Lecture 4. ARIMA Models and Properties

Zhonghai Lu

KTH Royal Institute of Technology

April 8, 2024

Table of contents

Autoregressive (AR) model or process

Autoregressive (AR) model or process

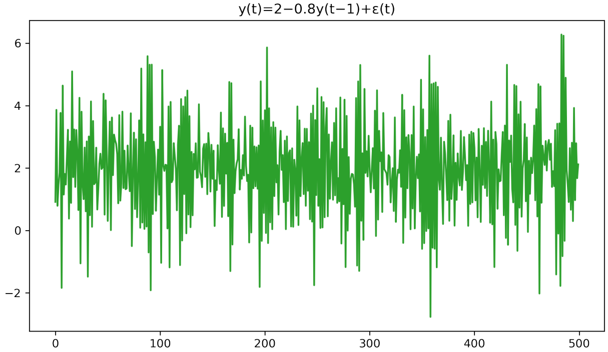
$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t$$

where ϵ_t is the white noise, p order of the model.
This is a multiple regression with **lagged values** of y_t as predictors.

AR(p) model

First-order AR(1) model

$$y_t = 2 - 0.8y_{t-1} + \epsilon_t$$
$$\epsilon_t \sim N(0,1), \quad T = 500.$$



AR(1) model

White noise (WN), Random Walk (RW)

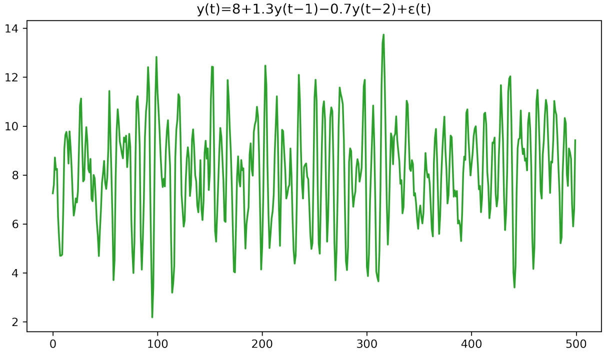
$$y_t = c + \phi_1 y_{t-1} + \epsilon_t$$

- When $\phi_1 = 0$, 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**
- When $\phi_1 < 0$, y_t tends to **oscillate between positive and negative values**.

AR(2) model

Second-order AR(2) model

$$y_t = 8 + 1.3y_{t-1} - 0.7y_{t-2} + \epsilon_t$$
$$\epsilon_t \sim N(0,1), \quad T = 500.$$



Stationarity condition

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

General condition of stationarity

Complex roots of the coefficient polynomial $1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p = 0$ lie outside the unit circle on the complex plane.

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

Strictly stationary vs. Weakly stationary process

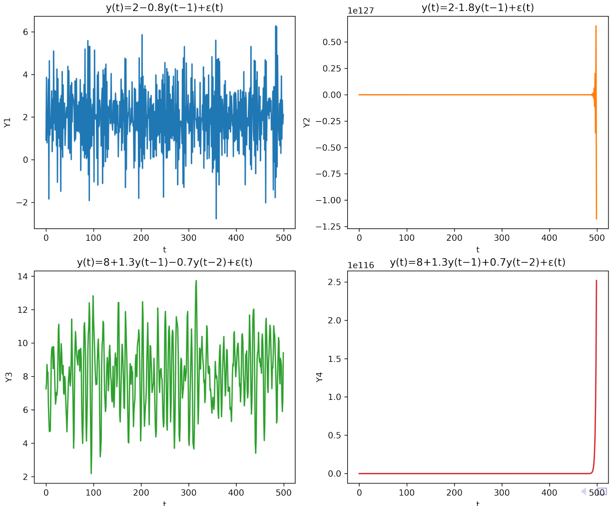
- Strictly stationary series: The joint distribution of a finite sub-sequence of random variables of the stochastic process remains the same as we shift it along the time axis, i.e., shift invariance.
- In reality, it is difficult to obtain and process its joint distributions.
- We usually consider **weakly stationary** series, which is described by its characteristic statistic using low-order moments such as mean, variance.

Weakly stationary

- A weakly stationary series $\{Y_t\}$ satisfies the following three conditions.
  1. $\forall t \in T$, second-order moment $EY_t^2 < \infty$
  2. $\forall t \in T$, first-order moment $EY_t = \mu$, μ is a constant.
  3. $\forall t, s \in T$, and $t + \Delta, s + \Delta \in T$, co-variance $cov(t, s) = cov(t + \Delta, s + \Delta)$
- Weakly stationary series means second-order moment stationary.
- From the third condition, we can derive that the series has a constant variance, $DY_t = cov(t, t) = cov(0, 0), \forall t \in T$.

Not all AR processes are stationary

The coefficients of an AR process determine whether the AR process is stationary or not.



Moving Average (MA) model or process

Moving Average (MA) model or process:

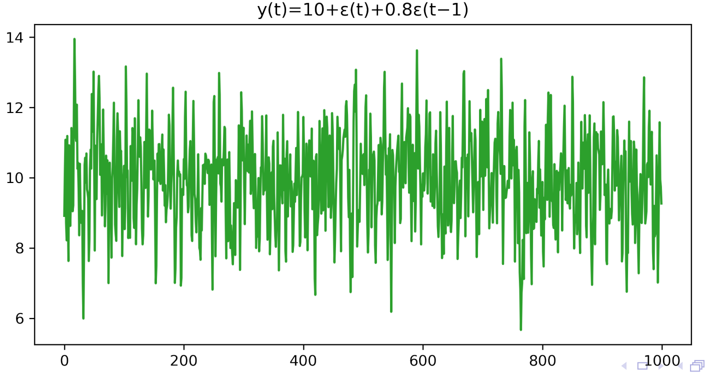
$$y_t = c + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \cdots + \theta_q\epsilon_{t-q},$$

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

MA(1) model

First-order MA(1) model

$$y_t = 10 + \epsilon_t + 0.8\epsilon_{t-1}$$
$$\epsilon_t \sim N(0, 1), \quad T = 1000.$$

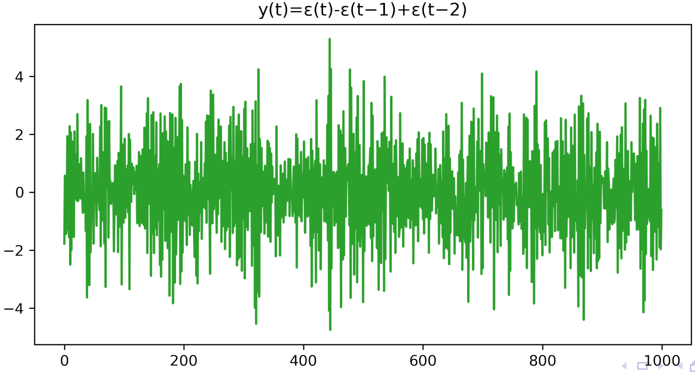


MA(2) model

Second-order MA(2) model

$$y_t = \epsilon_t - \epsilon_{t-1} + \epsilon_{t-2}$$

$$\epsilon_t \sim N(0,1), \quad T = 1000.$$



MA(infinite)

It is possible to write any stationary $AR(p)$ process as an $MA(\infty)$ process.
Example: AR(1)

$$\begin{aligned}
y_t &= \phi_1 y_{t-1} + \epsilon_t \\
&= \phi_1(\phi_1 y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= \phi_1^2 y_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \\
&= \phi_1^3 y_{t-3} + \phi_1^2 \epsilon_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \\
&= \cdots
\end{aligned}$$

Provided $-1 < \phi_1 < 1$:

$$y_t = \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_1^2 \epsilon_{t-2} + \phi_1^3 y_{t-3} + \cdots$$

Invertibility

- Any $MA(q)$ process can be written as an $AR(\infty)$ 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 is the counterpart to stationarity for the MA part of an ARMA process.
- Invertibility of an ARIMA model is equivalent to forecastability of an ETS (Exponential Smoothing) model.

Invertibility

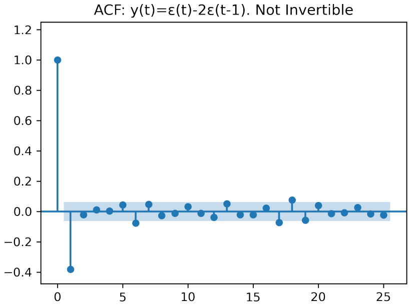
General condition for invertibility

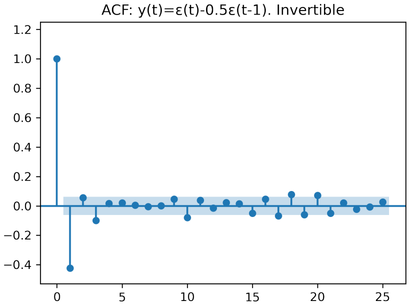
Complex roots of the coefficient polynomial $1 + \theta_1 z + \theta_2 z^2 + \cdots + \theta_q z^q = 0$ lie outside the unit circle on the complex plane.

- For $q = 1$: $-1 < \theta_1 < 1$.
- For $q = 2$: $-1 < \theta_2 < 1 \quad \theta_2 + \theta_1 > -1 \quad \theta_1 - \theta_2 < 1$.
- More complicated conditions hold for $q \geq 3$.
- Estimation software takes care of this.

Not all MA processes are invertible

- The coefficients of a MA process determine whether the MA process is invertible or not.
- Different MA(q) processes may generate the same ACF graphs.
- MA Invertibility sets constraints on the coefficients, ensuring a one-to-one mapping from ACF graph to its corresponding MA model.





ARMA models

Autoregressive Moving Average models:

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

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

Series transformation: Differencing

- Differencing is often used to transform a non-stationary series into a stationary series.
- The differenced series is the change between each observation in the original series.

$$y'_t = y_t - y_{t-1}$$

- The differenced series has $T - 1$ values since it cannot calculate a difference y'_1 for the first observation.

Random walk

- A random-walk series is a non-stationary series.

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

  where c is a drift constant.

- It can be turned into a stationary series by first-order differencing.

$$y'_t = y_t - y_{t-1} = \epsilon'_t$$

Second-order differencing

- Occasionally the differenced series is still not stationary and it may need to difference the data a second time.

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

- y_{t-2} has $T - 2$ values.
- In practice, it is almost never necessary to go beyond second-order differences.

Seasonal differencing

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

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

  where m is the length of a season.
- For monthly data, $m = 12$. For quarterly data, $m = 4$

Seasonal plus first differencing

- For seasonal series, seasonally differenced series is closer to being stationary.
- Remaining non-stationarity can be removed with further first difference.
- Example: If $y'_t = y_t - y_{t-12}$ denotes a seasonally differenced series, then twice-differenced series does actually:

$$y^*_t = y'_t - y'_{t-1} = (y_t - y_{t-12}) - (y_{t-1} - y_{t-13}) = y_t - y_{t-1} - y_{t-12} + y_{t-13}$$

Seasonal differencing vs first differencing

- Which differencing to be used first or does it matter?
  Yes. 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.
- When both seasonal and first differences are applied, it makes no difference which is done first—the result will be the same.
- It is important that if differencing is used, the differences are interpretable.
  E.g. taking lag 3 differences for yearly data results in a model which cannot be sensibly interpreted.

Backshift notation

- A very useful notation 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:**
$$B(By_t) = B^2 y_t = y_{t-2}.$$
- For monthly data, if we wish to shift attention to "the same month last year," then B^{12} is used, and the notation is $B^{12} y_t = y_{t-12}$.

Backshift notation

- The backward shift operator is convenient for describing the process of differencing.
- A first difference can be written as

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

- Note that a first difference is represented by $(1 - B)$. Similarly, if second-order differences (i.e., first differences of first differences) have to be computed, then:

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

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

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

$$
\begin{aligned}
(1 - B)(1 - B^m)y_t &= (1 - B - B^m + B^{m+1})y_t \\
&= y_t - y_{t-1} - y_{t-m} + y_{t-m-1}.
\end{aligned}
$$

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

ARIMA models

Autoregressive Moving Average models:

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

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

Autoregressive Integrated Moving Average models:

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

ARIMA models

Autoregressive Integrated Moving Average

ARIMA(p, d, q) model

AR: p = order of the autoregressive part
I: d = order of differencing
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 constant
- AR(p): ARIMA($p, 0, 0$)
- MA(q): ARIMA($0, 0, q$)

Backshift notation for ARIMA

- ARMA(p, q) model:

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

  or $(1 - \phi_1 B - \cdots - \phi_p B^p) y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\epsilon_t$

- ARIMA(1,1,1) model:

$$(1 - \phi_1 B)(1 - B) y_t = c + (1 + \theta_1 B)\epsilon_t$$

  where $(1 - \phi_1 B) y_t$ is AR(1), $(1 - B) y_t$ is the first-order difference, $(1 + \theta_1 B)\epsilon_t$ is MA(1)

- Write out: $y_t = c + y_{t-1} + \phi_1 y_{t-1} - \phi_1 y_{t-2} + \theta_1 \epsilon_{t-1} + \epsilon_t$

ARIMA(p,d,q) models

Intercept form

$$(1 - \phi_1 B - \cdots - \phi_p B^p) y'_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q) \epsilon_t$$

Mean form

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(y'_t - \mu) = (1 + \theta_1 B + \cdots + \theta_q B^q) \epsilon_t$$

- $y'_t = (1 - B)^d y_t$
- μ is the mean of y'_t
- $c = \mu(1 - \phi_1 - \cdots - \phi_p)$.

Seasonal ARIMA models

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

where $m =$ number of observations per season (e.g. year).

Seasonal ARIMA models

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

$$(1-\phi_1 B)(1-\phi_1 B^4)(1-B)(1-B^4)y_t = (1+\theta_1 B)(1+\theta_1 B^4)\epsilon_t$$

Non-seasonal AR(1)

Seasonal AR(1)

Non-seasonal difference

Seasonal difference

Non-seasonal MA(1)

Seasonal MA(1)

Seasonal ARIMA models

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

$$(1-\phi_1 B)(1-\phi_1 B^4)(1-B)(1-B^4)y_t = (1+\theta_1 B)(1+\theta_1 B^4)\epsilon_t.$$

All the factors can be multiplied out and the general model written as follows:

$$\begin{aligned} y_t = {} & (1+\phi_1)y_{t-1} - \phi_1 y_{t-2} + (1+\phi_1)y_{t-4} \\ & - (1+\phi_1+\phi_1+\phi_1\phi_1)y_{t-5} + (\phi_1+\phi_1\phi_1)y_{t-6} \\ & - \phi_1 y_{t-8} + (\phi_1+\phi_1\phi_1)y_{t-9} - \phi_1\phi_1 y_{t-10} \\ & + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_1\epsilon_{t-4} + \theta_1\theta_1\epsilon_{t-5}. \end{aligned}$$

Examples

The US Census Bureau uses the following models most often:

$\text{ARIMA}(0,1,1)(0,1,1)_m$	with log transformation
$\text{ARIMA}(0,1,2)(0,1,1)_m$	with log transformation
$\text{ARIMA}(2,1,0)(0,1,1)_m$	with log transformation
$\text{ARIMA}(0,2,2)(0,1,1)_m$	with log transformation
$\text{ARIMA}(2,1,2)(0,1,1)_m$	with no transformation

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, $\cdots$
- 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.

ARMA in statsmodels

Time-series analysis (tsa) in statsmodels.

- Generate samples using ARMA process.
  `https://www.statsmodels.org/dev/generated/statsmodels.tsa.arima_process.arma_generate_sample.html`
- Generate samples and property test (stationarity, invertibility)
  `https://www.statsmodels.org/dev/generated/statsmodels.tsa.arima_process.ArmaProcess.html`

Demo[1]

- AR model generation and stationarity test in statsmodels
  `http://localhost:8888/notebooks/IL2233VT22/Lec4_arima/AR_models.ipynb`
- MA model generation and invertibility test in statsmodels
  `http://localhost:8888/notebooks/IL2233VT22/Lec4_arima/MA_models.ipynb`

[1] Demo code is only for teacher.

Summary

- The coefficients of AR(p) models determine their stationarity.
- The coefficients of MA(q) models determine their invertibility.
- ARMA(p, q) models assume stationary data.
- ARIMA(p, d, q) models use differencing operation to make data stationary.
- Seasonal ARIMA(p, d, q)(P, D, Q)m models consider seasonality as one major characteristic to be captured for stationarity and model construction.

References

- Hyndman, R. J., Athanasopoulos, G. (2018). ”Forecasting: principles and practice”, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2. `https://otexts.com/fpp3/`
- Box, G. E. P.; Jenkins, G. M.; Reinsel, G. C. (2008). Time Series Analysis, Forecasting and Control (4th ed.). Hoboken, NJ: Wiley. ISBN 9780470272848.
- Brockwell, Peter; Davis, Richard (2009). Time Series: Theory and Methods (2nd ed.). New York: Springer. ISBN 9781441903198.