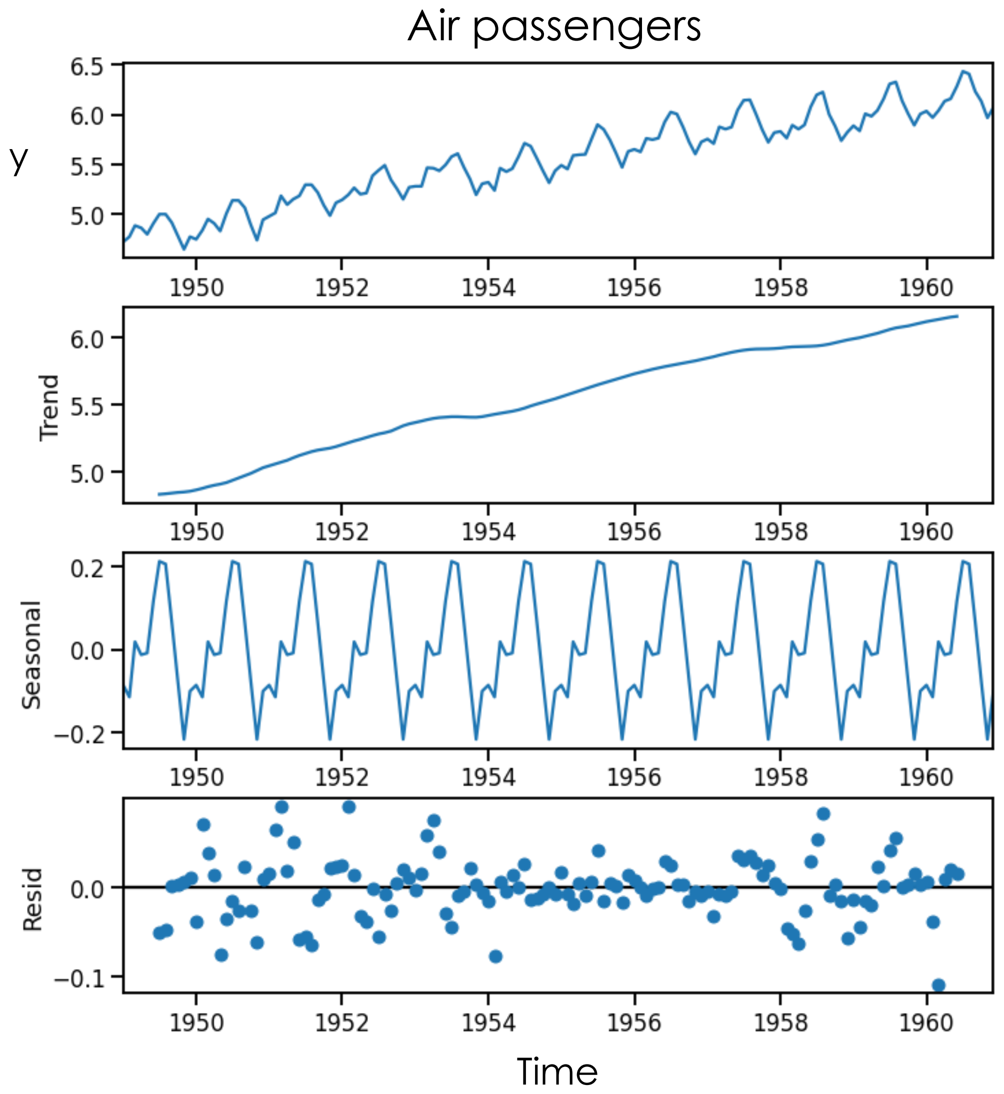
Autoregressive (AR) processes

Lag features

Motivation

- We will want to show how tools we introduce later (e.g., lag plots, correlation functions) behave for time series with various properties.
- Three properties already covered in the course so far are trend, seasonality, and white noise.
- We shall introduce a new property of a time series in this lecture:
  - Autoregressive (AR) property



Scope

- AR processes are a large topic covered in the theory of time series analysis and is broader topic than just forecasting. In this section on lag features the following is in and out of scope:

In scope

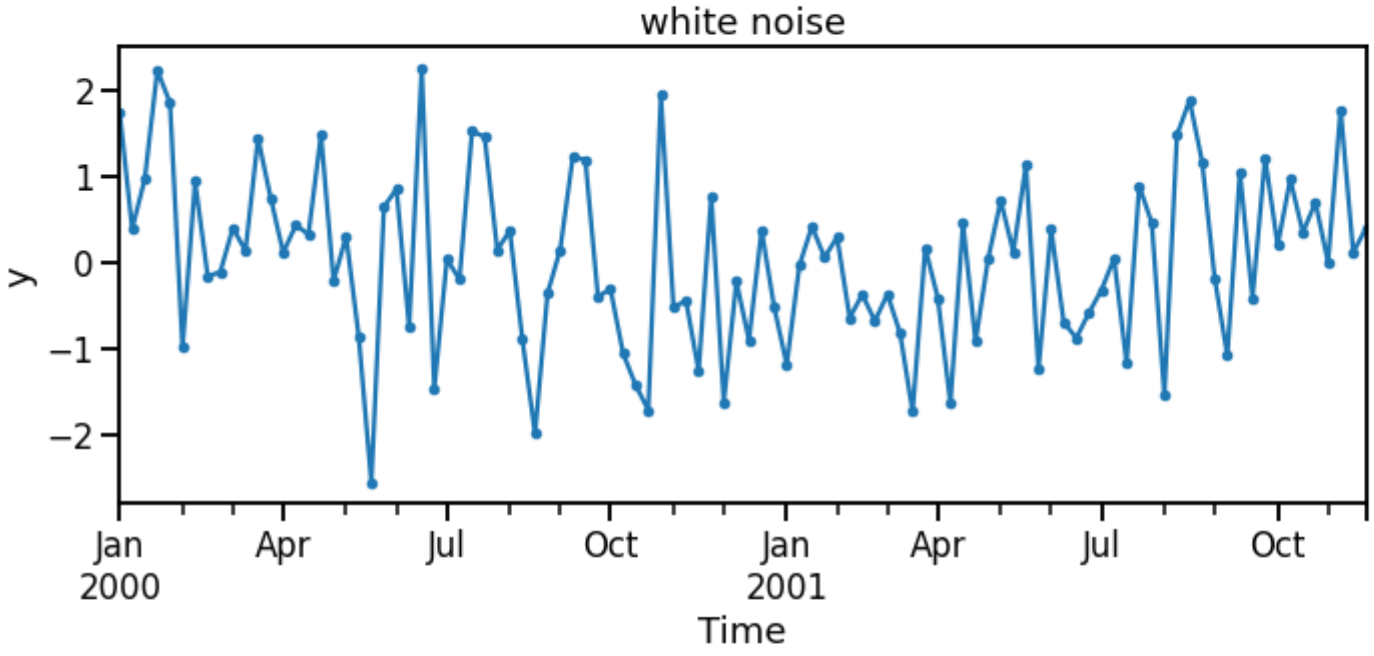
- Discuss the definitions of an AR process.
- Provide an intuition about the behaviour of AR processes.

Out of scope

- Mathematical proofs, derivations, and theorems about AR processes.

White noise

- $y_t = \epsilon_t$ where $\epsilon_t \sim N(0,1)$
- White noise has no predictive information in past values as there is no correlation at any two points in time.



Autoregressive (AR) processes

- Let's generate a time series y_t which is determined only from the previous value y_{t-1}, a constant c, and some white noise ϵ_t:

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

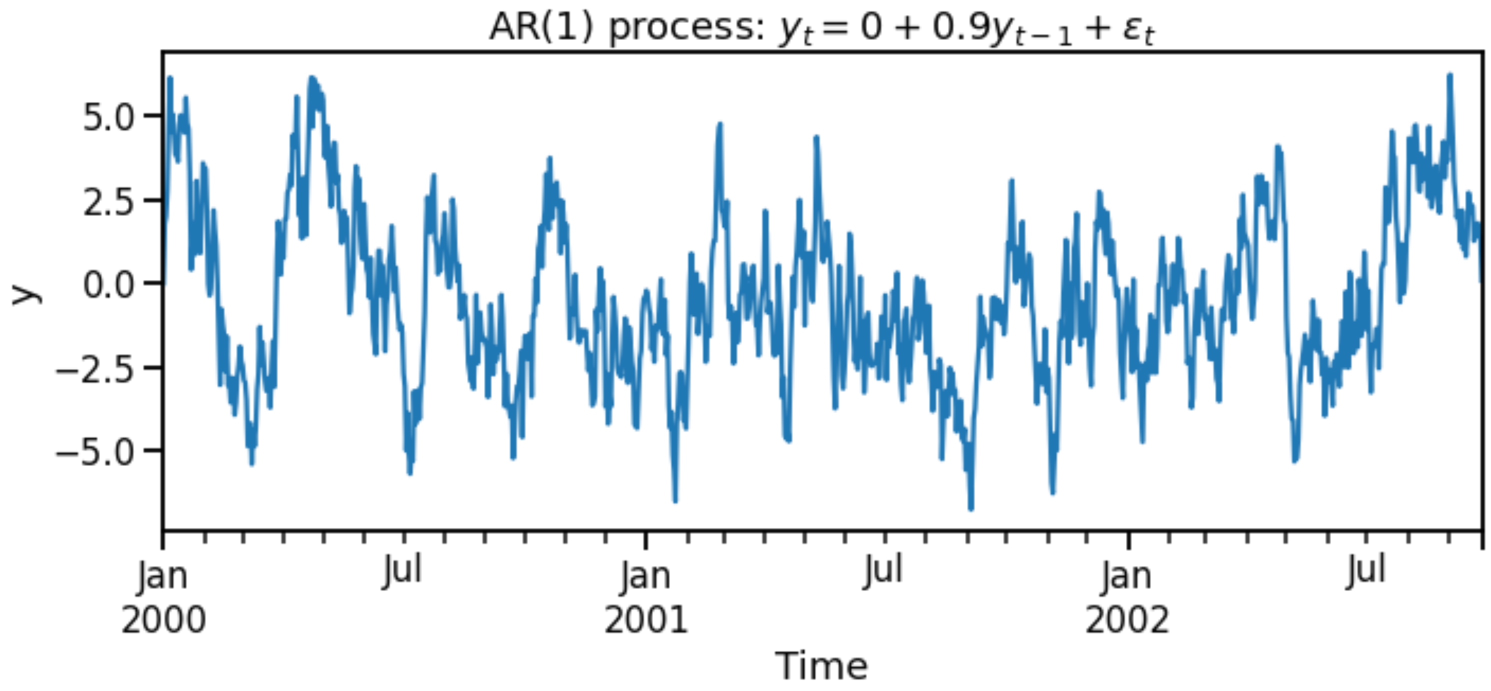
- This is known as an AR(1) process as it depends only on a lag of 1.

```
# Initial parameters
num_timesteps = 1000  # Length of time series we want
phi = 0.9
c = 0
ts = pd.date_range(start="2000-01-01", periods=num_timesteps, freq="D")

# Generate time series
y = np.zeros(num_timesteps)
for t in range(1, num_timesteps):
    noise = np.random.normal()
    y[t] = c + phi * y[t - 1] + noise

AR1 = pd.DataFrame(data={"y": y}, index=ts)
```

Autoregressive (AR) processes

- Let's generate a time series y_t which is determined only from the previous value y_{t-1}, a constant c, and some white noise ϵ_t:

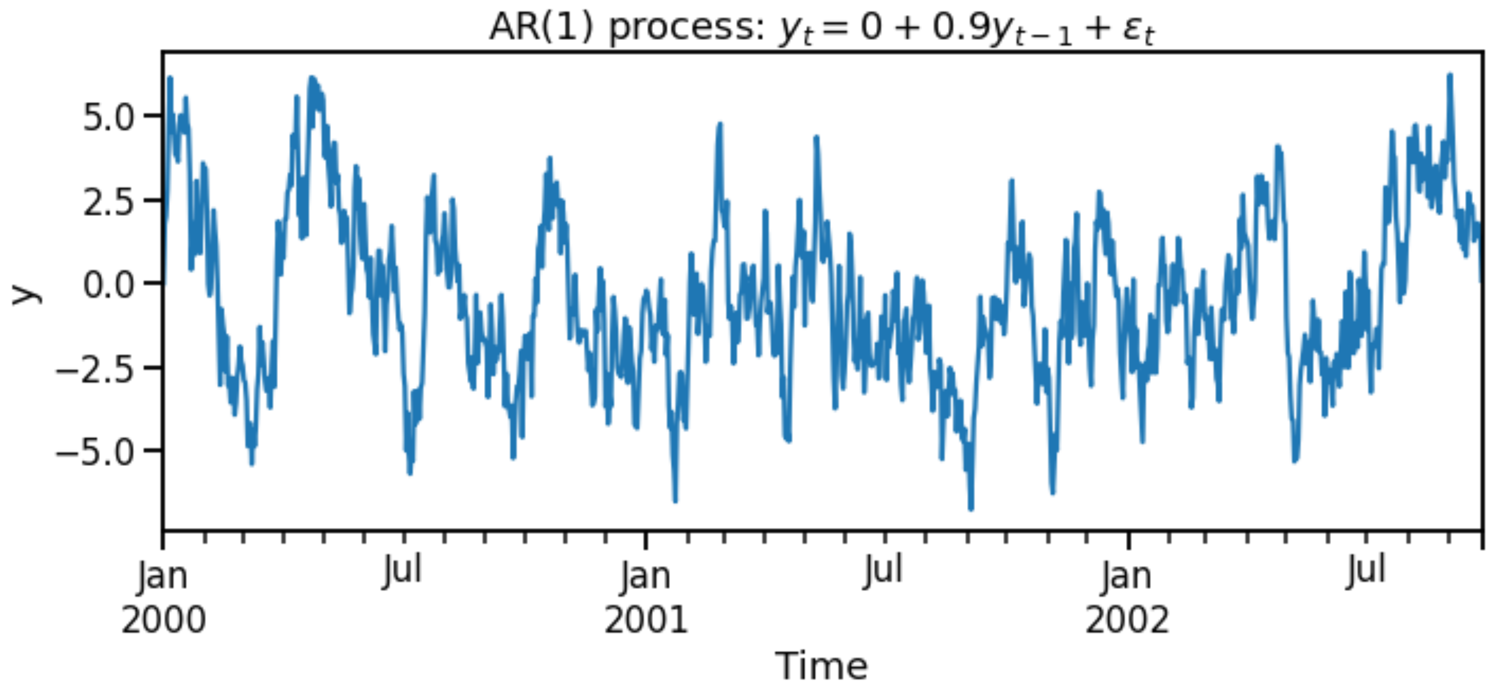
$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$



$$\phi_1 = 0.9$$
$$c = 0$$

Autoregressive (AR) processes

- We are now going to look at how the AR process behaves for different values of ϕ_1 and c.

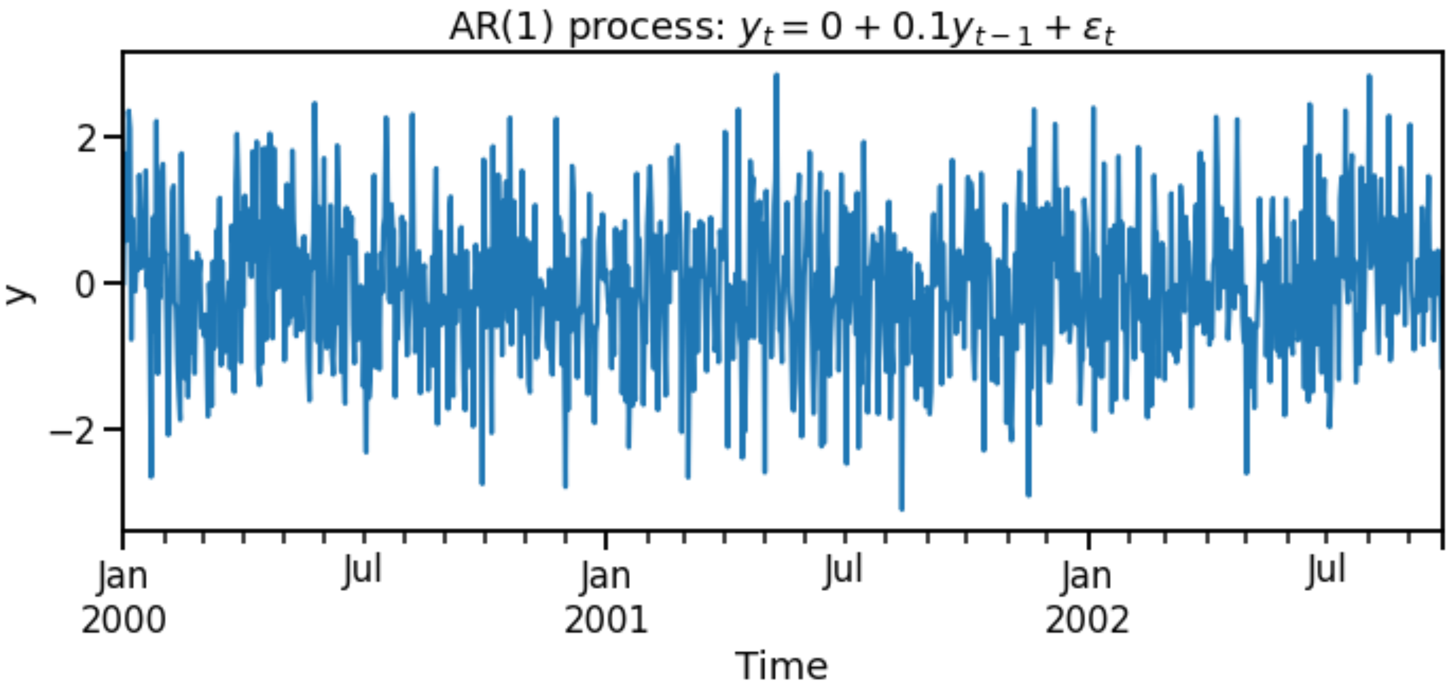


$$\phi_1 = 0.9$$
$$c = 0$$

Autoregressive (AR) processes

- Lag component coefficient is now smaller, hence noise component makes it look noisier.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

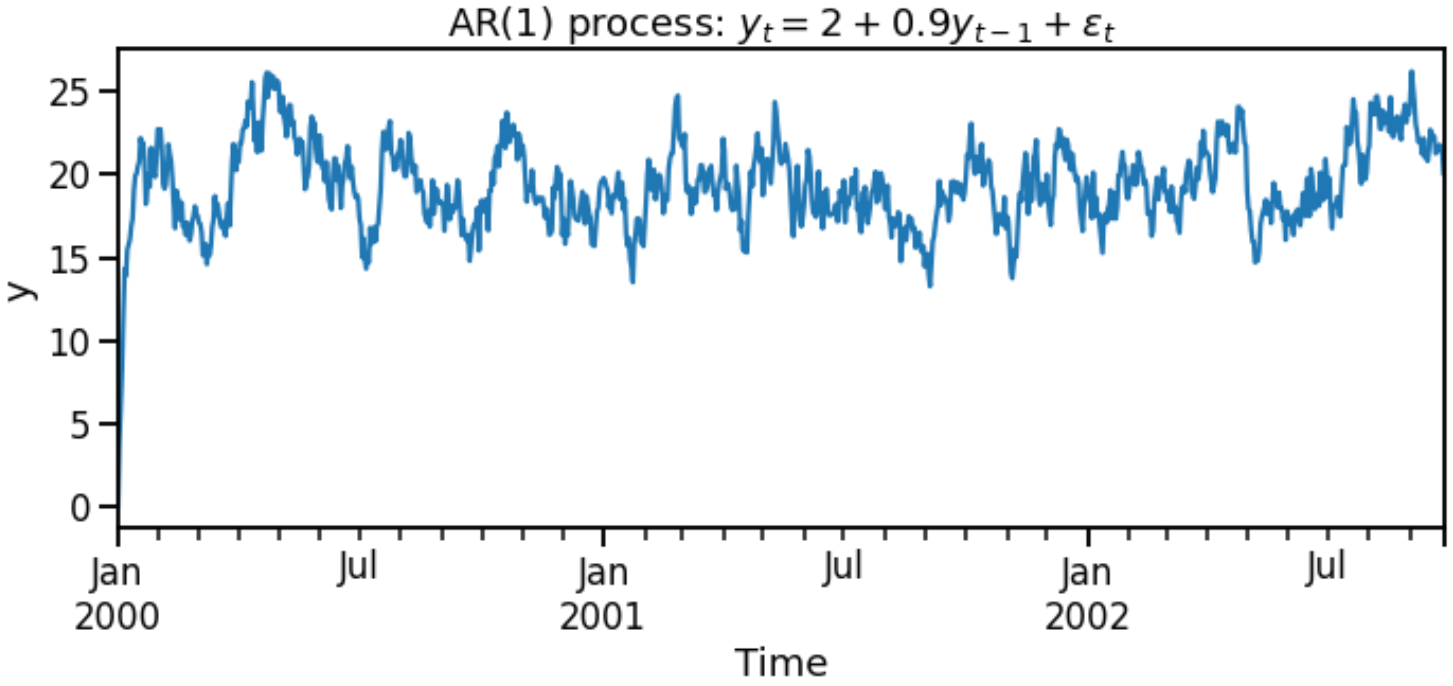


$$\phi_1 = 0.1$$
$$c = 0$$

Autoregressive (AR) processes

- If $c \neq 0$ we see the time series settle at a new baseline. This follows from the fact that the mean of an AR(1) process when $|\phi_1| < 1$ is given by [1]:

$$\mu = \frac{c}{1 - \phi_1}$$



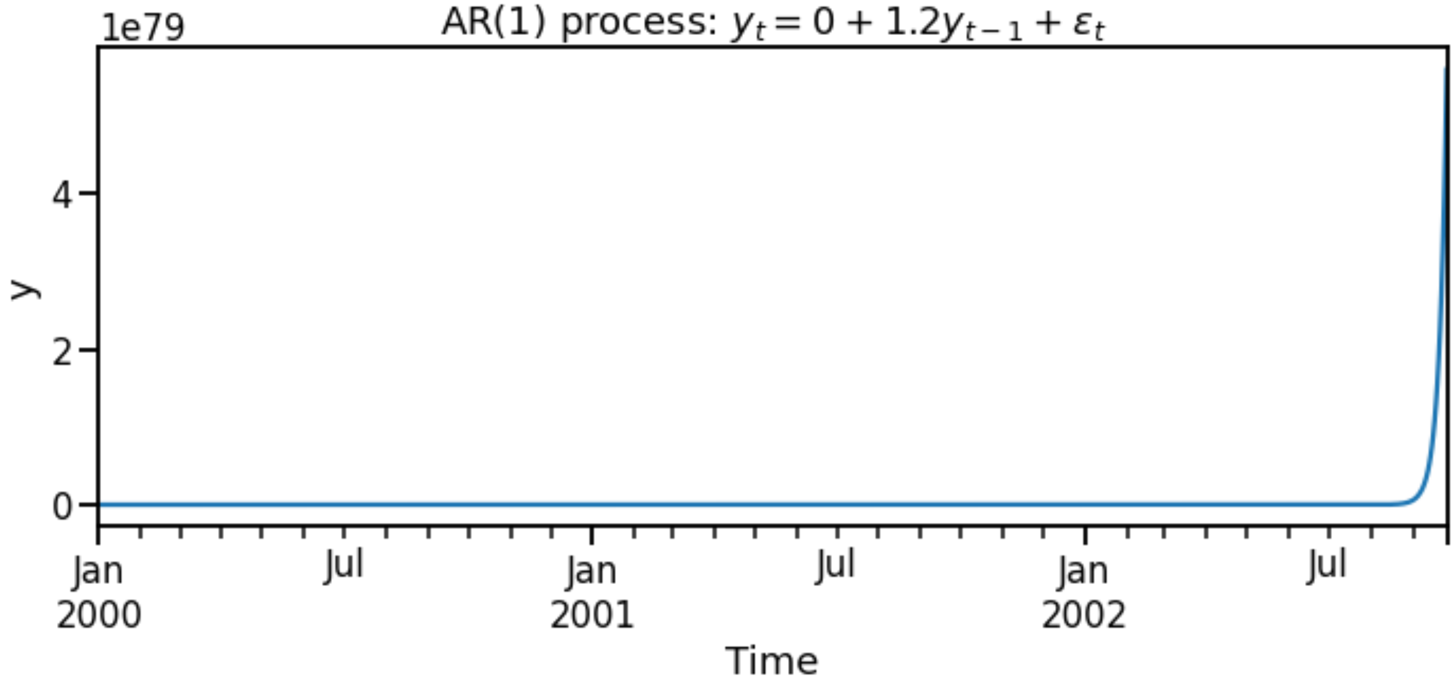
$$\phi_1 = 0.9$$
$$c = 2$$

[1] - George, E. P. "Box. Time series analysis: forecasting and control." (1970).

Autoregressive (AR) processes

- If $\phi_1 > 1$ we see the time series grow exponentially. Let's explain why this happens with an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$



$\phi_1 = 1.2$

$c = 0$

Autoregressive (AR) processes

- If $\phi_1 > 1$ we see the time series grow exponentially. Let's explain why this happens with an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	
2	
3	
4	
5	

$$\phi_1 = 1.2$$
$$c = 0$$

Autoregressive (AR) processes

- If $\phi_1 > 1$ we see the time series grow exponentially. Let's explain why this happens with an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	0.30
2	
3	
4	
5	

$$1.2 \times 0 + 0.3$$

$$\phi_1 = 1.2$$
$$c = 0$$

Autoregressive (AR) processes

- If $\phi_1 > 1$ we see the time series grow exponentially. Let's explain why this happens with an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	0.30
2	1.26
3	
4	
5	

$1.2 \times 0.3 + 0.9$

$\phi_1 = 1.2$
$c = 0$

Autoregressive (AR) processes

- If $\phi_1 > 1$ we see the time series grow exponentially. Let's explain why this happens with an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	0.30
2	1.26
3	2.41
4	
5	

$1.2 \times 1.26 - 0.2$

$\phi_1 = 1.2$
$c = 0$

Autoregressive (AR) processes

- If $\phi_1 > 1$ we see the time series grow exponentially. Let's explain why this happens with an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	0.30
2	1.26
3	2.41
4	3.19
5	

$1.2 \times 2.41 + 0.3$

$\phi_1 = 1.2$
$c = 0$

Autoregressive (AR) processes

- If $\phi_1 > 1$ we see the time series grow exponentially. Let's explain why this happens with an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	0.30
2	1.26
3	2.41
4	3.19
5	3.43

$1.2 \times 3.19 \; - 0.4$

$\phi_1 = 1.2$
$c = 0$

Autoregressive (AR) processes

- If $\phi_1 > 1$ we see the time series grow exponentially. Let's explain why this happens with an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	0.30
2	1.26
3	2.41
4	3.19
5	3.43

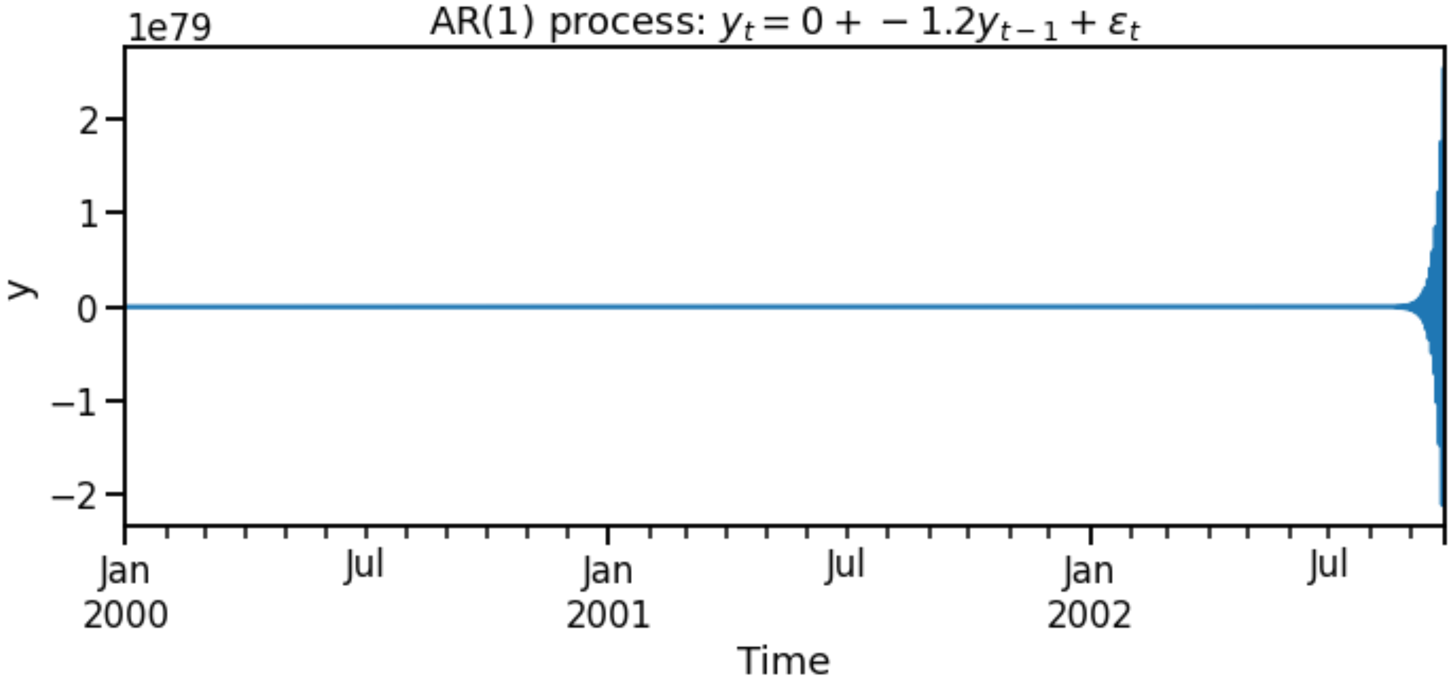
y grows by 20% on average each iteration. Hence, this results in exponential growth which comes from the fact that $\phi_1 > 1$.

$$\phi_1 = 1.2$$
$$c = 0$$

Autoregressive (AR) processes

- If $\phi_1 < -1$ we see the time series grow exponentially but oscillates from a negative to positive value over each iteration. Let's explain this through an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$



$$\phi_1 = -1.2$$
$$c = 0$$

Autoregressive (AR) processes

- If $\phi_1 < -1$ we see the time series grow exponentially but oscillates from a negative to positive value over each iteration. Let's explain this through an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	
2	
3	
4	
5	

$$\phi_1 = -1.2$$
$$c = 0$$

Autoregressive (AR) processes

- If $\phi_1 < -1$ we see the time series grow exponentially but oscillates from a negative to positive value over each iteration. Let's explain this through an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	-0.30
2	
3	
4	
5	

$$-1.2 \times 0 - 0.3$$

$$\phi_1 = -1.2$$
$$c = 0$$

Autoregressive (AR) processes

- If $\phi_1 < -1$ we see the time series grow exponentially but oscillates from a negative to positive value over each iteration. Let's explain this through an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	-0.30
2	1.26
3	
4	
5	

$-1.2\times-0.3 + 0.9$

$$\phi_1 = -1.2$$
$$c = 0$$

Autoregressive (AR) processes

- If $\phi_1 < -1$ we see the time series grow exponentially but oscillates from a negative to positive value over each iteration. Let's explain this through an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	-0.30
2	1.26
3	-1.72
4	
5	

$-1.2 \times 1.26 \ - 0.2$

$$\phi_1 = -1.2$$
$$c = 0$$

Autoregressive (AR) processes

- If $\phi_1 < -1$ we see the time series grow exponentially but oscillates from a negative to positive value over each iteration. Let's explain this through an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	-0.30
2	1.26
3	-1.72
4	2.36
5	

$-1.2 \times -1.72 + 0.3$

$\phi_1 = -1.2$
$c = 0$

Autoregressive (AR) processes

- If $\phi_1 < -1$ we see the time series grow exponentially but oscillates from a negative to positive value over each iteration. Let's explain this through an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	-0.30
2	1.26
3	-1.72
4	2.36
5	-3.23

$-1.2 \times 2.36 - 0.4$

$\phi_1 = -1.2$

$c = 0$

Autoregressive (AR) processes

- If $\phi_1 < -1$ we see the time series grow exponentially but oscillates from a negative to positive value over each iteration. Let's explain this through an example.

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

Time	y
0	0.00
1	-0.30
2	1.26
3	-1.72
4	2.36
5	-3.23

y still grows by 20% on average each iteration. However, the negative sign means that the result oscillates from positive to negative on each iteration.

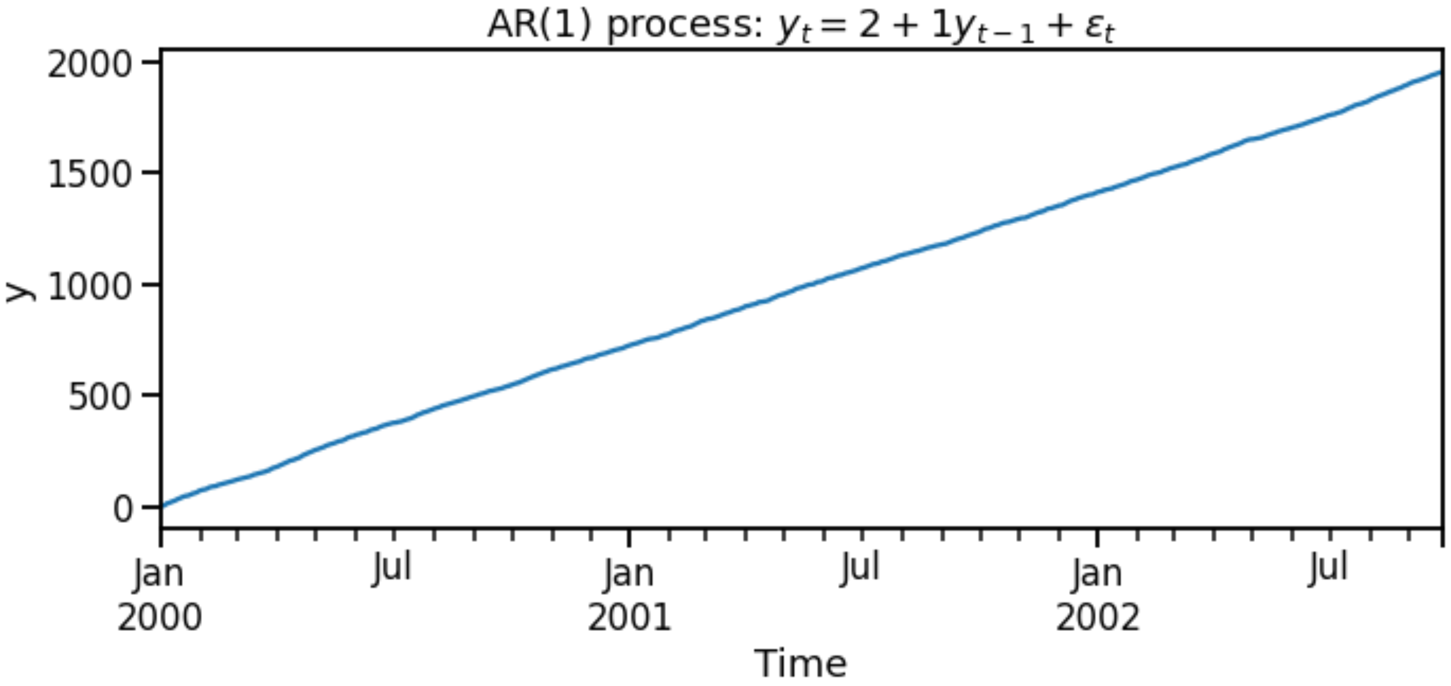
Hence, we still see exponential growth but with oscillations which comes from the fact that $\phi_1 < -1$.

$$\phi_1 = -1.2$$
$$c = 0$$

Autoregressive (AR) processes

- If $\phi_1 = 1$ and $|c| > 0$ we see the time series grow linearly. This occurs because when $\phi_1 = 1$ all we are doing is adding a factor c and some noise ϵ to the previous time step.

$$y_t = c + 1 * y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$



$\phi_1 = 1$
$c = 2$

AR(1) summary

$$y_t = c + \phi_1 y_{t-1} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

- This is an interesting time series because it is designed to be correlated to the most recent lag (i.e., lag of 1).
- $|\phi_1| > 1$: exponential growth, time series is not stationary.
- $\phi_1 = 1$ and $c > 0$: the time series grows linearly, time series is not stationary.
- $|\phi_1| < 1$: the time series varies around a mean value, time series is stationary.
- $\phi_1 < 0$: the time series oscillates between positive and negative values, stationary if $|\phi_1| < 1$.
- We can use an AR(1) process to generate time series where future values are correlated to past values. This allows us to test methods which identify whether a lag of 1 is helpful or not.

AR(p) process

- What about time series which depend on more lags? An AR(p) process depends on p lags and is defined as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t \text{ where } \epsilon_t \sim N(0,1)$$

- Recall that an AR(1) process requires $|\phi_1| < 1$ to be stationary. For an AR(p) process there are much more complex requirements on all the coefficients to ensure the process is stationary (see references for more detail).

- These time series are interesting because by design they depend on multiple lagged values.

- Hence, an AR(p) process is a good test case for methods that select which set of lags are important or not. This is how we use AR processes in the course!

- We will see them in action in the notebooks.

References

1. George, E. P. "Box. Time series analysis: forecasting and control." (1970).
2. Brockwell and Davis, 2010. Introduction to Time Series and Forecasting, 2nd edition.

Summary

An autoregressive (AR) process is a class of time series where future values depend on past values and white noise.

An AR process is determined by previous values and therefore will be correlated to lag values of itself. Hence, lag features should help predict an AR process.

AR processes provide time series which we can use to test methods which identify helpful lags.