香港城市大學
City University of Hong Kong

IS6400:
Business Data Analytics

Time Series II: AutoReg

專業 創新 胸懷全球
Professional · Creative
For The World

Before class

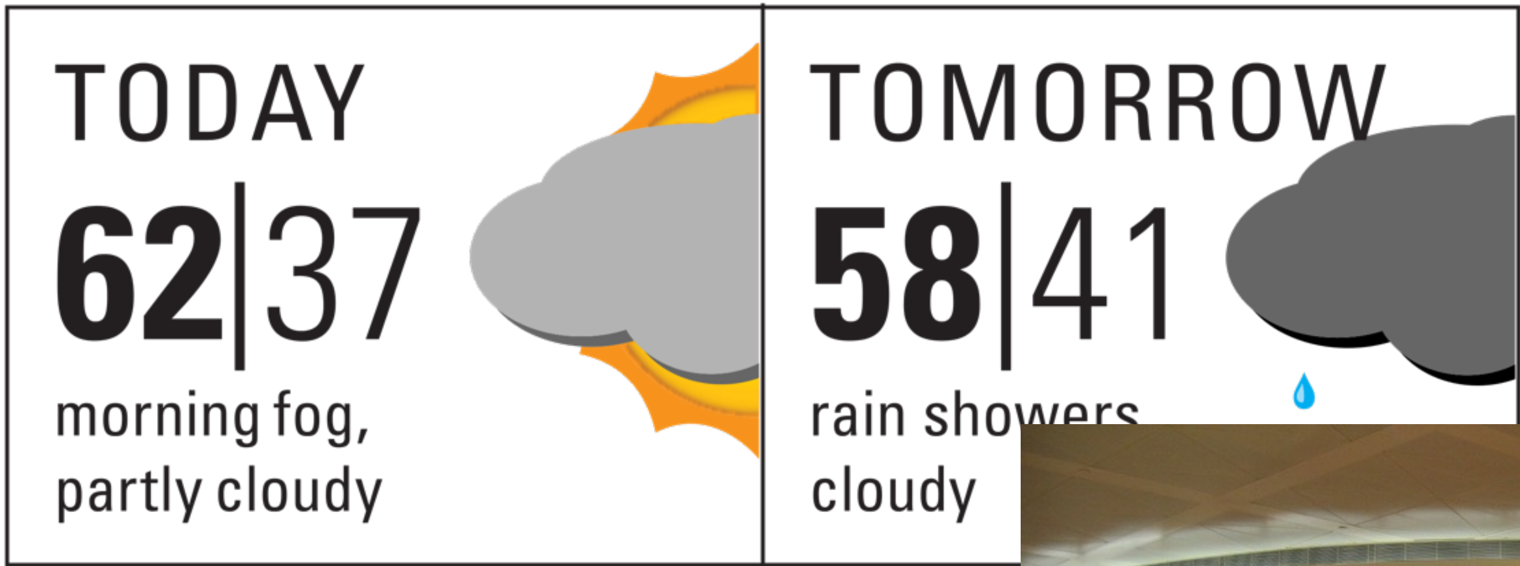
- Homework is available!

- Week 6 redo:
  - Same materials but with slides
  - Recordings are available

- When using Google Colab, we need to update "statsmodels" (code included in the notebook):

```
# if running in Google Colab, please run the following line
# !pip install --upgrade statsmodels
```

Forecasting



What we will talk about

- Time series features
- Exponential Smoothing
- AutoRegresison

Week 7

Week 8

Our journey begins with…

- What is time series data?

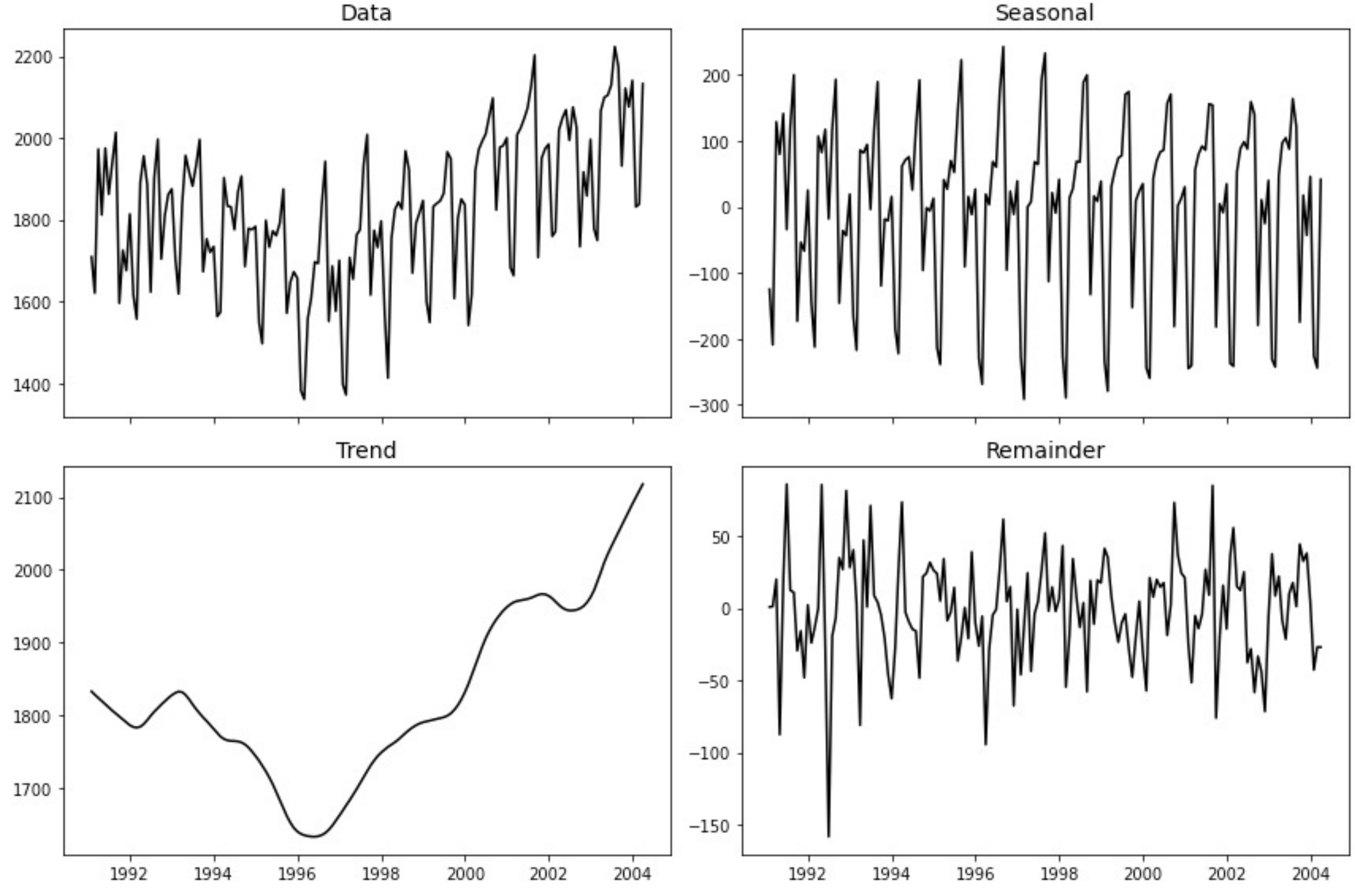
A summary of smoothing

Trend Component	Seasonal Component		
	N	A	M
	(None)	(Additive)	(Multiplicative)
N (None)	(N,N)	(N,A)	(N,M)
A (Additive)	(A,N)	(A,A)	(A,M)
A_d (Additive damped)	(A_d,N)	(A_d,A)	(A_d,M)

From Table 8.5 of Hyndman and Athanasopoulos (2018)

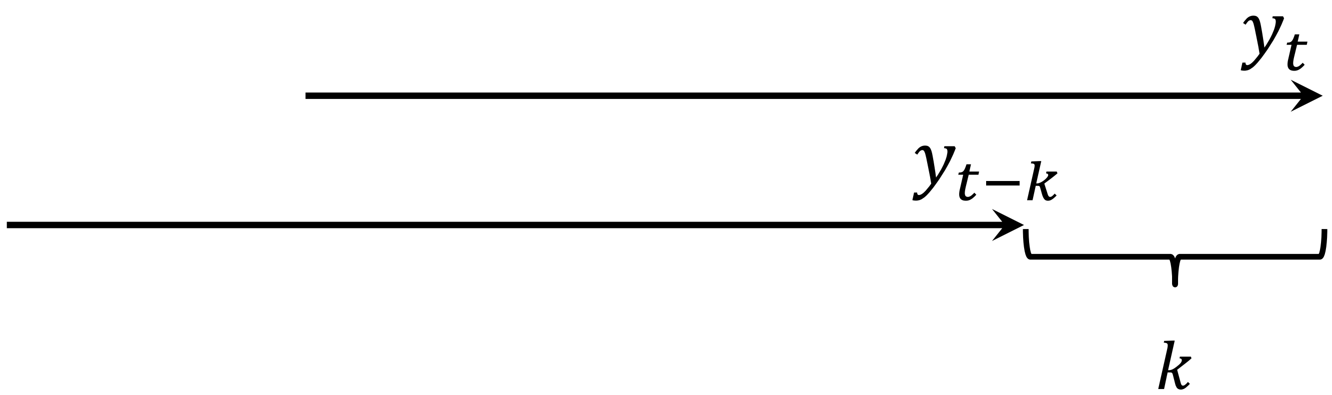
What features did we talk about?

- Trend + Seasonality

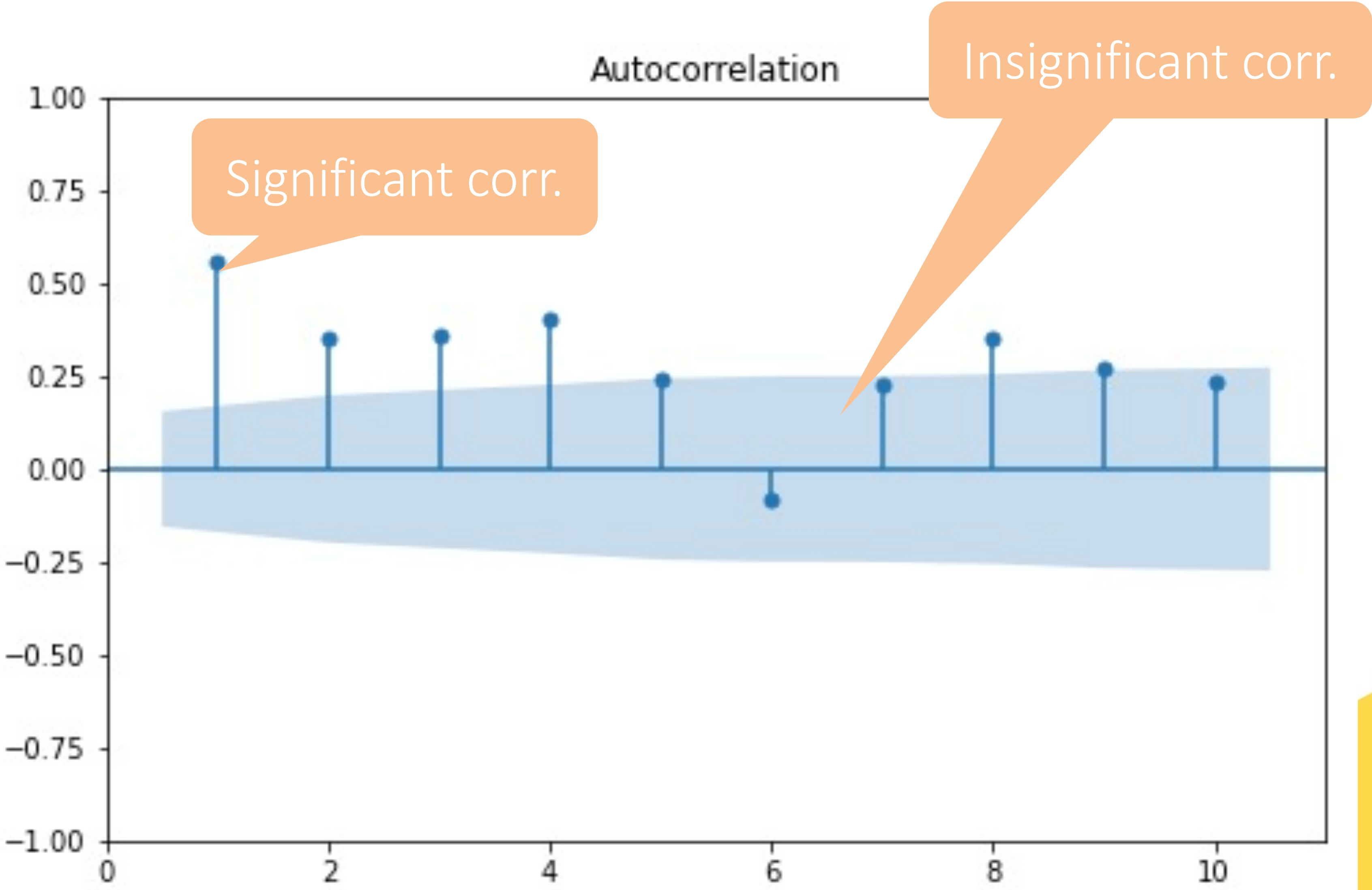


What features did we talk about?

- Autocorrelation:
  - Correlation with lagged self



Autocorrelation Diagram

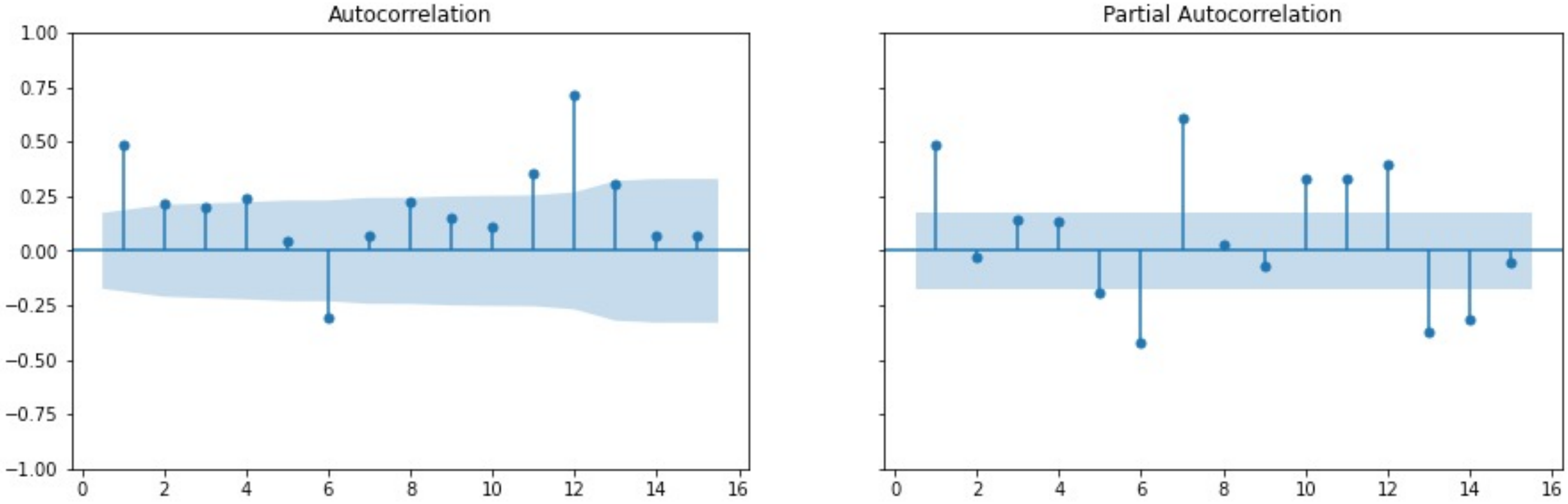


Partial Autocorrelation

- Partial autocorrelation of lag k:
  - Correlation between y_t and y_{t-k}
  - Controlling for prior correlations

- Example
  - Lag-1 partial autocorr is the same as Lag-1 autocorr.
  - Lag-2 partial autocorr is the correlation between y_t and y_{t-2} controlling for autocorr between y_t and y_{t-1}

Partial Autocorrelation Diagram



Our very first AutoReg model: AR(p)

Constant "trend": like an intercept

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

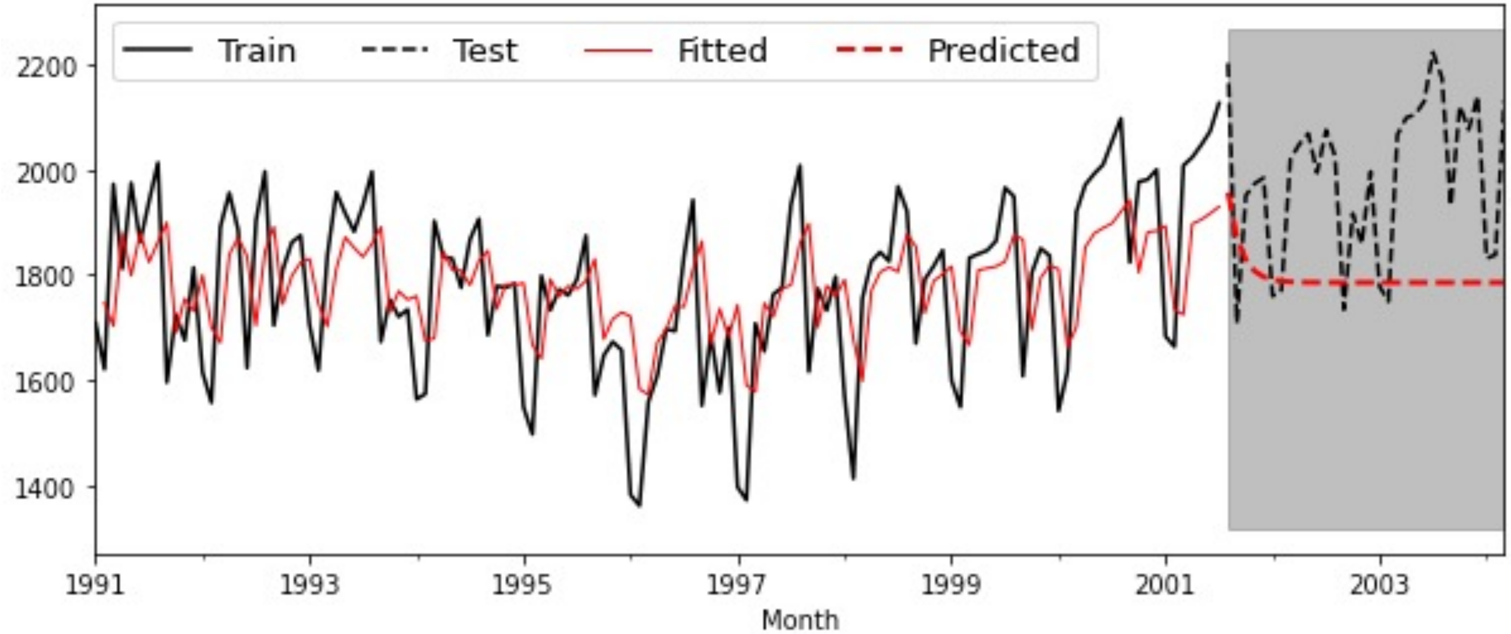
$$= const + \epsilon_t + \sum_{i=1}^{p} \phi_i y_{t-i}$$

Isn't this familiar? Do you remember the simple exp smoothing method?

$$\phi_i = \alpha(1-\alpha)^i$$

$$\hat{y_{t+1}} = \alpha y_t + \alpha(1-\alpha) y_{t-1} + \alpha(1-\alpha)^2 y_{t-2} + \ldots$$

Try it out: Set p to 1



		lo. Observations:	127
Model:	AutoReg(1)	Log Likelihood	-805.182
Method:	Conditional MLE	S.D. of innovations	144.228
Date:	Tue, 01 Mar 2022	AIC	1616.364
Time:	16:29:05	BIC	1624.873
Sample:	02-28-1991	HQIC	1619.821
	- 07-31-2001		

	coef	std err	z	P>|z|	[0.025	0.975]
const	889.4105	140.756	6.319	0.000	613.535	1165.286
Ridership.L1	0.5018	0.079	6.368	0.000	0.347	0.656

Parameter Selection: Pick a p

- The interpretations of PACF and AR model coefficients are the same…

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

- The correlation between lagged variables controlling for others…

Then we can just pick a number by looking at the PACF

Intuition for the AR identification

- How do we interpret linear regression coefficients?

$$\mathrm{IQ} = 5.95 \times \mathrm{MomHighSchool} + 0.56 \times \mathrm{MomIQ} + 25.73$$

Every one unit increase in "MomIQ" is correlated with 0.56 unit increase in "IQ" controlling for "MonHighSchool"

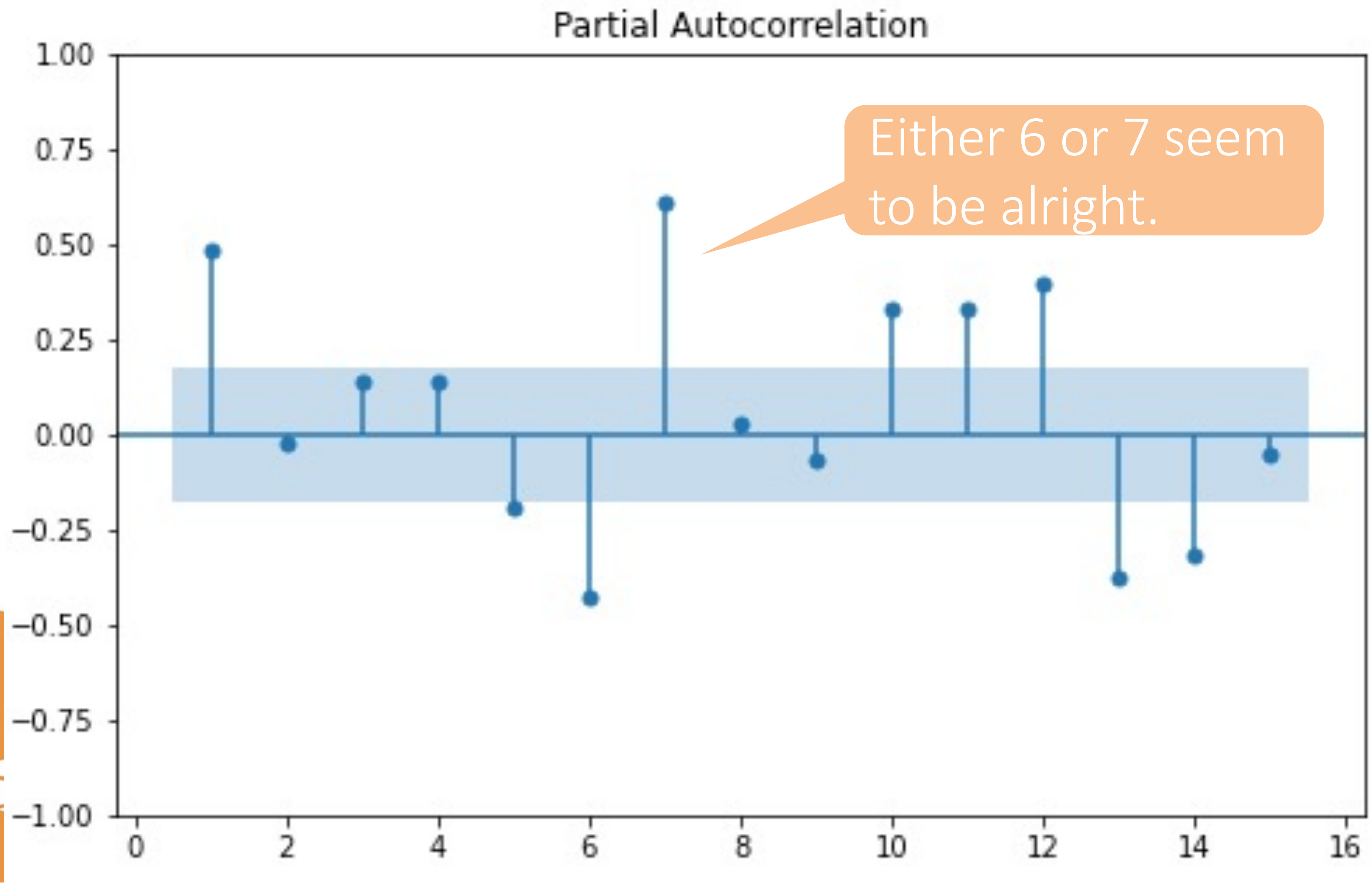
- Same for the AR model!

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

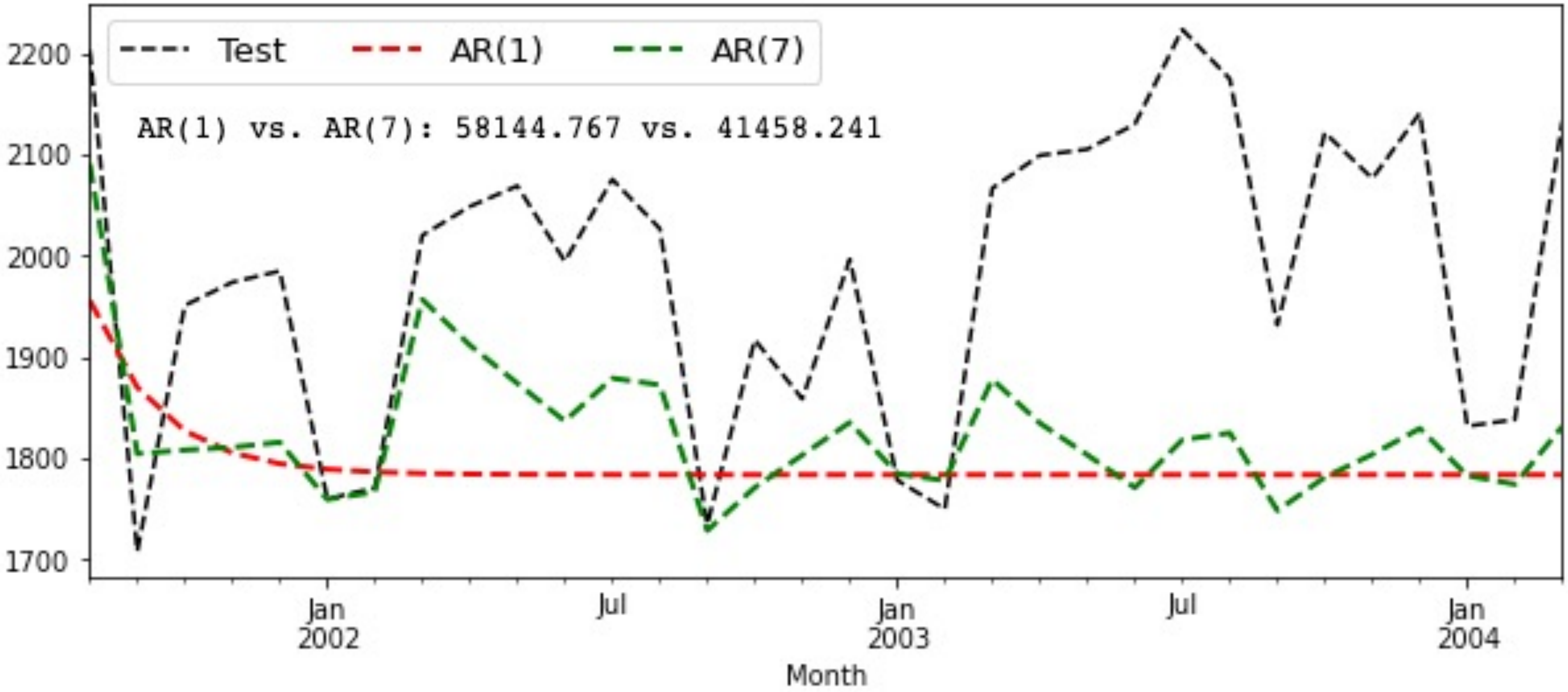
This is just the same as PACF for lag-p

Correlation between lag-p time series y_{t-p} and the time series value y_t controlling for y_{t-1}, y_{t-2}, …, which are lower orders of lagged values…

Parameter Selection: Pick a p



Let's see if it gets better with $p = 7$



Moving Average: MA(q)

$$y_t = const + \epsilon_t + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

Intuition: we can forecast future time series values based on our past mistakes

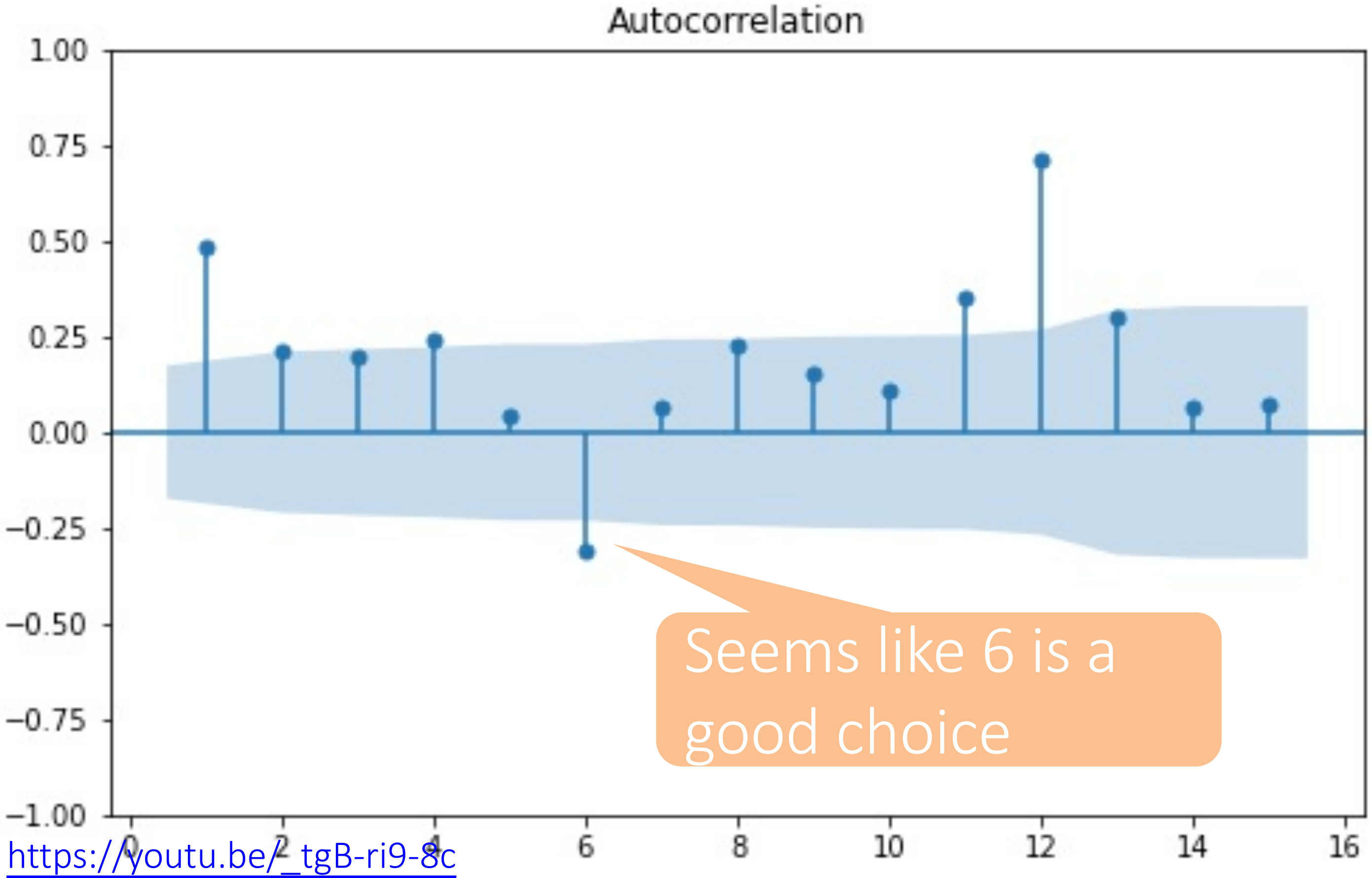
$$\begin{aligned} y_t &= \phi_1 y_{t-1} + \epsilon_t + const \\ &= \phi_1(\phi_1 y_{t-2} + \epsilon_{t-1}) + \epsilon_t + const \\ &= \phi_1^2 y_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t + const \\ &= \phi_1^2(\phi_1 y_{t-3} + \epsilon_{t-2}) + \phi_1 \epsilon_{t-1} + \epsilon_t + const \\ &= \phi_1^3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

The equivalence between AR(1) and MA(∞).

This is optional.

Parameter Selection: Pick a q

- For this model, we need to look at the ACF (autocorrelation):



Great explanation: https://youtu.be/_tgB-ri9-8c

Intuition for MA identification

$$y_t = const + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \theta_3\epsilon_{t-3} + \cdots + \theta_{t-q}\epsilon_{t-q}$$

$\epsilon_j \sim N(0, \sigma_j^2)$;
$j \in \{t-1, \ldots, t-q\}$

This tells us that the predictors in this MA model are independent!

θ_j

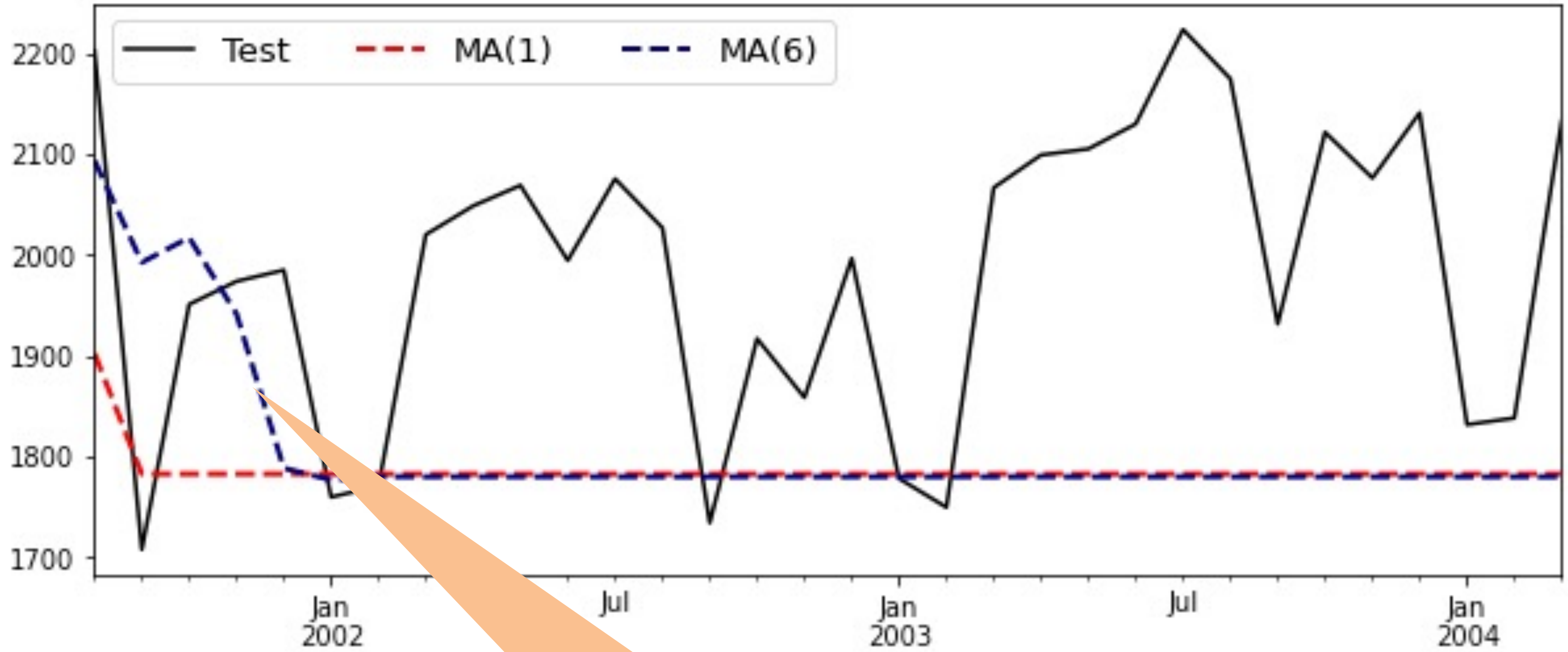
$\epsilon_j = y_j - \hat{y}_j$

$Corr(y_t, y_j)$

The coefficient is essentially the correlation between these two lagged series, which is therefore ACF!

Parameter Selection: Pick a q

- Compare performance with different parameters for the MA model.



Slightly better.

CityU

ARMA: The combination of AR and MA

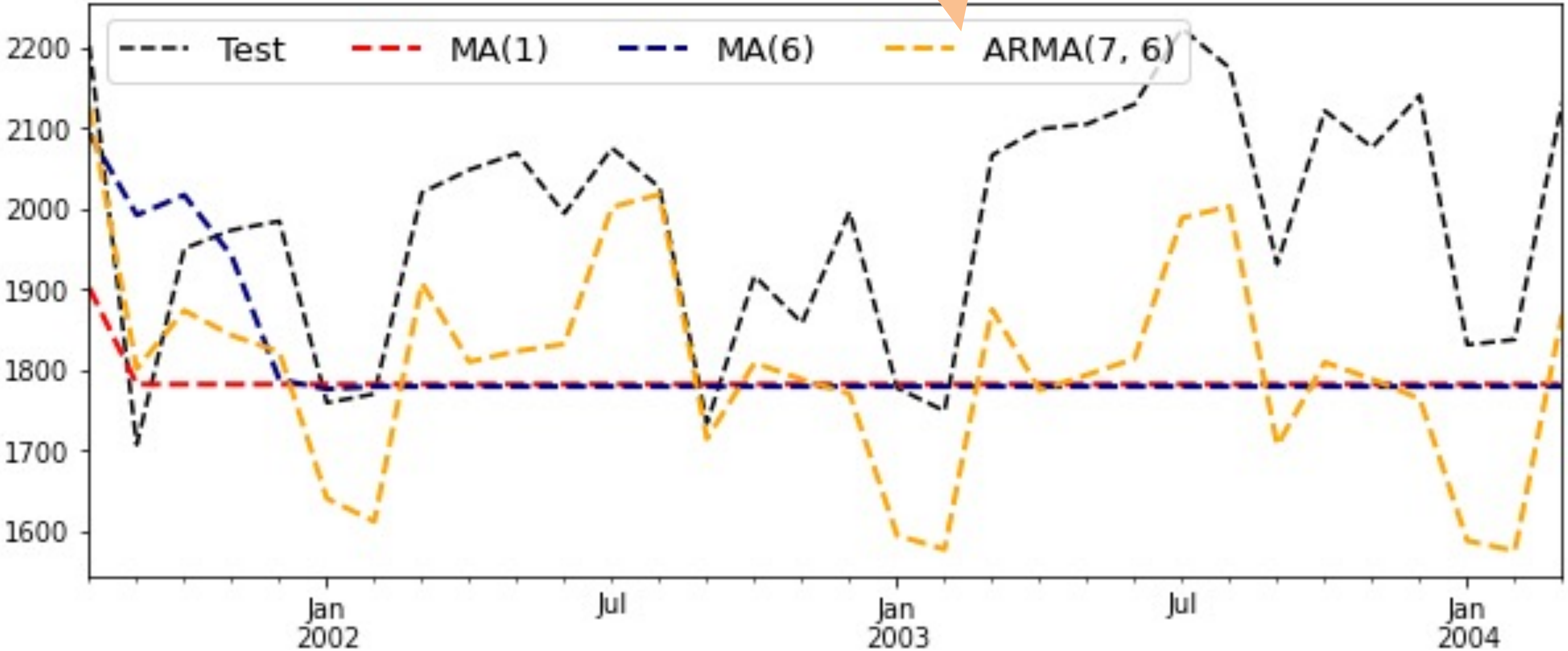
Intuition: we can forecast future time series values based on past patterns.

$$y_t = const + \epsilon_t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

Intuition: we can forecast future time series values based on past patterns.

Use $q = 6; p = 7$

Much better!



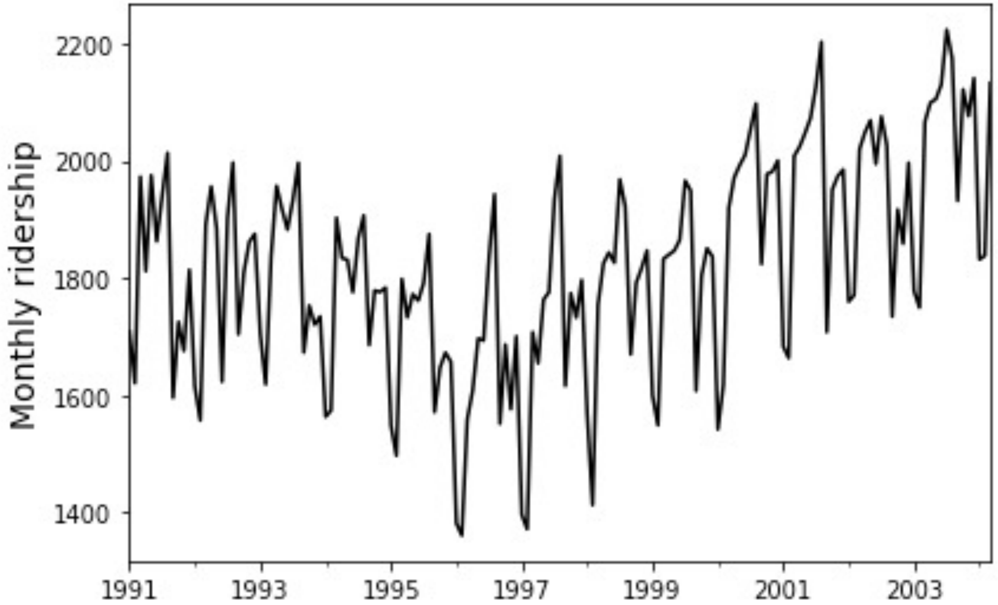
Well, what else can we add to ARMA?

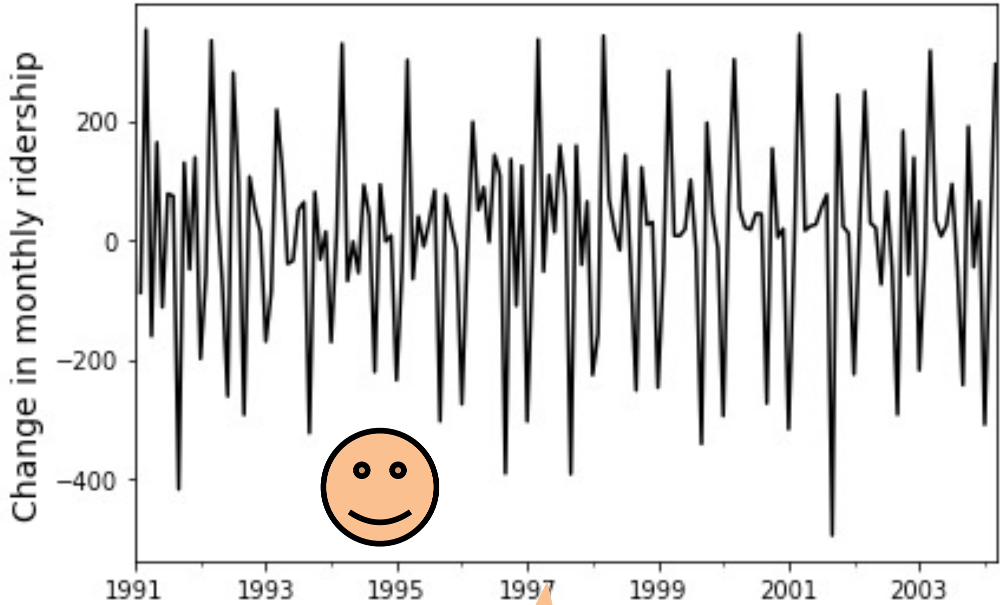
- Did we talk about trend? Not yet!

- Is our dataset stationary?
  - In fact, most real-world time series datasets are non-stationary!

Stationarity and Differencing

Which one is stationary?





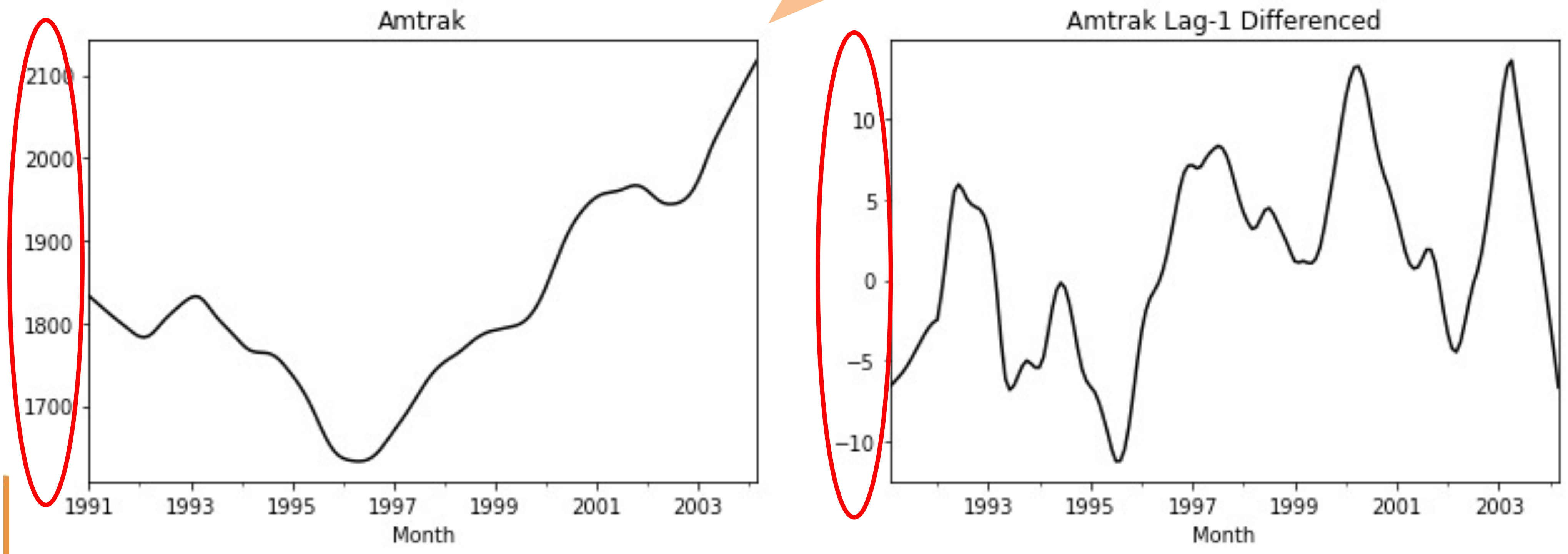

```
Before differencing: 0.010
After differencing: 0.100
```

Check p-value of the KPSS test!

Lag-1 differencing: De-trend!

Compare the trend component

Trends by STL decomposition



How about we add this to ARMA?

- AR-I-MA, where "I" is for "integrated", the reverse of differencing

$$y_t = const + \epsilon_t + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

$$\downarrow$$

Differenced time series values

$$y'_t = const + \epsilon_t + \sum_{i=1}^{p} \phi_i y'_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j}$$

ARIMA: The Final Model

- Three parameters ARIMA(p, d, q):
  - The AR order p
  - The differencing order d
  - The MA order q

- In Python, it is easy to use:

```
arima = sm.tsa.SARIMAX(amtrak_train, order=(p, d, q), trend='c').fit(disp=False)
```

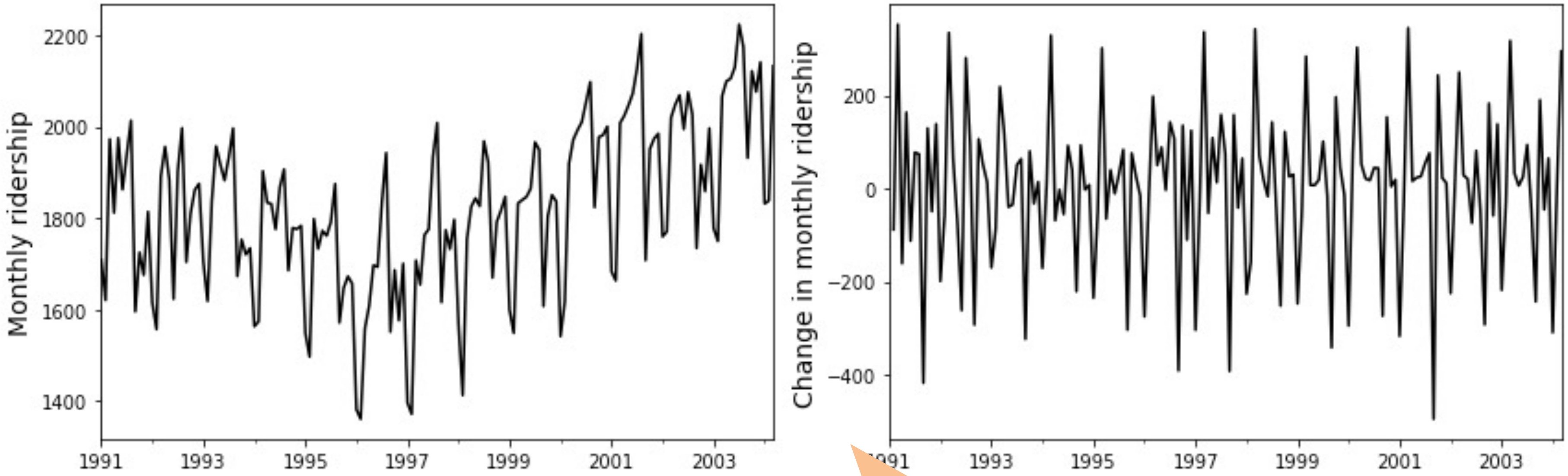
ARIMA: Parameter Selection

- The AR order p
  - Look at PACF cutoff.

Sometimes we may even need double differencing!

- The differencing order d
  - Do we need to de-trend?
- The MA order q
  - Look at ACF cutoff.

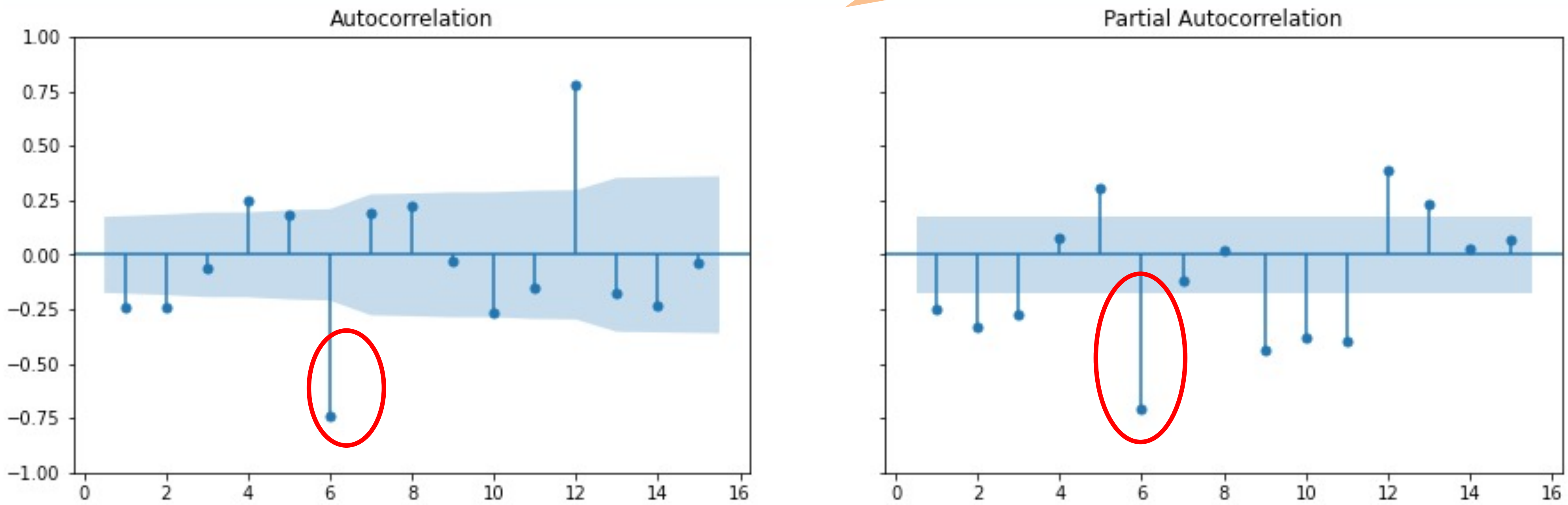
ARIMA identification in practice



What does this figure tell us?
(which parameter(s) can we decide?)

ARIMA identification in practice

The ACF and PACF of the differenced (if any) time series

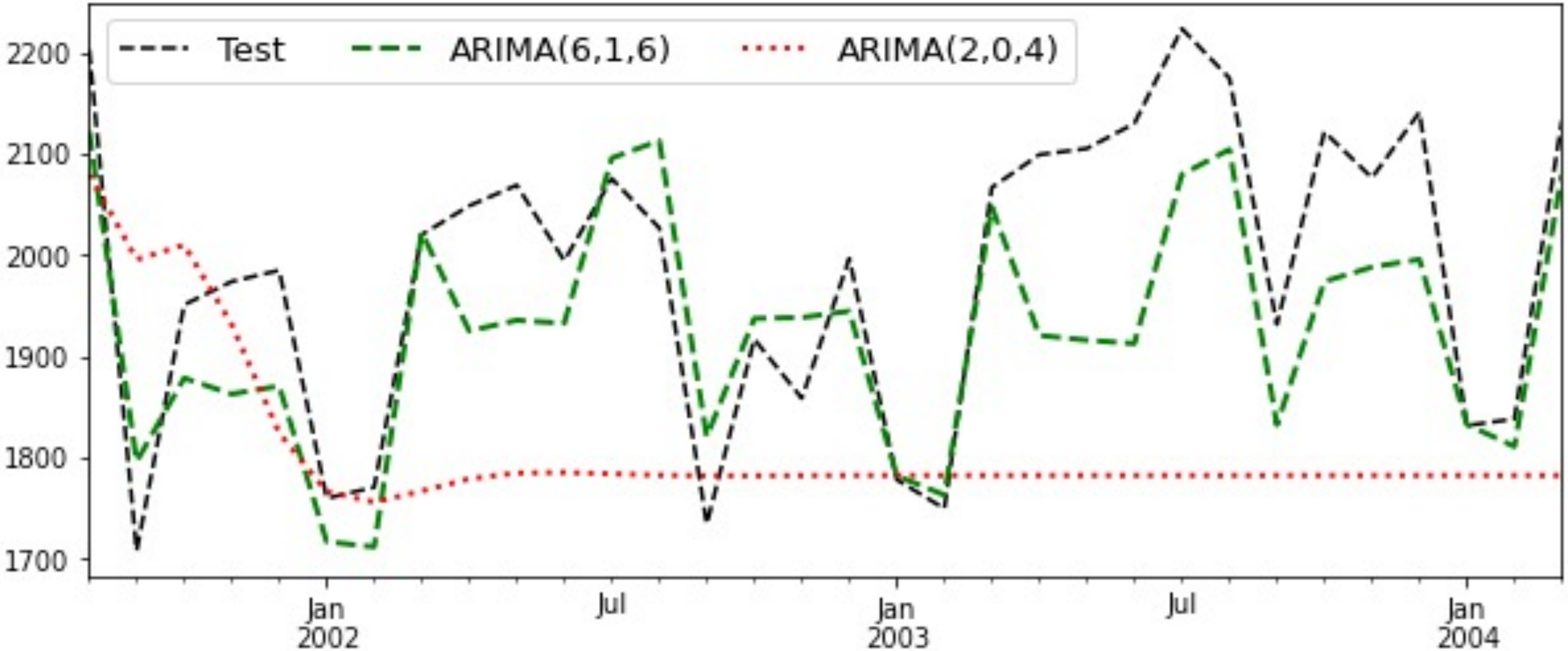


What does this figure tell us?
(which parameter(s) can we decide?)

ARIMA Demo

Hand-picked model

Arbitrarily determined parameters



Advanced Topics: Seasonal ARIMA

- What we have so far: ARIMA(p, d, q)

- What if there is any seasonality?
  - Add more parameters!
  - How?

Advanced Topics: Seasonal ARIMA

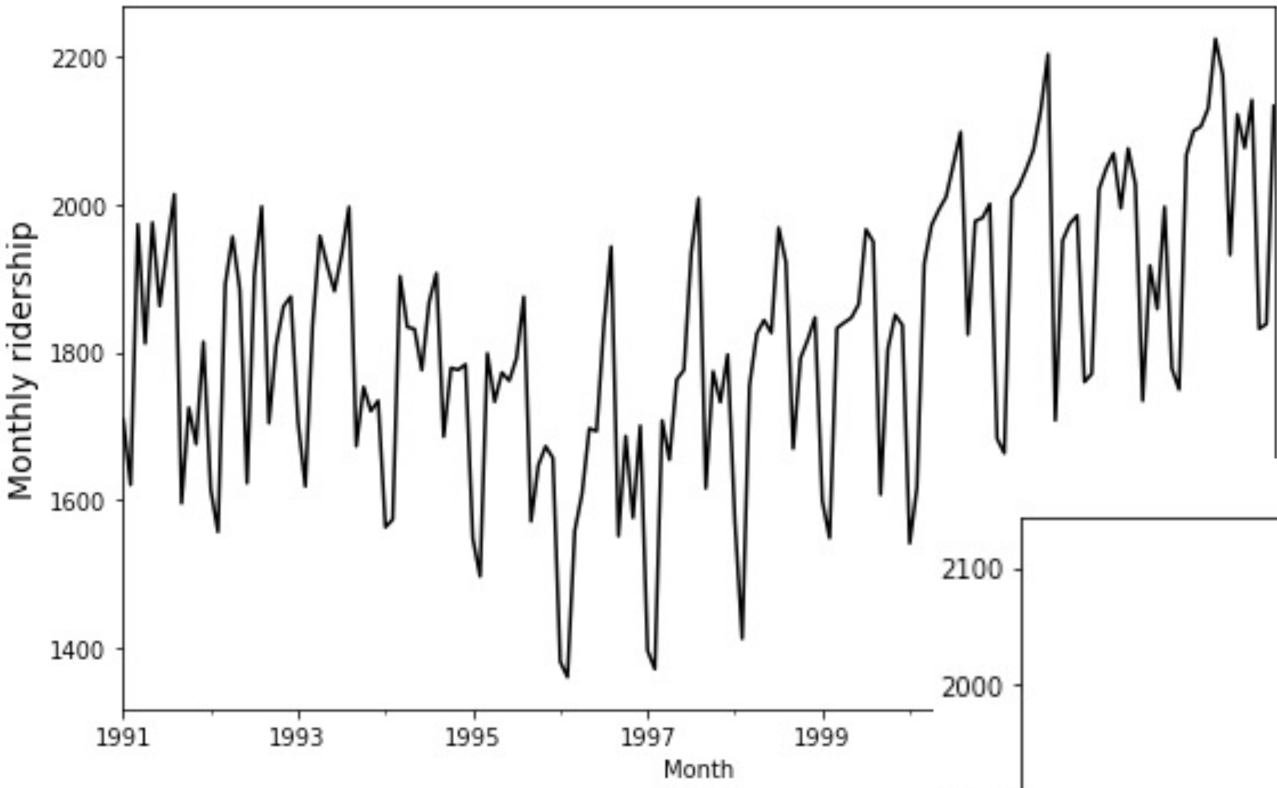
- Add seasonal parameters to ARIMA(p, d, q)

- The meta model:

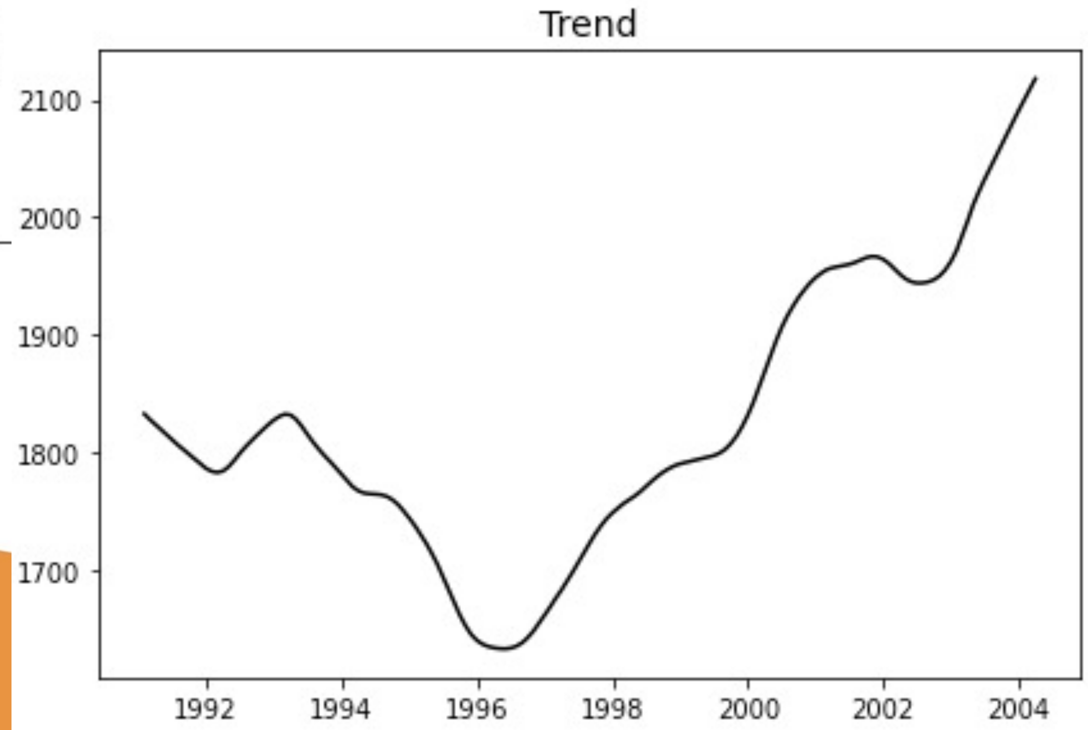
$$\text{ARIMA}(p, d, q)(P, D, Q)_m$$

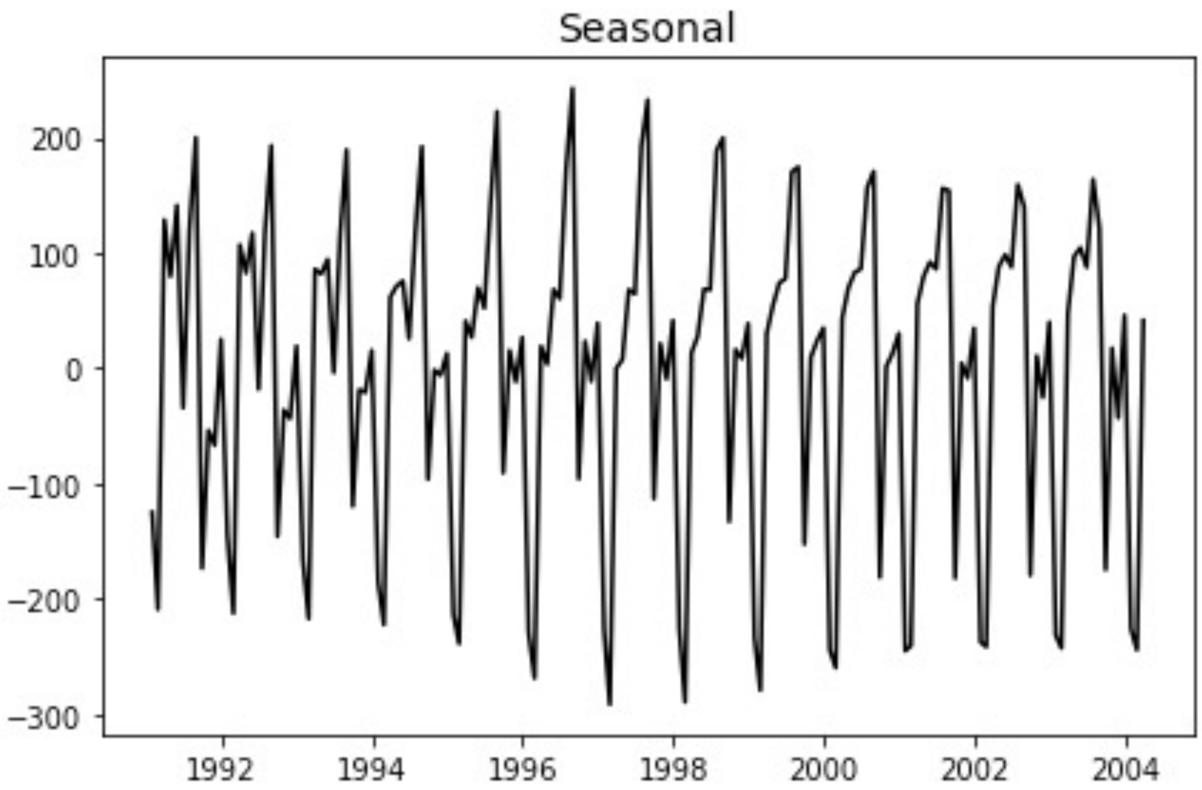
These are similar to the previous parameters but focus on seasonal components!

Again: the Amtrak ridership data

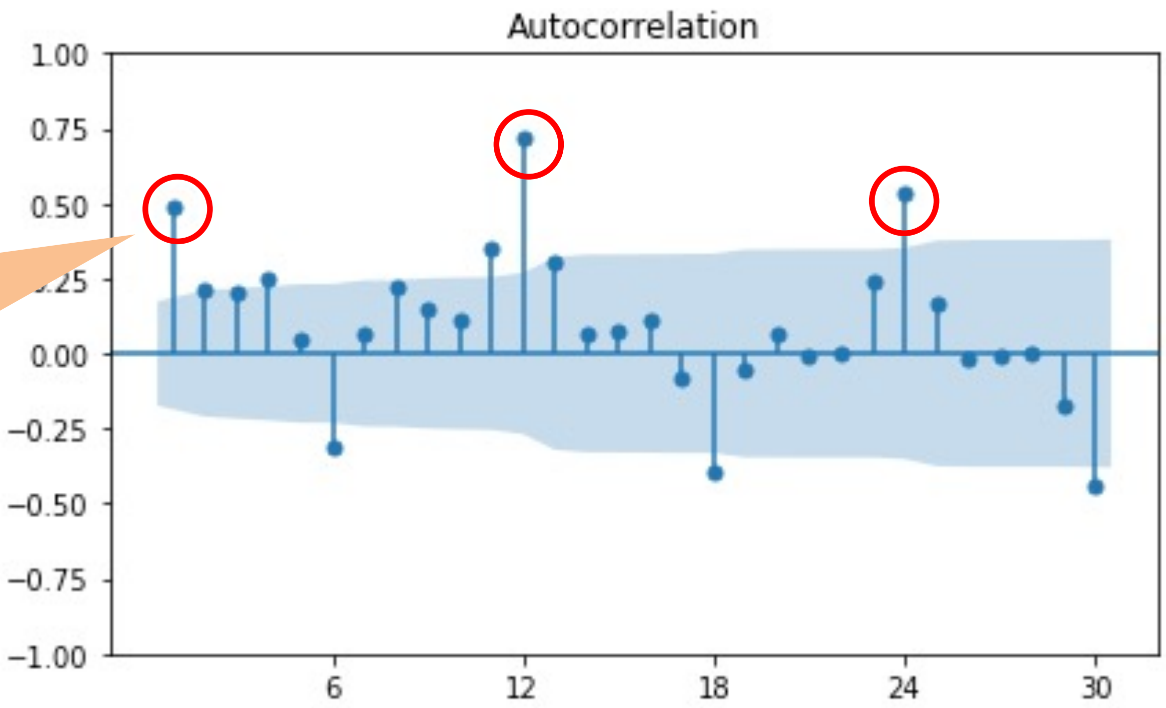


There is a trend… We can de-trend by lag-1 differencing…





There seems to be having a 12-month periodic cycle, i.e., seasonality



We could do lag-m differencing to de-seasonality!

Over-differencing

- Currently, we find a need for double differencing:
  - At the level of lag-1 to de-trend
  - At the level of lag-12 to de-seasonality

- But we do not want to do over-differencing.

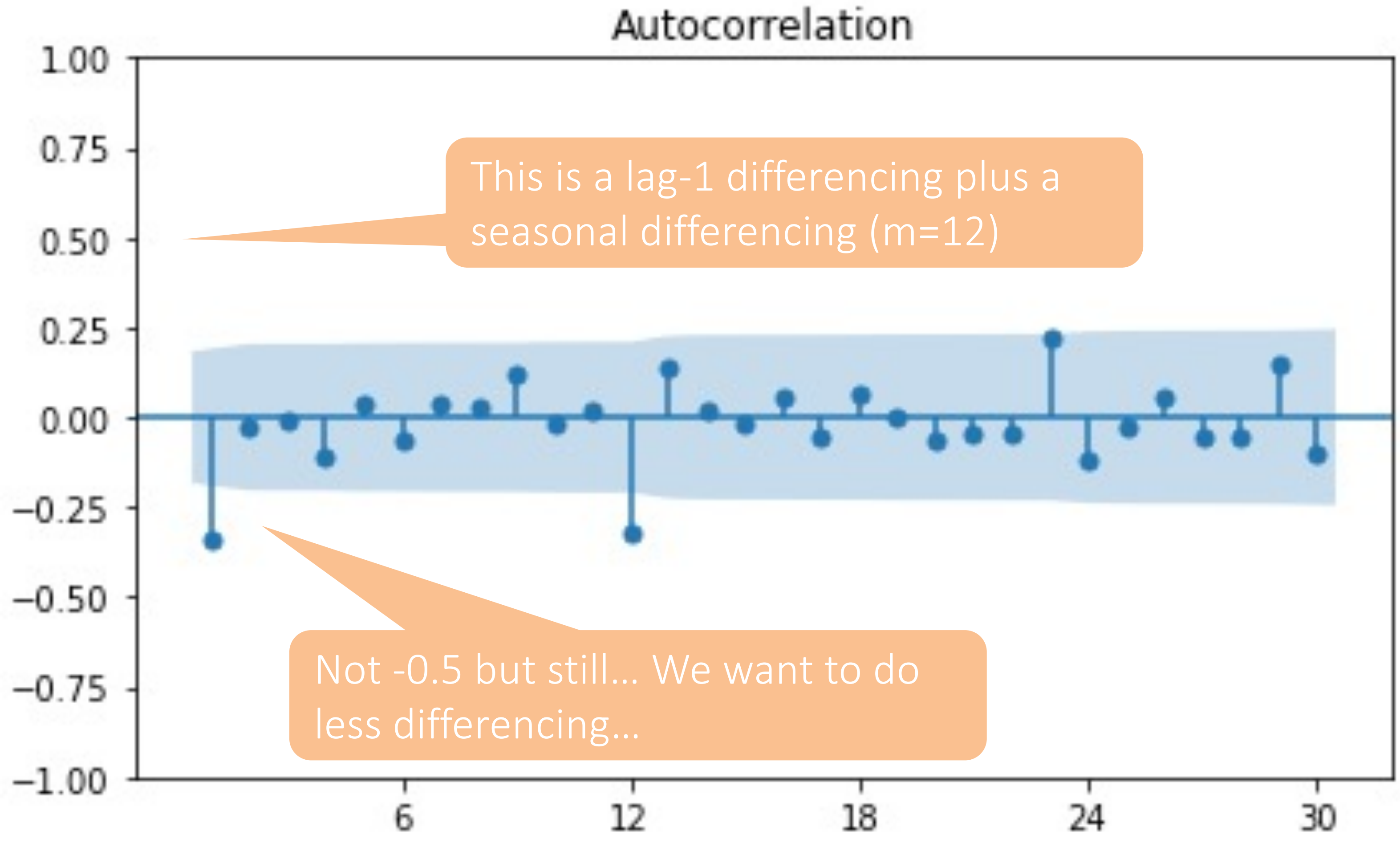
If the lag-1 autocorrelation is -0.5 or more negative, the series may be over-differenced.

BEWARE OF OVERDIFFERENCING!!

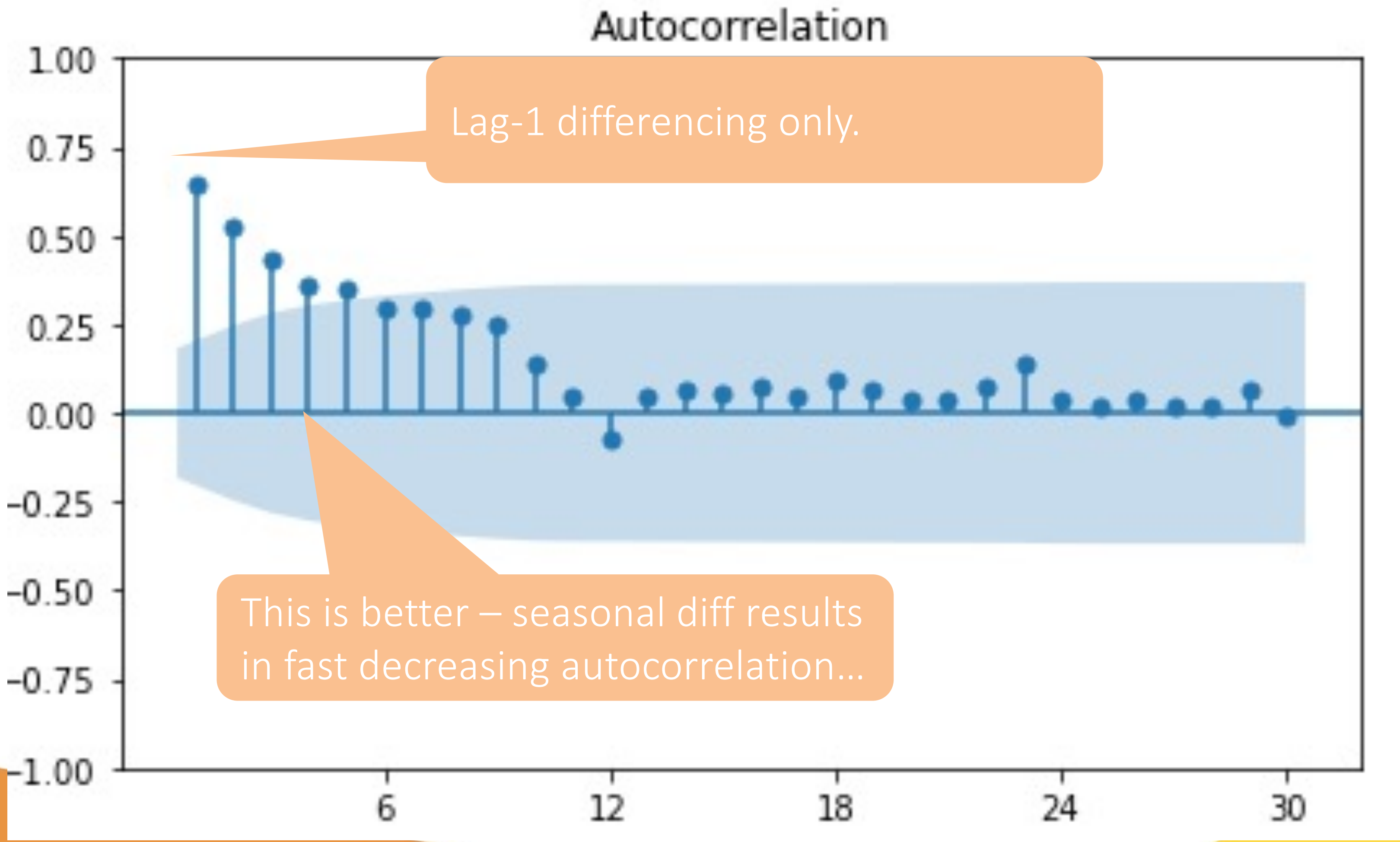
https://people.duke.edu/~rnau/411arim2.htm

CityU

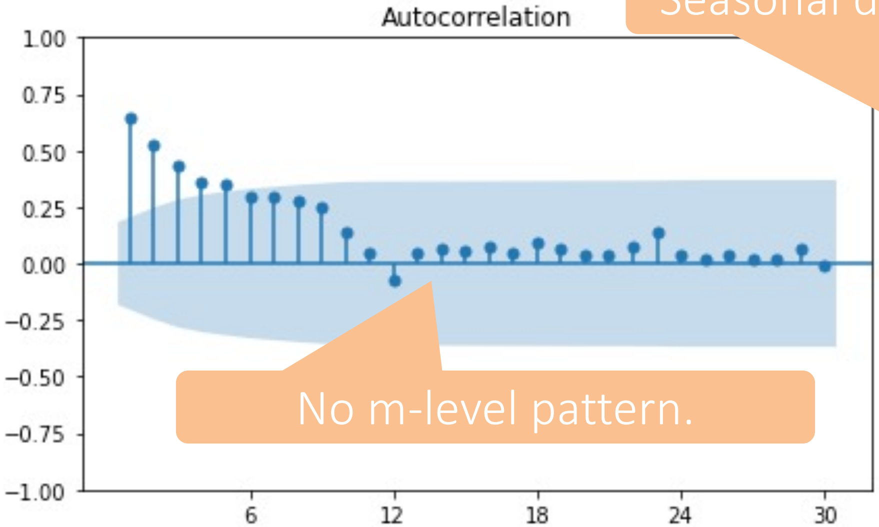
Over-differencing?

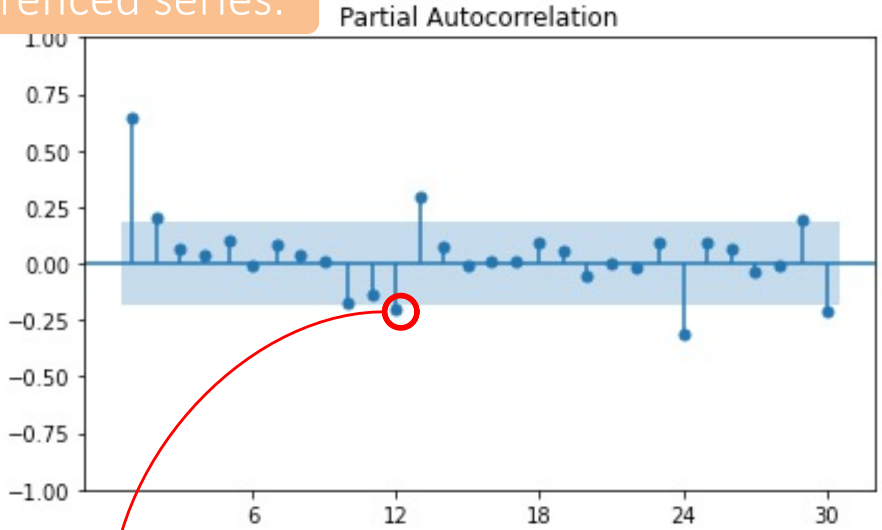


Now let’s just do seasonal diff…



Seasonal differenced series.





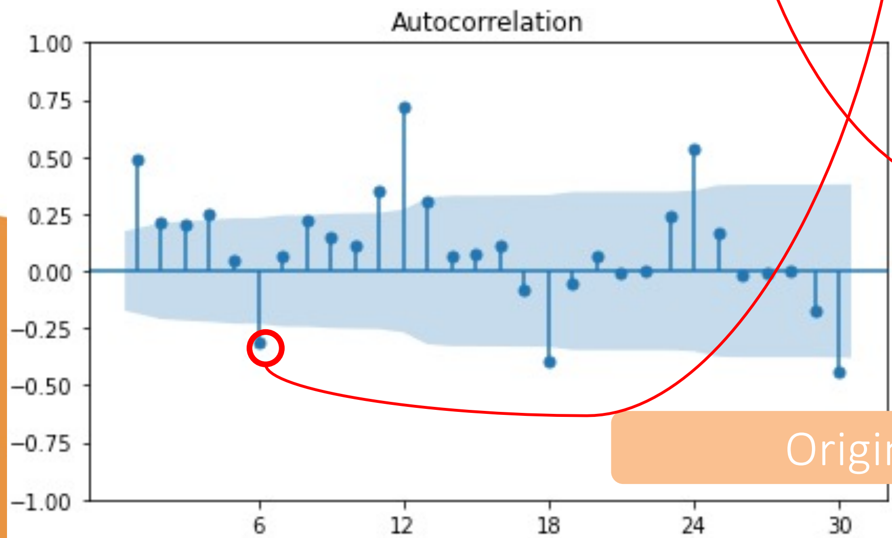
No m-level pattern.

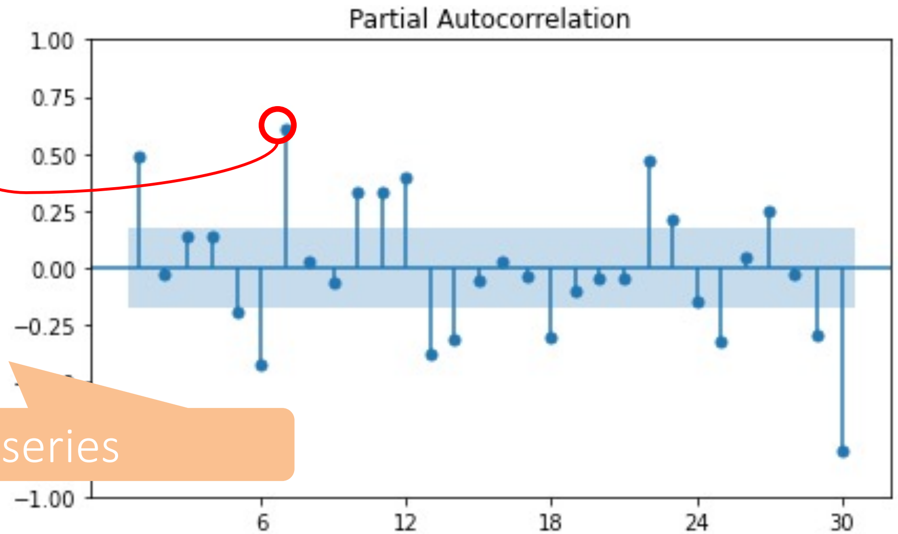
$$\text{ARIMA}(p, d, q)(P, D, Q)_m$$

What is “m” here?

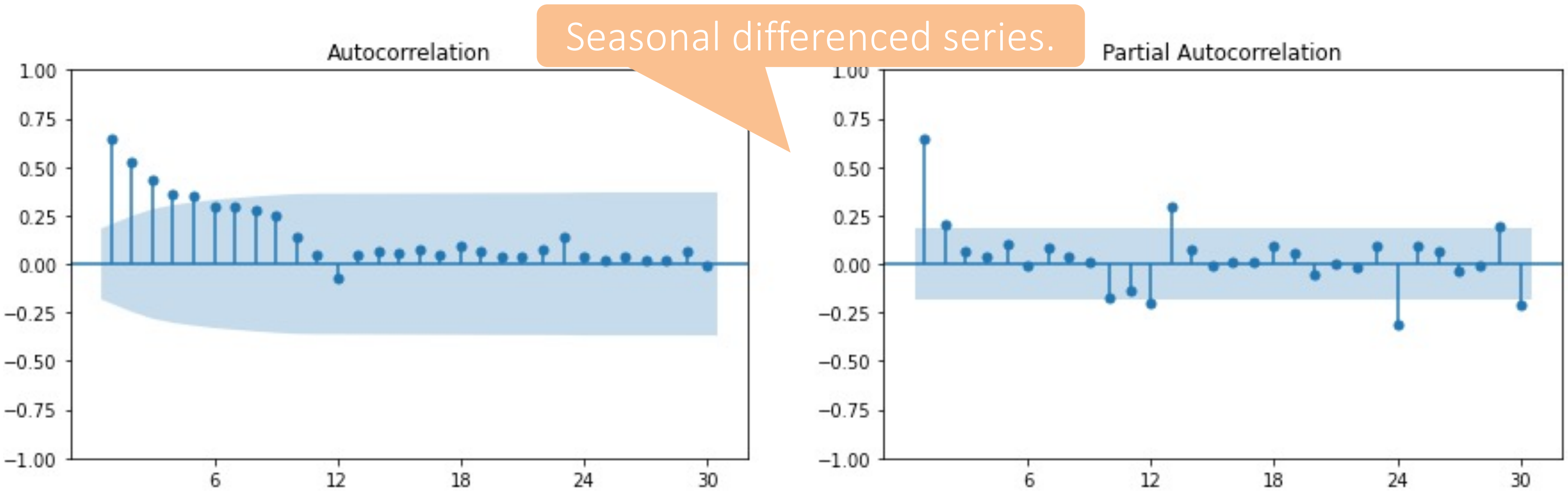
Any diff?

Any seasonal diff?





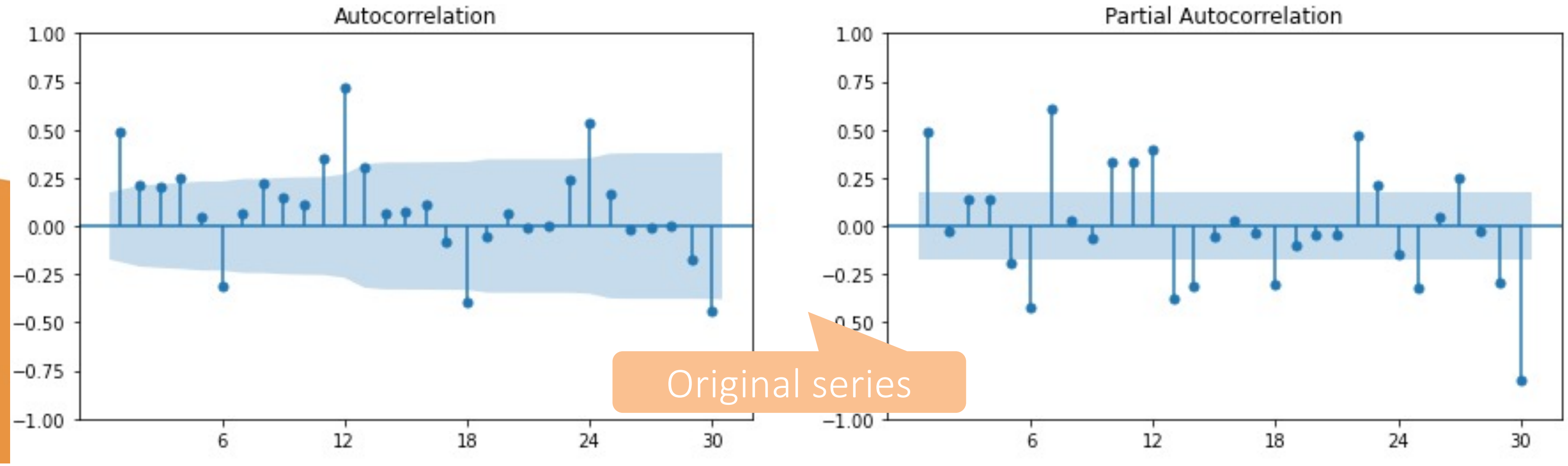
Original series



$$\text{ARIMA}(p,d,q)(P,D,Q)_m$$

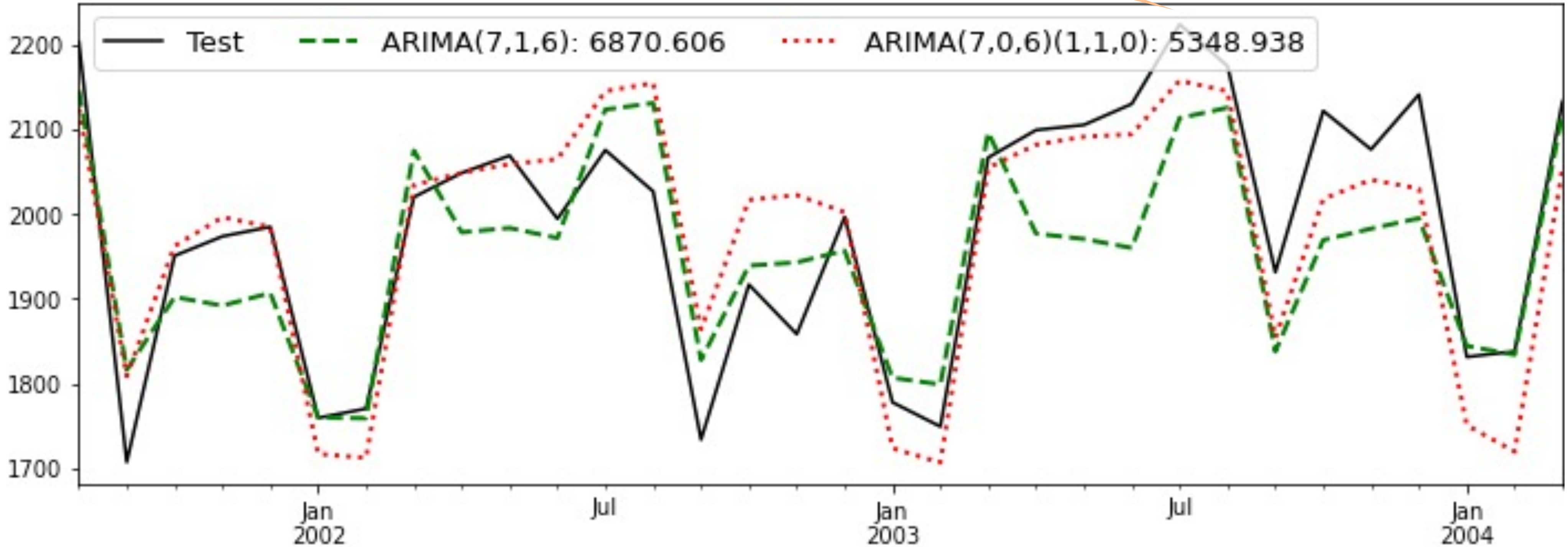
$$\downarrow$$

$$\text{ARIMA}(7,0,6)(1,1,0)_{12}$$



Is it good?

A lower MSE!



Advanced Topics: Vector AutoReg

- When we have multiple time series values, what should we do?

- They may affect each other: Endogenous

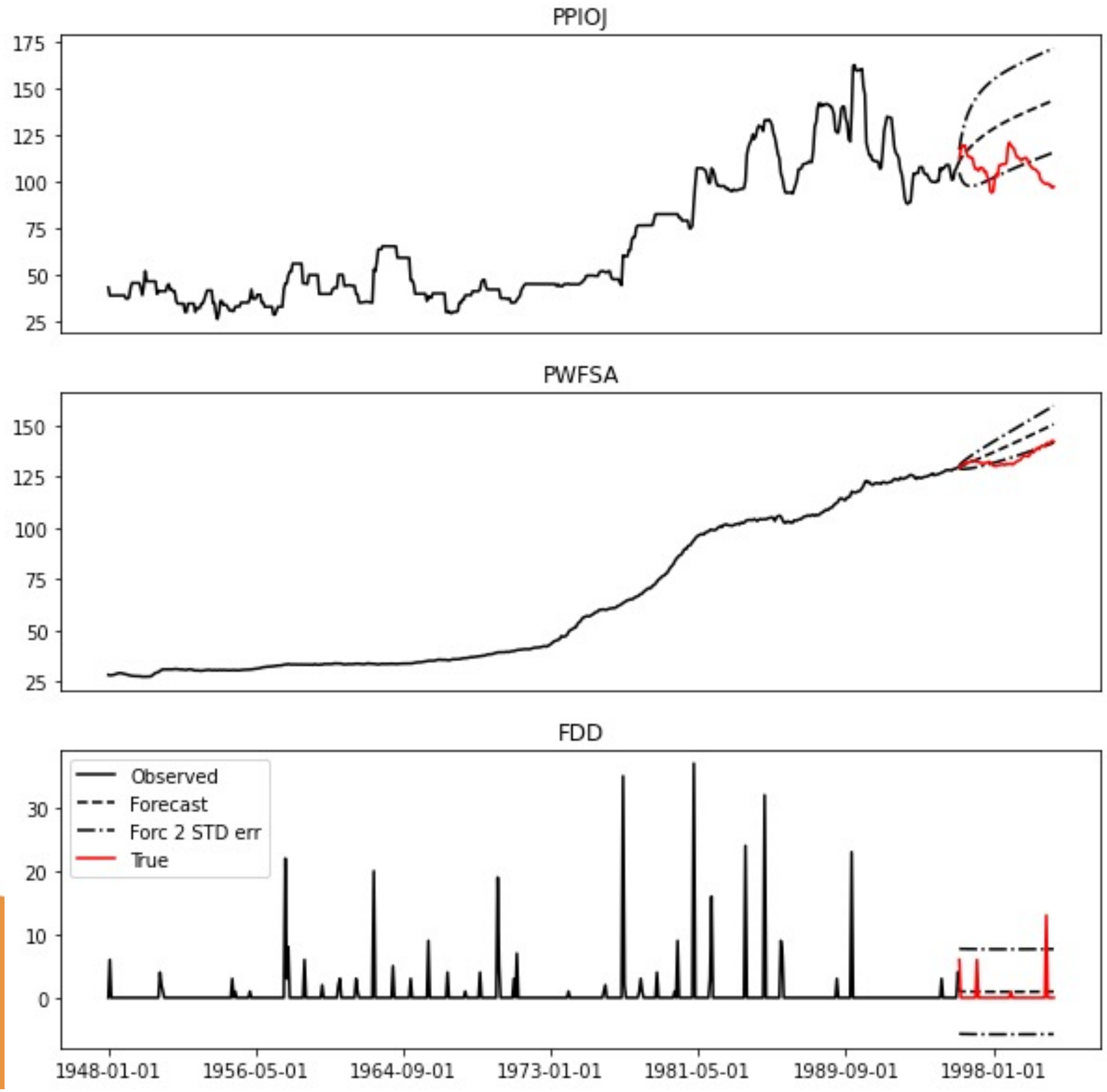
$$
\begin{aligned}
y_t &= const_1 + \phi_{11} y_{t-1} + \phi_{12} z_{t-1} + \epsilon_{1t} \\
z_t &= const_1 + \phi_{21} y_{t-1} + \phi_{21} z_{t-1} + \epsilon_{2t}
\end{aligned}
$$

```python
N_train = int(0.9 * oj_data.shape[0])
N_test = oj_data.shape[0] - N_train
oj_train = oj_data.iloc[:N_train].copy(deep=True)
```

d in 3ms, finished 00:07:41 2022-03-03

```python
var = sm.tsa.VAR(oj_train).fit(maxlags=2)
```

Easy to do with *statsmodels*!



Suggested References

- Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3.

- Nau, R. (n.d.) Statistical forecasting: notes on regression and time series analysis https://people.duke.edu/~rnau/411home.htm

- Business analytics using forecasting: https://youtube.com/playlist?list=PLoK4oIB1jeK0LHLbZW3DTT05e4srDYxFq

香港城市大學
City University of Hong Kong

The End

專業 創新 胸懷全球
Professional · Creative
For The World