Lecture-19-20
Course: Applied Data Science

Autoregressive Integrated Moving Average (ARIMA)

By
Dr. Sibarama Panigrahi
Assistant Professor, Department of Computer Sc. & Engineering
National Institute of Technology, Rourkela, Odisha, 769008, India
Mobile No.: +91-7377302566
Email: panigrahis[at]nitrkl[dot]ac[dot]in
panigrahi[dot]sibarama[at]gmail[dot]com

Stationarity

– In the most intuitive sense, ***stationarity means that the statistical properties*** of a process generating a time series do not change over time.

– *It does not mean that the series does not change over time*, **just that the *way* it changes does not itself change over time.**

– **The algebraic equivalent is thus a linear function, perhaps, and not a constant one;**
  - the value of a linear function changes as x grows, but the *way* it changes remains constant
  - it has a constant slope; one value that captures that rate of change.

Why Stationarity is Important

- Stationary processes are easier to analyze.
- The stationary processes are a sub-class of a wider family of possible models of reality.
- This sub-class is much easier to model and investigate.
- It hints that such processes should be possible to predict, as the *way* they change is predictable.

Strong Stationarity

- It requires the shift-invariance (in time) of the finite-dimensional distributions of a stochastic process.
- This means that the distribution of a finite sub-sequence of random variables of the stochastic process remains the same as we shift it along the time index axis.

Formally, the discrete stochastic process $\boldsymbol{X}=\{x_i\ ;\ i\in\mathbb{Z}\}$ is *stationary* if

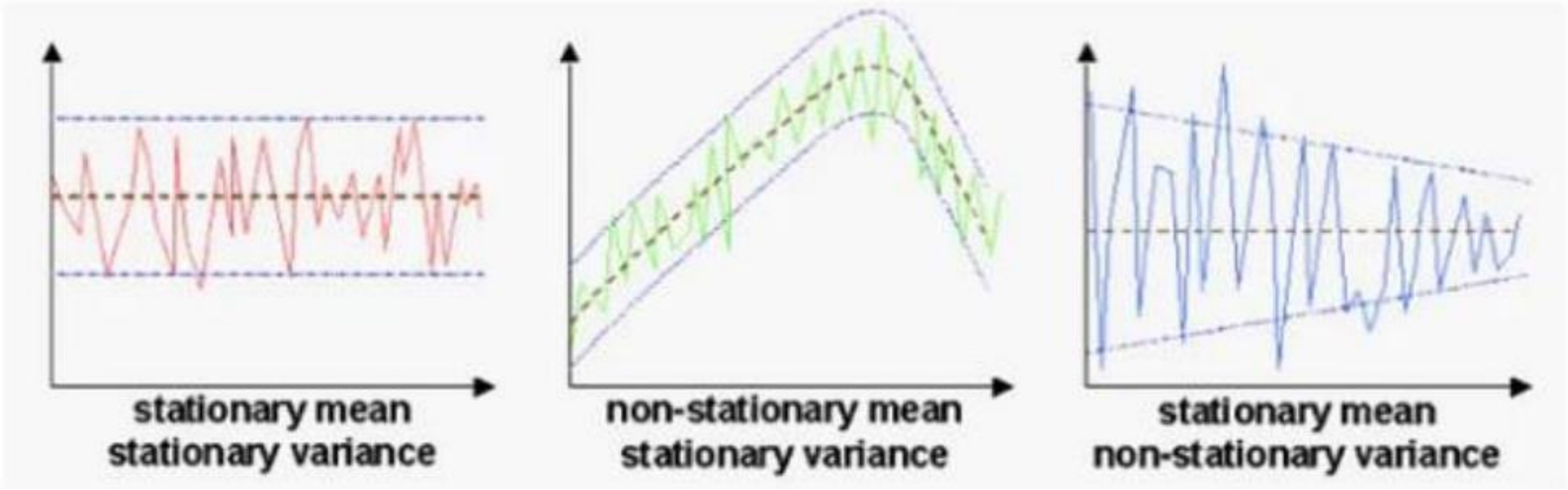
$$F_X(x_{t_{1+\tau}}, \ldots, x_{t_{n+\tau}}) = F_X(x_{t_1}, \ldots, x_{t_n})$$

Weak Stationarity

– Weak stationarity only requires the shift-invariance (in time) of the first moment and the cross moment (the auto-covariance).

– This means the process has the same mean at all time points, and that the covariance between the values at any two time points, t and $t-k$, depend only on k, the difference between the two times, and not on the location of the points along the time axis.

– Their properties are contrasted nicely with those of their counterparts in the following Figure.



Stationarity

- **Trend Stationary**: A time series that does not exhibit a trend.

- **Seasonal Stationary**: A time series that does not exhibit seasonality.

Non-Stationary Time Series

- Observations from a **non-stationary time series** show *seasonal effects, trends, and other structures that depend on the time index*.

- *Summary statistics like the mean and variance do change over time*, providing a drift in the concepts a model may try to capture.

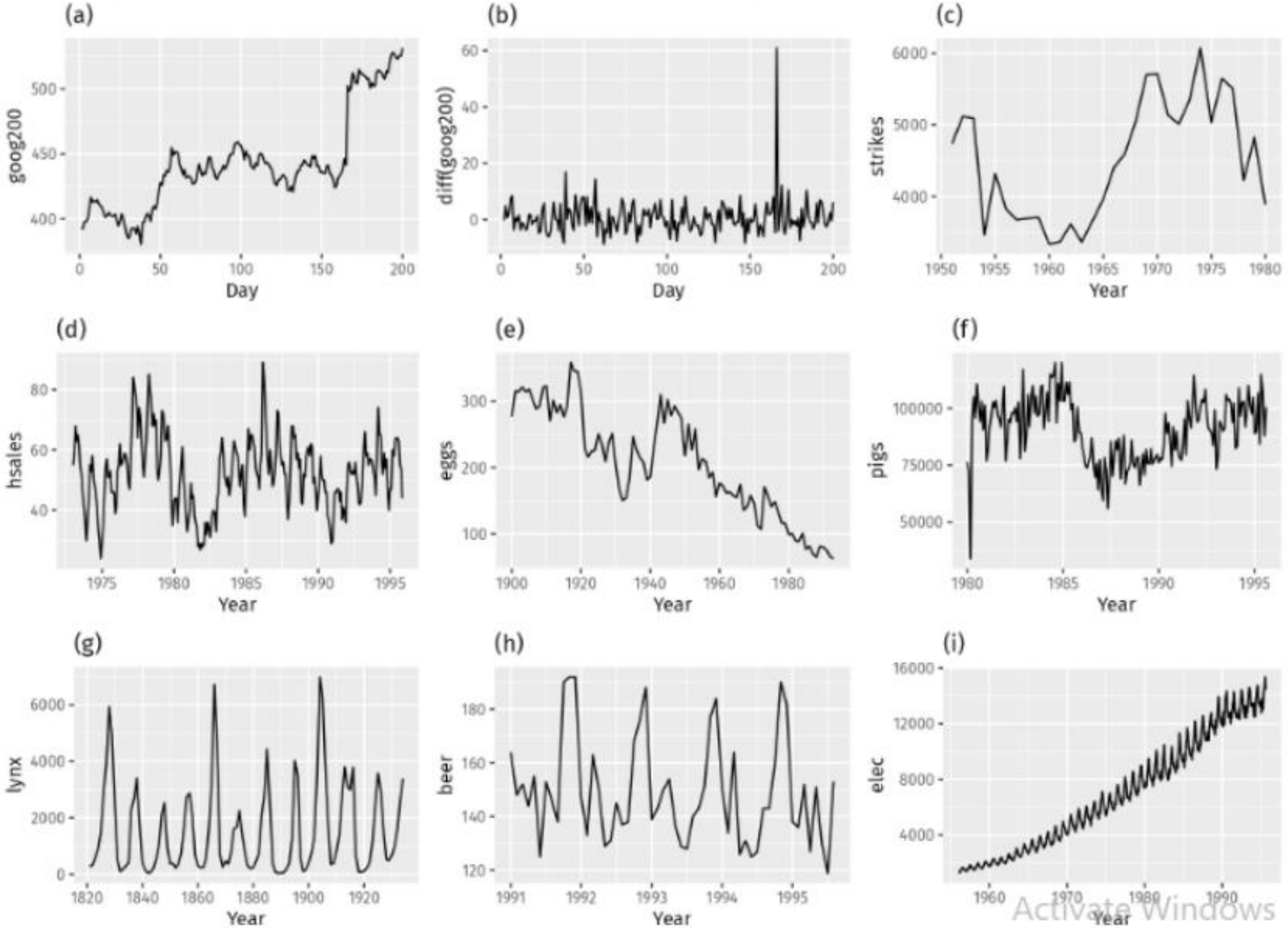
- *Classical time series analysis and forecasting methods are concerned with making non-stationary time series data stationary by identifying and removing trends and removing seasonal effects.*

Checks for Stationarity

- There are many methods to check whether a time series (direct observations, residuals, otherwise) is stationary or non-stationary.
  - **Look at Plots**: You can review a time series plot of your data and visually check if there are any obvious trends or seasonality.
  - **Summary Statistics**: You can review the summary statistics for your data for seasons or random partitions and check for obvious or significant differences.
  - **Statistical Tests**: You can use statistical tests to check if the expectations of stationarity are met or have been violated.
    - **Augmented Dickey-Fuller test**

Stationarity

- Which of the following is stationary (a,g):
  - Seasonality rules out (d), (h) and (i). Trends and changing levels rules out series (a), (c), (e), (f) and (i). Increasing variance also rules out (i)



Stationarity

Definition

If $\{y_t\}$ is a stationary time series, then for all s, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on t.

A stationary series is:

- roughly horizontal
- constant variance
- no patterns predictable in the long-term

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Stationarity

Definition

If $\{y_t\}$ is a stationary time series, then for all s, the distribution of $(y_t, \ldots, y_{t+s})$ does not depend on t.

- Transformations help to stabilize the variance.
- For ARIMA modelling, we also need to stabilize the mean.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Non-stationarity in the mean

Identifying non-stationary series

- time plot.
- The ACF of stationary data drops to zero relatively quickly
- The ACF of non-stationary data decreases slowly.
- For non-stationary data, the value of r_1 is often large and positive.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Differencing

- Differencing helps to stabilize the mean.
- The differenced series is the change between each observation in the original series: $y'_t = y_t - y_{t-1}$.
- The differenced series will have only $T - 1$ values since it is not possible to calculate a difference y'_1 for the first observation.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Second-order differencing

Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time:

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

- y''_t will have $T - 2$ values.
- In practice, it is almost never necessary to go beyond second-order differences.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Seasonal differencing

A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

$$y'_t = y_t - y_{t-m}$$

where m = number of seasons.

- For monthly data $m = 12$.
- For quarterly data $m = 4$.
- Seasonally differenced series will have $T - m$ obs.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Seasonal differencing

When both seasonal and first differences are applied…

- it makes no difference which is done first—the result will be the same.
- If seasonality is strong, we recommend that seasonal differencing be done first because sometimes the resulting series will be stationary and there will be no need for further first difference.

It is important that if differencing is used, the differences are interpretable.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Interpretation of differencing

- first differences are the change between one observation and the next;
- seasonal differences are the change between one year to the next.

But taking lag 3 differences for yearly data, for example, results in a model which cannot be sensibly interpreted.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Unit root tests

Statistical tests to determine the required order of differencing.

1. Augmented Dickey Fuller test: null hypothesis is that the data are non-stationary and non-seasonal. H_0: non-stationary
2. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test: null hypothesis is that the data are stationary and non-seasonal. H_0: stationary
3. Other tests available for seasonal data.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Backshift notation

A very useful notational device is the backward shift operator, B, which is used as follows:

$$By_t = y_{t-1}$$

In other words, B, operating on y_t, has the effect of **shifting the data back one period**.

Two applications of B to y_t **shifts the data back two periods**:

$$B(By_t) = B^2y_t = y_{t-2}$$

For monthly data, if we wish to shift attention to "the same month last year", then B^{12} is used, and the notation is $B^{12}y_t = y_{t-12}$.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Backshift notation

The backward shift operator is convenient for describing the process of *differencing*.

A first-order difference can be written as

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t$$

Similarly, if second-order differences (i.e., first differences of first differences) have to be computed, then:

$$y''_t = y_t - 2y_{t-1} + y_{t-2} = (1 - B)^2 y_t$$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Backshift notation

- Second-order difference is denoted $(1-B)^2$.
- *Second-order difference* is not the same as a *second difference*, which would be denoted $1-B^2$;
- In general, a dth-order difference can be written as

$$(1-B)^d y_t$$

- A seasonal difference followed by a first difference can be written as

$$(1-B)(1-B^m)y_t$$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Backshift notation

The "backshift" notation is convenient because the terms can be multiplied together to see the combined effect.

$$(1 - B)(1 - B^m)y_t = (1 - B - B^m + B^{m+1})y_t$$
$$= y_t - y_{t-1} - y_{t-m} + y_{t-m-1}.$$

For monthly data, $m = 12$ and we obtain the same result as earlier.

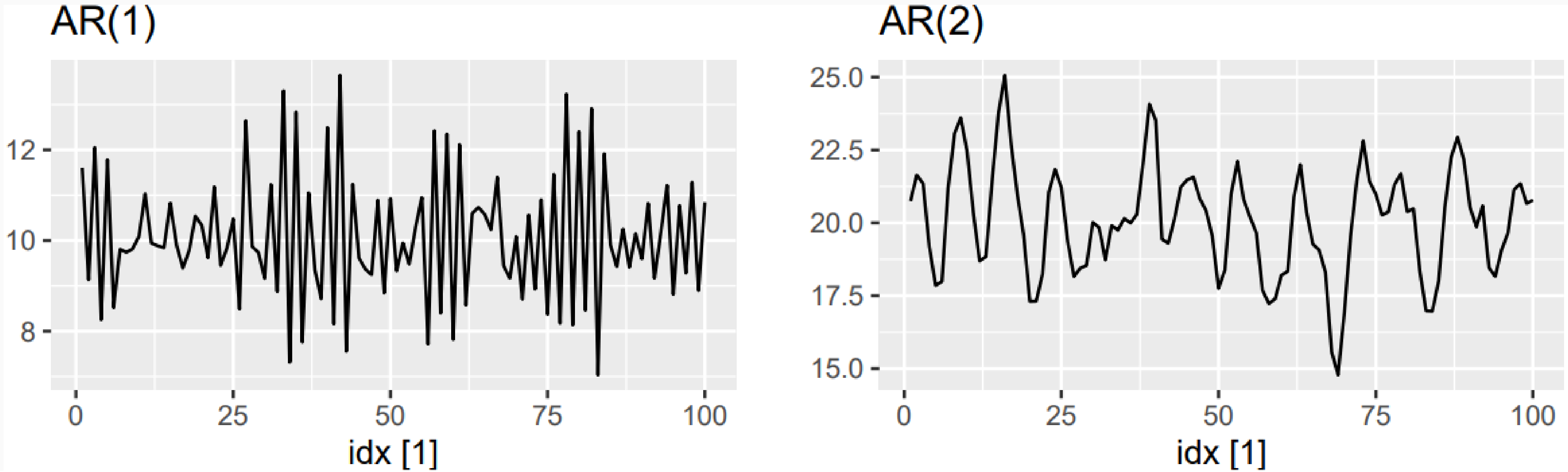
Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Autoregressive models

Autoregressive model - AR(p):

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t,$$

where ε_t is white noise. This is a multiple regression with lagged values of y_t as predictors.

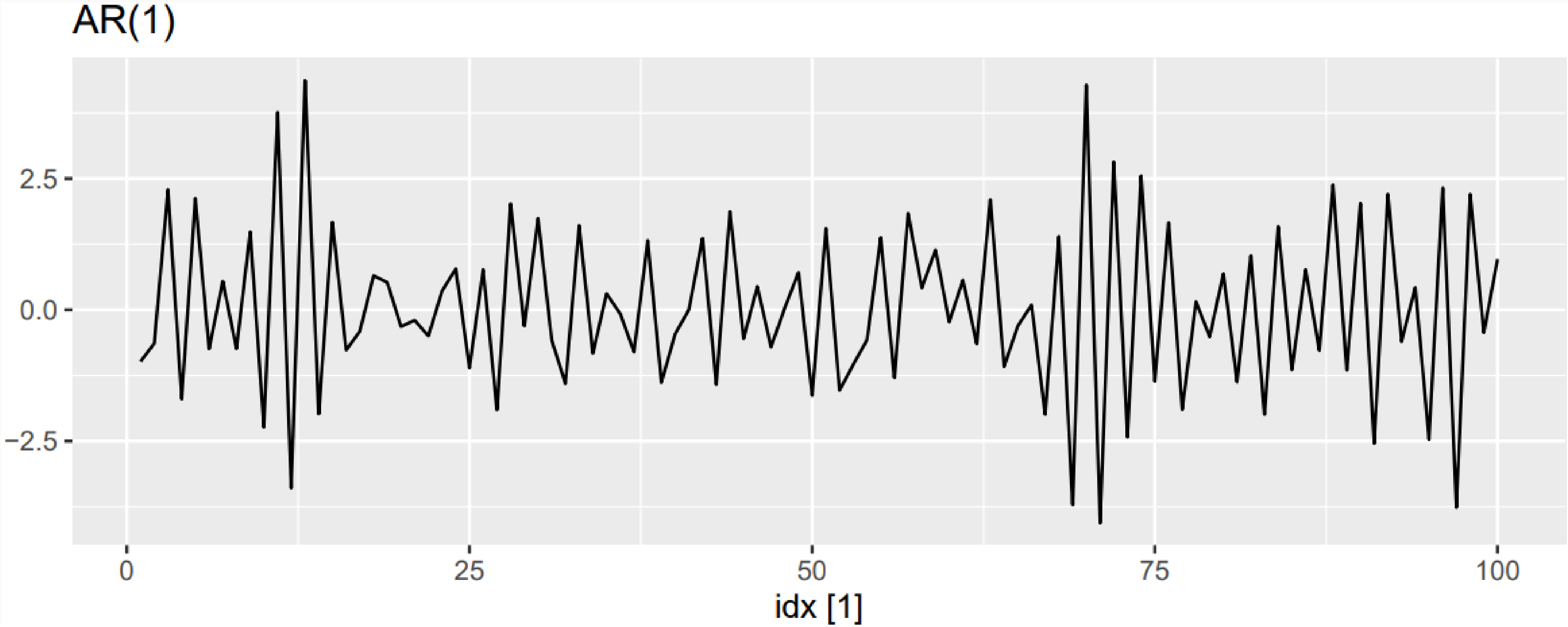


Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

AR(1) model

$$y_t = -0.8y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim N(0, 1), \quad T = 100.$$



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

AR(1) model

$$y_t = \phi_1 y_{t-1} + \varepsilon_t$$

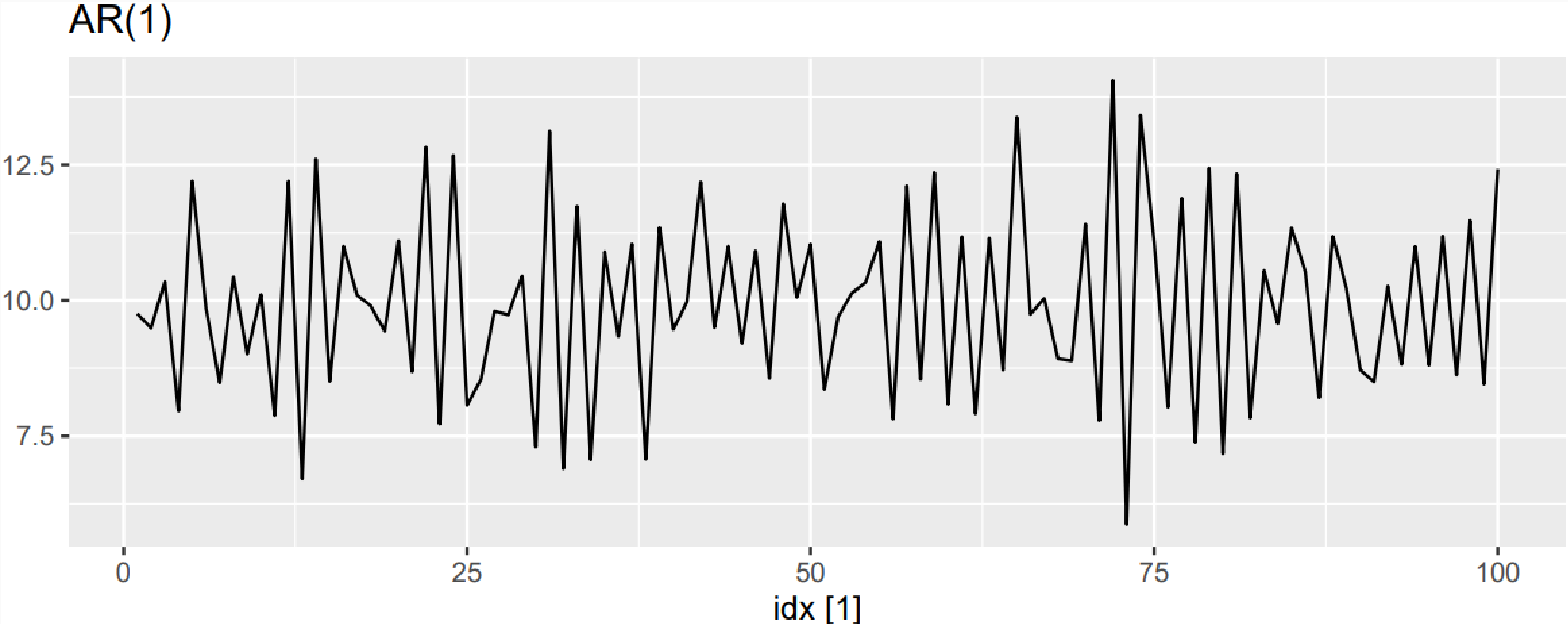
- When $\phi_1 = 0$, y_t is equivalent to a WN
- When $\phi_1 = 1$, y_t is equivalent to a RW
- We require $|\phi_1| < 1$ for stationarity. The closer ϕ_1 is to the bounds the more the process wanders above or below it's unconditional mean (zero in this case).
- When $\phi_1 < 0$, y_t tends to oscillate between positive and negative values.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

AR(1) model including a constant

$$y_t = 18 - 0.8y_{t-1} + \varepsilon_t$$

$$\varepsilon_t \sim N(0, 1), \quad T = 100.$$



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

AR(1) model including a constant

$$y_t = c + \phi_1 y_{t-1} + \varepsilon_t$$

- When $\phi_1 = 0$ and $c = 0$, y_t is equivalent to WN;
- When $\phi_1 = 1$ and $c = 0$, y_t is equivalent to a RW;
- When $\phi_1 = 1$ and $c \neq 0$, y_t is equivalent to a RW with drift;

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

AR(1) model including a constant

- c is related to the mean of y_t.
- Let $E(y_t) = \mu$
- $\mu = c + \phi_1\mu$
- $\mu = \frac{c}{1-\phi_1}$
- `ARIMA()` takes care of whether you need a constant or not, or you can overide it.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Stationarity conditions

We normally restrict autoregressive models to stationary data, and then some constraints on the values of the parameters are required.

General condition for stationarity

Complex roots of $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$ lie outside the unit circle on the complex plane.

- For $p = 1$: $-1 < \phi_1 < 1$.
- For $p = 2$:
  $-1 < \phi_2 < 1 \qquad \phi_2 + \phi_1 < 1 \qquad \phi_2 - \phi_1 < 1$.
- More complicated conditions hold for $p \geq 3$.
- Estimation software takes care of this.

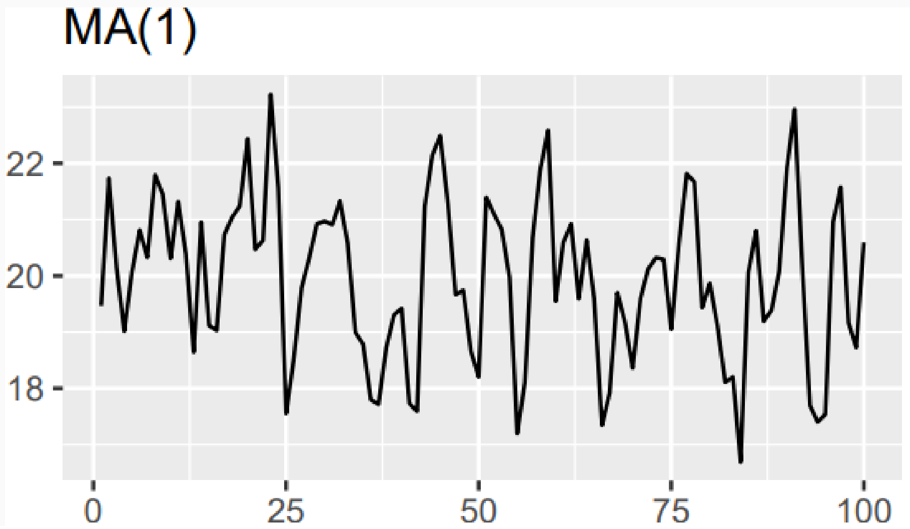
Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

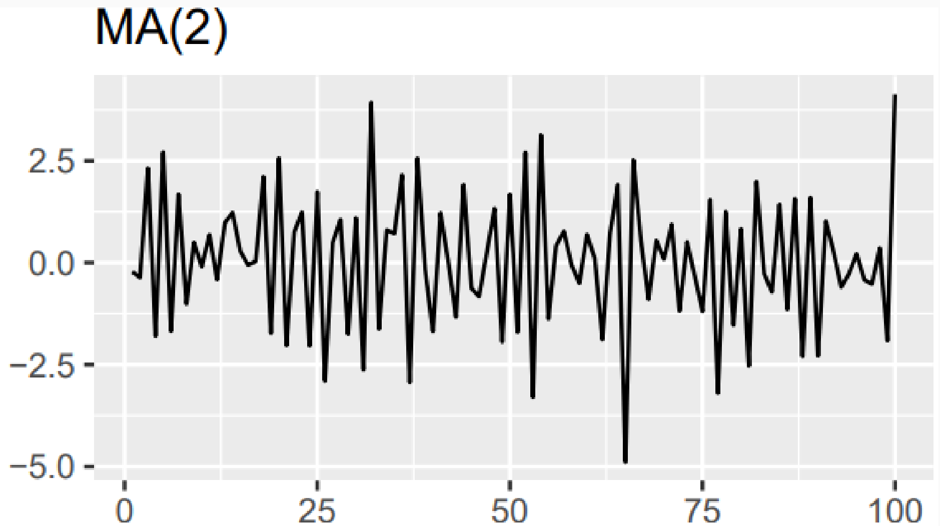
Moving Average (MA) models

Moving Average (MA) models:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q},$$

where ε_t is white noise. This is a multiple regression with ***past errors*** as predictors. *Don't confuse this with moving average smoothing!*



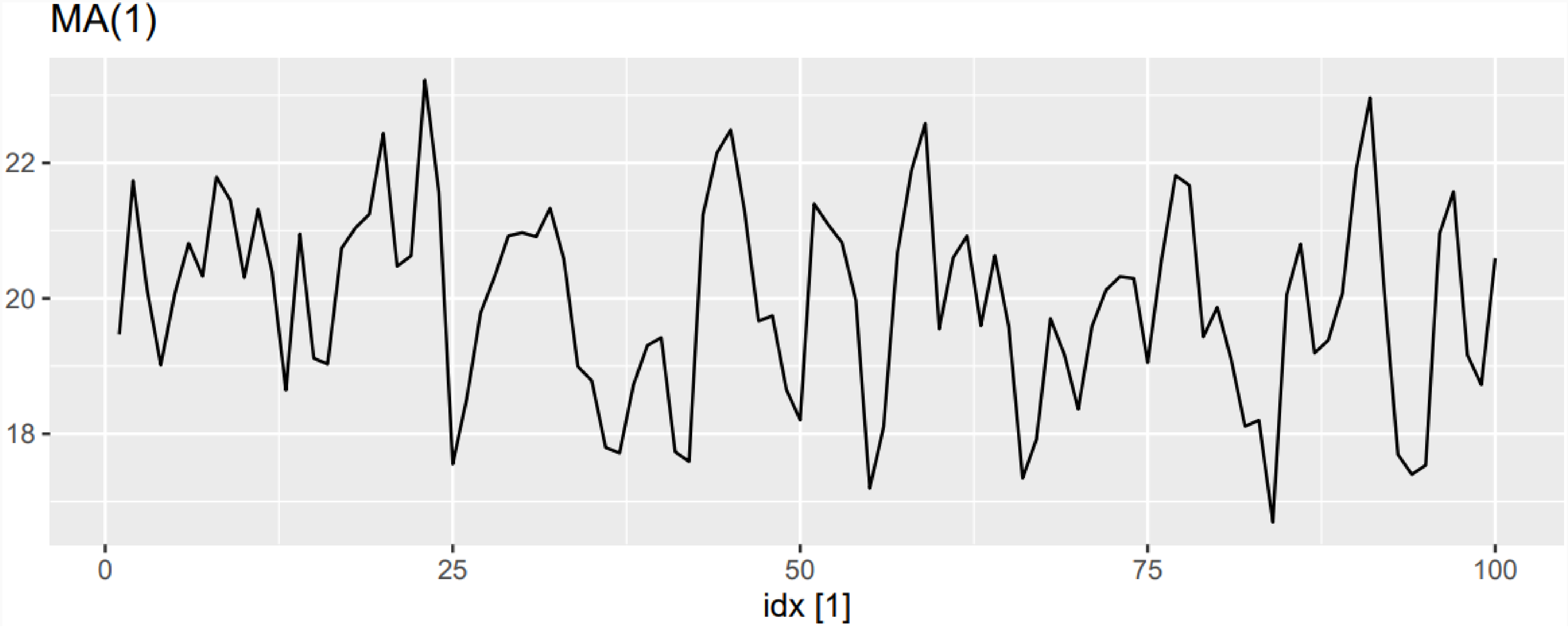


Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

MA(1) model

$$y_t = 20 + \varepsilon_t + 0.8\varepsilon_{t-1}$$

$$\varepsilon_t \sim N(0, 1), \quad T = 100.$$



Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

MA(∞) models

It is possible to write any stationary AR(p) process as an MA(∞) process.

Example: AR(1)

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \varepsilon_t \\
&= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \phi_1^2 y_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&= \phi_1^3 y_{t-3} + \phi_1^2 \varepsilon_{t-2} + \phi_1 \varepsilon_{t-1} + \varepsilon_t \\
&\dots
\end{aligned}
$$

Provided $-1 < \phi_1 < 1$:

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Invertibility

- Any MA(q) process can be written as an AR(∞) process if we impose some constraints on the MA parameters.
- Then the MA model is called "invertible".
- Invertible models have some mathematical properties that make them easier to use in practice.
- Invertibility of an ARIMA model is equivalent to forecastability of an ETS model.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

ARIMA models

AR: autoregressive (lagged observations as inputs)
I: integrated (differencing to make series stationary)
MA: moving average (lagged errors as inputs)

An ARIMA model is rarely interpretable in terms of visible data structures like trend and seasonality. But it can capture a huge range of time series patterns.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

ARIMA models

Autoregressive Moving Average models:

$$y_t = c + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t.$$

- Predictors include both **lagged values of y_t and lagged errors.**
- Conditions on AR coefficients ensure stationarity.
- Conditions on MA coefficients ensure invertibility.

Autoregressive Integrated Moving Average models

- Combine ARMA model with **differencing**.
- $(1 - B)^d y_t$ follows an ARMA model.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

ARIMA models

Autoregressive Integrated Moving Average models

ARIMA(p, d, q) model

AR: p = order of the autoregressive part

I: d = degree of first differencing involved

MA: q = order of the moving average part.

- White noise model: ARIMA(0,0,0)
- Random walk: ARIMA(0,1,0) with no constant
- Random walk with drift: ARIMA(0,1,0) with const.
- AR(p): ARIMA(p,0,0)
- MA(q): ARIMA(0,0,q)

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Backshift notation for ARIMA

- ARMA model:

$$y_t = c + \phi_1 B y_t + \cdots + \phi_p B^p y_t + \varepsilon_t + \theta_1 B \varepsilon_t + \cdots + \theta_q B^q \varepsilon_t$$

$$\text{or} \quad (1 - \phi_1 B - \cdots - \phi_p B^p) y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q) \varepsilon_t$$

- ARIMA(1,1,1) model:

$$\underbrace{(1 - \phi_1 B)}_{\text{AR(1)}} \quad \underbrace{(1 - B) y_t}_{\text{First difference}} = c + \underbrace{(1 + \theta_1 B) \varepsilon_t}_{\text{MA(1)}}$$

Expand: $y_t = c + y_{t-1} + \phi_1 y_{t-1} - \phi_1 y_{t-2} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Maximum likelihood estimation

Having identified the model order, we need to estimate the parameters $c, \phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$.

- MLE is very similar to least squares estimation obtained by minimizing

$$\sum_{t-1}^{T} e_t^2$$

- The `ARIMA()` function allows CLS or MLE estimation.
- Non-linear optimization must be used in either case.
- Different software will give different estimates.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Information criteria

Akaike's Information Criterion (AIC):

$$\text{AIC} = -2\log(L) + 2(p + q + k + 1),$$

where L is the likelihood of the data, $k = 1$ if $c \neq 0$ and $k = 0$ if $c = 0$.

Corrected AIC:

$$\text{AICc} = \text{AIC} + \frac{2(p + q + k + 1)(p + q + k + 2)}{T - p - q - k - 2}.$$

Bayesian Information Criterion:

$$\text{BIC} = \text{AIC} + [\log(T) - 2](p + q + k + 1).$$

Good models are obtained by minimizing either the AIC, AICc or BIC.
Our preference is to use the AICc.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Point Forecasting

1. Rearrange ARIMA equation so y_t is on LHS.
2. Rewrite equation by replacing t by $T + h$.
3. On RHS, replace future observations by their forecasts, future errors by zero, and past errors by corresponding residuals.

Start with $h = 1$. Repeat for $h = 2, 3, \ldots$.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Point Forecasting

ARIMA(3,1,1) forecasts: Step 1

$$(1 - \hat{\phi}_1 B - \hat{\phi}_2 B^2 - \hat{\phi}_3 B^3)(1 - B) y_t = (1 + \hat{\theta}_1 B)\varepsilon_t,$$

$$\left[1 - \hat{\phi}_1 B - \hat{\phi}_2 B^2 - \hat{\phi}_3 B^3 - B + \hat{\phi}_1 B^2 + \hat{\phi}_2 B^3 + \hat{\phi}_3 B^4\right] y_t = (1 + \hat{\theta}_1 B)\varepsilon_t$$

$$\left[1 - (1 + \hat{\phi}_1) B + (\hat{\phi}_1 - \hat{\phi}_2) B^2 + (\hat{\phi}_2 - \hat{\phi}_3) B^3 + \hat{\phi}_3 B^4\right] y_t = (1 + \hat{\theta}_1 B)\varepsilon_t$$

$$y_t - (1 + \hat{\phi}_1) y_{t-1} + (\hat{\phi}_1 - \hat{\phi}_2) y_{t-2} + (\hat{\phi}_2 - \hat{\phi}_3) y_{t-3} + \hat{\phi}_3 y_{t-4} = \varepsilon_t + \hat{\theta}_1 \varepsilon_{t-1}$$

$$y_t = (1 + \hat{\phi}_1) y_{t-1} - (\hat{\phi}_1 - \hat{\phi}_2) y_{t-2} - (\hat{\phi}_2 - \hat{\phi}_3) y_{t-3} - \hat{\phi}_3 y_{t-4} + \varepsilon_t + \hat{\theta}_1 \varepsilon_{t-1}$$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Point forecasts (h=1)

$$y_t = (1+\hat{\phi}_1)y_{t-1} - (\hat{\phi}_1 - \hat{\phi}_2)y_{t-2} - (\hat{\phi}_2 - \hat{\phi}_3)y_{t-3} - \hat{\phi}_3 y_{t-4} + \varepsilon_t + \hat{\theta}_1\varepsilon_{t-1}$$

ARIMA(3,1,1) forecasts: Step 2

$$y_{T+1} = (1+\hat{\phi}_1)y_T - (\hat{\phi}_1 - \hat{\phi}_2)y_{T-1} - (\hat{\phi}_2 - \hat{\phi}_3)y_{T-2} - \hat{\phi}_3 y_{T-3} + \varepsilon_{T+1} + \hat{\theta}_1\varepsilon_T$$

ARIMA(3,1,1) forecasts: Step 3

$$\hat{y}_{T+1|T} = (1+\hat{\phi}_1)y_T - (\hat{\phi}_1 - \hat{\phi}_2)y_{T-1} - (\hat{\phi}_2 - \hat{\phi}_3)y_{T-2} - \hat{\phi}_3 y_{T-3} + \hat{\theta}_1 e_T$$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Point forecasts (h=2)

$$y_t = (1 + \hat{\phi}_1)y_{t-1} - (\hat{\phi}_1 - \hat{\phi}_2)y_{t-2} - (\hat{\phi}_2 - \hat{\phi}_3)y_{t-3} - \hat{\phi}_3 y_{t-4} + \varepsilon_t + \hat{\theta}_1\varepsilon_{t-1}$$

ARIMA(3,1,1) forecasts: Step 2

$$y_{T+2} = (1 + \hat{\phi}_1)y_{T+1} - (\hat{\phi}_1 - \hat{\phi}_2)y_T - (\hat{\phi}_2 - \hat{\phi}_3)y_{T-1} - \hat{\phi}_3 y_{T-2} + \varepsilon_{T+2} + \hat{\theta}_1\varepsilon_{T+1}.$$

ARIMA(3,1,1) forecasts: Step 3

$$\hat{y}_{T+2|T} = (1 + \hat{\phi}_1)\hat{y}_{T+1|T} - (\hat{\phi}_1 - \hat{\phi}_2)y_T - (\hat{\phi}_2 - \hat{\phi}_3)y_{T-1} - \hat{\phi}_3 y_{T-2}.$$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Prediction intervals

Assuming $\varepsilon_t \sim N(0, \sigma^2)$

95% prediction interval

$$\hat{y}_{T+h|T} \pm 1.96\sqrt{v_{T+h|T}}$$

where $v_{T+h|T}$ is estimated forecast variance.

- $v_{T+1|T} = \hat{\sigma}^2$ for all ARIMA models
- Multi-step prediction intervals for ARIMA(0,0,q):

$$y_t = \varepsilon_t + \sum_{i=1}^{q} \hat{\theta}_i \varepsilon_{t-i}.$$

$$v_{T|T+h} = \hat{\sigma}^2 \left[1 + \sum_{i=1}^{h-1} \hat{\theta}_i^2\right], \qquad \text{for } h = 2, 3, \ldots.$$

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Prediction intervals

- Prediction intervals increase in size with forecast horizon.
- Prediction intervals can be difficult to calculate by hand
- Calculations assume residuals are uncorrelated and normally distributed.
- Prediction intervals tend to be too narrow.
  - the uncertainty in the parameter estimates has not been accounted for.
  - model uncertainty has not been accounted for.
  - the ARIMA model assumes historical patterns will not change during the forecast period.
  - the ARIMA model assumes uncorrelated future errors.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Seasonal ARIMA models

$$\text{ARIMA} \quad \underbrace{(p, d, q)}_{\substack{\uparrow \\ \text{Non-seasonal part} \\ \text{of the model}}} \quad \underbrace{(P, D, Q)_m}_{\substack{\uparrow \\ \text{Seasonal part of} \\ \text{of the model}}}$$

where m = number of observations per year.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Seasonal ARIMA models

E.g., ARIMA(1, 1, 1)(1, 1, 1)$_4$ model (without constant)

$$(1-\phi_1 B)(1-\Phi_1 B^4)(1-B)(1-B^4)y_t = (1+\theta_1 B)(1+\Theta_1 B^4)\varepsilon_t.$$

(Non-seasonal AR(1))

(Seasonal AR(1))

(Non-seasonal difference)

(Seasonal difference)

(Non-seasonal MA(1))

(Seasonal MA(1))

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Seasonal ARIMA models

The seasonal part of an AR or MA model will be seen in the seasonal lags of the PACF and ACF.

$ARIMA(0,0,0)(0,0,1)_{12}$ will show:

- a spike at lag 12 in the ACF but no other significant spikes.
- The PACF will show exponential decay in the seasonal lags; that is, at lags 12, 24, 36, ...

$ARIMA(0,0,0)(1,0,0)_{12}$ will show:

- exponential decay in the seasonal lags of the ACF
- a single significant spike at lag 12 in the PACF.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

How does ARIMA() work for seasonal models?

A seasonal ARIMA process

$$\Phi(B)\phi(B)(1-B)^d(1-B)^D y_t = c + \Theta(B)\theta(B)\varepsilon_t$$

Need to select appropriate orders: d, D, p, q, P, Q, and whether to include the intercept c.

Hyndman and Khandakar (JSS, 2008) algorithm:

- Select no. differences d via KPSS test and D using seasonal strength.
- Select p, q, P, Q and c by minimising AICc.
- Use stepwise search to traverse model space.

Source: Forecasting Principles and Practice, 3rd Edition, Rob. J. Hyndman and George Athanasopoulos

Thank You

For Your Valuable Time.