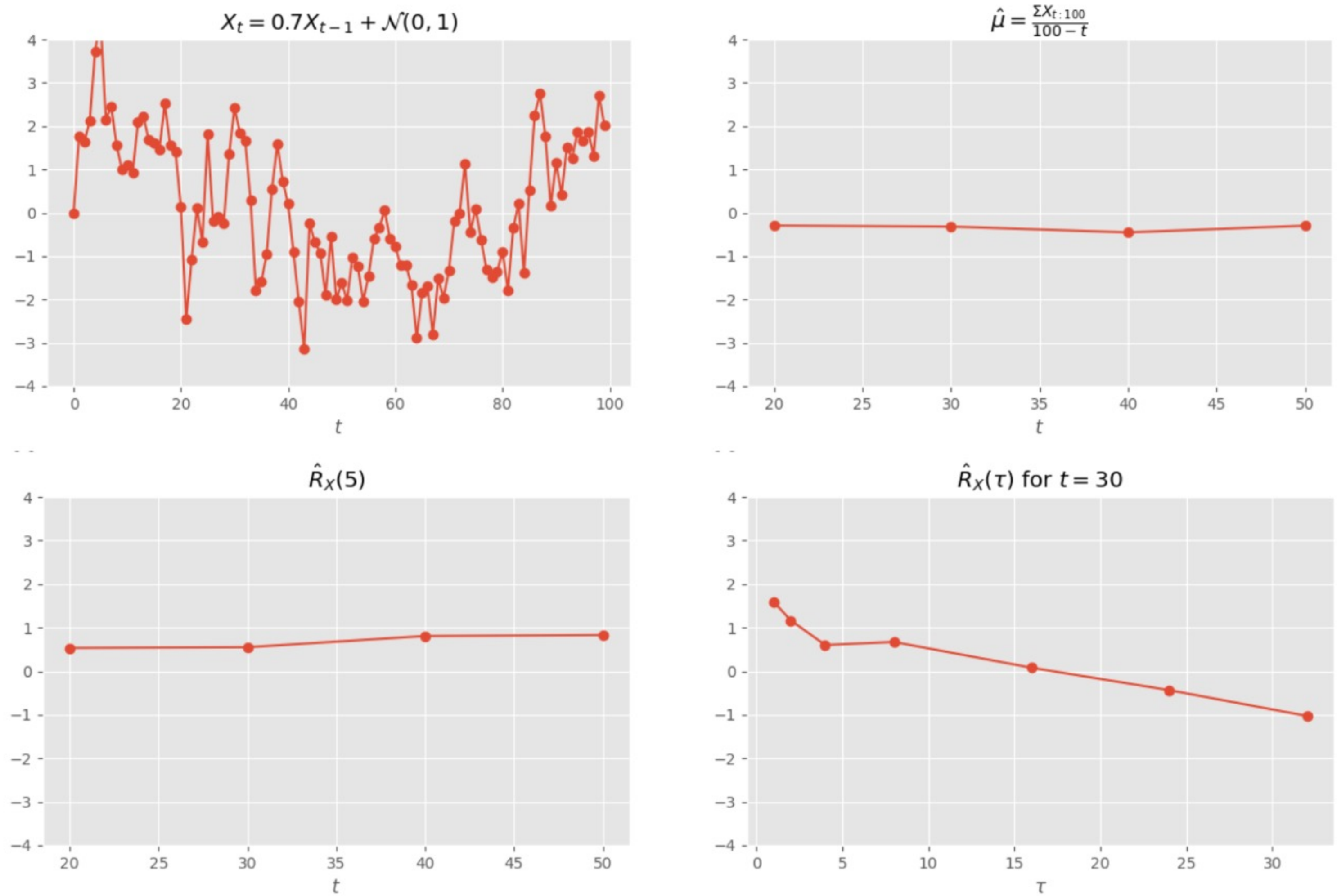
- Process

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + w_t$$

- w_t is white

- Interpretation

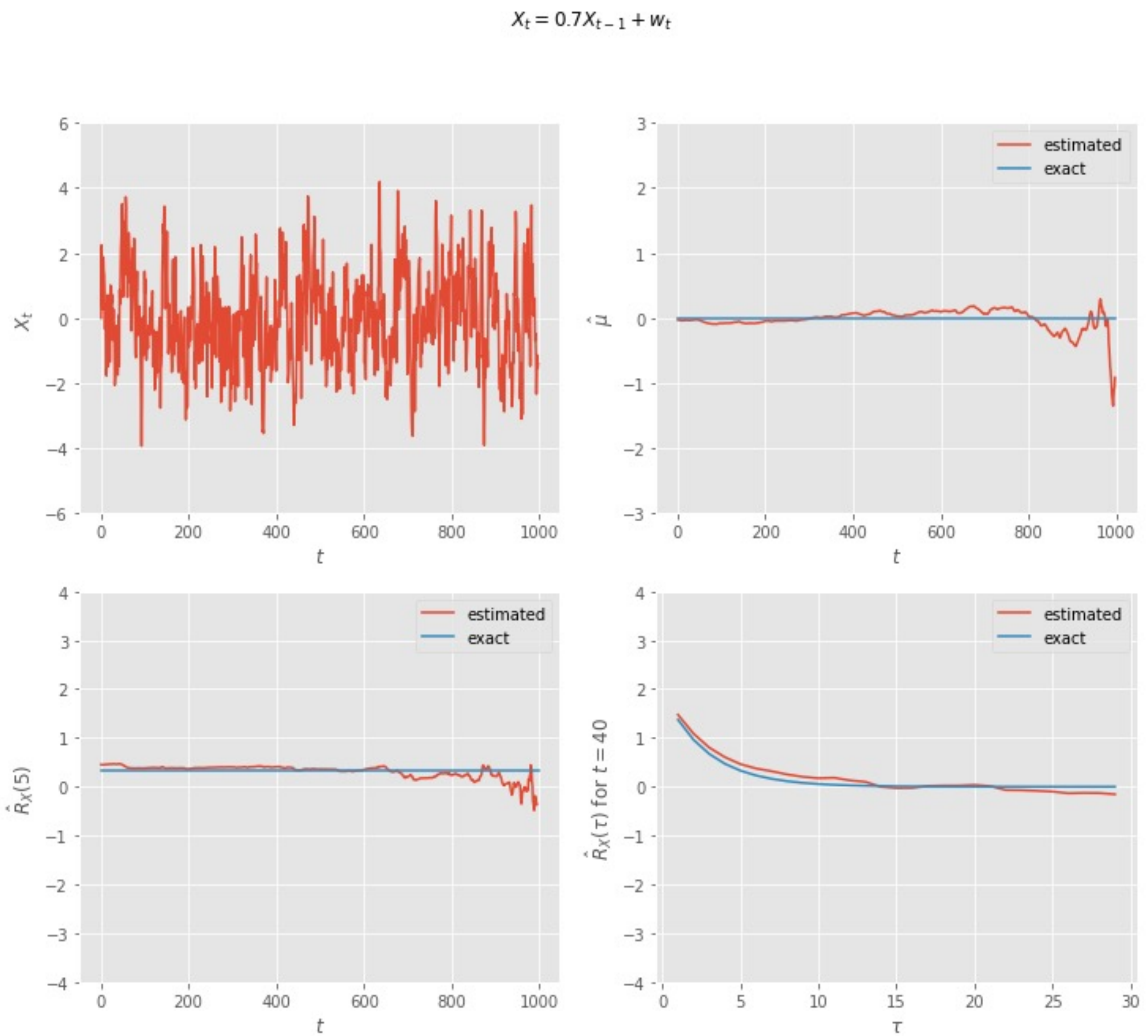
Example AR(1)



Massachusetts Institute of Technology

MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

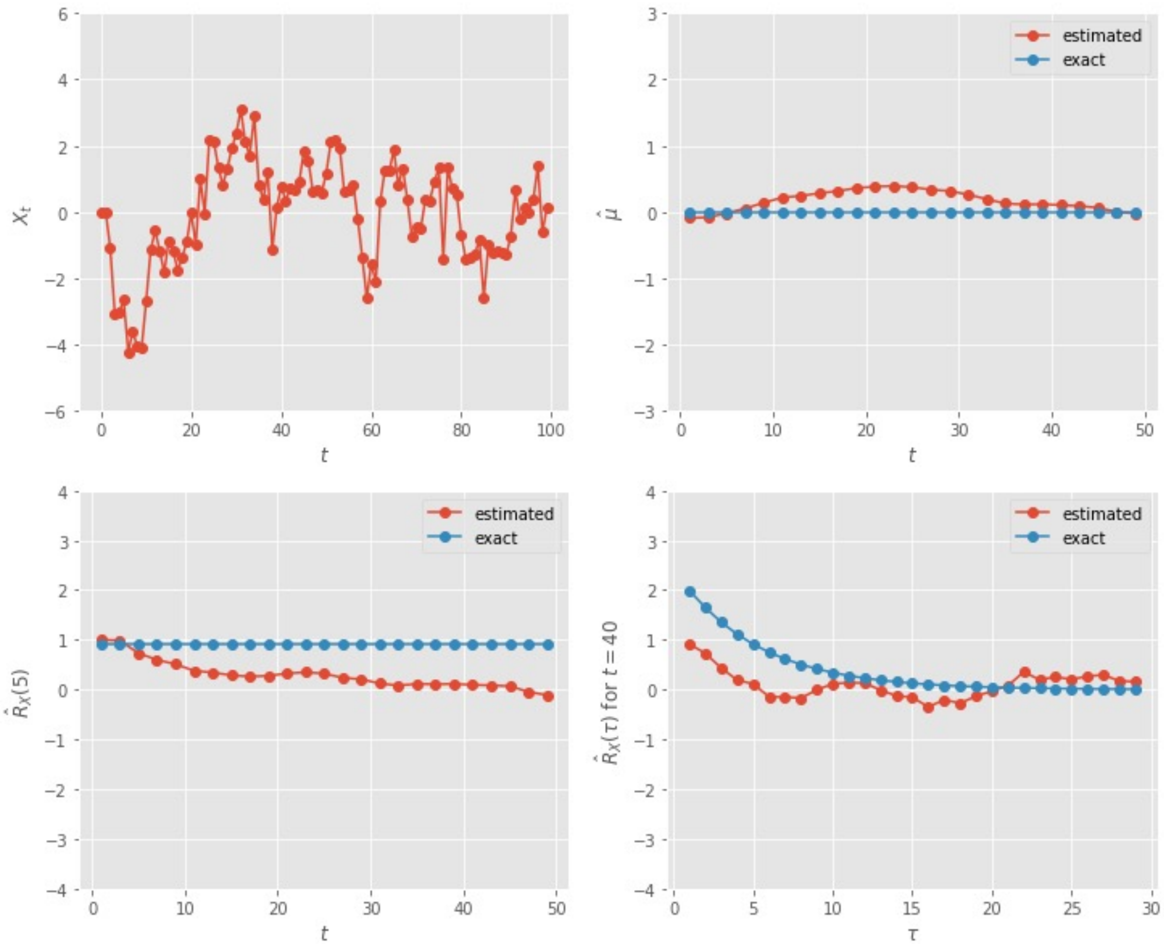
Example: AR(1)



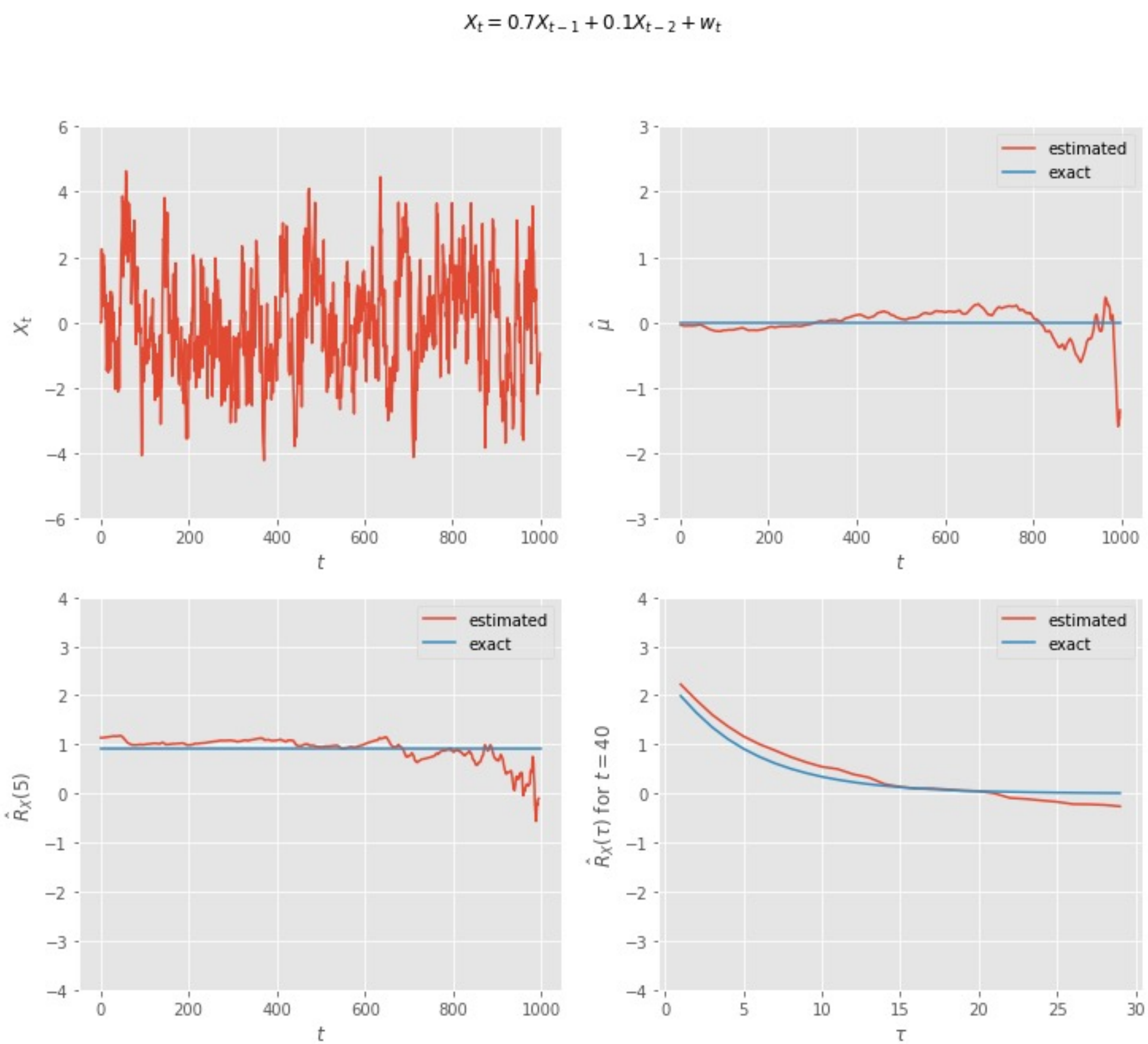
$X_t = 0.7X_{t-1} + w_t$
X_t
t
$\hat{\mu}$
estimated
exact
$\hat{R}_X(5)$
$\hat{R}_X(\tau)$ for $t = 40$
τ

Example: AR(2)

$$X_t = 0.7X_{t-1} + 0.1X_{t-2} + w_t$$



Example: AR(2)



$X_t = 0.7X_{t-1} + 0.1X_{t-2} + w_t$
X_t
t
$\hat{\mu}$
t
estimated
exact
$\hat{R}_X(5)$
t
estimated
exact
$\hat{R}_X(\tau)$ for $t = 40$
τ
estimated
exact

When is AR(p) Stationary

- Response needs to be stable

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + w_t$$

- Root of the polynomial $(1 - a_1 z - a_2 z^2 \; \ldots . -a_p \, z^p \;)$ are inside the unit disc.

- What about the Random Walk: $X_t \; - X_{t-1}$ is stationary!

- Notion can be generalized!

Massachusetts Institute of Technology

IDSS MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Autocovariance of AR(p)

- How do you compute the exact Autocovariance of

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + w_t$$

- It follows that:

$$\mathbf{E}(X_{t-\tau}X_t) = \sum_{i=1}^{p} a_i \mathbf{E}(X_{t-\tau}X_{t-i}) + \mathbf{E}(X_{t-\tau}w_t)$$

- Conclusion

$$R_X(\tau) = \sum_{i=1}^{p} a_i R_X(\tau - i) + \sigma^2 \delta(\tau)$$

Example: AR(2)

- Recall: $R_X(\tau) = R_X(-\tau)$

- Solve for $R_X(0), R_X(1), R_X(2)$ from

$$R_X(0) = a_1 R_X(1) + a_2 R_X(2) + \sigma^2$$

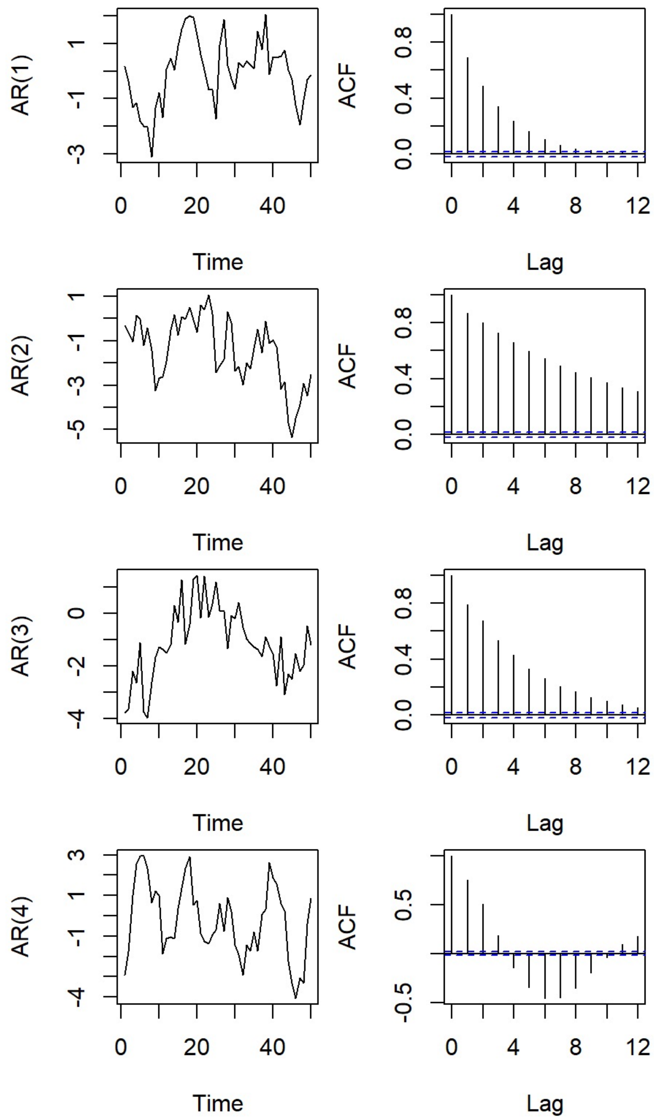
$$R_X(1) = a_1 R_X(0) + a_2 R_X(1)$$

$$R_X(2) = a_1 R_X(1) + a_2 R_X(0)$$

- Compute the rest

$$R_X(\tau) = \sum_{i=1}^{p} a_i R_X(\tau - i), \tau \geq 3$$

Geometric Shape of ACF for AR(p)



Massachusetts Institute of Technology

IDSS
MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Moving Average MA(q)

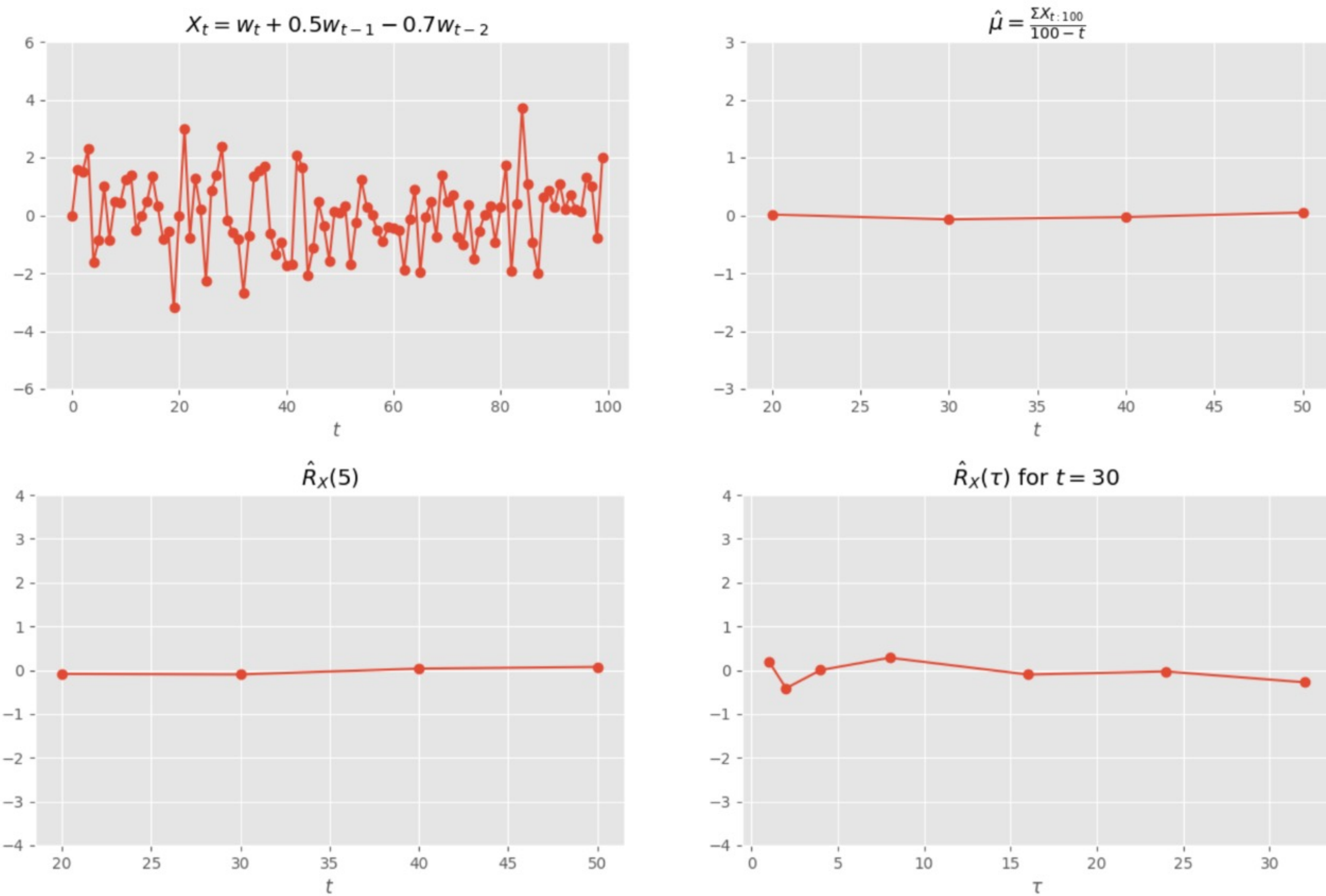
- Process

$$X_t = \sum_{i=0}^{q} b_i w_{t-i}$$

- w_t is white

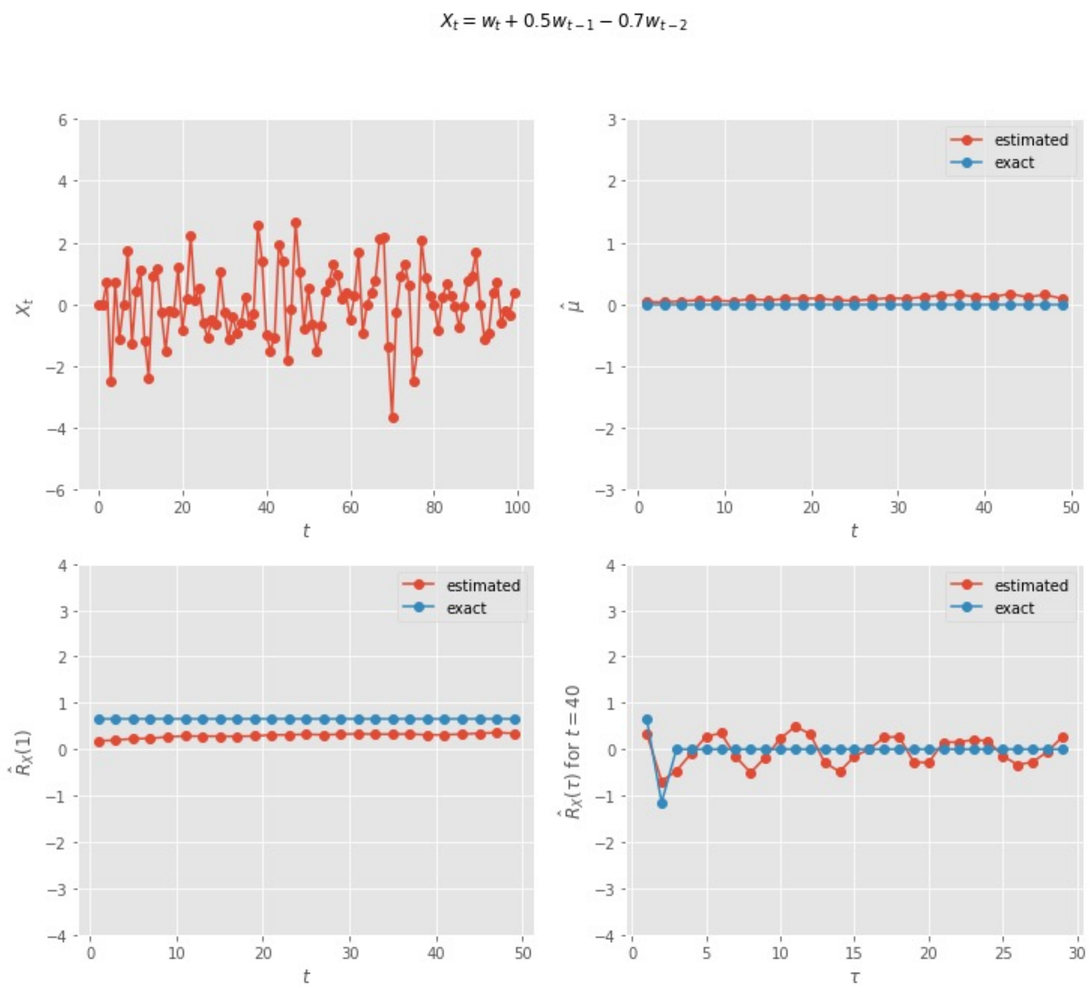
- Interpretation

Example MA(2)



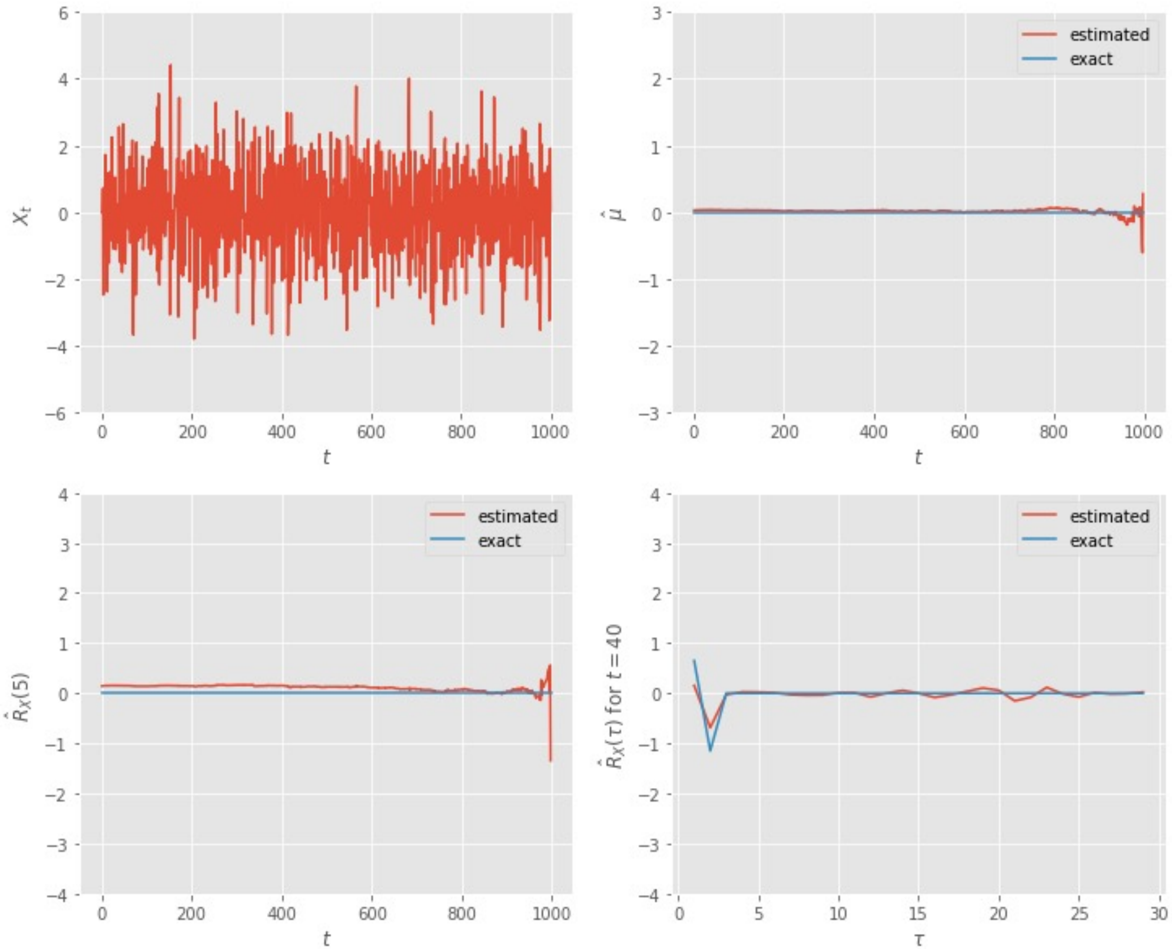
$X_t = w_t + 0.5w_{t-1} - 0.7w_{t-2}$
$\hat{\mu} = \frac{\Sigma X_{t:100}}{100-t}$
$\hat{R}_X(5)$
$\hat{R}_X(\tau)$ for $t = 30$
t
τ

Sample vs Actual



More Samples

$$X_t = w_t + 0.5w_{t-1} - 0.7w_{t-2}$$



Autocovariance of MA

- How do you compute the exact Autocovariance of

$$X_t = \sum_{i=0}^{q} b_i w_{t-i}$$

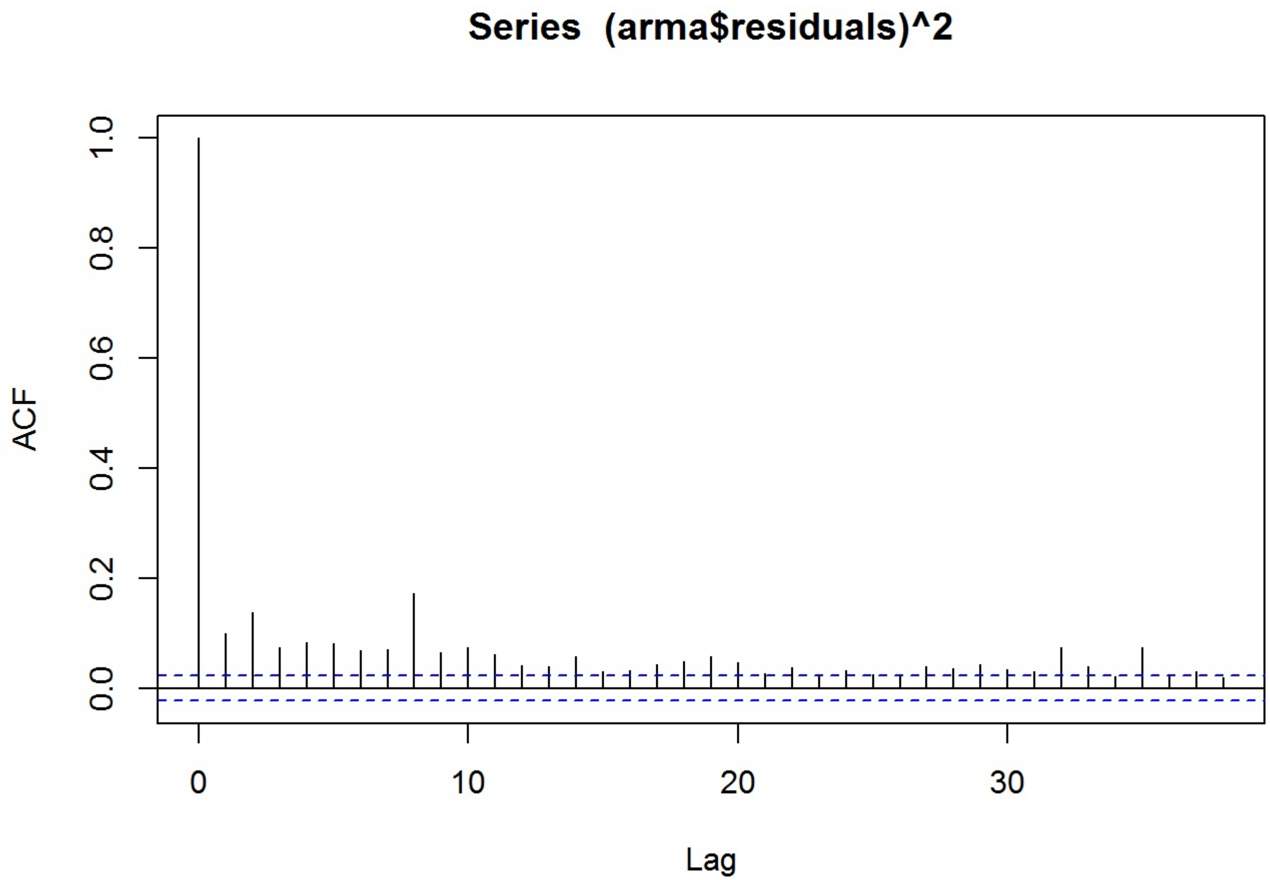
- It follows that:

$$\mathbf{E}(X_t X_{t+\tau}) = \sum_{i=0}^{q} \sum_{j=0}^{q} b_i b_j \mathbf{E}(w_{t-i} W_{t+\tau-j})$$

- Conclusion: simple convolution with finite support

$$R_X(\tau) = \sigma^2 \sum_{j=0}^{q} b_j b_{j-\tau}$$

Finite lag Shape of ACF for MA



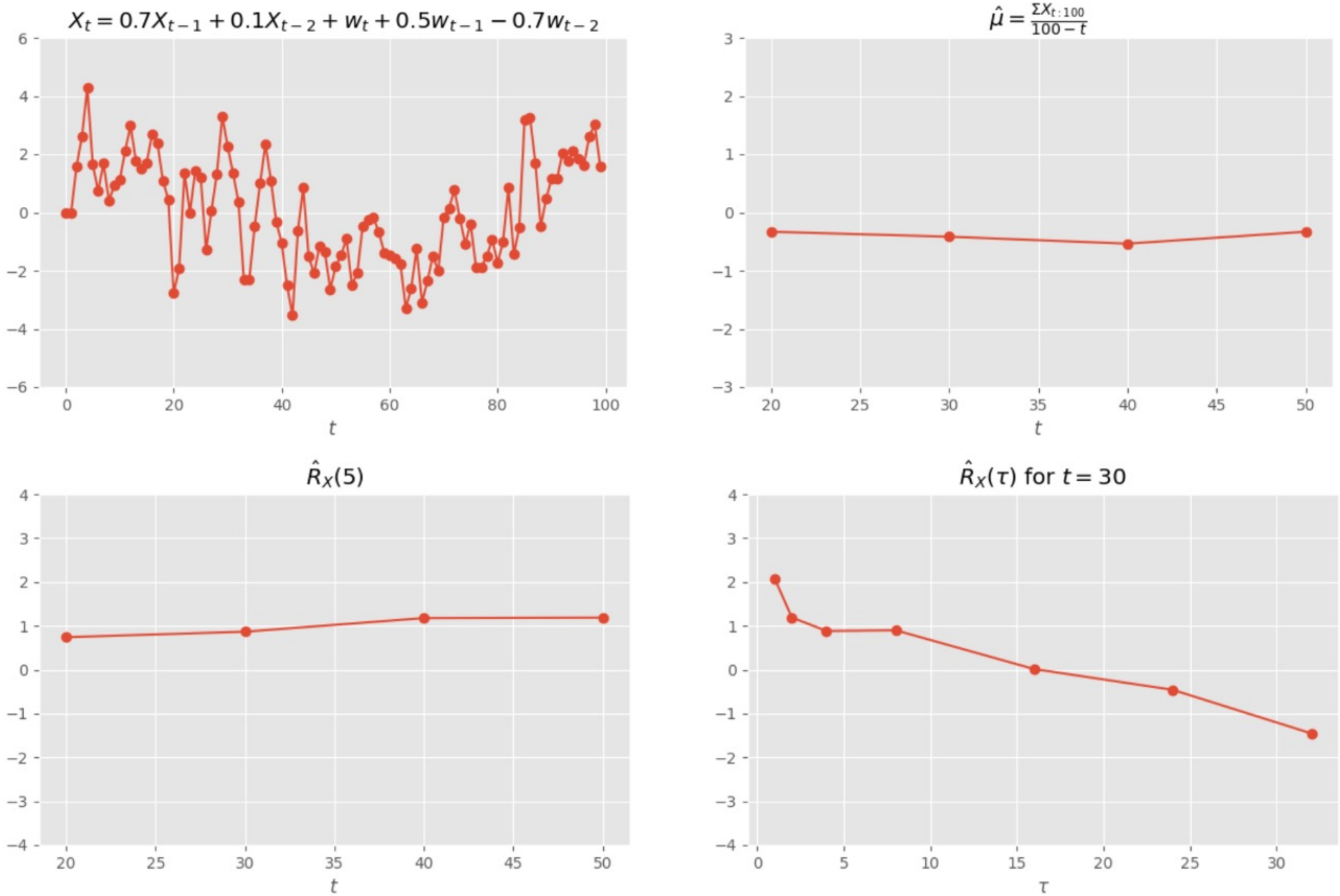
Autoregressive Moving Average Process (ARMA)

- We can combine the two processes

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + \sum_{j=0}^{q} b_j w_{t-j}$$

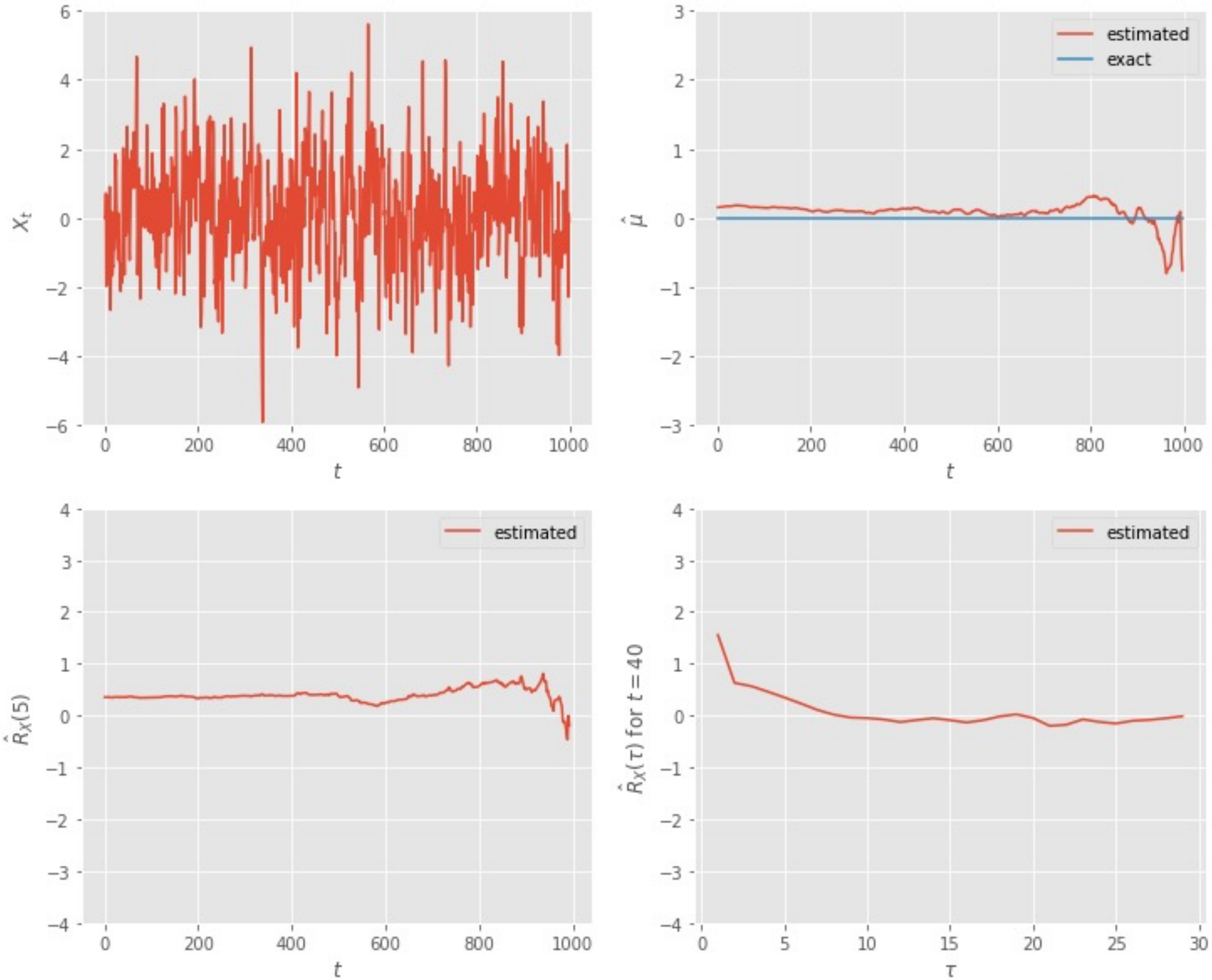
- Superposition of different processes

Example: ARMA



ARMA: More Samples

$$X_t = 0.7X_{t-1} + 0.1X_{t-2} + w_t + 0.5w_{t-1} - 0.7w_{t-2}$$



Integrated Processes

- Extensions of Random Walks
- X_t is not stationary
- Difference process may be
  - First order: $X_t - X_{t-1}$
  - Second Order: $X_t - 2X_{t-1} + \mathrm{X}_{\mathrm{t}-2}$
  - Higher order

ARI (Autoregressive Integrated)

- $A(z) = \ (1 - a_1 z - a_2 z^2 \ \ \ldots . - a_p \ z^p \ )$
  - $A(1) = 0$

- $A(z) = (1 - z)\ \tilde{A}\ (\mathrm{z})$

- Process is not stationary!

- $\tilde{X}_{\mathrm{t}} \ \equiv X_t \ - X_{t-1}$ is stationary:

$$\tilde{A}\ (\mathrm{z})\tilde{X}_{\mathrm{t}} = w_t$$

ARI (Autoregressive Integrated)

- $A(z) = (1 - a_1 z - a_2 z^2 \; \ldots . - a_p \, z^p \;)$
- Higher order: $A(z) = (1 - z)^2 \tilde{A}\,(\mathrm{z})$
- Process is not stationary!
- $\tilde{X}_\mathrm{t} \equiv X_t - 2\,X_{t-1} + X_{t-2}$ is stationary:

$$\tilde{A}\,(\mathrm{z})\tilde{X}_\mathrm{t} = w_t$$

- Generalizes to ARMA (ARIMA)

Massachusetts Institute of Technology

IDSS MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Questions

Part III: Learning Time Series

Learning a Time Series Model

- Data to Models
  - Autoregressive (AR) models
  - Moving Average (MA)

- AR learning Looks like a standard Least Squares

  - What's the catch?
  - What can we learn?

- How about other models

- Example: Consumer Price Data

Learning AR(p)

- Data

$$X_0,\, X_1,\, X_2, \ldots X_N$$

- Model (known p)

$$X_t = \sum_{i=1}^{p} a_i X_{t-i} + w_t$$

- Algorithm: Define

$$\mathbf{x} = (x_{p+1},\, x_{p+2},\, \ldots x_N)' \qquad \mathbf{a} = (a_1,\, a_2,\, \ldots a_p)'$$

AR(p)

- Define

$$\mathbf{x} = (x_{p+1},\, x_{p+2},\, \ldots x_N)' \qquad \mathbf{a} = (a_1,\, a_2,\, \ldots a_p)'$$

$$\mathbf{A} = \begin{pmatrix} x_p & x_{p-1} & \cdots & x_1 \\ x_{p+1} & x_p & \cdots & x_2 \\ \vdots & \vdots & \vdots & \vdots \\ x_{N-1} & x_{N-1} & \cdots & x_{N-m} \end{pmatrix}.$$

- Minimize the mismatch by selecting $\mathbf{a}$

$$\min_{\mathbf{a}} \|\mathbf{x} - A\mathbf{a}\|$$

Example:

- Generate 100 data points according to

$$x_t = .7\, x_{t-1} + .1\, x_{t-2} + \mathrm{w_t}$$

- Fit an AR(1) model
  - Answer: $a = .8704$

- Fit an AR(2) model
  - Answer $a_1 = 0.7448 \;\; and\; a_2 = .1742$

- Observations?

Order Estimation

- Derive multiple estimates and choose based on error on cross validation data
- Add penalty to model complexity (MDL, AIC)
- Use the Autocovariance function (ACF) as a guidance!
  - Can we do better than ACF?

Partial Autocovariance Function (PACF)

- $X_t \quad | X_{t+1} , X_{t+2} , \ldots, X_{t+k-1} | \; X_{t+k}$

- Project X_{t+1} and X_{t+k} on the variables in between:

$$P_t \quad and \quad P_{t+k}$$

- $\gamma(k) = \boldsymbol{E}\,(X_t - P_t)(X_{t+k} - P_{t+k}\,)$

- For AR(p): $\quad \gamma(k) = 0 \;\; \forall \;\; k \geq p$

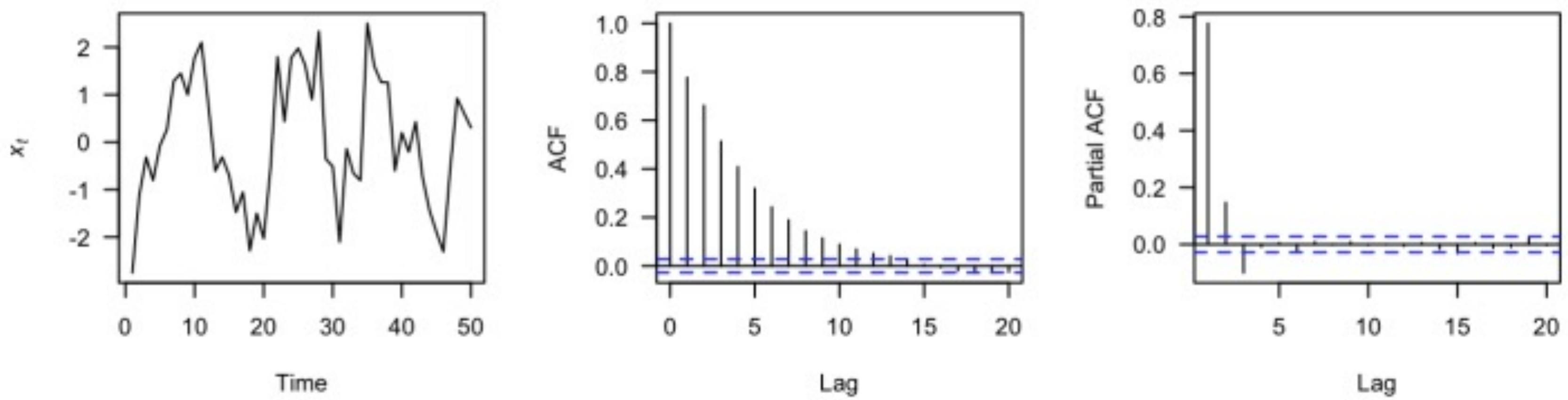
- *Use to estimate order!*

Massachusetts Institute of Technology

IDSS MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Example: AR(3)

$$X_t = .7\,X_{t-1} + .2\,X_{t-2} - .1\,X_{t-3}$$

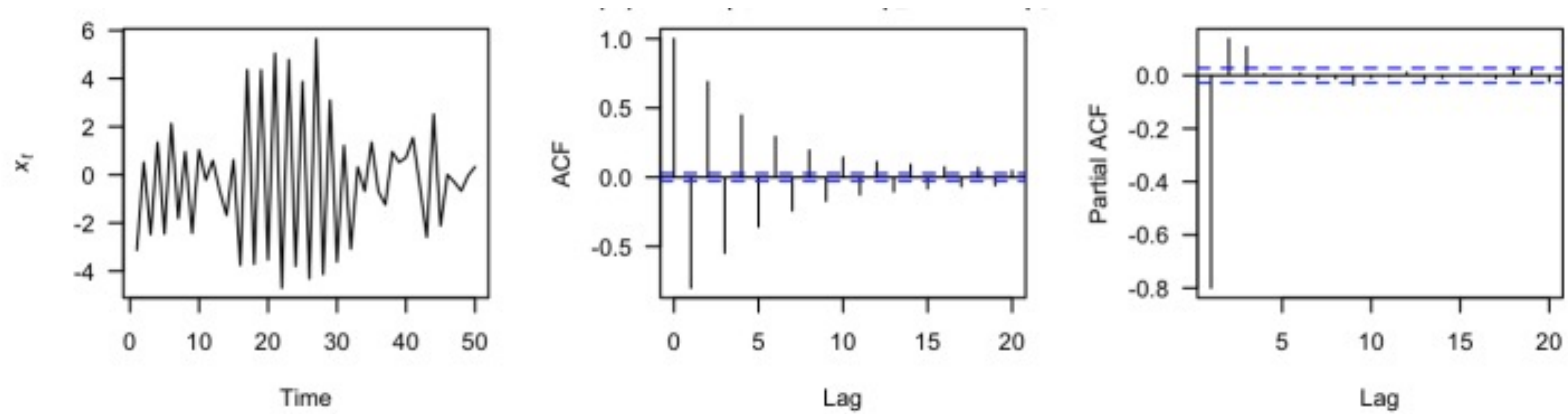


Massachusetts Institute of Technology

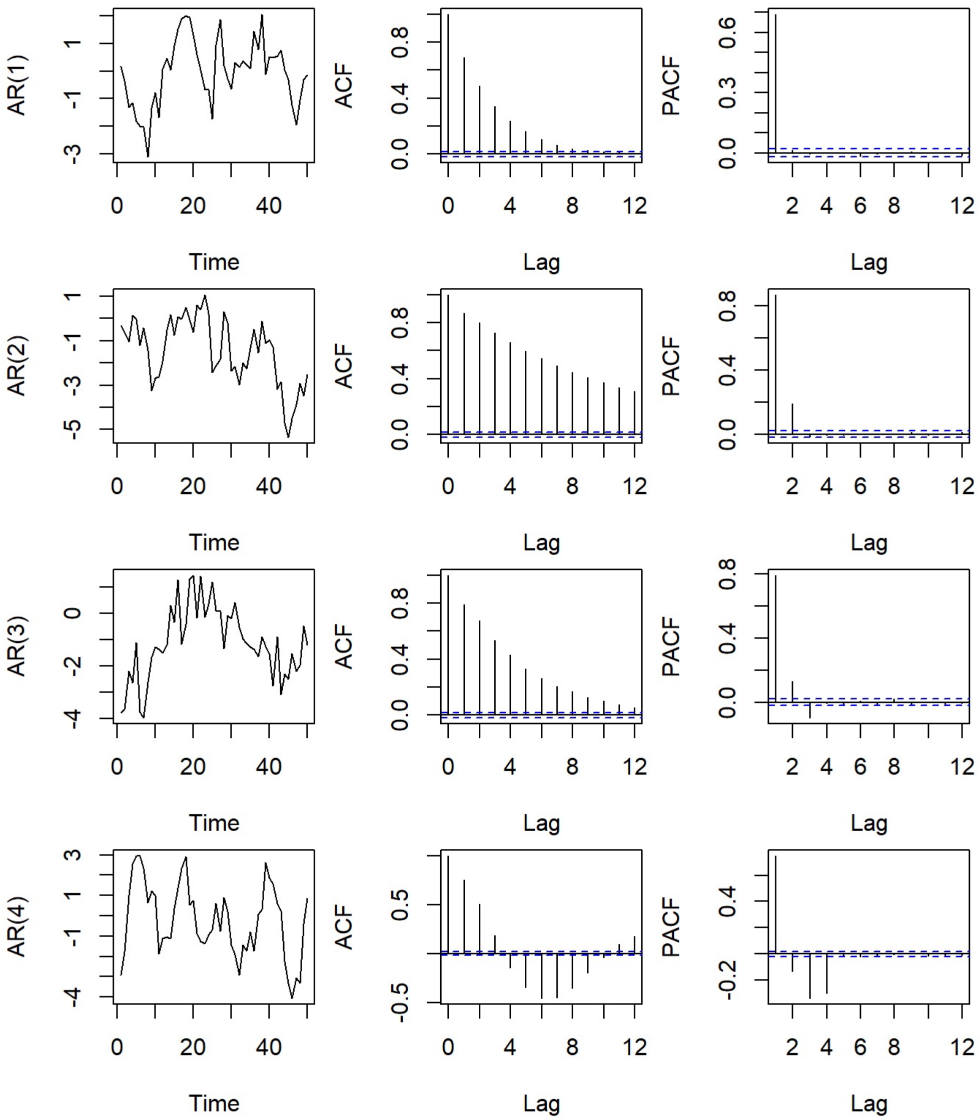
IDSS
MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Example: AR(3)

$$X_t = -.7\, X_{t-1} + \, .2\, X_{t-2} + .1\, X_{t-3}$$



Geometric Shape of ACF for AR(p)



Massachusetts Institute of Technology

IDSS
MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Learning a MA model

- Data

$$X_0, X_1, X_2, \dots X_N$$

- Model (known p)

$$X_t = \sum_{i=0}^{p} b_i w_{t-i}$$

- Recall

$$R_X(\tau) = \mathbf{E}(X_t X_{t+\tau}) = \sum_{j=0}^{p} b_j b_{j-\tau} \sigma^2$$

Learning a MA model

- Compute the empirical Autocovariance $\hat{R}_X(\tau)$

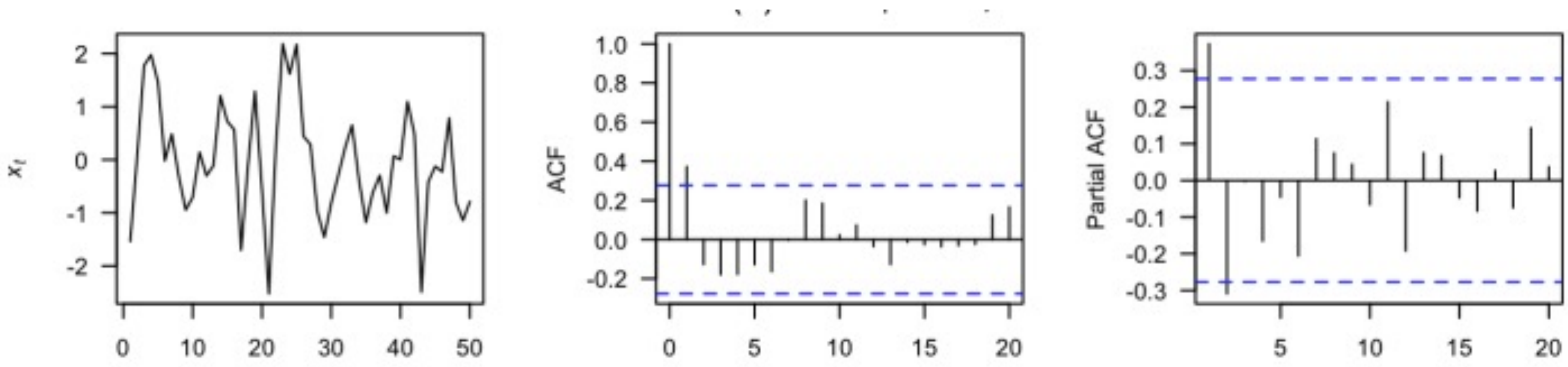
- Solve an inverse problem:

$$\hat{R}_X(\tau) = \sigma^2 \sum_{j=0}^{p} b_j b_{j-\tau}$$

- Quadratic Equation!

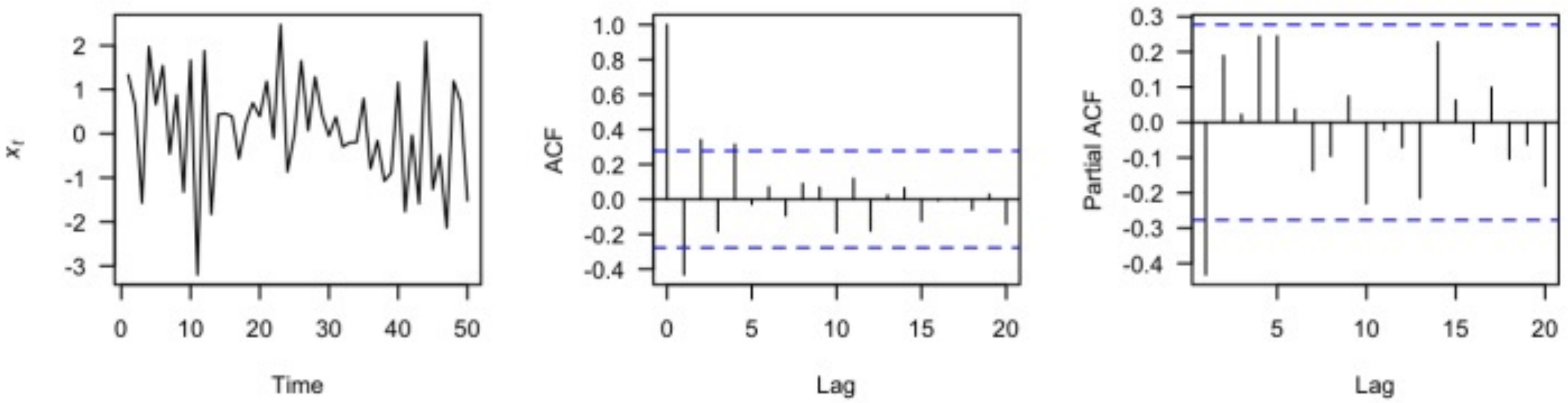
Example: MA(2)

$$X_t = .7\, W_{t-1} + W_t$$



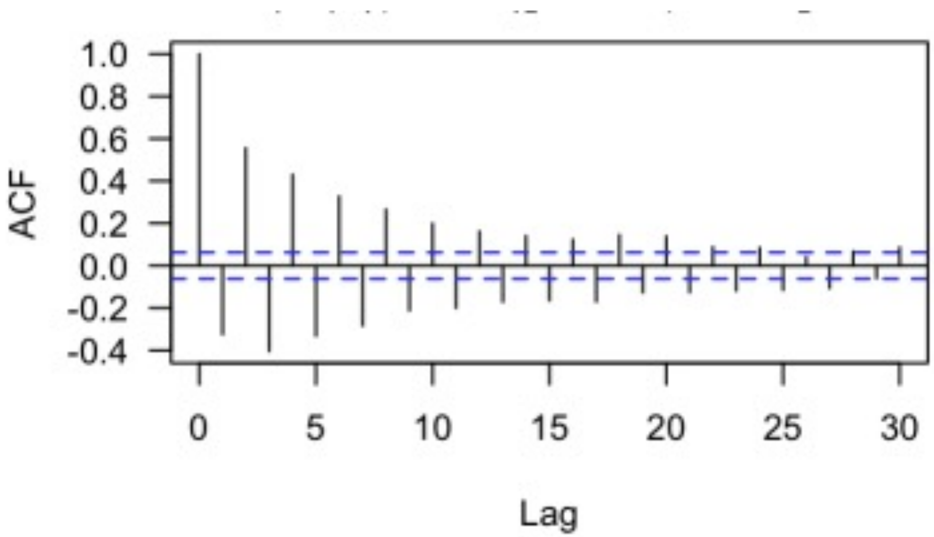
Example: MA(3)

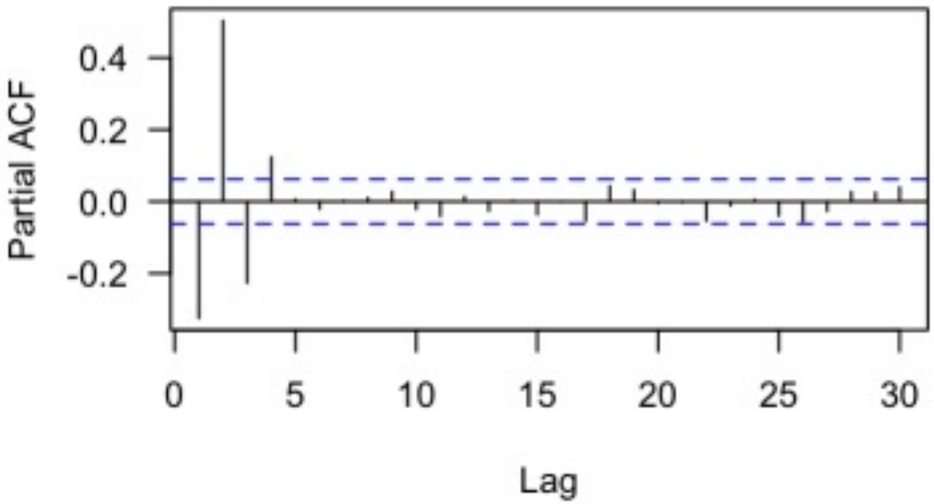
$$X_t = W_t - .7\,W_{t-1} + .2\,W_{t-2} + .1\,W_{t-3}$$



Example: ARMA(3)

$$X_t = .7\,X_{t-1} + \;.2\,X_{t-2} + .7\,W_t + \;.2\,W_{t-1}$$





ACF vs PACF

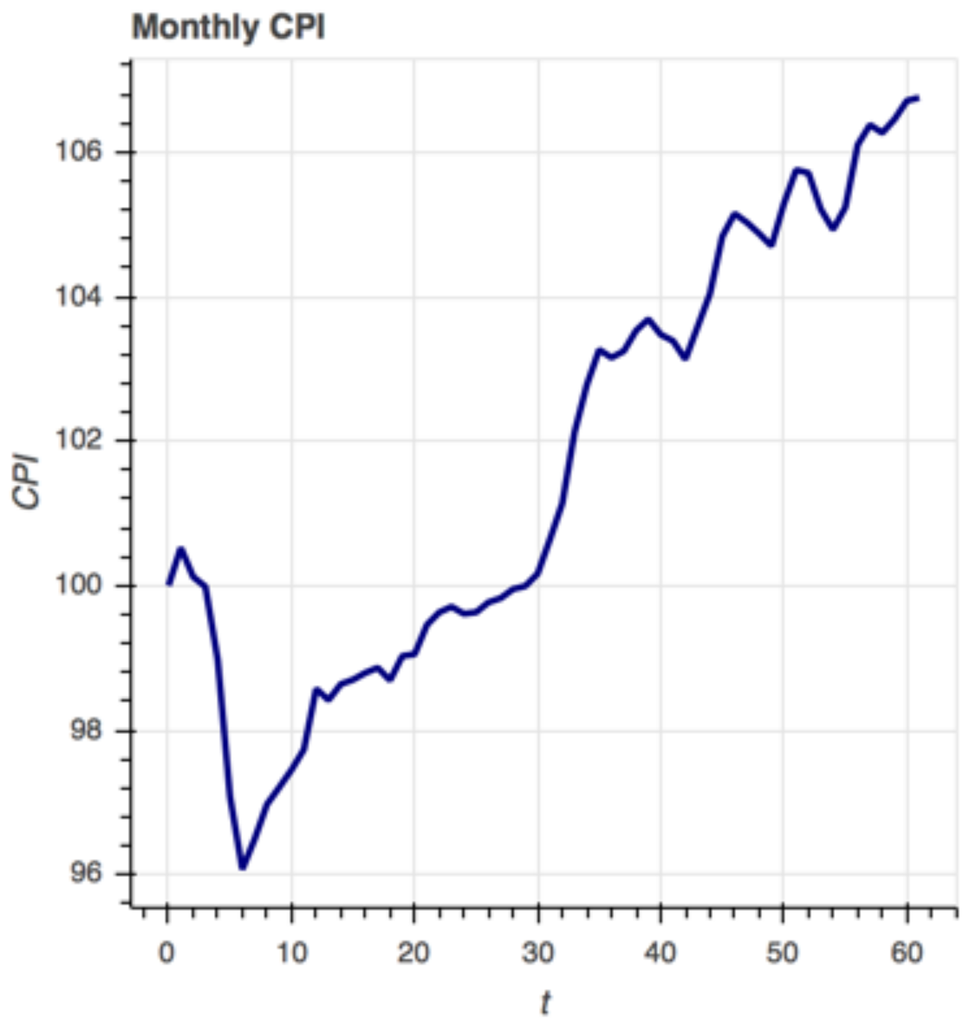
	ACF	PACF
AR(p)	decays	zero for $h > p$
MA(q)	zero for $h > q$	decays
ARMA(p,q)	decays	decays

Billion Price Data

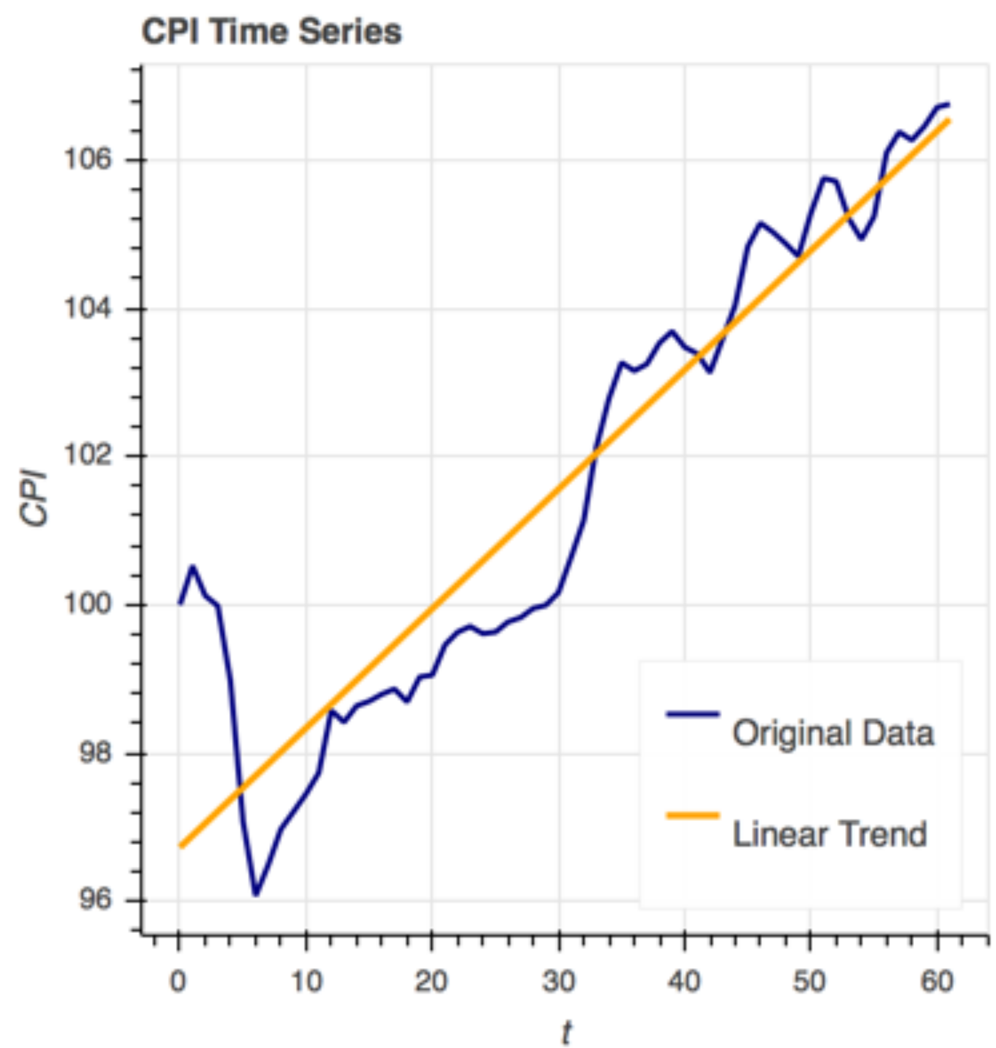
- The goal of this problem is to analyze the PriceStats data from the MIT Billion Prices Project, provided by Professor Rigobon.

- Consumer Price Index Data: (consumer price index, the price of a ”market basket of consumer goods and services” - a proxy for inflation)

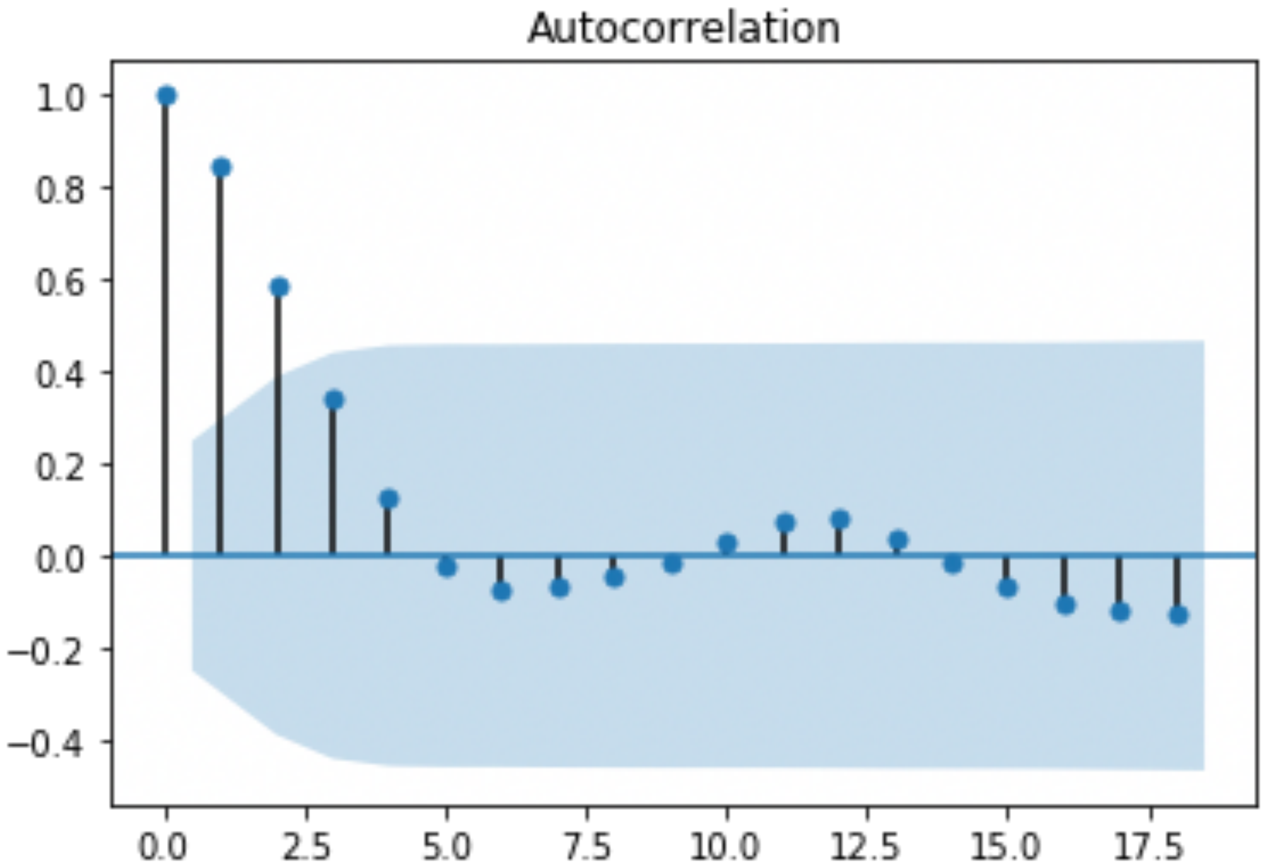
Consumer Price Index



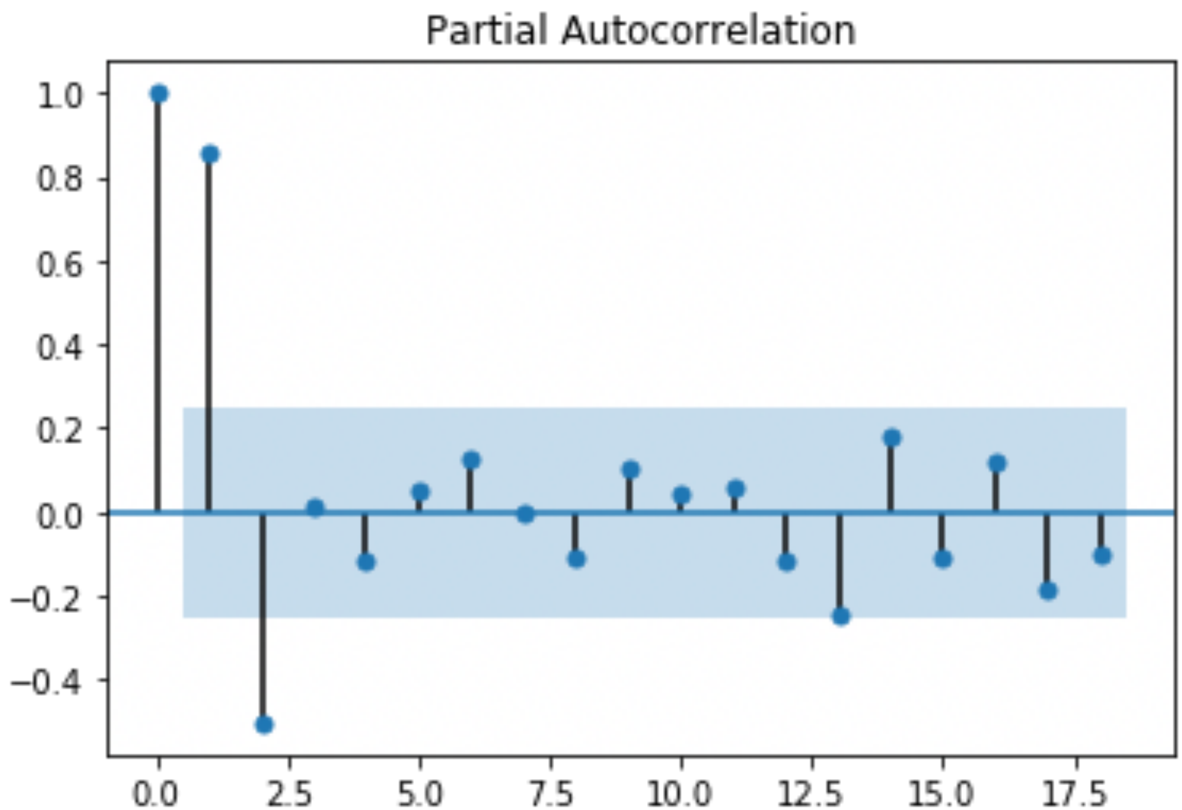
Identify Trends in the CPI data



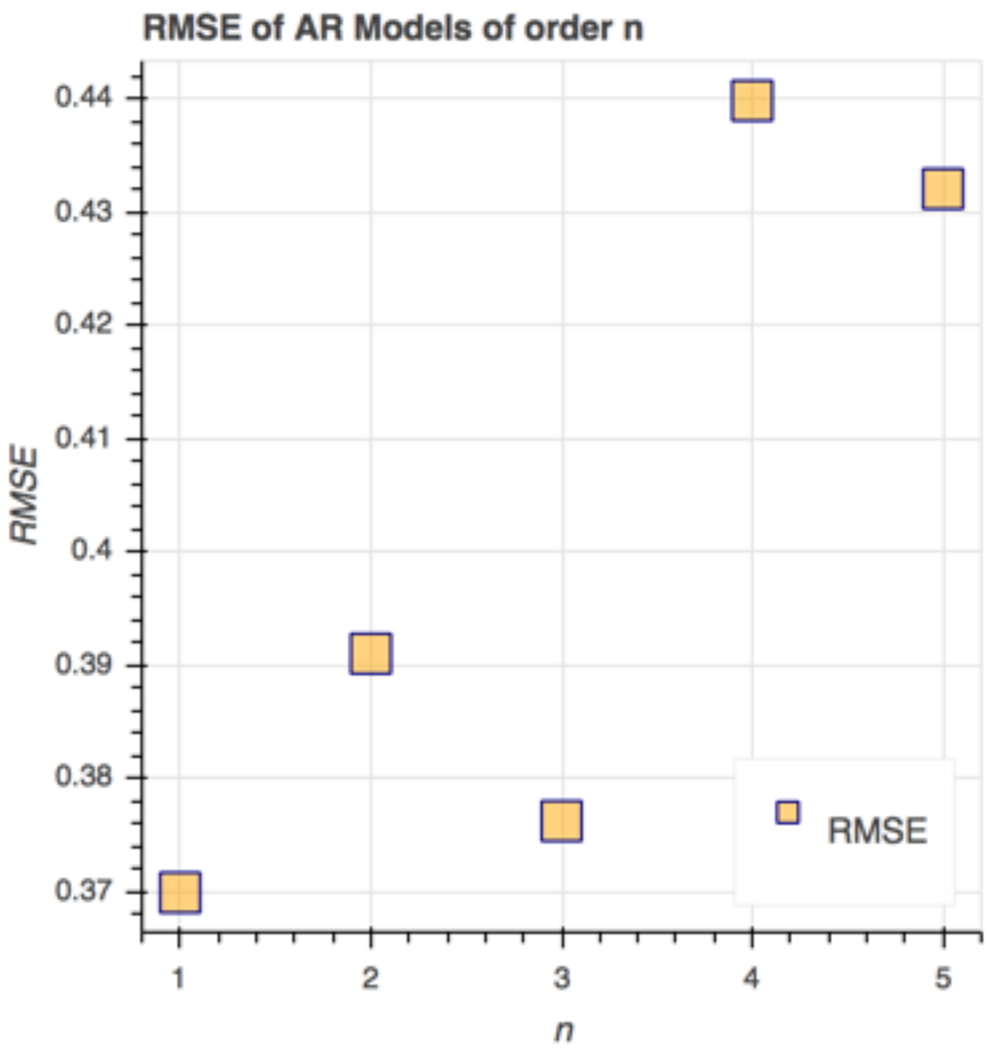
Empirical Autocovariance



Empirical Partial Autocovariance



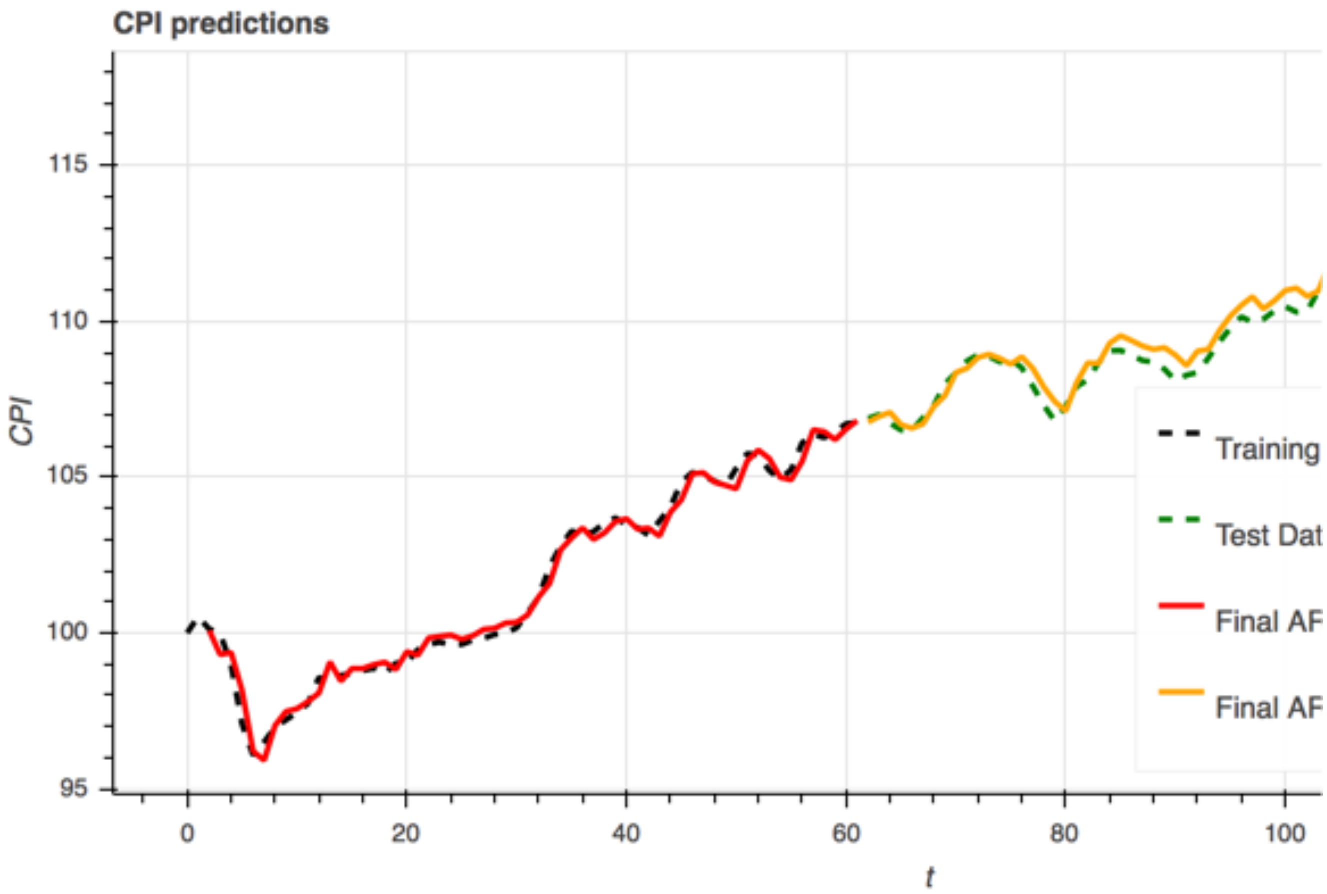
AR(1) or AR(2)



Massachusetts Institute of Technology

IDSS MIT INSTITUTE FOR DATA, SYSTEMS, AND SOCIETY

Prediction of AR(2)



Thank You