INSOFE
Inspire...Educate...Transform.






Inspire…Educate…Transform.

Supervised Models

Time series

Dr. Dakshinamurthy V Kolluru

President – INSOFE

October 23, 2016

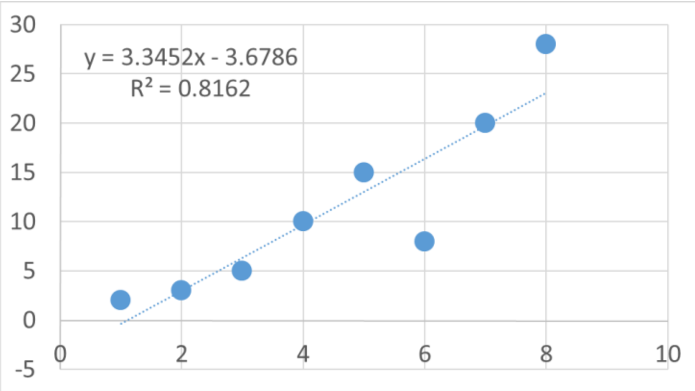
Why Modeling

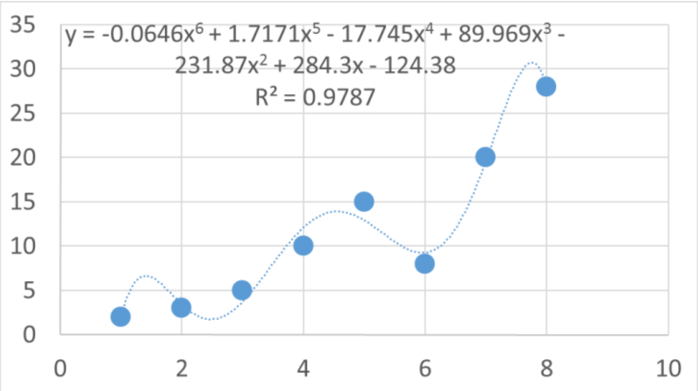
- In any business, there are some easy-to-measure metrics
  - Age; Gender; Income; Education level; etc.

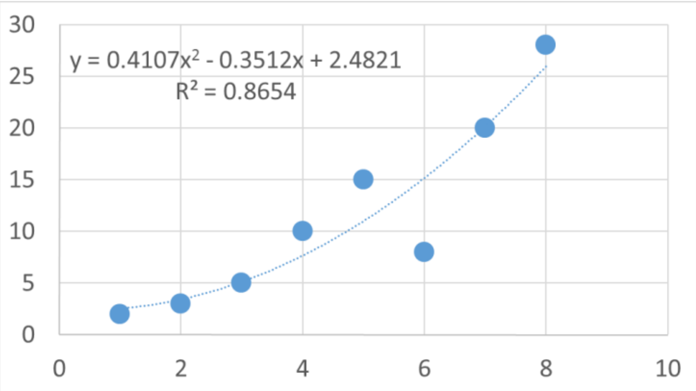
  and a difficult-to-measure metric
  - Amount of loan to give; Will she buy or not; How many days will he stay in the hospital; etc.

- Supervised learning is about computing the latter using the former

Saving ourselves from chasing R^2





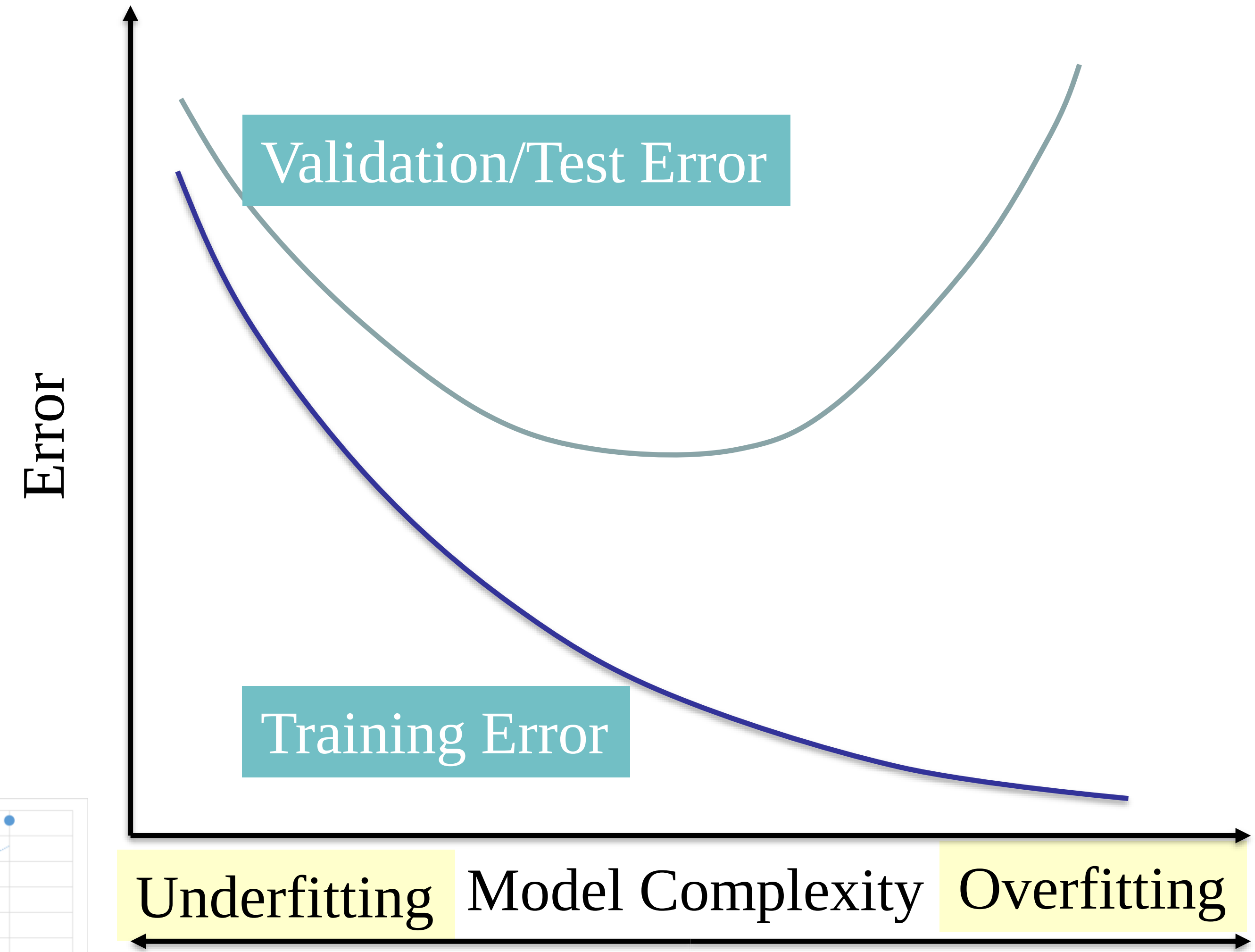


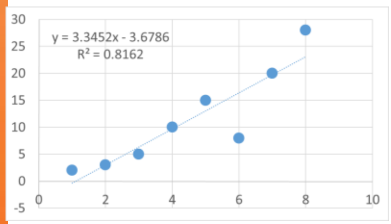
Too Simple a Model Underfit	Too Complex a Model Overfit	Right Model Reasonable fit

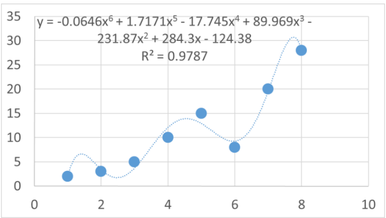
The Ultimate Test of Model Accuracy

- Holdout set: Split data into train, validation and test sets (in 70:20:10 or 60:20:20, etc. ratios), and **ensure model performance is similar**.
  - Training Set: For fitting a model
  - Validation Set: For selecting a model based on estimated prediction errors
  - Test Set: For assessing selected model’s performance on “new” data
- k-fold cross-validation: Same as holdout but useful when the data size is small.

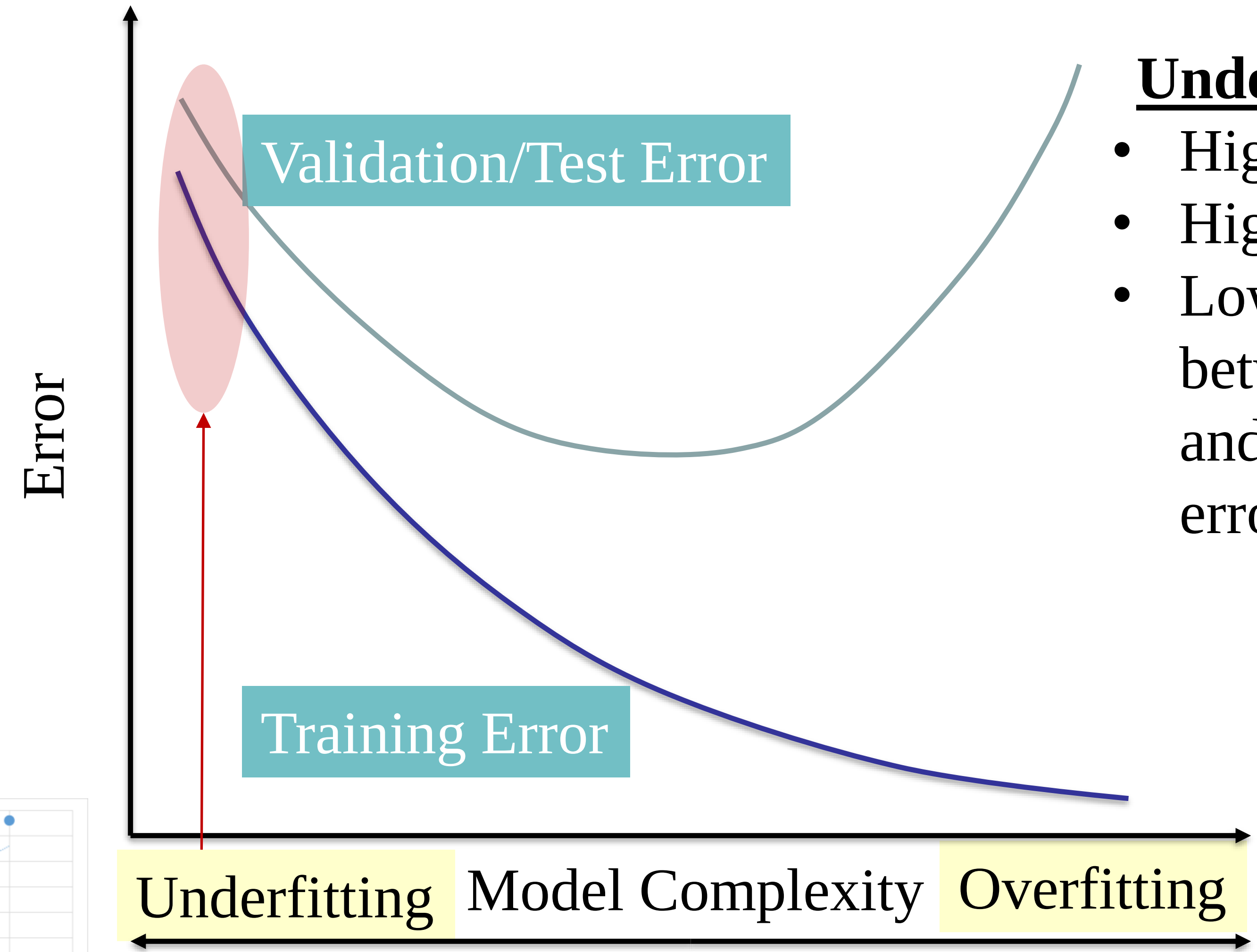
Bias-Variance Tradeoff and Underfitting vs Overfitting

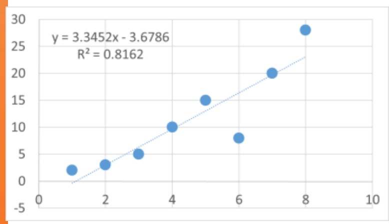






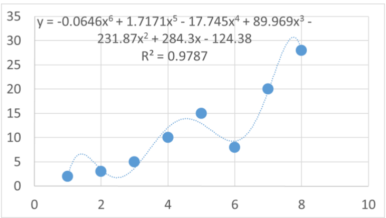
Diagnosing Bias and Variance



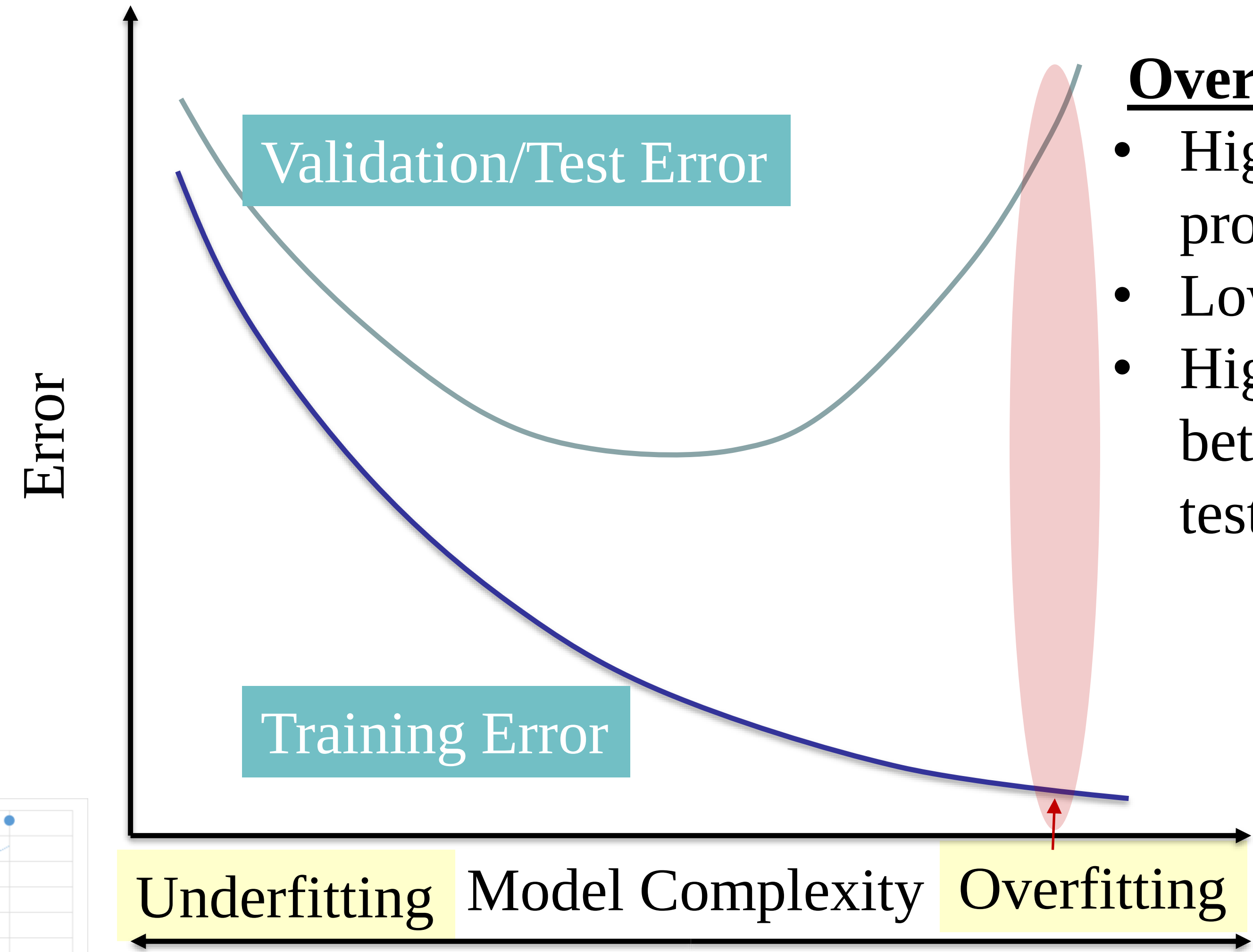


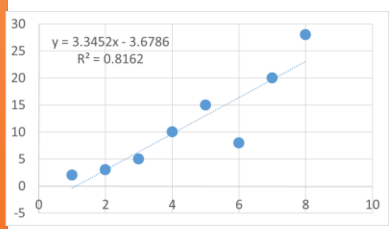
<u>Underfitting (Bias)</u>

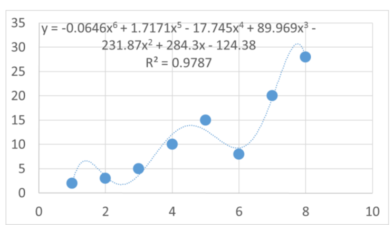
- High Bias problem
- High training error
- Low difference between training and test/validation errors



Diagnosing Bias and Variance







<u>Overfitting (Variance)</u>

- High Variance problem
- Low training error
- High difference between training and test/validation errors

Bias-Variance Tradeoff

Ways of detecting and minimizing Bias and Variance

Outliers and Influential Observations can cause statistical bias. Can be identified using various methods like Box plots, points outside $\pm 2 \; or \; \pm 3$ standard deviations/errors, residual plots, etc.

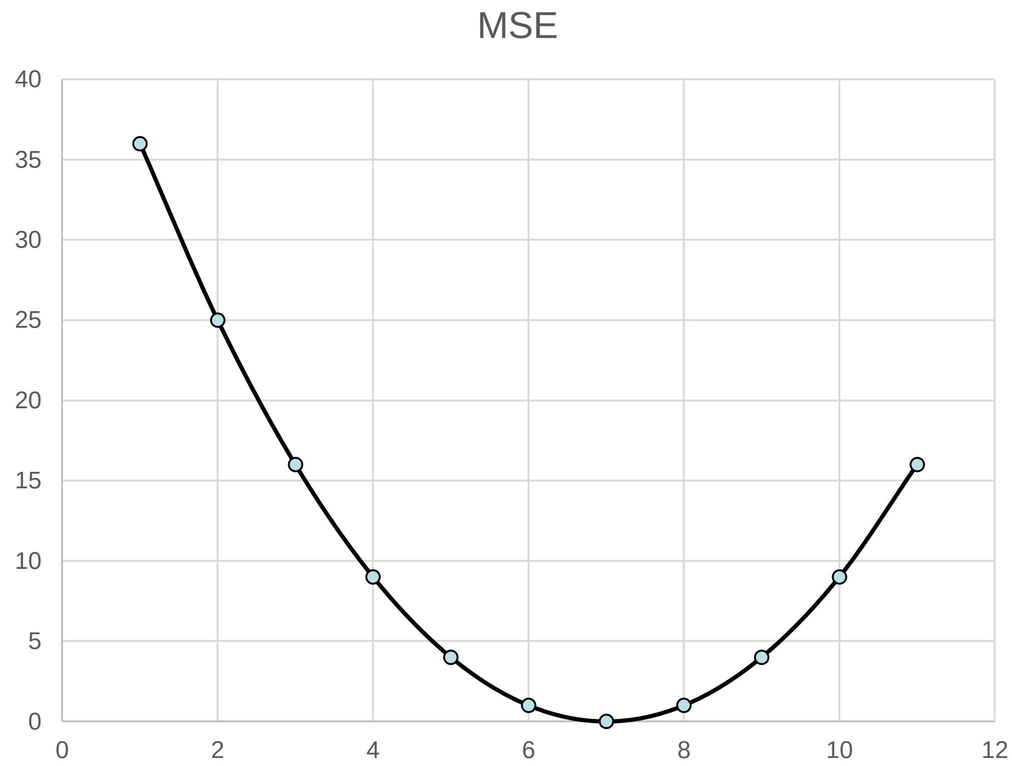
Bias cannot be corrected by increasing training sample size.

Variance or standard error can be minimized by increasing training sample size.

Bagging (bootstrap aggregating) techniques *(taught later in the program)* can be used to minimize errors.

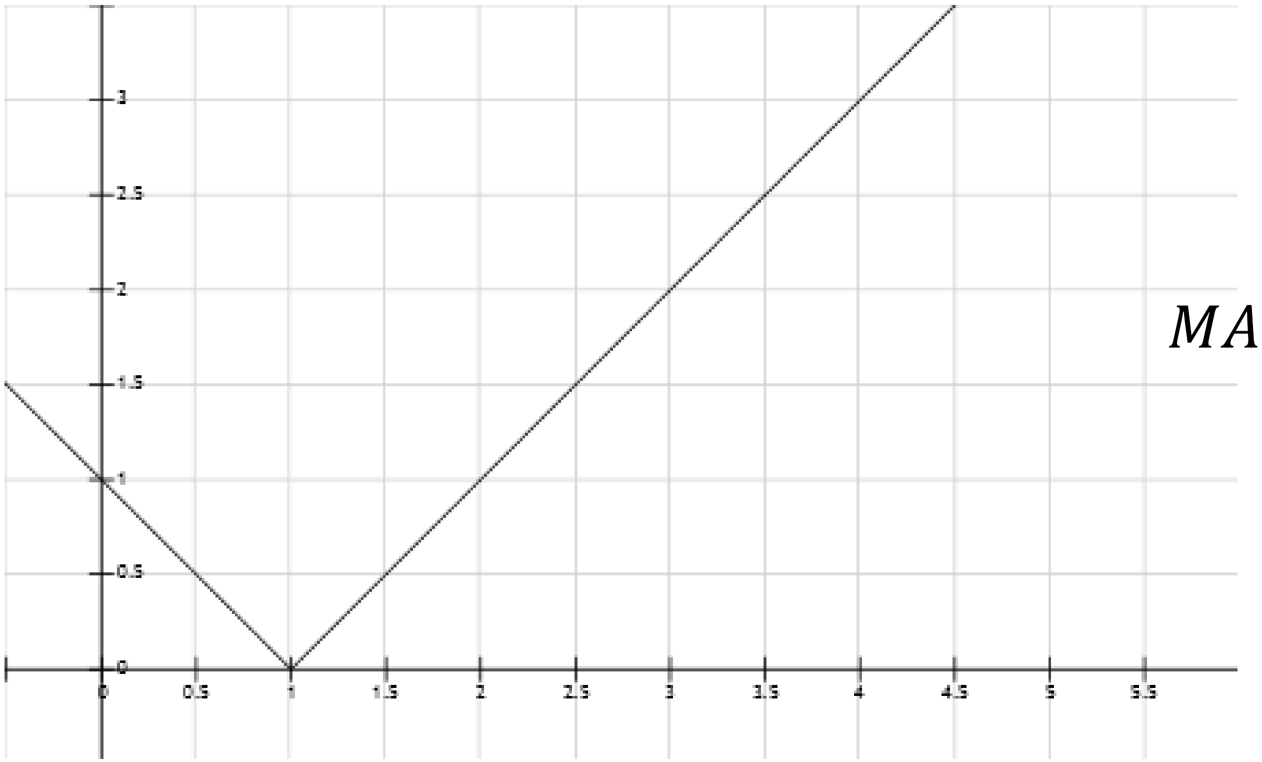
FORECASTING

MSE



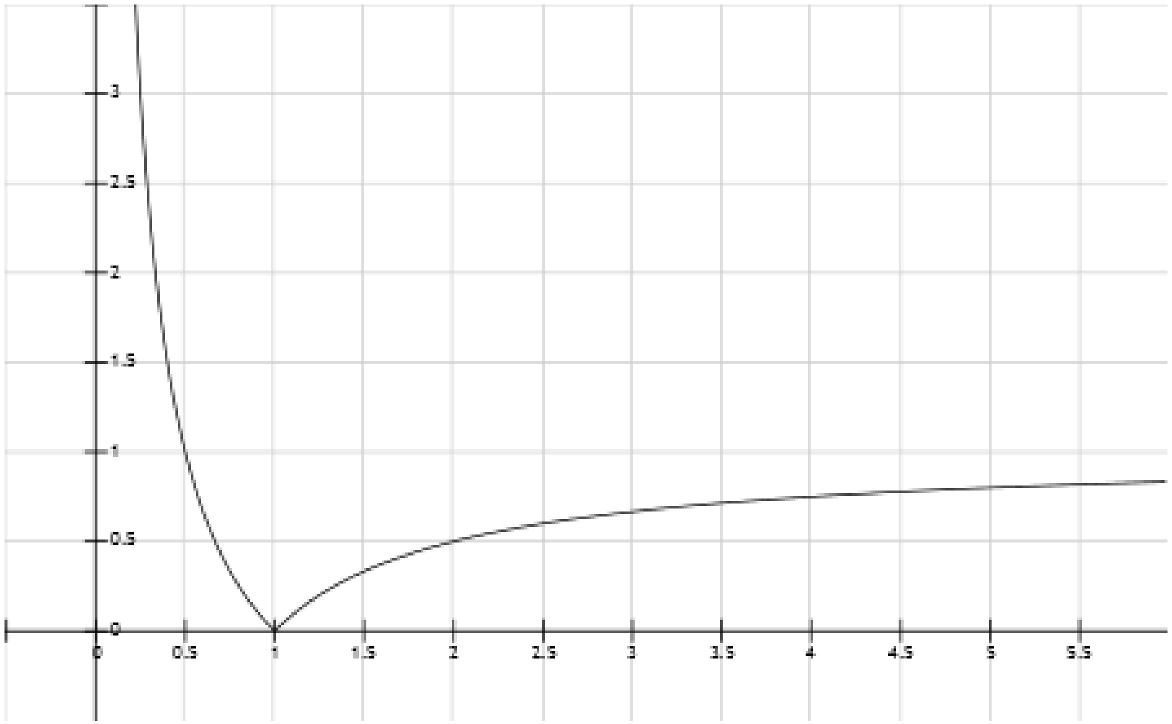
$$MSE = \frac{\sum_{i=1}^{n}(P_i - A_i)^2}{n}$$

MAE



$$MAE = \frac{\sum_{i=1}^{n} |P_i - A_i|}{n}$$

MAPE



$$MAPE = \frac{\sum_{i=1}^{n} \frac{|P_i - A_i|}{A_i}}{n}$$

NMSE

$$NMSE = \frac{MSE\ of\ developed\ model}{MSE\ of\ naive\ model}$$

TIME SERIES

Why time series

- Causal independent variables are
  - Unknown to us
  - Not available
  - Might not fit the data well
  - Difficult to forecast

Typical time series

$$\check{y}_{t+1} = f(y_t, y_{t-1}, y_{t-2} \cdots)$$

$$+ f(x_1, x_2, x_3 \ldots)$$

IMPORTANT CONCEPTS

Autocorrelation (ACF) and PACF

- ACF: nth lag of ACF is the correlation between a day and n days before that
- PACF: The same with all internal correlations are removed

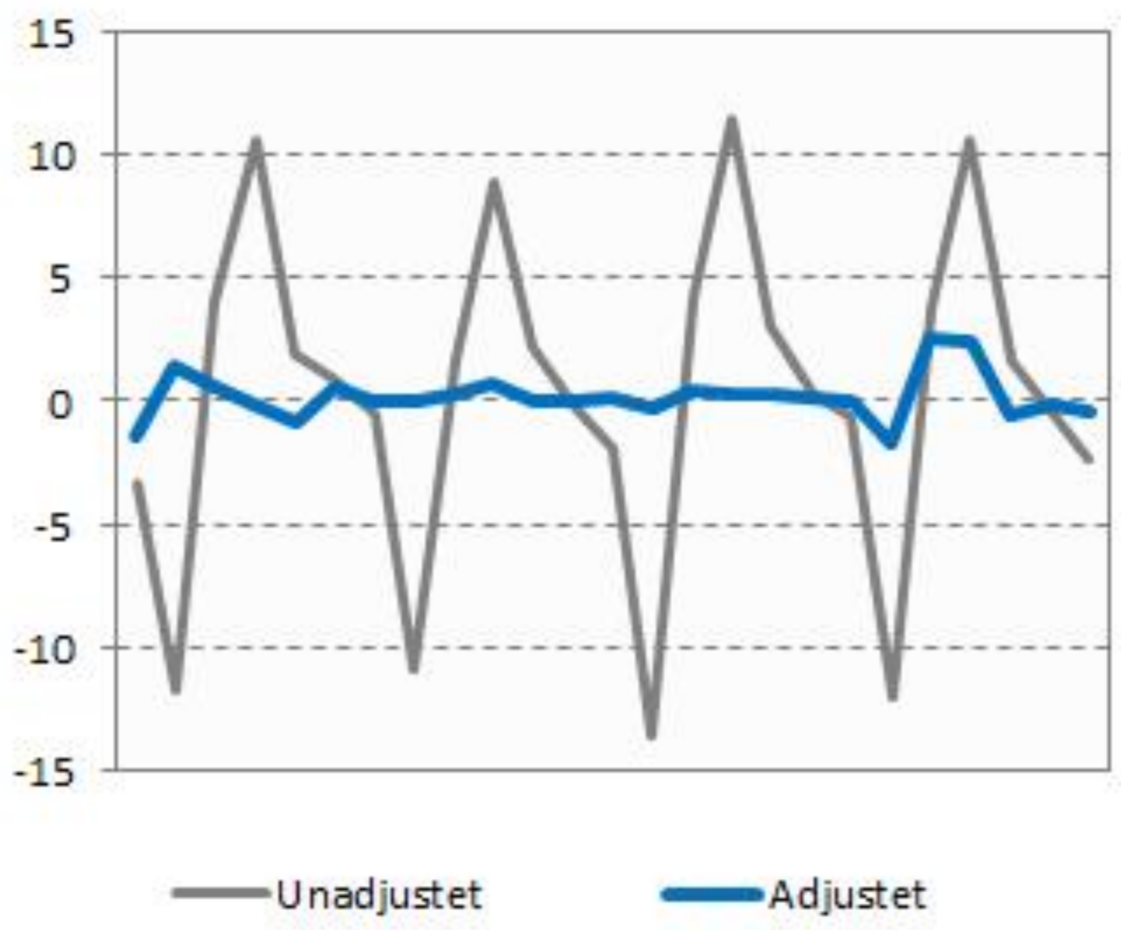
Components of time series

- Trend

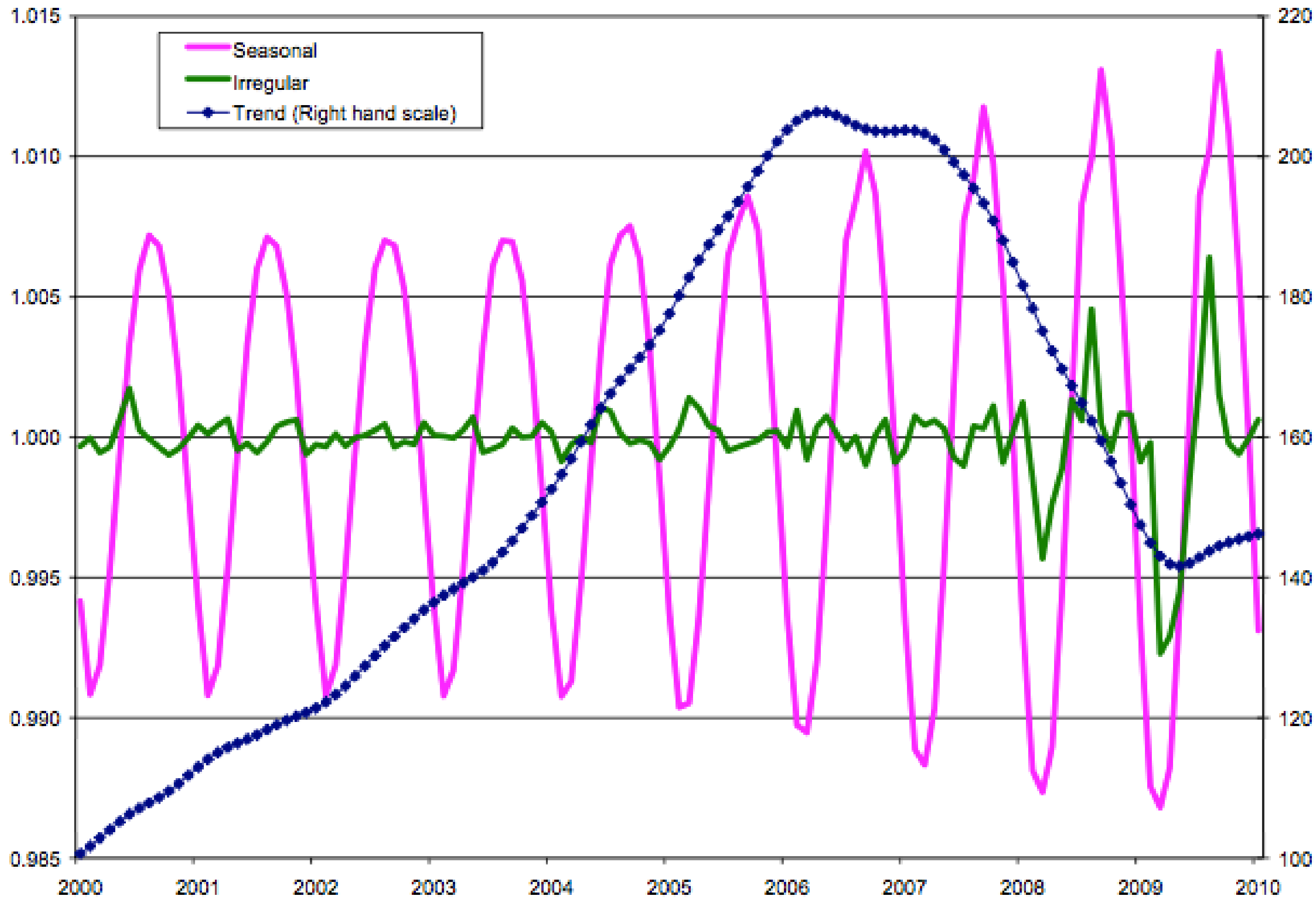
- Seasonality/Cyclicality

- Random component

Seasonality



Trend, seasonality and randomness



CSE 7202c

ACF and PACF for trend, seasonality and randomness

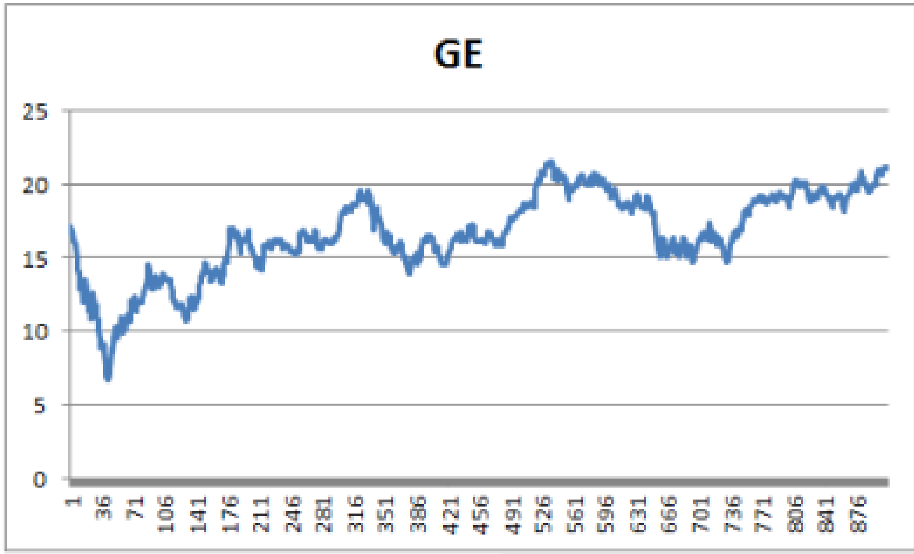
- They vary significantly

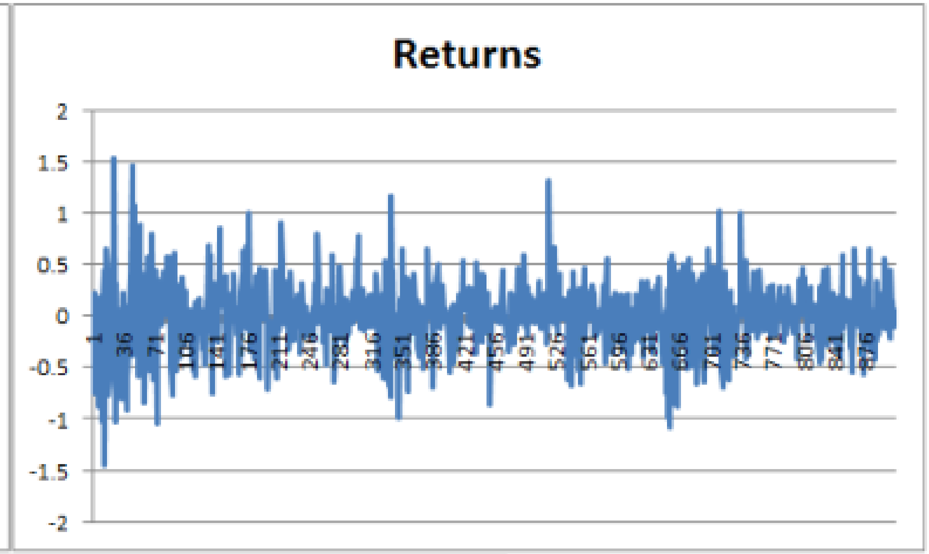
- Let us R!

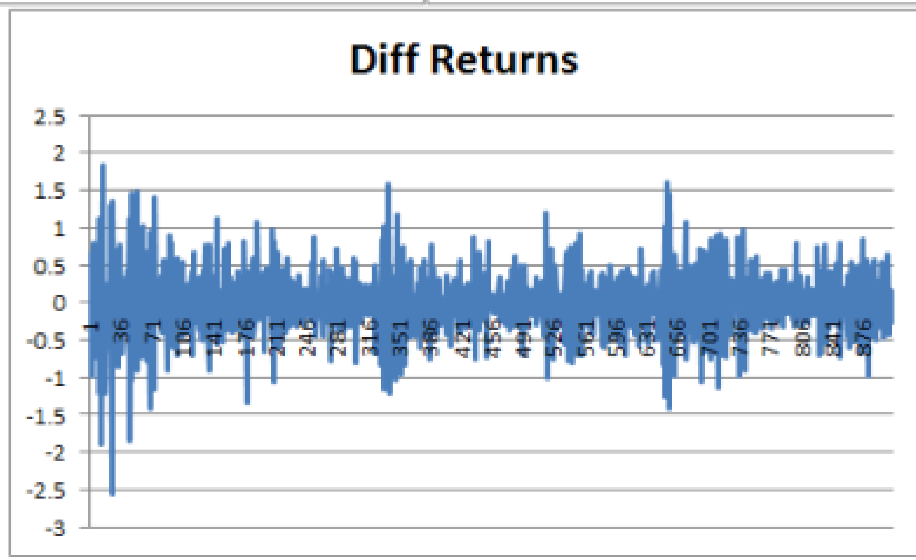
Stationary and Non-Stationary

- Stationary data has a constant mean
- If the data is stationary, forecasting is easier!
- Differencing to convert non-stationary to stationary

Removing trend from data







ACF and PACF of stationary and non-stationary

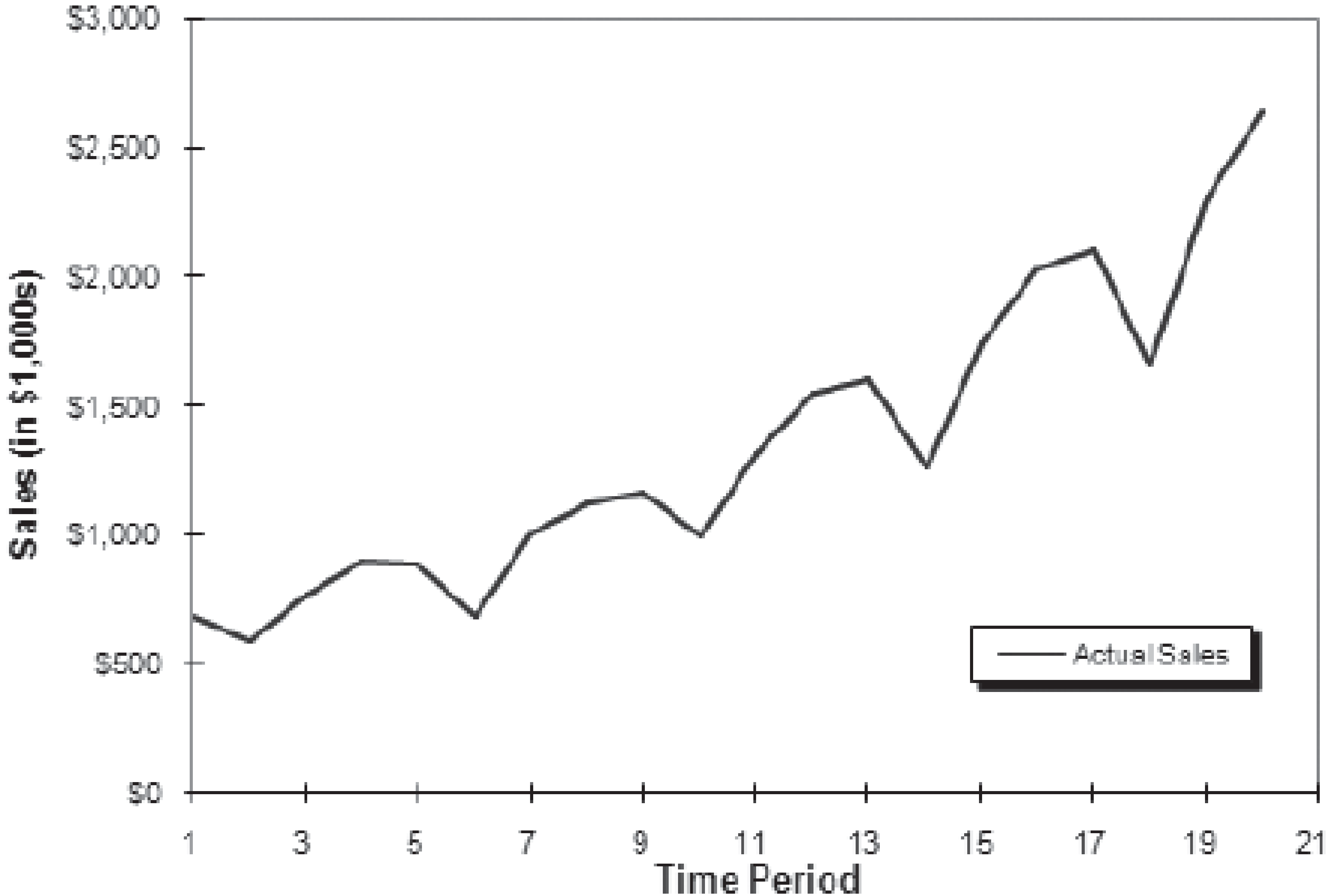
- Non-stationary series have an ACF that remains significant for half a dozen or more lags, rather than quickly declining to zero.
- You must difference such a series until it is stationary before you can identify the process.

A CRUDE WAY OF SOLVING TIME SERIES (CURVE FITTING)

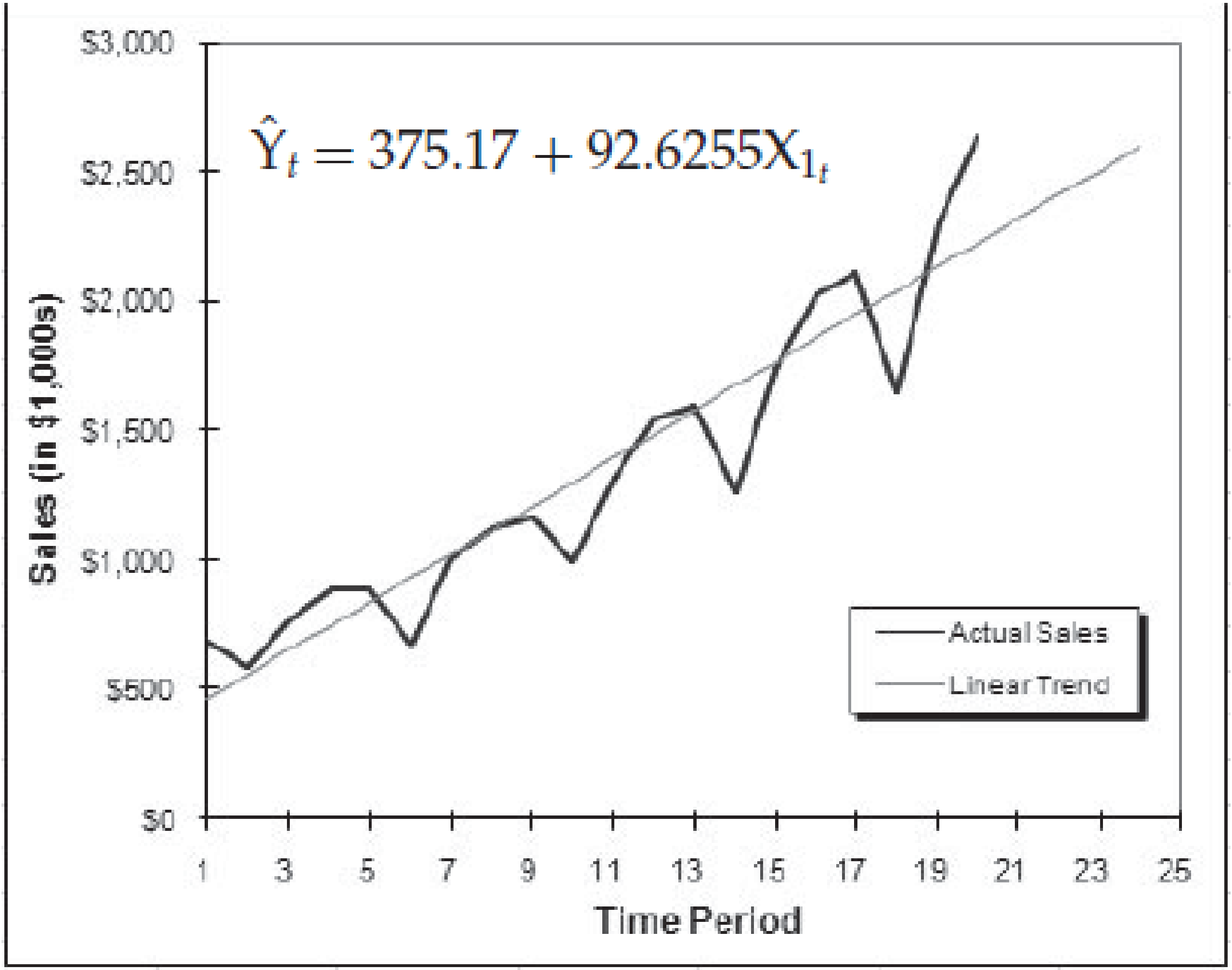
Regression on time

- Use when trend is the most pronounced

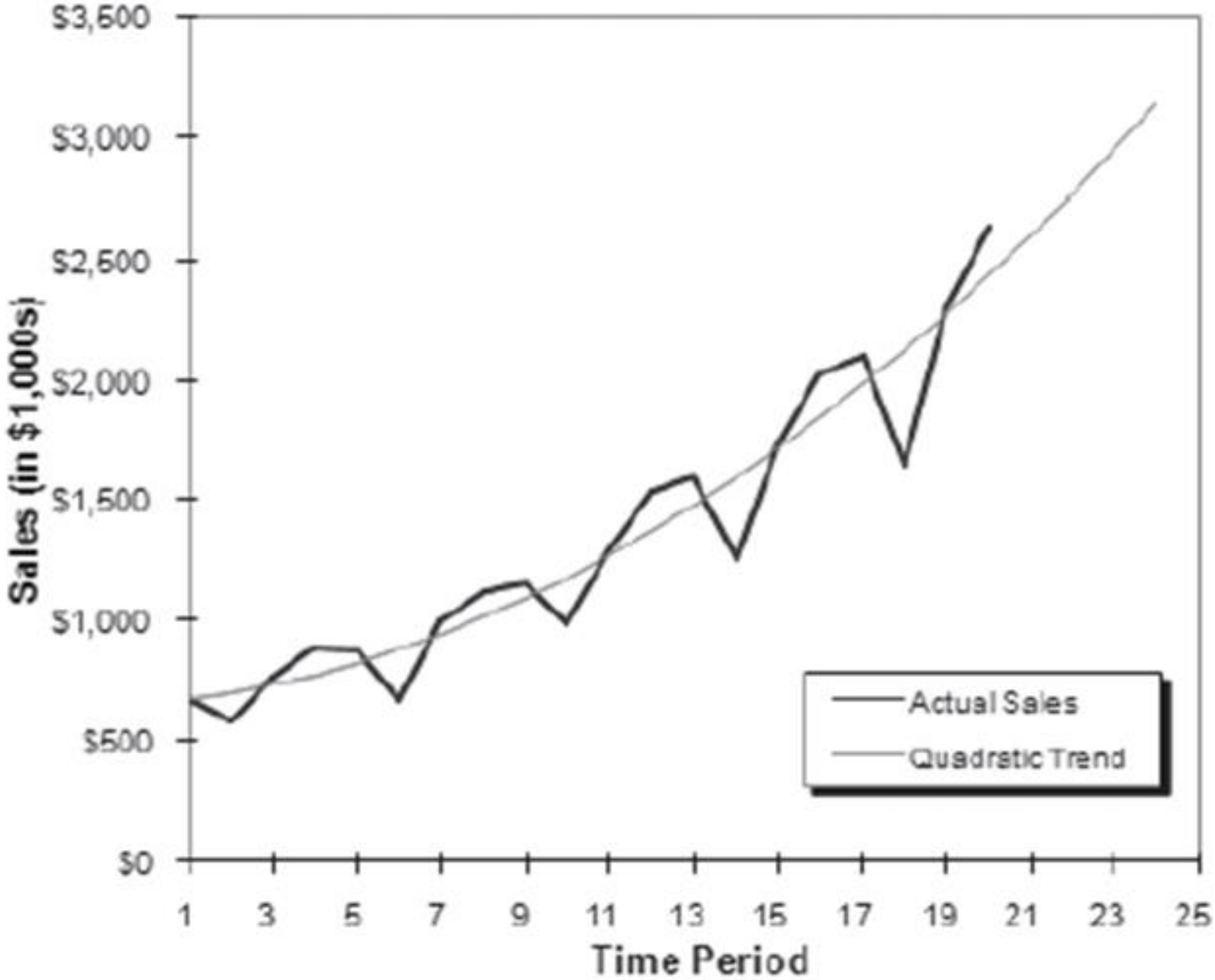
- ACF decays exponentially and PACF has very few spikes


$3,000
$2,500
$2,000
$1,500
$1,000
$500
$0
Sales (in $1,000s)
1
3
5
7
9
11
13
15
17
19
21
Time Period
Actual Sales

Regression analysis



Quadratic trend



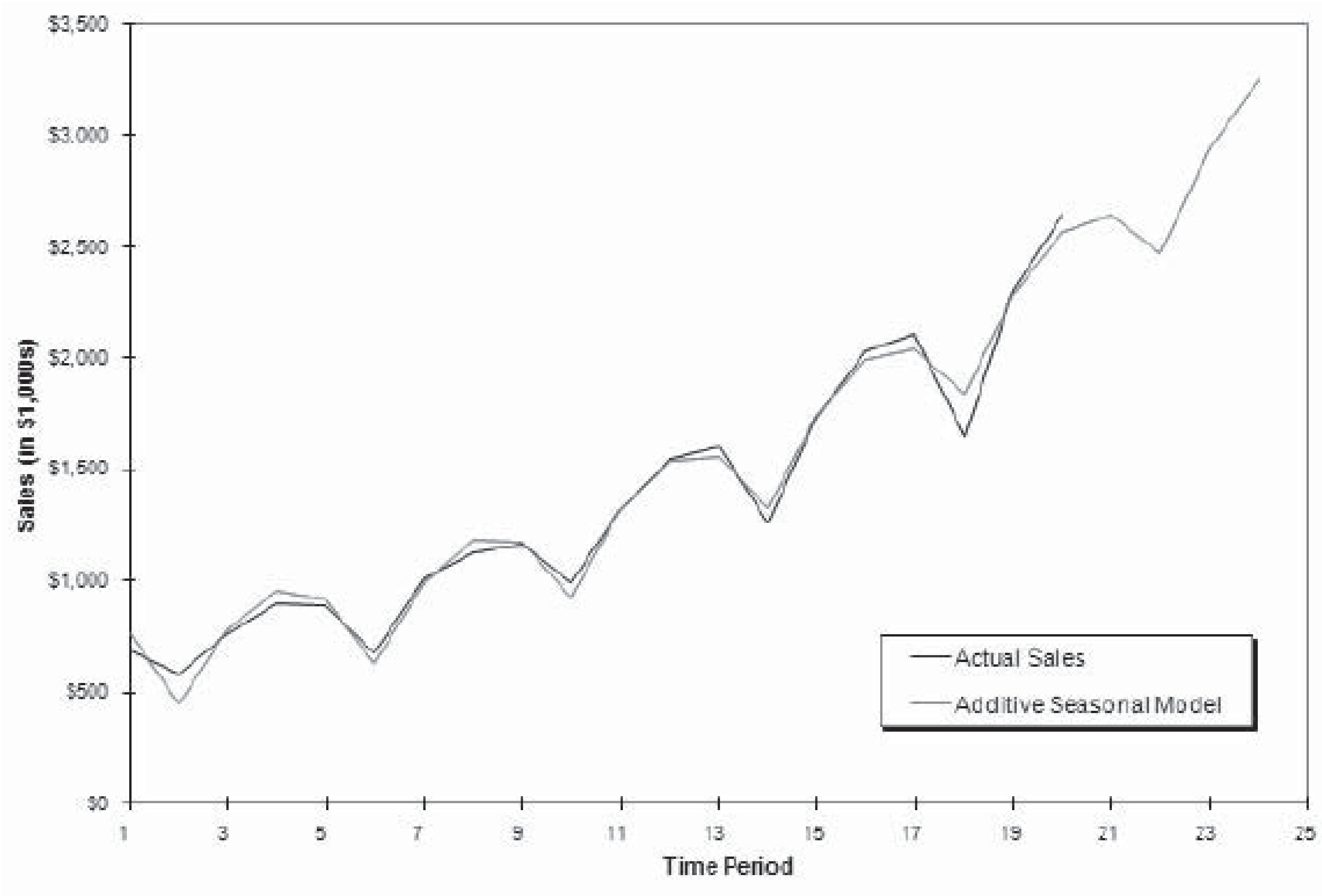
Seasonal regression models

	Value of		
Quarter	X_{3_t}	X_{4_t}	X_{5_t}
1	1	0	0
2	0	1	0
3	0	0	1
4	0	0	0

$$Y_t = \beta_0 + \beta_1 X_{1_t} + \beta_2 X_{2_t} + \beta_3 X_{3_t} + \beta_4 X_{4_t} + \beta_5 X_{5_t} + \varepsilon_t$$

where, $X1_t = t$ and $X2_t = t^2$.

Seasonal regression models



Another crude way of incorporating seasonality

- Take the trend prediction and actual prediction

- Depending on additive or multiplicative model compute the deviation and map it as seasonality effect for each prediction

Case

Year	Quarter	Time variable (this is created)	Revenues
2008	I	1	10.2
	II	2	12.4
	III	3	14.8
	IV	4	15
	I	5	11.2
	II	6	14.3
	III	7	18.4
	IV	8	18

```
Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5595 -0.9384  0.4405  1.3265  1.9286 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  10.0393     1.5531   6.464  0.00065 ***
x             0.9440     0.3076   3.069  0.02196 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 

Residual standard error: 1.993 on 6 degrees of freedom
Multiple R-squared: 0.6109,     Adjusted R-squared: 0.5461 
F-statistic: 9.422 on 1 and 6 DF,  p-value: 0.02196
```

Seasonality: Multiplicative

Time	Observed values TSI (assuming no impact of cyclicality)	Predicted values (per the regression) T	SI = TSI/T
1	10.2	10.983	0.929
2	12.4	11.927	1.040
3	14.8	12.871	1.150
4	15	13.815	1.086
5	11.2	14.759	0.759
6	14.3	15.703	0.911
7	18.4	16.647	1.105
8	18	17.591	1.023

Quarterly seasonality

Time	Average seasonality factor
Q1	0.844
Q2	0.975
Q3	1.127
Q4	1.054

Computations

- Trend $Y_9 = 10.039 + 0.944(9) = 18.535$

- Corrected for seasonality and randomness: $18.535 * 0.844 = 15.643$

Issues with regressing on time

- It is too much of a curve fit For a statistician to sleep well!

- If there is no trend or if seasonality and fluctuations are more important than trend, then the coefficients behave wierdly

TIME SERIES: MORE ROBUST ANALYSES

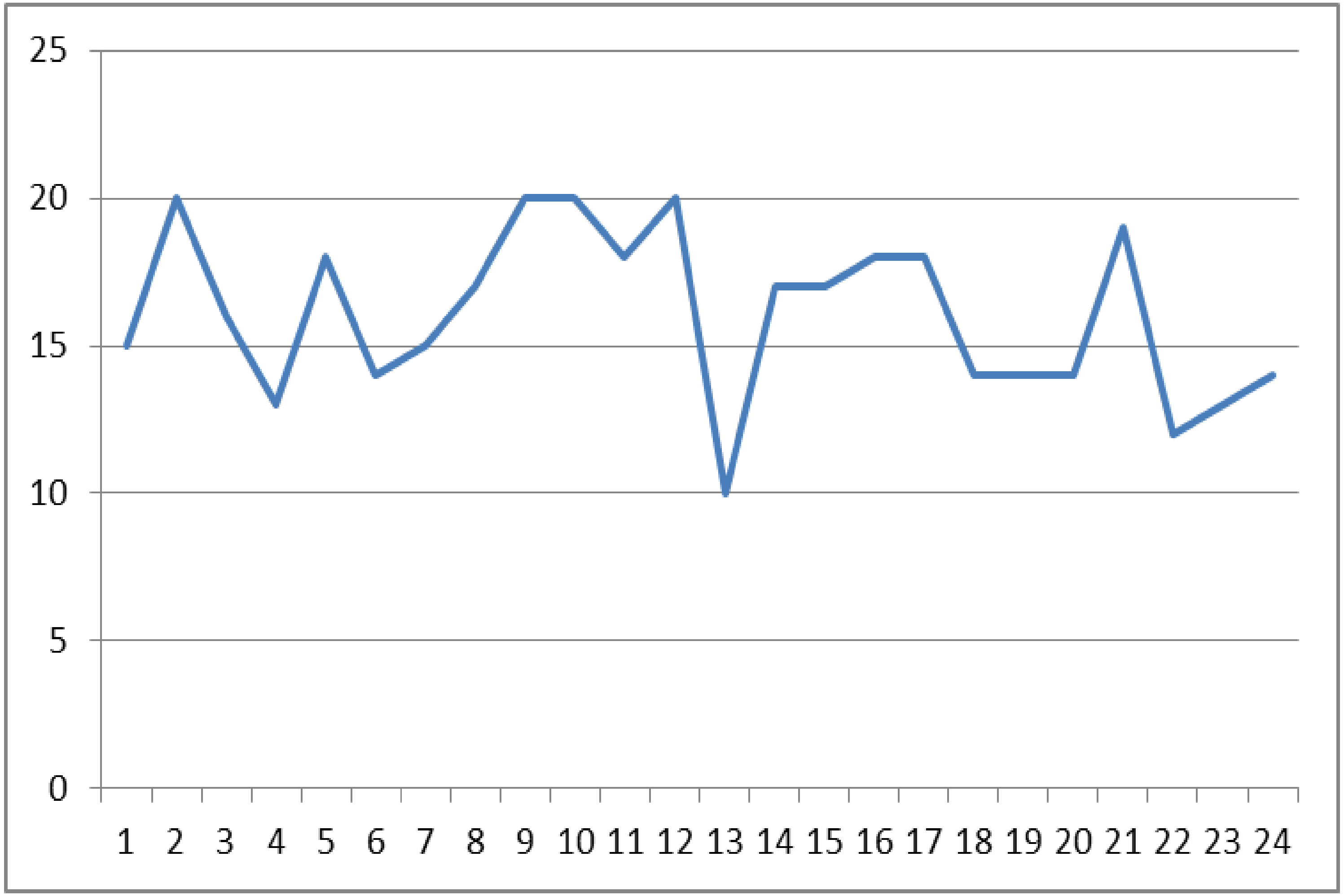
Identifying moving average processes

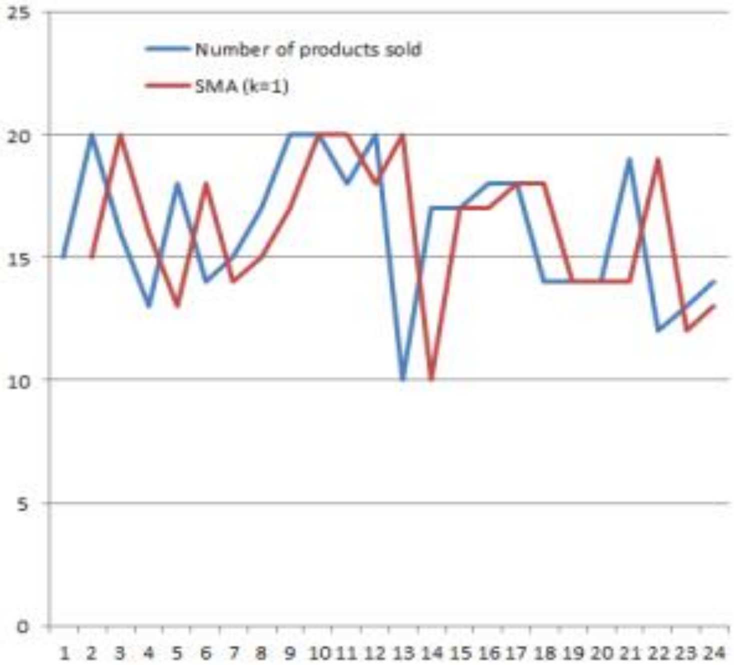
- We use different techniques for different processes
  - Random stationary
  - Seasonal
  - Trend
- First we need to identify them

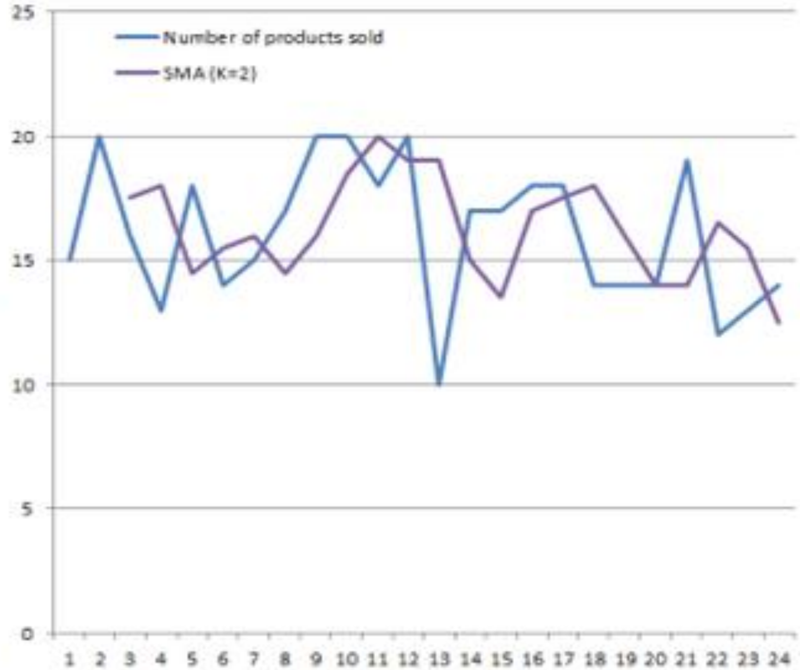
Stationary model: Case 1

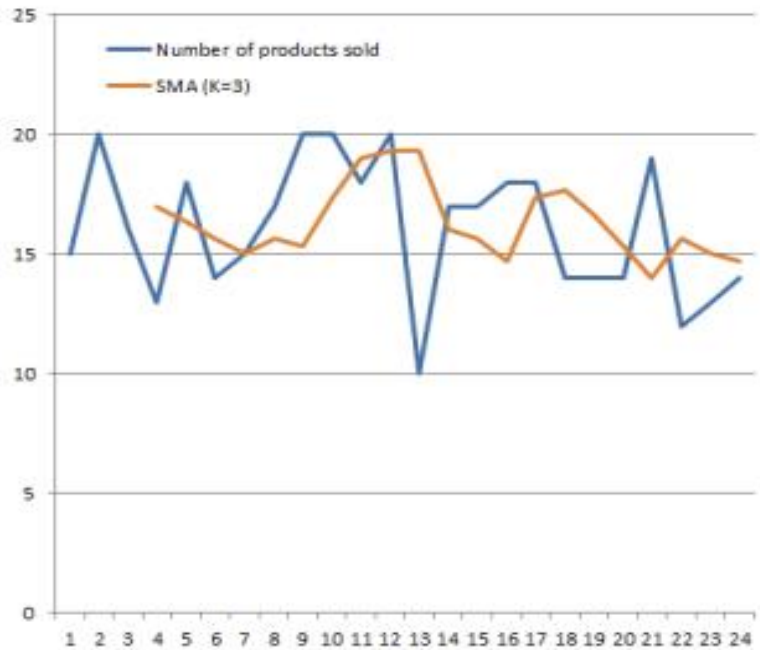
Number of products sold	SMA (k=1)	Error	SMA (K=2)	Error	SMA (K=3)	Error
15						
20	15	5				
16	20	4	17.5	1.5		
13	16	3	18	5	17	4
18	13	5	14.5	3.5	16.333333	1.666667
14	18	4	15.5	1.5	15.666667	1.666667
15	14	1	16	1	15	0
17	15	2	14.5	2.5	15.666667	1.333333
20	17	3	16	4	15.333333	4.666667
20	20	0	18.5	1.5	17.333333	2.666667
18	20	2	20	2	19	1
20	18	2	19	1	19.333333	0.666667
10	20	10	19	9	19.333333	9.333333
17	10	7	15	2	16	1
17	17	0	13.5	3.5	15.666667	1.333333
18	17	1	17	1	14.666667	3.333333
18	18	0	17.5	0.5	17.333333	0.666667
14	18	4	18	4	17.666667	3.666667
14	14	0	16	2	16.666667	2.666667
14	14	0	14	0	15.333333	1.333333
19	14	5	14	5	14	5
12	19	7	16.5	4.5	15.666667	3.666667
13	12	1	15.5	2.5	15	2
14	13	1	12.5	1.5	14.666667	0.666667
		2.913043		**2.681818**		**2.492063**

Moving Averages









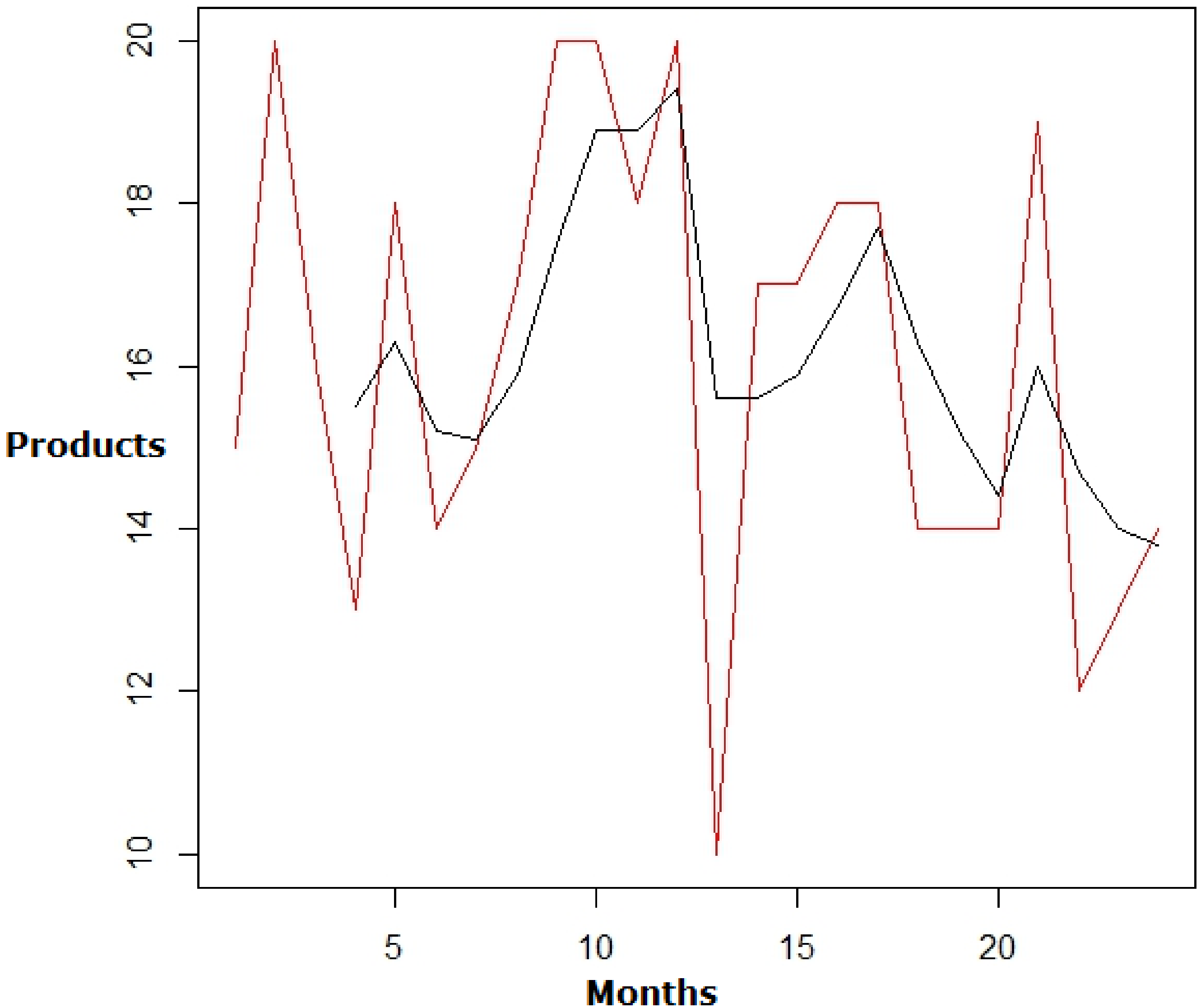
Only decision point is K

Weighted moving average

$$\hat{Y}_{t+1} = w_1 Y_t + w_2 Y_{t-1} + \cdots + w_k Y_{t-k+1}$$

- Typically we choose a time period of moving average and weights are chosen such that the error is minimized

WMA



Exponential smoothing

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t)$$

Above equation indicates that the predicted value for time period t + 1 (y_{t+1}) is equal to the predicted value for the previous period (y_t) plus an adjustment for the error made in predicting the previous period's value ($\alpha(Y_t - y_t)$).

The parameter α can assume any value between 0 and 1 ($0 \leq \alpha \leq 1$).

Exponential smoothing in other ways

$$\widehat{Y_{t+1}} = \widehat{Y_t} + \alpha(Y_t - \widehat{Y_t})$$

$$= \alpha Y_t + (1 - \alpha)\, \widehat{Y_t}$$

$$\widehat{Y_{t+1}} = Y_t - (1 - \alpha)\left(Y_t - \widehat{Y_t}\right)$$

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1 - \alpha) Y_{t-1} + \alpha(1 - \alpha)^2\, Y_{t-2} + \cdots + \alpha(1 - \alpha)^n\, Y_{t-n} + \cdots$$

LET'S EXPLORE HOW ALPHA CHANGES

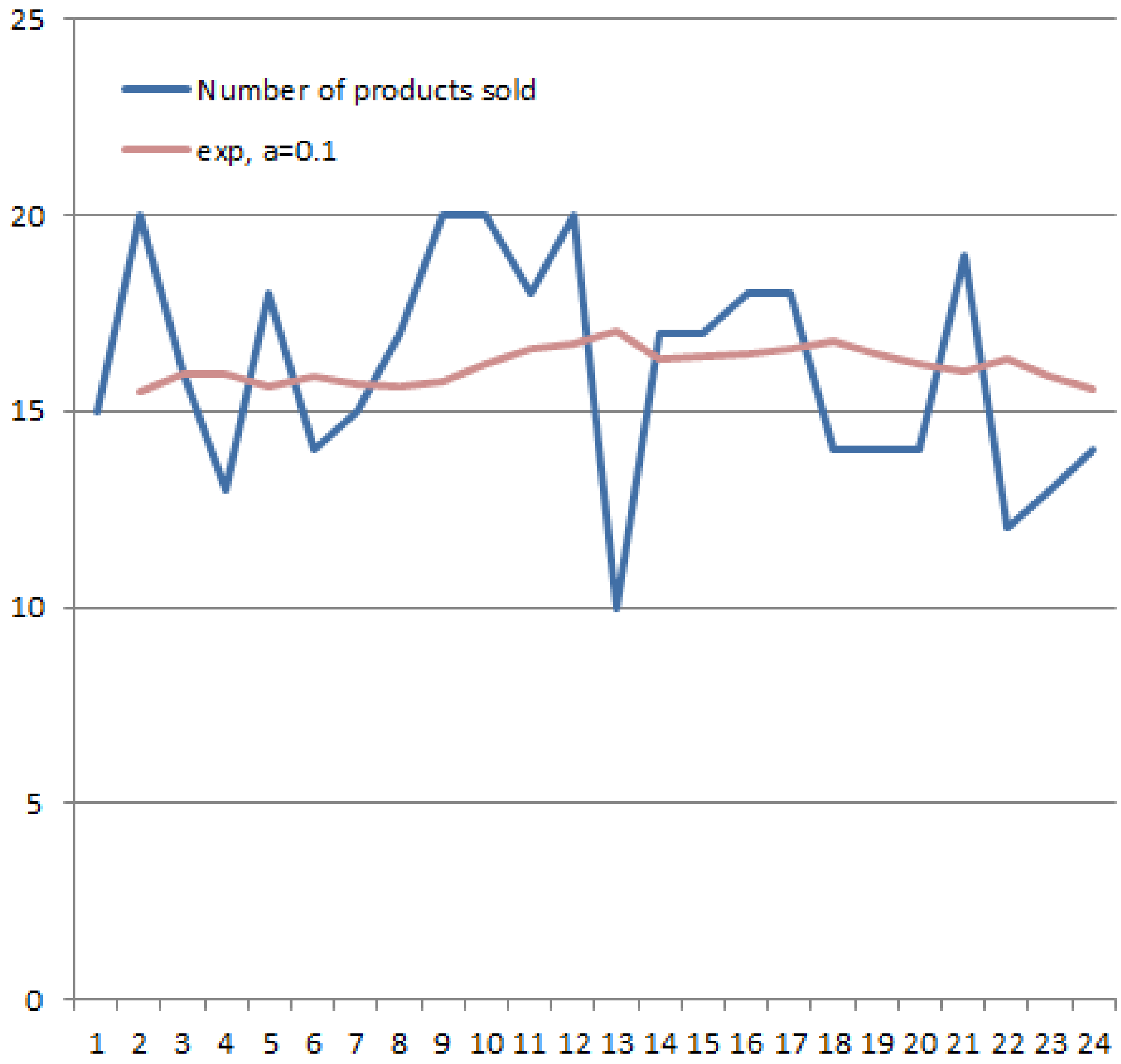
understanding exponential smoothing

- Forecast
  - Interpolation between previous *forecast* and previous *observation*
  - Previous *forecast* plus fraction of previous error
  - Previous *observation* minus fraction 1- of previous error
  - *Exponentially weighted (i.e. discounted)*

Exponential smoothing

- Y at t+1
- Y at t+2
- All future predictions are same! This is in accordance with stationary assumption

EMA


Number of products sold
exp, a=0.1
25
20
15
10
5
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Box–Jenkins methodology

- Model identification and model selection.
- Parameter estimation.
- Model checking
- http://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/The_Box-Jenkins_Method.pdf

Model selection

SHAPE	INDICATED MODEL
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, rest are essentially zero	Moving average model, order identified by where plot becomes zero.
Decay, starting after a few lags	Mixed autoregressive and moving average model.
All zero or close to zero	Data is essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.

In practice

- There are techniques that automate model selection

ARIMA(p,d,q) model

- p is the number of autoregressive terms (a linear regression of the current value of the series against one or more prior values of the series)
- d is the number of non-seasonal differences, (d is the order of the differencing used to make the time series stationary)

- q is the number of lagged forecast errors in the prediction equation. q is the order of the moving average model (a linear regression of the current value of the series against the white noise or random shocks of one or more prior values of the series).

p,d,q

- (0,0,0): without growth

- ARIMA(0,1,0) = random walk with growth;

$$\hat{Y}(t) - Y(t-1) = \mu$$

ARIMA(1,1,0)

- Differenced first-order autoregressive model

$$\hat{Y}(t) - Y(t-1) = \mu + \phi(Y(t-1) - Y(t-2))$$

which can be rearranged to

$$\hat{Y}(t) = \mu + Y(t-1) + \phi(Y(t-1) - Y(t-2))$$

ARIMA(0,1,1)

- Without constant = simple exponential smoothing

$$\widehat{Y_t} - Y_{t-1} = -\theta e_{t-1}$$

A "mixed" model-- ARIMA(1,1,1)

$$\hat{Y}(t) = \mu + Y(t-1) + \phi\,(Y(t-1) - Y(t-2)) - \theta\, e(t-1)$$

Time series Ensemblers

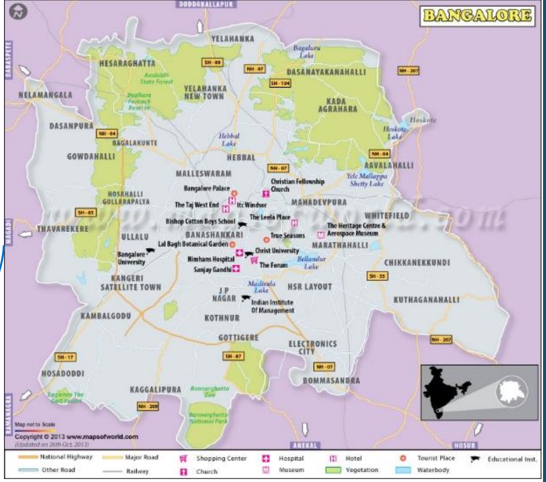
$$\hat{Y}_t = b_0 + b_1 F_{1_t} + b_2 F_{2_t} + b_3 F_{3_t}$$

ARIMAX

- Causal + Time series analysis

INSOFE

Inspire...Educate...Transform.



HYDERABAD

Office and Classrooms

Plot 63/A, Floors 1&2, Road # 13, Film Nagar,
Jubilee Hills, Hyderabad - 500 033
+91-9701685511 (Individuals)
+91-9618483483 (Corporates)

Social Media

Web:	http://www.insofe.edu.in
Facebook:	https://www.facebook.com/insofe
Twitter:	https://twitter.com/Insofeedu
YouTube:	http://www.youtube.com/InsofeVideos
SlideShare:	http://www.slideshare.net/INSOFE
LinkedIn:	http://www.linkedin.com/company/international-school-of-engineering

BENGALURU

Office

Incubex, #728, Grace Platina, 4th Floor, CMH Road,
Indira Nagar, 1st Stage, Bengaluru – 560038
+91-9502334561 (Individuals)
+91-9502799088 (Corporates)

Classroom

KnowledgeHut Solutions Pvt. Ltd., Reliable Plaza,
Jakkasandra Main Road, Teacher's Colony, 14th Main
Road, Sector – 5, HSR Layout, Bengaluru - 560102