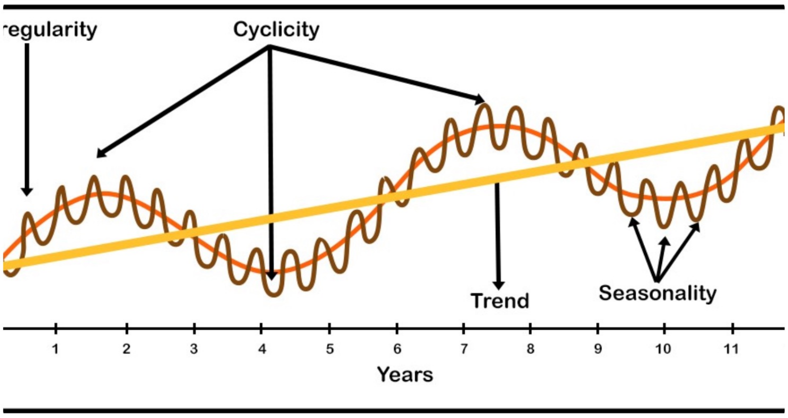
Time Series Patterns

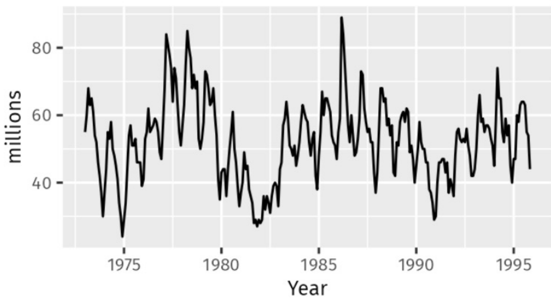
1



1

Time Series

2



Strong seasonality/seasonal behavior within each year
Strong cyclic behavior
No apparent trend

2

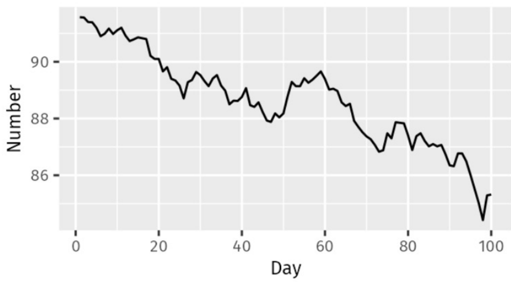
Time Series

3

3

Time Series

4

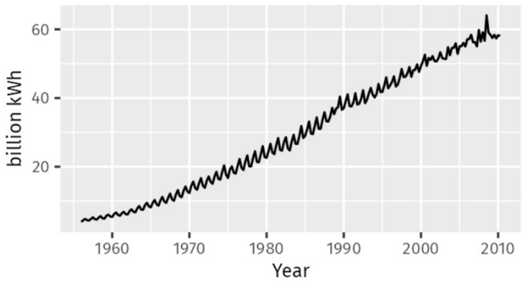


No seasonality
Downward trend

4

Time Series

5

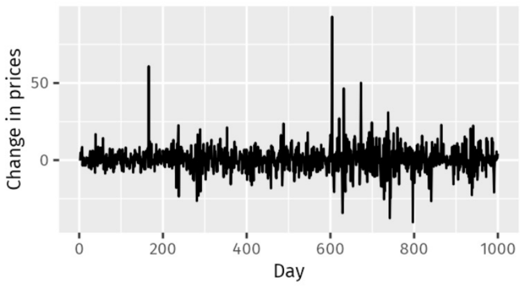


Strong upward/increasing trend
Strong seasonality
No cyclic behavior

5

Time Series

6

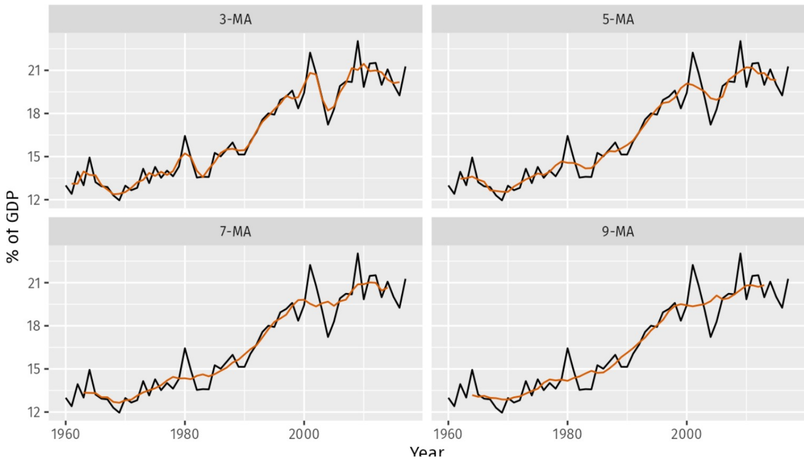


No trend, seasonality or cyclic behaviour
Appears to be random fluctuations

6

Time Series: Moving Averages

7

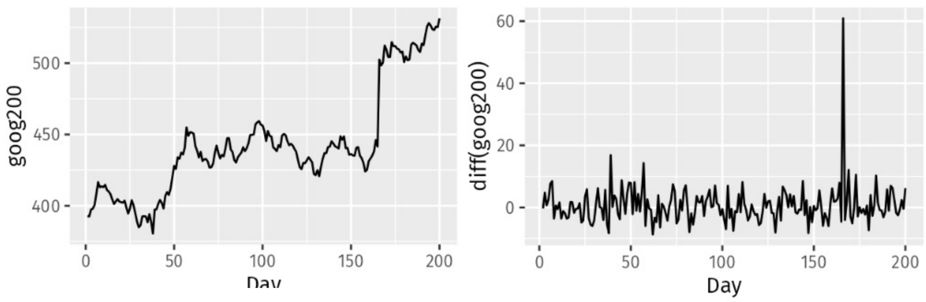


7

Stationarity and Differencing

8

An effective way to stabilize the mean across time



8

Stationarity and Differencing

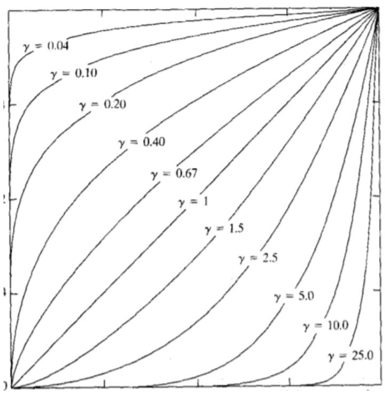
Second-order differencing

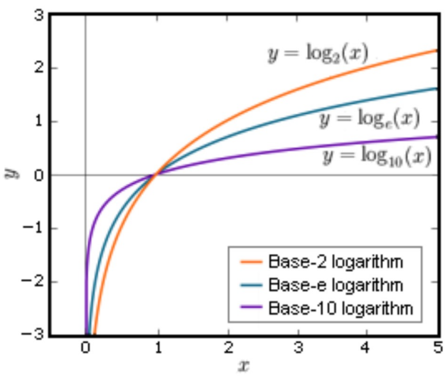
$$\begin{aligned} y''_t &= y'_t - y'_{t-1} \\ &= (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \end{aligned}$$

In this case, we would model the change in the changes of the original data.

Stationarity Through Power Transformation

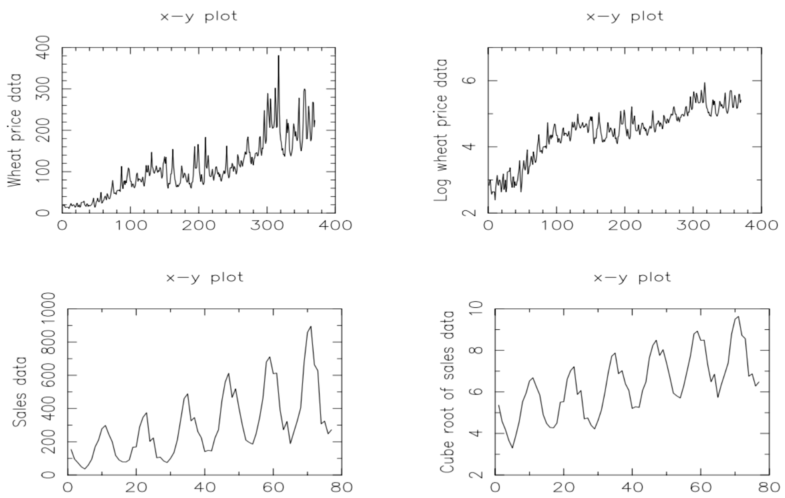
An effective way to stabilize the variance across time is to apply a power transformation (square root, cube root, etc)





Stationarity Through Power Transformation

11

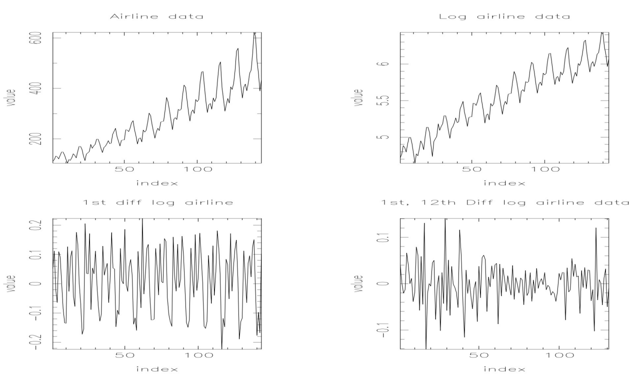


11

Stationarity Through Log Transformation and Differencing

12

Log to stabilize the variance, the 1st difference of the log to eliminate the linear trend, the 12th difference of the 1st difference of the log to eliminate the seasonality



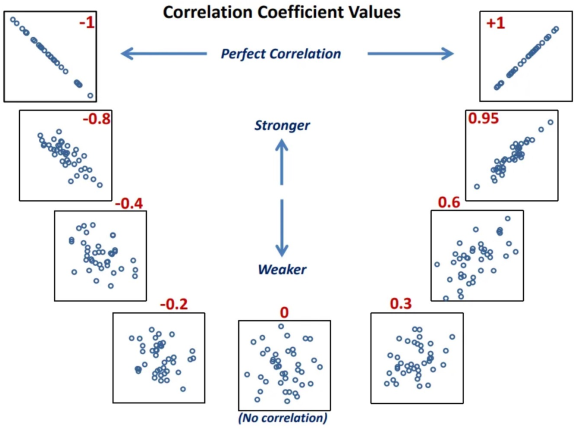
Seasonal Differencing

12

Correlation

13

Pearson Correlation Coefficient



13

Autocorrelation

14

Also known as Serial Correlation/Lagged Correlation

Measures the relationship between a variable's current values and its historical/past values

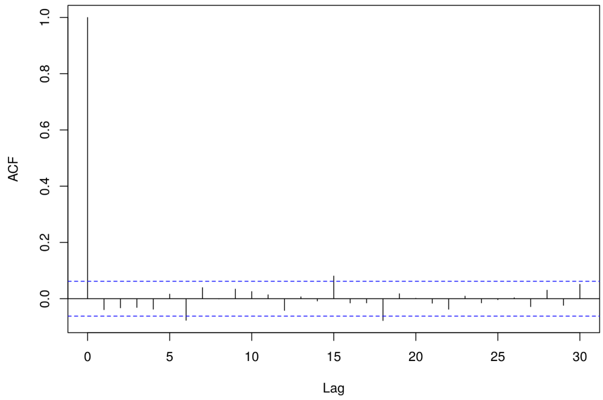
$$r_k = \frac{\sum_{t=k+1}^{T} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{T} (y_t - \bar{y})^2}$$

t	x_t	x_{t-1} (lag 1 value)
1	13	*
2	14	13
3	8	14
4	10	8
5	16	10

14

White Noise ACF

15



It provides a confirmation that we have eliminated any remaining correlation from the residuals *and thus have a good model fit*.

15

Partial Autocorrelation

16

Suppose we are regressing a variable Y on other variables X1, X2, and X3, the partial correlation between Y and X3 is the amount of correlation between Y and X3 that is not explained by their common correlations with X1 and X2.

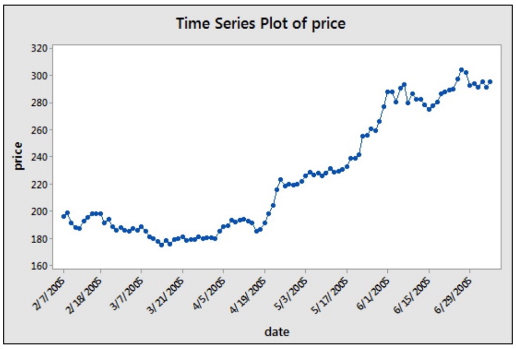
$$\frac{\text{Covariance}(y, x_3|x_1, x_2)}{\sqrt{\text{Variance}(y|x_1, x_2)\text{Variance}(x_3|x_1, x_2)}}$$

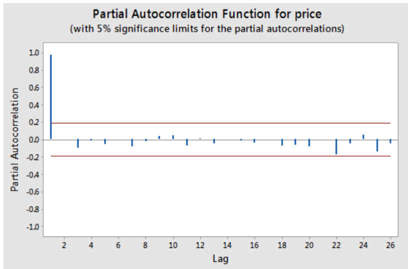
PAC measures correlation between X_s and X_t with the effect of “everything in the middle” removed

16

Partial Autocorrelation

17

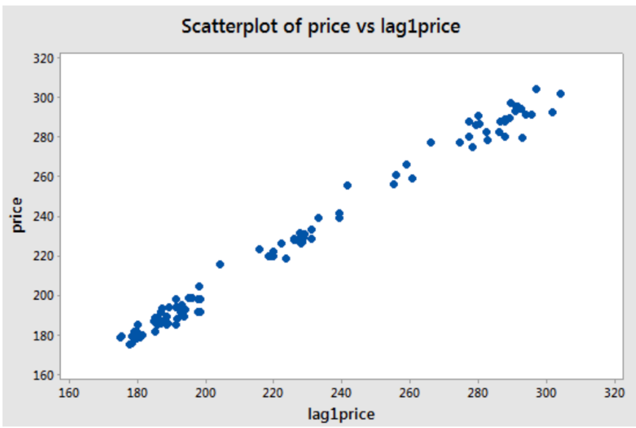




17

Partial Autocorrelation

18

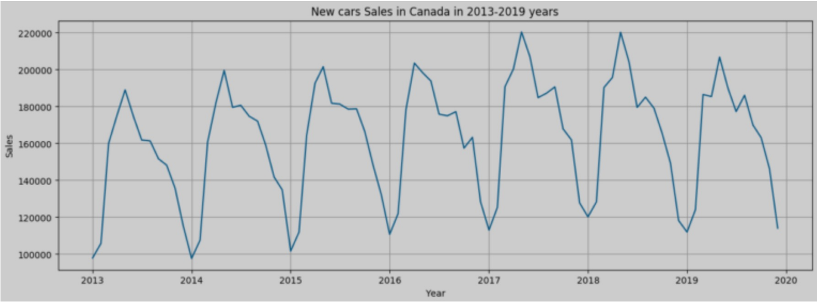


18

AR Model

19

- A specific type of regression model
  - Predicts a value for a variable based on past values of the same variable
  - Current observation is regressed using a set of previous observations (weighted sum of its own lagged values)



19

AR Model

20

- A specific type of regression model
  - Predicts a value for a variable based on past values of the same variable

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + w_t$$

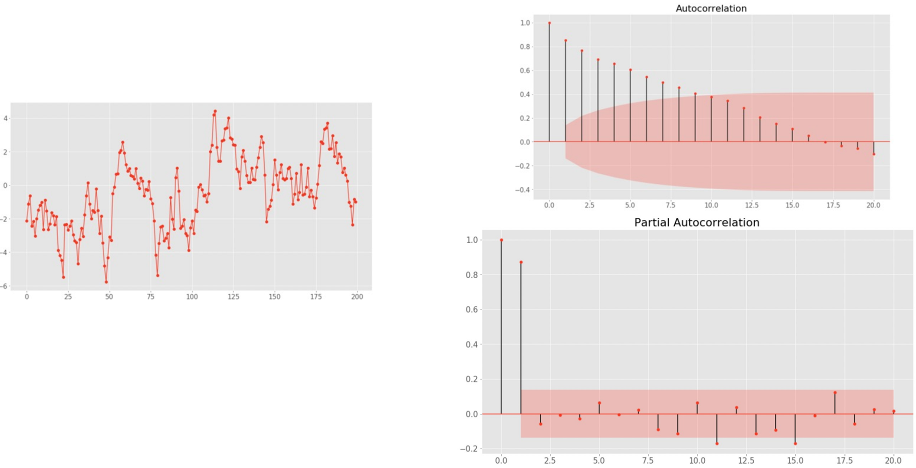
  - An important property of AR(p) models: For $k > p$, theoretical partial autocorrelation function is 0
    - Identification of an AR model is best done with PACF

20

AR(1) Model: Simulated Examples

21

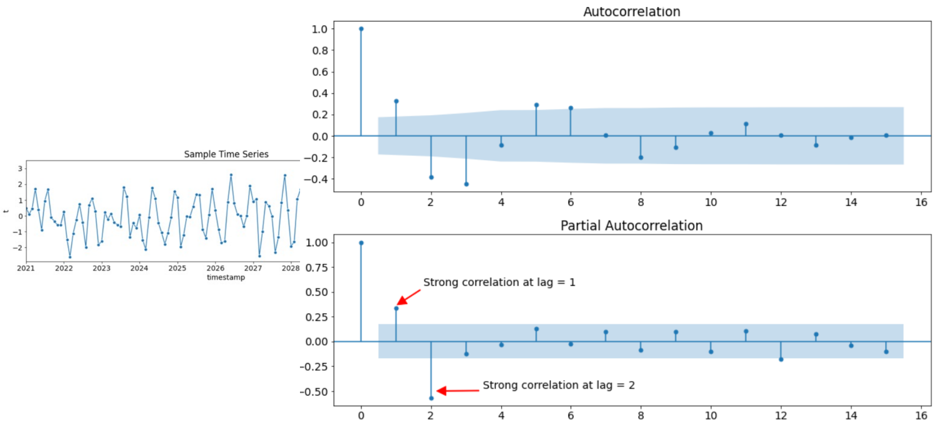
$$X_t = 0.9X_{t-1} + w_t$$



21

AR(2) Model: Simulated Examples

22



22

MA Model: MA(1)

23

Index (t)	$\hat{y}_t$	ε_t	y_t
1	100	4	104
2	102	-2	100
3	99	2	101
4	101	6	107
5	103	3	106
6	101.5	2.5	104

Prediction: $\hat{y}_t = \phi_0 + \phi_1 \varepsilon_{t-1}$

MA(2): $\hat{y}_t = \phi_0 + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2}$

$y_t = \phi_0 + \phi_1 \varepsilon_{t-1} + \phi_2 \varepsilon_{t-2} + \varepsilon_t$

23

MA Model

24

- MA model describes a time series by weighted sum of lagged residuals

- An important property of MA(q) models
- Nonzero autocorrelations for the first q lags, and zero correlations for all lags $k > q$

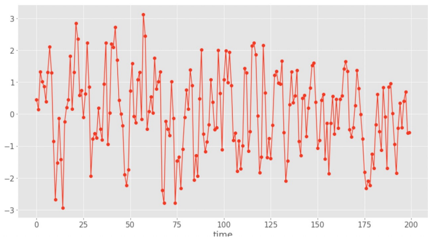
- Identification of an MA model is best done with ACF rather than PACF

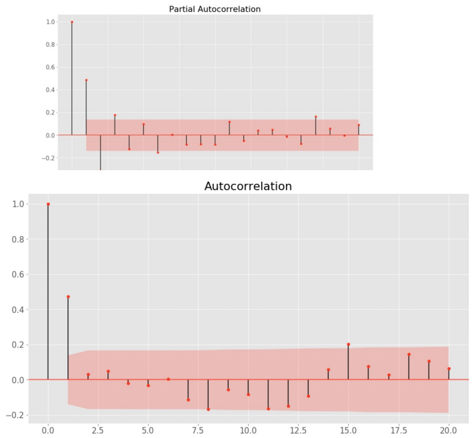
24

MA(1) Model: Simulated Examples

25

$$X_t = w_t + 0.8 \times w_{t-1}$$



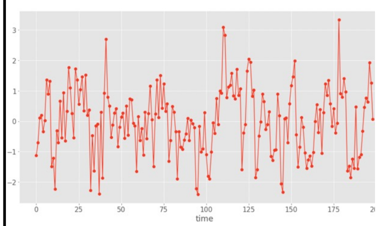


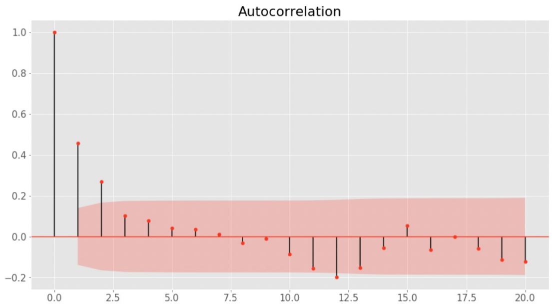
25

MA(2) Model: Simulated Examples

26

$$X_t = w_t + 0.5 \times w_{t-1} + 0.3 \times w_{t-2}$$



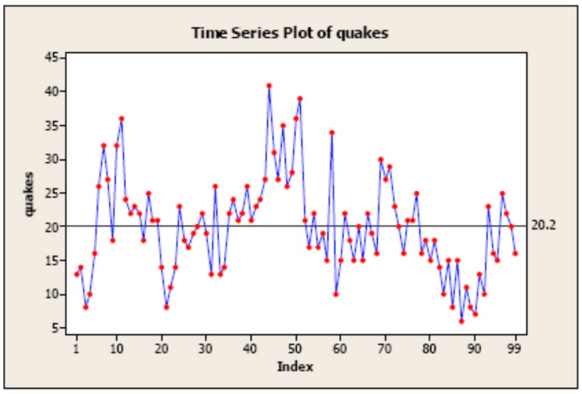


26

Time Series Modeling

27

Annual number of earthquakes in the world with seismic magnitude over 7.0

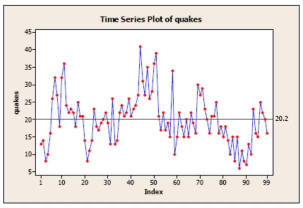


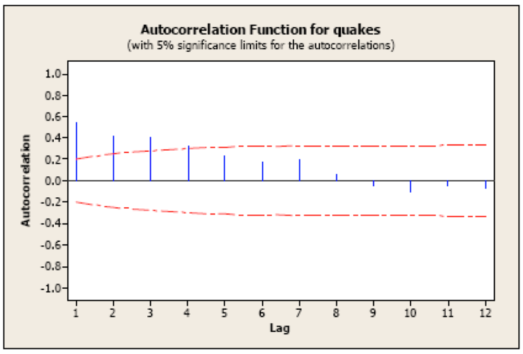
27

Time Series Modeling

28

Annual number of earthquakes in the world with seismic magnitude over 7.0



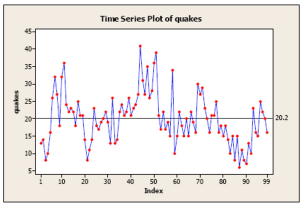


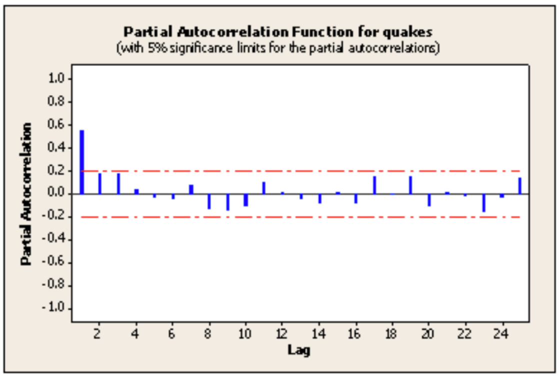
28

Time Series Modeling

29

Annual number of earthquakes in the world with seismic magnitude over 7.0





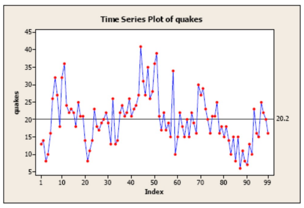
Suggests a possible AR(1) model for this time series data

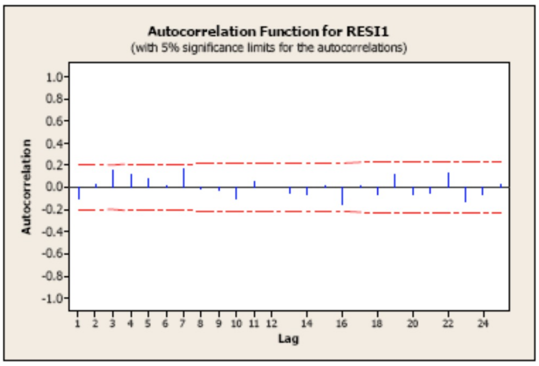
29

Time Series Modeling

30

Annual number of earthquakes in the world with seismic magnitude over 7.0





ACF of the residuals; nothing is significant

30

ARMA Model

31

- A combination of Autoregressive and Moving Average models

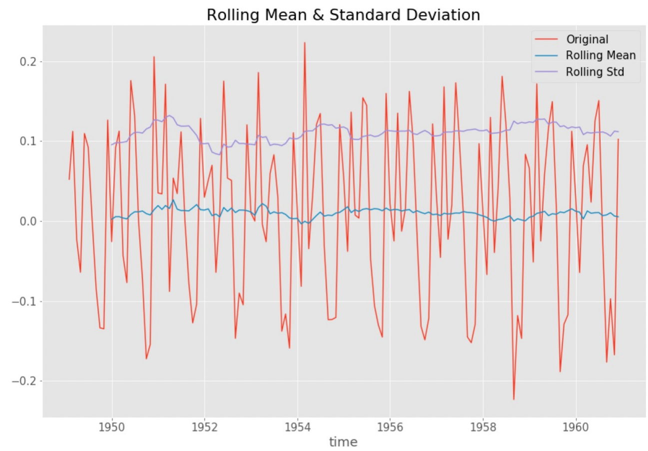
$$X_t = w_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j w_{t-j}$$

	$AR(p)$	$MA(q)$	$ARMA(p, q)$
ACF	Tails off	Cuts off after lag q	Tails off
PACF	Cuts off after lag p	Tails off	Tails off

31

ARIMA Model

32

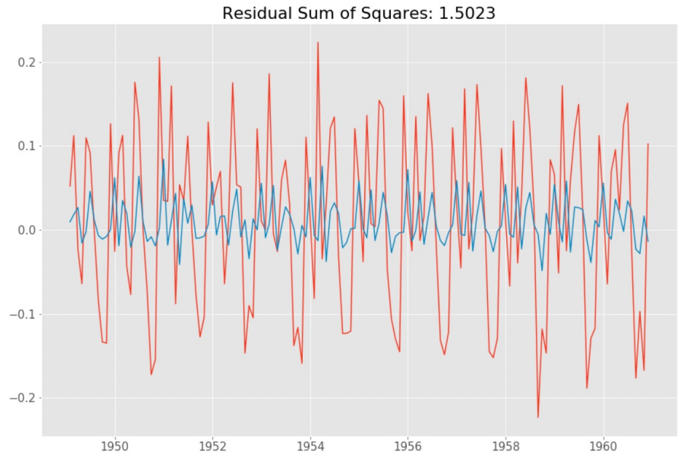


First order differencing over logarithm of passengers

32

ARIMA(2,0,0) Model

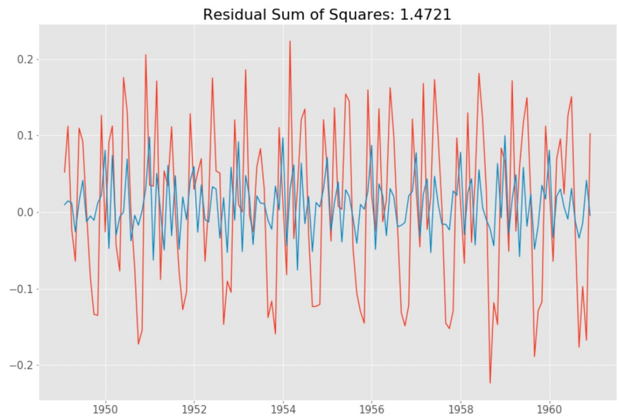
33



33

ARIMA(0,0,2) Model

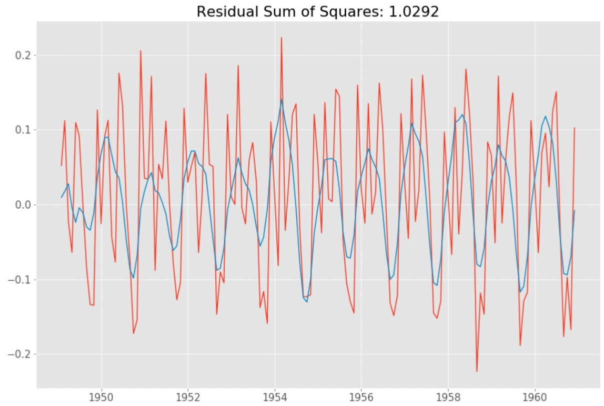
34



34

ARIMA(2,0,2) Model

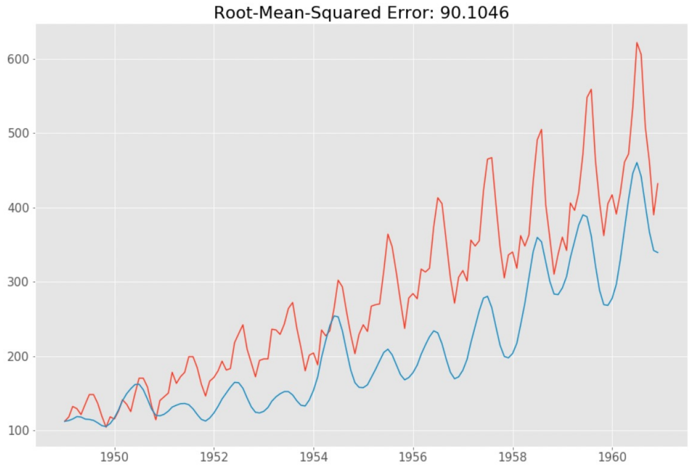
35



35

ARIMA(2,0,2) Model with Inverse Transformation

36



36

Modelling Procedure: Summary

37

- Plot the data and identify any patterns
- If necessary, transform the data to stabilize the variance
- If the data is non-stationary, take first differences of the data until the data are stationary
- Examine the ACF and PACF: Is an ARIMA(p,d,0) or ARIMA(0,d,q) model appropriate?
- Try your chosen model(s), and use the AIC to search for a better model
- Check the residuals from your chosen model by plotting the ACF of the residuals. If they do not look like white noise, try a modified model
- Once the residuals look like white noise, calculate forecasts

- For steps 3-5, use Auto ARIMA algorithm

  Uses a stepwise search to traverse the model space and identifies the optimal model

37

Information Criteria

38

- A measure of quality of a statistical model, which attempts to accurately represent the process that generated the time series
- Information Criteria estimates the information lost by a model
- Akaike information criterion (AIC)
- Bayesian Information Criterion (BIC)
- They consist of a goodness-of-fit term plus a penalty to control over-fitting

38

Thank You!

39