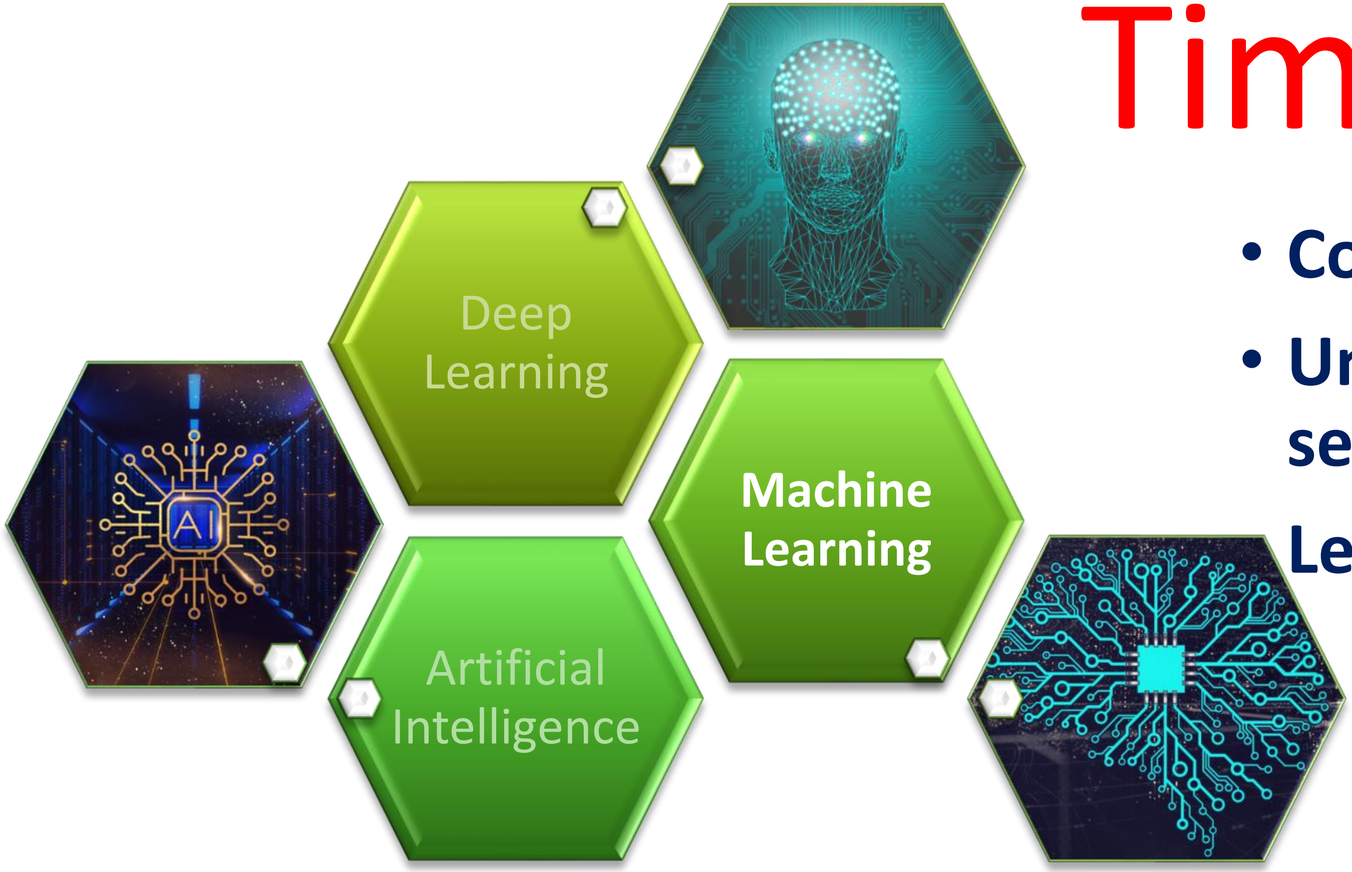


Time Series

- **Course Code: CSE471**
- **Unit 3: Statistical Models for time series**

Lecture1

Statistical models for time Series

- Linear statistical models for time series are related to linear regression but account for the correlations that arise between data points in the same time series.
- The specific models we will discuss are:
    - Autoregressive (AR) models,
    - Moving average (MA) models, and
    - Autoregressive integrated moving average (ARIMA) models
    - Vector autoregression (VAR)

Why not Linear Regression?

Here are some reasons why linear regression cannot always be used for time series forecasting:

- Linear regression assumes that the relationship between the independent and dependent variables is linear. However, time series data often exhibits non-linear relationships. For example, the demand for a product may increase exponentially over time. In this case, linear regression would not be able to accurately forecast the demand.
- Linear regression assumes that the errors are independent and identically distributed (iid). However, time series data often exhibits autocorrelation, which means that the errors are not independent. This can make it difficult for linear regression to accurately forecast the data.
- **Linear regression is not able to capture the seasonality of time series data**. Seasonality refers to the regular patterns that occur in time series data. For example, the demand for electricity may increase in the summer and decrease in the winter. Linear regression is not able to capture these patterns, which can lead to inaccurate forecasts.

In general, linear regression is a good choice for forecasting time series data that exhibits a linear relationship and does not have a lot of seasonality. However, if the data does not meet these conditions, then linear regression may not be the best choice.

Statistical Methods Developed for Time Series

Autoregressive Models

- The autoregressive (AR) model relies on the intuition that the past predicts the future and so a time series process in which the value at a point in time *t* is a function of the series values at earlier points in time.

The simplest AR model, an AR(1) model, describes a system as follows:

$$y_t = b_0 + b_1 \times y_{t-1} + e_t$$

The value of the series at time t is a function of a constant b_0, its value at the previous time step multiplied by another constant $b_1 \times y_{t-1}$ and an error term that also varies with time e_t. This error term is assumed to have a constant variance and a mean of 0. We denote an autoregressive term that looks back only to the immediately prior time as an AR(1) model because it includes a lookback of one lag.

We can calculate both the expected value of y_t and its variance, given y_{t-1}, if we know the value of b_0 and b_1. See Equation 6-1.[3]

$$\textit{Equation 6-1. } E(y_t \mid y_{t-1}) = b_0 + b_1 \times y_{t-1} + e_t$$

$$Var(y_t \mid y_{t-1}) = Var(e_t) = Var(e)$$

The generalization of this notation allows the present value of an AR process to depend on the p most recent values, producing an AR(p) process.

We now switch to more traditional notation, which uses ϕ to denote the autoregression coefficients:

$$y_t = \phi_0 + \phi_1 \times y_{t-1} + \phi_2 \times y_{t-2} + \ldots + \phi_p \times y_{t-p} + e_t$$

We can determine the conditions for an AR model to be stationary from the definition of stationarity. We continue our focus on the simplest AR model, AR(1) in Equation 6-2.

Equation 6-2. $y_t = \phi_0 + \phi_1 \times y_{t-1} + e_t$

We assume the process is stationary and then work "backward" to see what that implies about the coefficients. First, from the assumption of stationarity, we know that the expected value of the process must be the same at all times. We can rewrite y_t per the equation for an AR(1) process:

$$E(y_t) = \mu = E(y_{t-1})$$

By definition, e_t has an expected value of 0. Additionally the phis are constants, so their expected values are their constant values. Equation 6-2 reduces on the lefthand side to:

$$E(y_t) = E(\phi_0 + \phi_1 \times y_{t-1} + e_t)$$

$$E(y_t) = \mu$$

And on the righthand side to:

$$\phi_0 + \phi_1 \times \mu + 0$$

This simplifies to:

$$\mu = \phi_0 + \phi_1 \times \mu$$

which in turn implies that (Equation 6-3).

Equation 6-3. $\mu = \frac{\phi_0}{1 - \phi_1}$

So we find a relationship between the mean of the process and the underlying AR(1) coefficients.

We can take similar steps to look at how a constant variance and covariance impose conditions on the ϕ coefficients. We begin by substituting the value of ϕ_0, which we can derive Equation 6-4 from Equation 6-3.

Equation 6-4. $\phi_0 = \mu \times (1 - \phi_1)$

into Equation 6-2:

$$y_t = \phi_0 + \phi_1 \times y_{t-1} + e_t$$

$$y_t = (\mu - \mu \times \phi_1) + \phi_1 \times y_{t-1} + e_t$$

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + e_t$$

If you inspect Equation 6-4, what should jump out at you is that it has very similar expressions on the lefthand and righthand sides, namely $y_t - \mu$ and $y_{t-1} - \mu$. Given that this time series is stationary, we know that the math at time $t - 1$ should be the same as the math at time t. We rewrite Equation 6-4 in the frame of a time one step earlier as Equation 6-5.

Equation 6-5. $y_{t-1} - \mu = \phi_1(y_{t-2} - \mu) + e_{t-1}$

We can then substitute this into Equation 6-4 as follows:

$$y_t - \mu = \phi_1(\phi_1(y_{t-2} - \mu) + e_{t-1}) + e_t$$

We rearrange for clarity in Equation 6-6.

Equation 6-6. $y_t - \mu = e_t + \phi_1(e_{t-1} + \phi_1(y_{t-2} - \mu))$

It should catch your eye that another substitution is possible in Equation 6-6 for $y_{t-2} - \mu$ using the same recursive substitution we used earlier, but instead of working on y_{t-1} we will work on y_{t-2}. If you do this substitution, the pattern becomes clear:

$$y_t - \mu = e_t + \phi_1(e_{t-1} + \phi_1(e_{t-2} + \phi_1(y_{t-3} - \mu)))$$
$$= e_t + \phi \times e_{t-1} + \phi^2 \times e_{t-2} + \phi^3 \times e_{t-3} + \text{(expressions still to be substituted)}$$

So we can conclude more generally that $y_t - \mu = \sum_{i=1}^{\infty} \phi_1^i \times e_{t-i}$.

In plain English, y_t minus the process mean is a linear function of the error terms.

This result can then be used to compute the expectation of $E[(y_t - \mu) \times e_{t+1}] = 0$ given that the values of e_t at different t values are independent. From this we can conclude that the covariance of y_{t-1} and e_t is 0, as it should be. We can apply similar logic to calculate the variance of y_t by squaring this equation:

$$y_t - \mu = \phi_1(y_{t-1} - \mu) + e_t$$

$$var(y_t) = \phi_1^2 var(y_{t-1}) + var(e_t)$$

Because the variance quantities on each side of the equation must be equal due to stationarity, such that $(var(y_t) = var(y_t - 1)$, this implies that:

$$var(y_t) = \frac{var(e_t)}{1 - \phi_1^2}$$

Given that the variance must be greater than or equal to 0 by definition, we can see that ϕ_1^2 must be less than 1 to ensure a positive value on the righthand side of the preceding equation. This implies that for a stationary process we must have $-1 < \phi_1 < 1$. This is a necessary and sufficient condition for weak stationarity.

Autoregressive models are remarkably flexible at handling a wide range of different time series patterns. The two series in Figure show series from an AR(1) model and an AR(2) model. Changing the parameters $\phi_1, \ldots, \phi_p$ results in different time series patterns. The variance of the error term ε_t will only change the scale of the series, not the patterns.

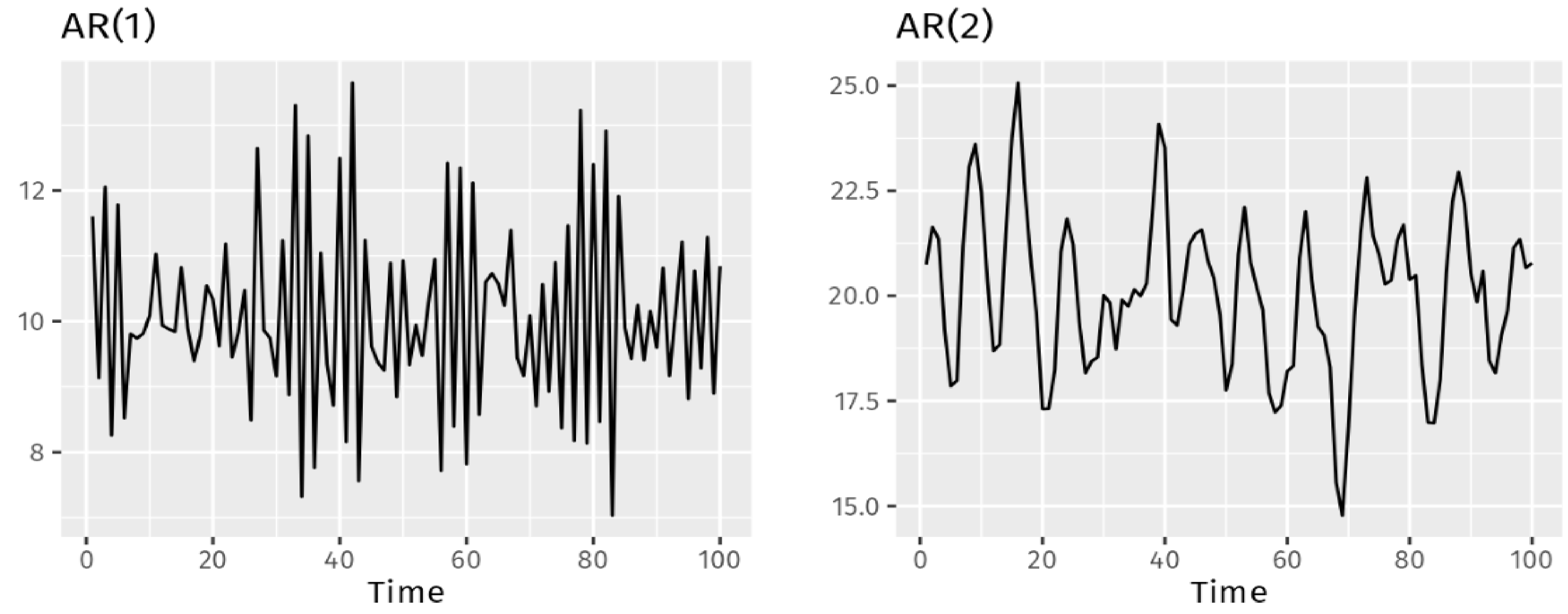


Figure : Two examples of data from autoregressive models with different parameters. Left: AR(1) with $y_t = 18 - 0.8y_{t-1} + \varepsilon_t$. Right: AR(2) with $y_t = 8 + 1.3y_{t-1} - 0.7y_{t-2} + \varepsilon_t$. In both cases, ε_t is normally distributed white noise with mean zero and variance one.

For an AR(1) model:

- when $\phi_1 = 0$, y_t is equivalent to white noise;
- when $\phi_1 = 1$ and $c = 0$, y_t is equivalent to a random walk;
- when $\phi_1 = 1$ and $c \neq 0$, y_t is equivalent to a random walk with drift;
- when $\phi_1 < 0$, y_t tends to oscillate around the mean.

We normally restrict autoregressive models to stationary data, in which case some constraints on the values of the parameters are required.

- For an AR(1) model: $-1 < \phi_1 < 1$.
- For an AR(2) model: $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.

How to Pick the Auto-Regressive Model?

PACF Plot

- PACF is the partial autocorrelation function that explains the partial correlation between the series and lags of itself.
- In simple terms, PACF can be explained using a linear regression where we predict y(t) from y(t-1), y(t-2), ...

How to Pick the Auto-Regressive Model?

YEAR	PROFIT
2010	76
2011	80
2012	90
2013	88
2014	81
2015	82
2016	94
2017	95
2018	93
2019	84
2020	82

How to Pick the Auto-Regressive Model?

$t = 2$

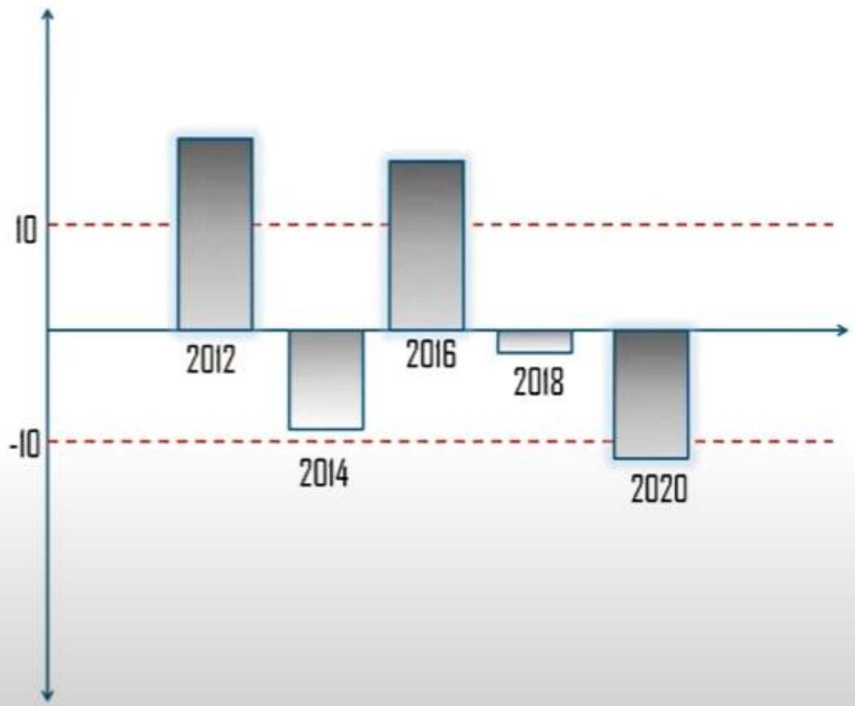
$y_t - y_{t-2}$

YEAR	PROFIT
2010	76
2011	80
2012	90
2013	88
2014	81
2015	82
2016	94
2017	95
2018	93
2019	84
2020	82

YEAR	LAG
2010	0
2012	14
2014	-9
2016	13
2018	-1
2020	-11

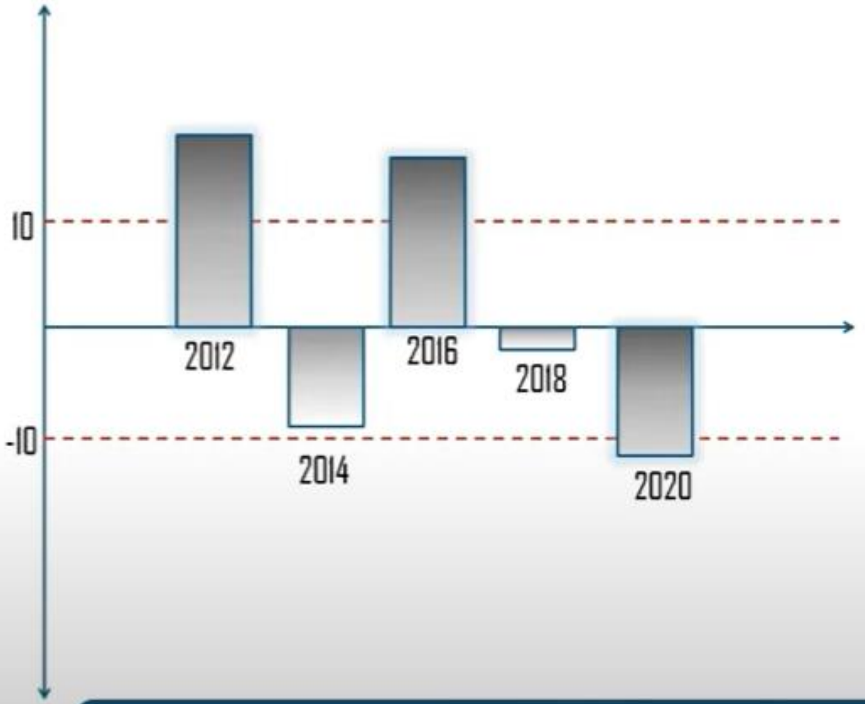
How to Pick the Auto-Regressive Model?

YEAR	LAG
2010	0
2012	14
2014	-9
2016	13
2018	-1
2020	-11



How to Pick the Auto-Regressive Model?

YEAR	LAG
2010	0
2012	14
2014	-9
2016	13
2018	-1
2020	-11



$$F_{2021} = c + \varphi y_{2012} + \varphi y_{2016} + \varphi y_{2020} + \varepsilon t$$

Moving Average (MA) Model

- A moving average (MA) model relies on a picture of a process in which the value at each point in time is a function of the recent past value “error” terms, each of which is independent from the others. We will review this model in the same series of steps we used to study AR models.

The model

A moving average model can be expressed similarly to an autoregressive model except that the terms included in the linear equation refer to present and past error terms rather than present and past values of the process itself. So an MA model of order *q* is expressed as:

$$y_t = \mu + e_t + \theta_1 \times e_{t-1} + \theta_2 \times e_{t-2} \ldots + \theta_q \times e_{t-q}$$

MA models are by definition weakly stationary without the need to impose any constraints on their parameters. This is because the mean and variance of an MA process are both finite and invariant with time because the error terms are assumed to be iid with mean 0. We can see this as:

$$E(y_t = \mu + e_t + \theta_1 \times e_{t-1} + \theta_2 \times e_{t-2} \ldots + \theta_q \times e_{t-q})$$
$$= E(\mu) + \theta_1 \times 0 + \theta_2 \times 0 + \ldots = \mu$$

For calculating the variance of the process, we use the fact that the e_t terms are iid and also the general statistical property that the variance of the sum of two random variables is the same as their individual variances plus two times their covariance. For iid variables the covariance is 0. This yields the expression:

$$Var(y_t) = (1 + \theta_1^2 + \theta_2^2 + \ldots. + \theta_q^2) \times \sigma_e^2$$

So both the mean and variance of an MA process are constant with time regardless of the parameter values.

Moving Average Basic Concepts

The following are proofs of properties found in Moving Averages Basic Concepts

Property 1: The mean of an MA(q) process is μ.

Proof:

$$E[y_i] = \mu + E[\varepsilon_i] + \theta_1 E[\varepsilon_{i-1}] + \cdots + \theta_q E[\varepsilon_{i-q}] = \mu + 0 + \theta_1 \cdot 0 + \cdots + \theta_q \cdot 0 = \mu$$

Property 2: The variance of an MA(q) process is

$$var(y_i) = \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2)$$

Proof:

$$\begin{aligned} var(y_i) &= 0 + var(\varepsilon_i) + \theta_1^2 var(\varepsilon_{i-1}) + \cdots + \theta_q^2 var(\varepsilon_{i-k}) = \sigma^2 + \theta_1^2\sigma^2 + \cdots + \theta_q^2\sigma^2 \\ &= \sigma^2(1 + \theta_1^2 + \cdots + \theta_q^2) \end{aligned}$$

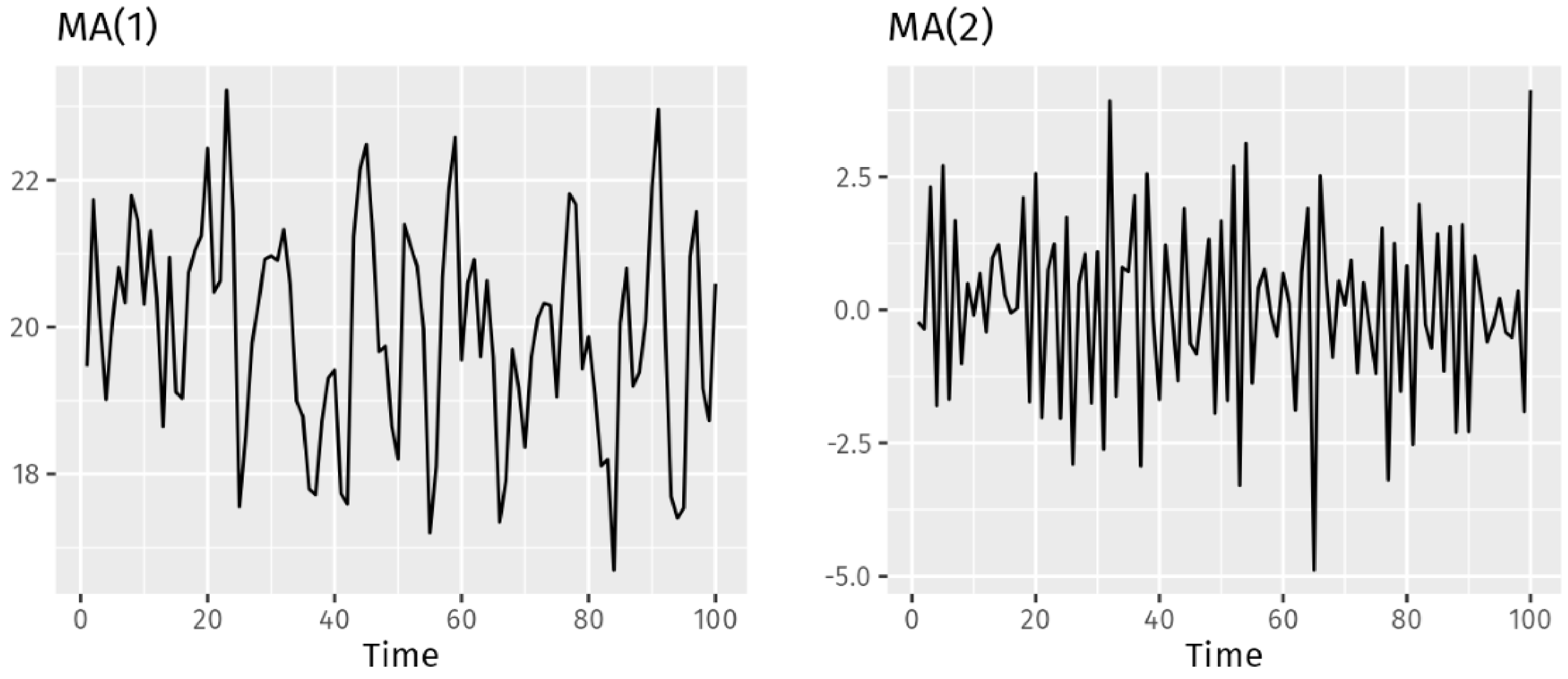


Figure Two examples of data from moving average models with different parameters. Left: MA(1) with $y_t = 20 + \varepsilon_t + 0.8\varepsilon_{t-1}$. Right: MA(2) with $y_t = \varepsilon_t - \varepsilon_{t-1} + 0.8\varepsilon_{t-2}$. In both cases, ε_t is normally distributed white noise with mean zero and variance one.

Figure shows some data from an MA(1) model and an MA(2) model. Changing the parameters $\theta_1, \dots, \theta_q$ results in different time series patterns. As with autoregressive models, the variance of the error term ε_t will only change the scale of the series, not the patterns.

Invertibility AR(1)→MA(∞)

It is possible to write any stationary AR(p) model as an MA(∞) model. For example, using repeated substitution, we can demonstrate this for an AR(1) model:

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

Provided $-1 < \phi_1 < 1$, the value of ϕ_1^k will get smaller as k gets larger. So eventually we obtain

$$
y_t = \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \cdots,
$$

an MA(∞) process.

The reverse result holds if we impose some constraints on the MA parameters. Then the MA model is called **invertible**. That is, we can write any invertible MA(q) process as an AR(∞)

Invertibility MA(1)→AR(∞)

For example, consider the MA(1) process, $y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$. In its AR(∞) representation, the most recent error can be written as a linear function of current and past observations:

$$\varepsilon_t = \sum_{j=0}^{\infty} (-\theta)^j y_{t-j}.$$

When $|\theta| > 1$, the weights increase as lags increase, so the more distant the observations the greater their influence on the current error. When $|\theta| = 1$, the weights are constant in size, and the distant observations have the same influence as the recent observations. As neither of these situations make much sense, we require $|\theta| < 1$, so the most recent observations have higher weight than observations from the more distant past. Thus, the process is invertible when $|\theta| < 1$.

The invertibility constraints for other models are similar to the stationarity constraints.

- For an MA(1) model: $-1 < \theta_1 < 1$.
- For an MA(2) model: $-1 < \theta_2 < 1, \;\; \theta_2 + \theta_1 > -1, \;\; \theta_1 - \theta_2 < 1$.