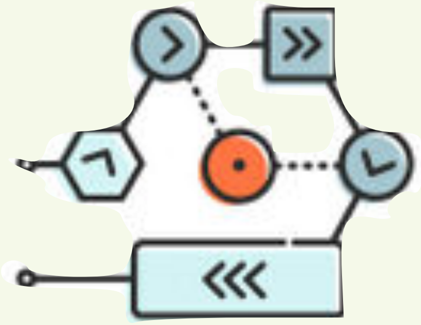
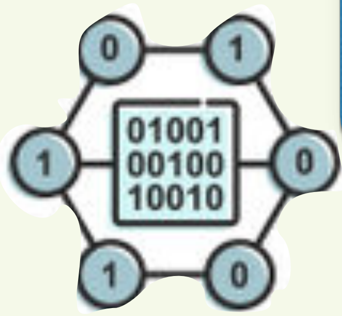


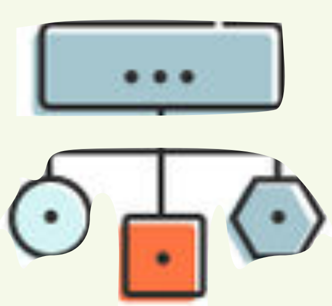
COURSE

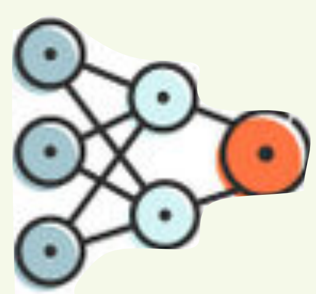
“APLICAÇÕES DE INTELIGÊNCIA ARTIFICIAL”

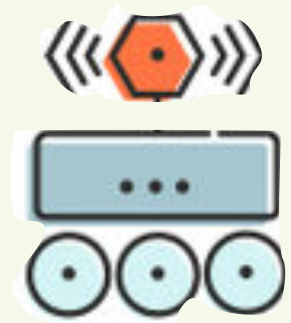
University of Aveiro

Mestrado em Eng. Computational









EXTRA INFO - this label will appear in slides with complementary information

SUMMARY: Time series methods of auto-regression type:

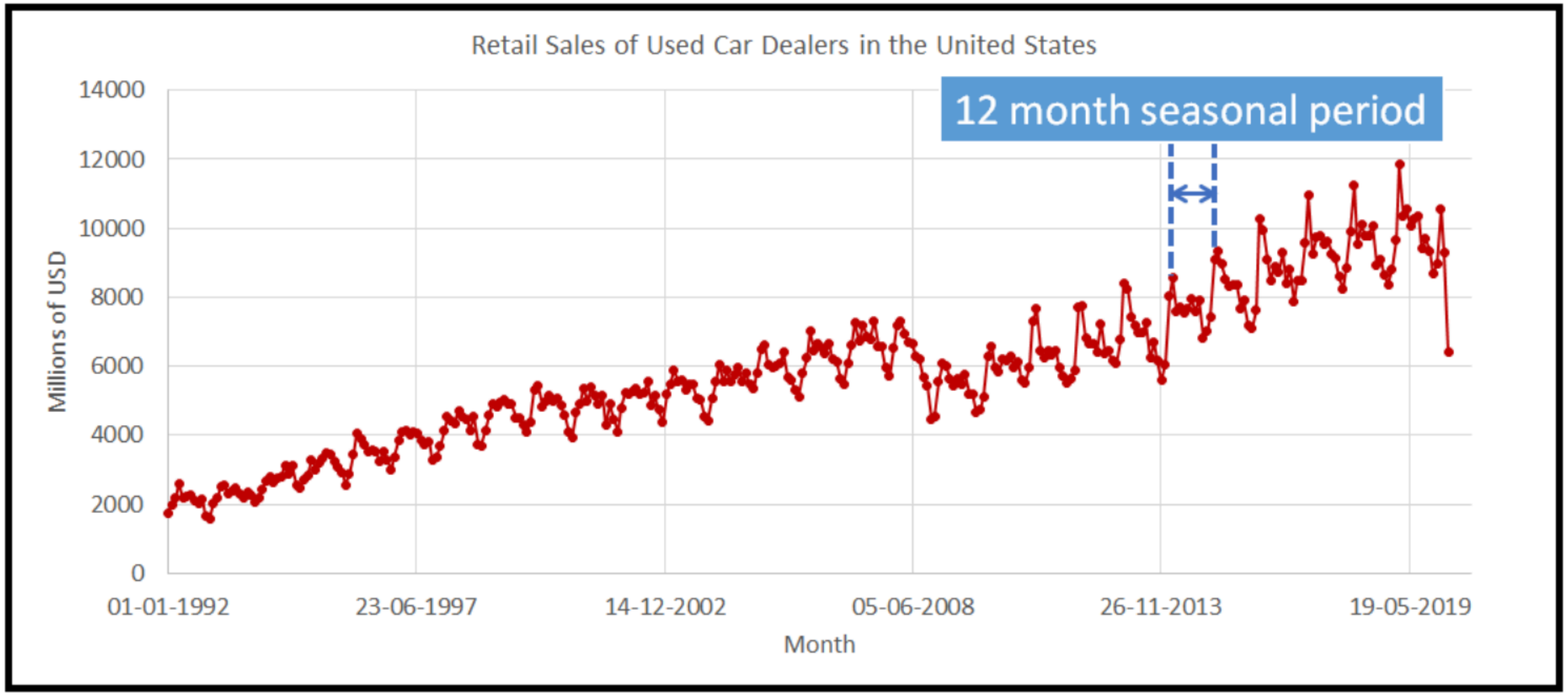
[T] Components of a time series. ACF and PACF. SARIMAX variant models;

[P] Determine the components.

EXTRA INFO - this label will appear in slides with complementary information

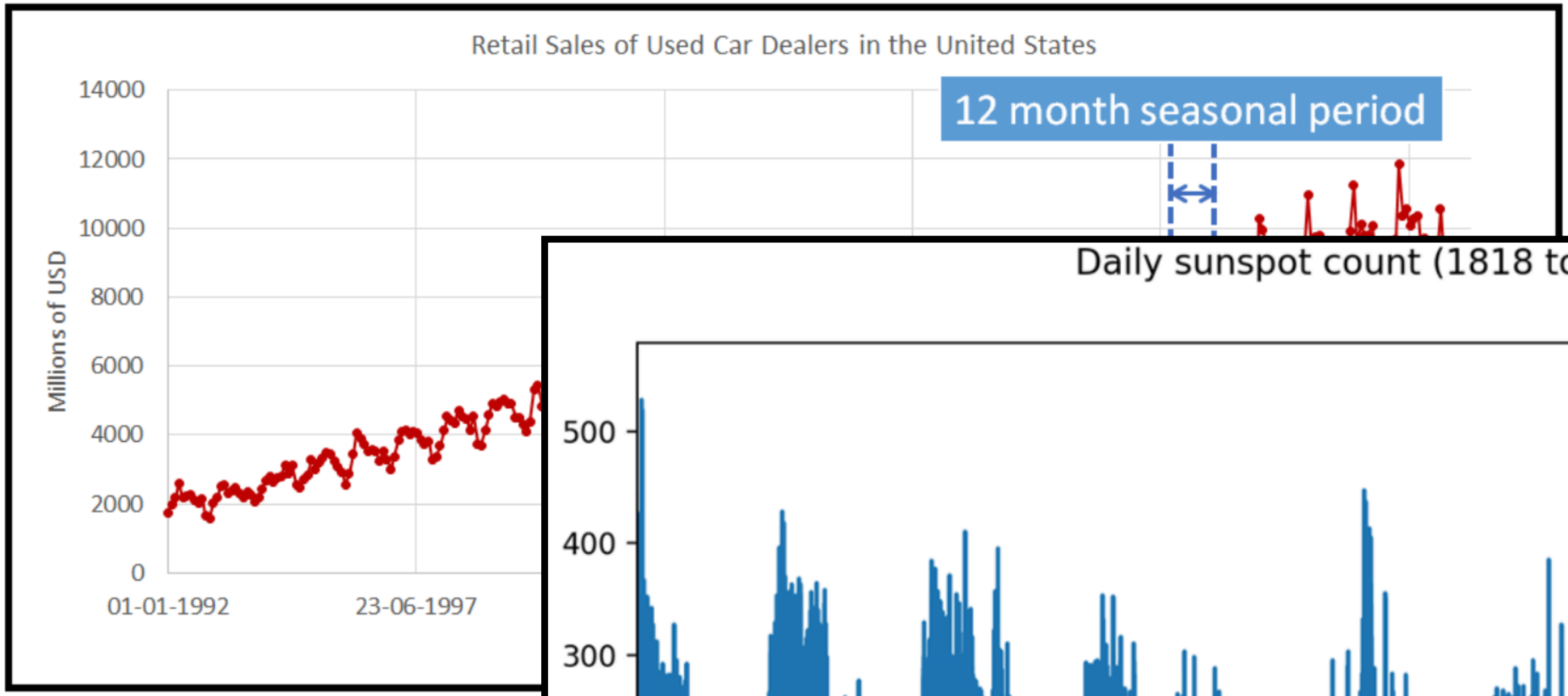
Time Series Components and Effects

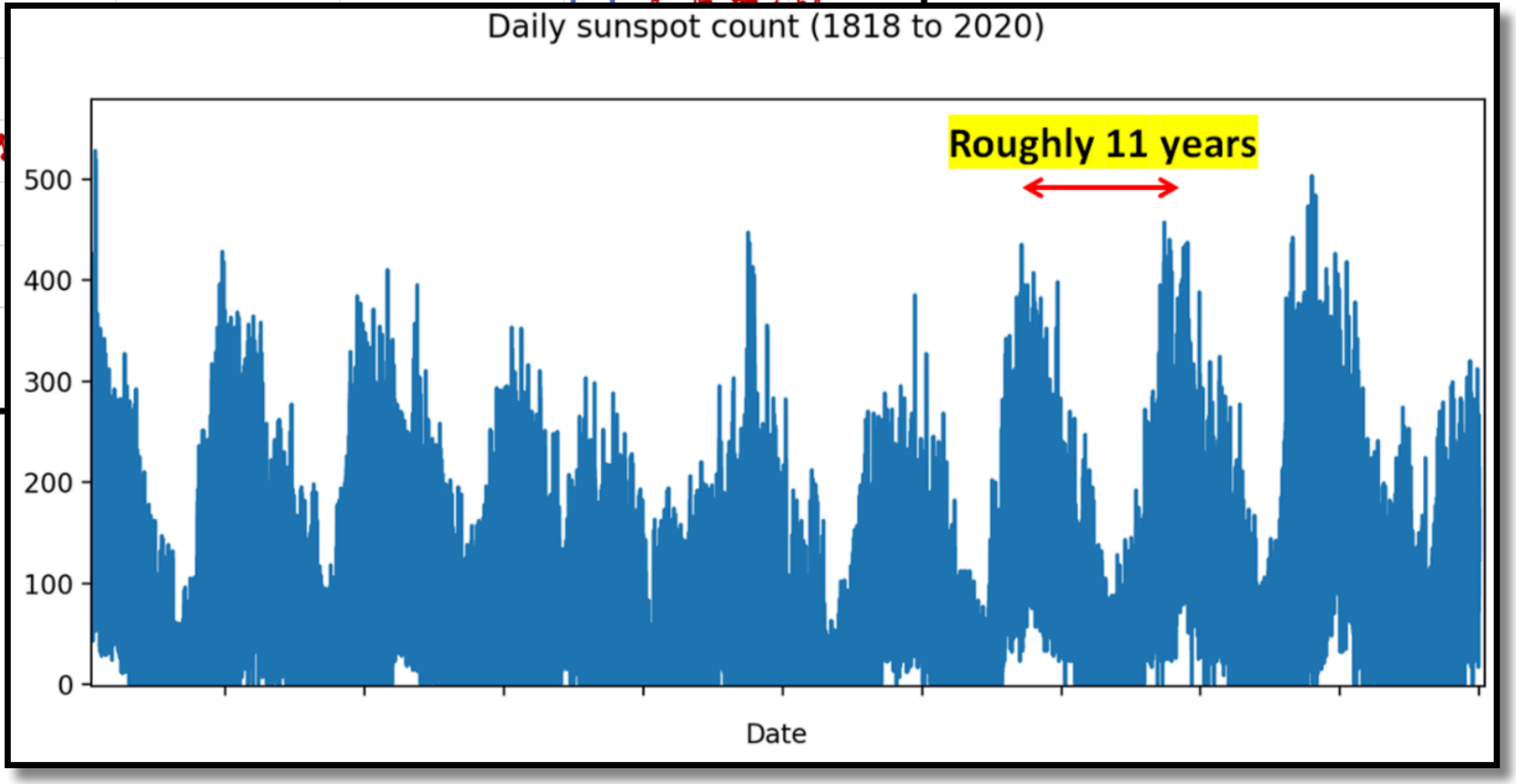
1. Seasonal component;



Time Series Components and Effects

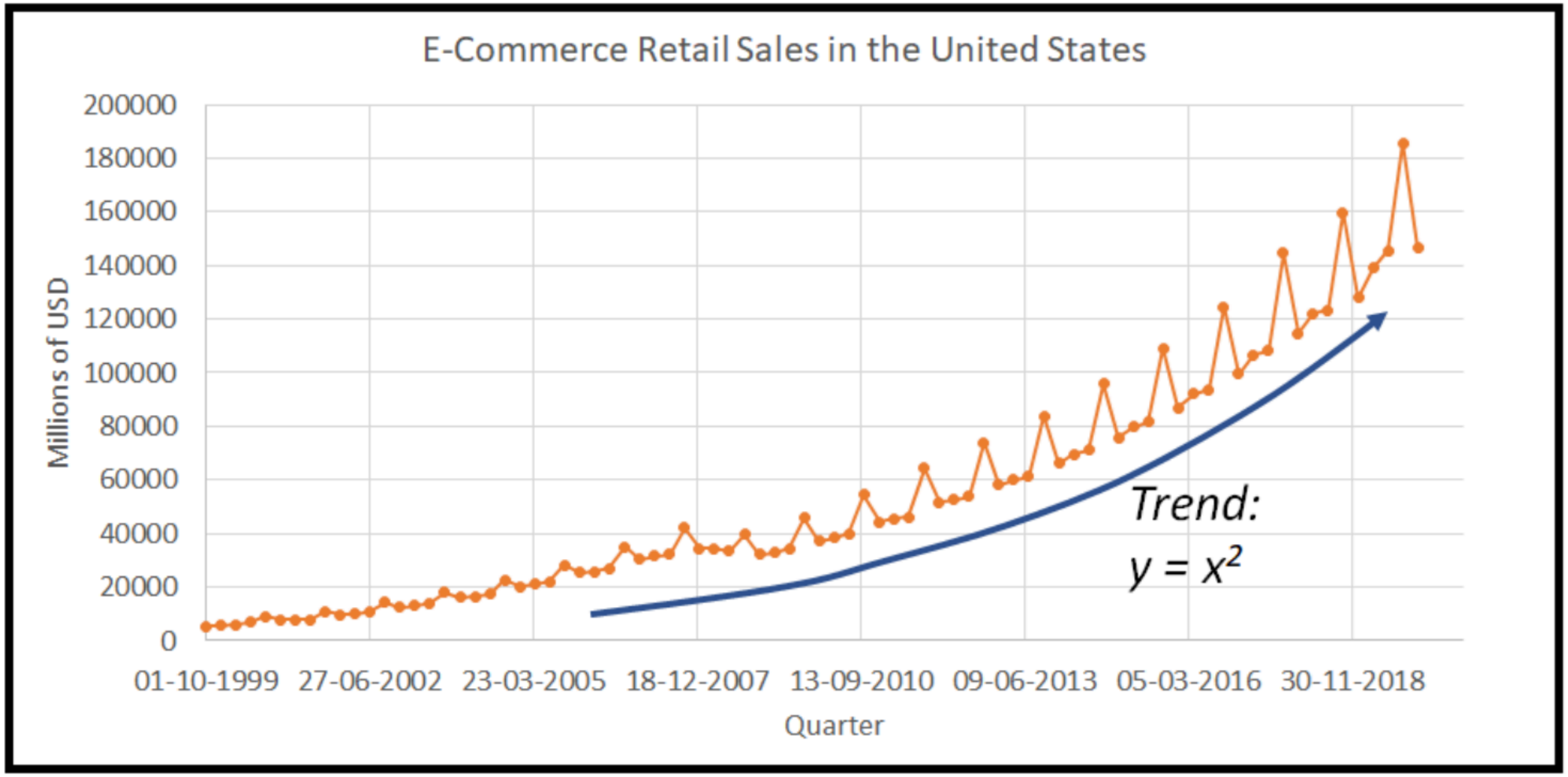
1. <u>Seasonal component</u>;





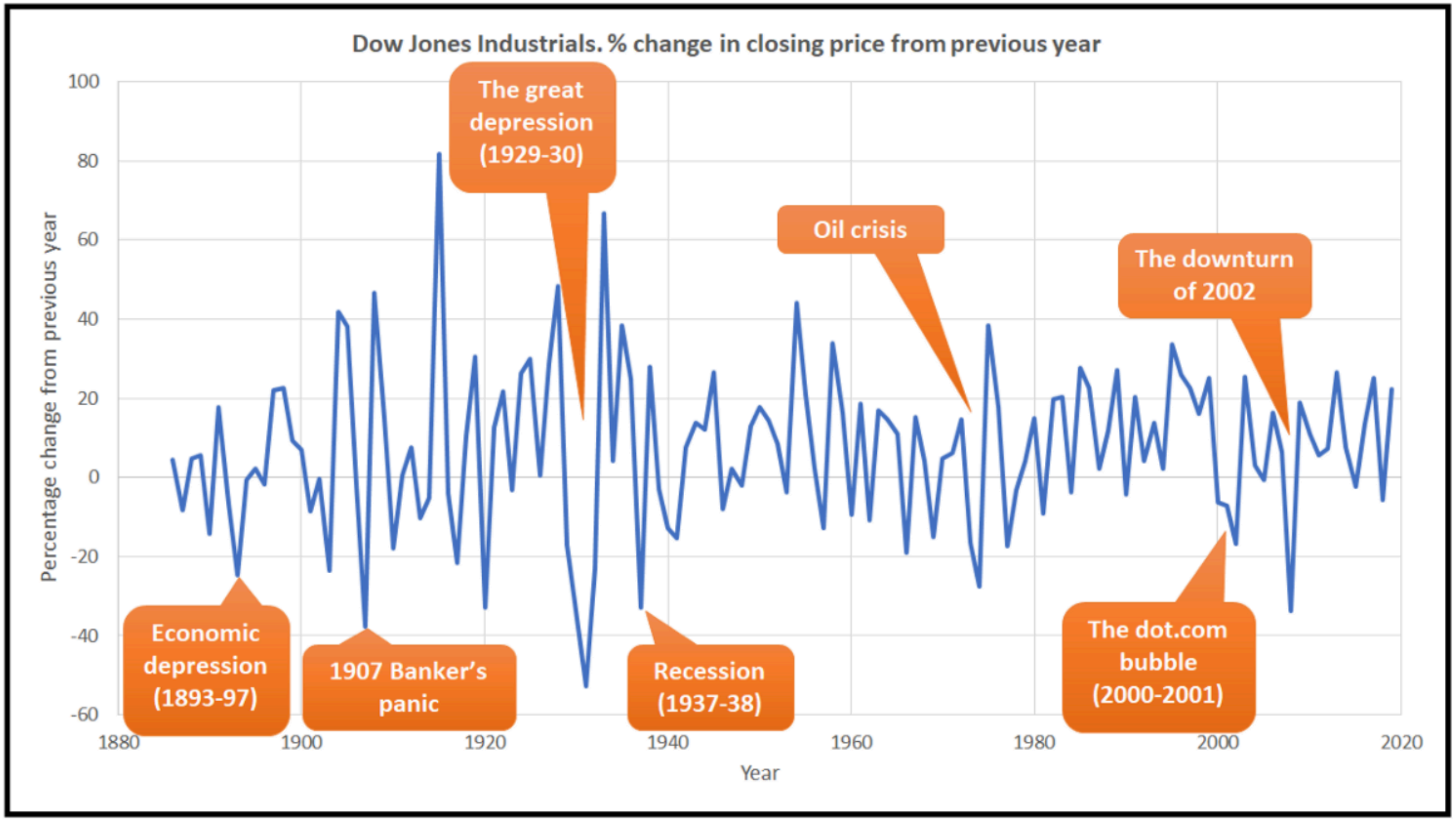
Time Series Components and Effects

1. Seasonal component;
2. Trend component;



Time Series Components and Effects

1. Seasonal component;
2. Trend component;
3. Cyclical components: manually isolated or diluted in other components;



Time Series Components and Effects

1. Seasonal component;
2. Trend component;
3. Cyclical component;
4. Noise component: the remaining values, after removing the value of 1 to 3.

Time Series Components and Effects

1. Seasonal component;
2. Trend component;
3. Cyclical component;
4. Noise component.

We have two possible effects combinations:

i. Additive combination: $y_i = t_i + s_i + n_i$

Time Series Components and Effects

1. Seasonal component;
2. Trend component;
3. Cyclical component;
4. Noise component.

We have two possible effects combinations:

i. Additive combination: $y_i = t_i + s_i + n_i$

ii. Multiplicative combination: $y_i = t_i \times s_i \times n_i$

Time Series Components and Effects

1. Seasonal component;
2. Trend component;
3. Cyclical component;
4. Noise component.

We have two possible effects combinations:

i. Additive combination: $y_i = t_i + s_i + n_i$

ii. Multiplicative combination: $y_i = t_i \times s_i \times n_i$

How can we calculate the components:

STEP 1: Identify the length of the seasonal period;

STEP 2: Isolate the trend;

STEP 3: Isolate the seasonality+noise;

STEP 4: Isolate the seasonality;

STEP 5: Isolate the noise.

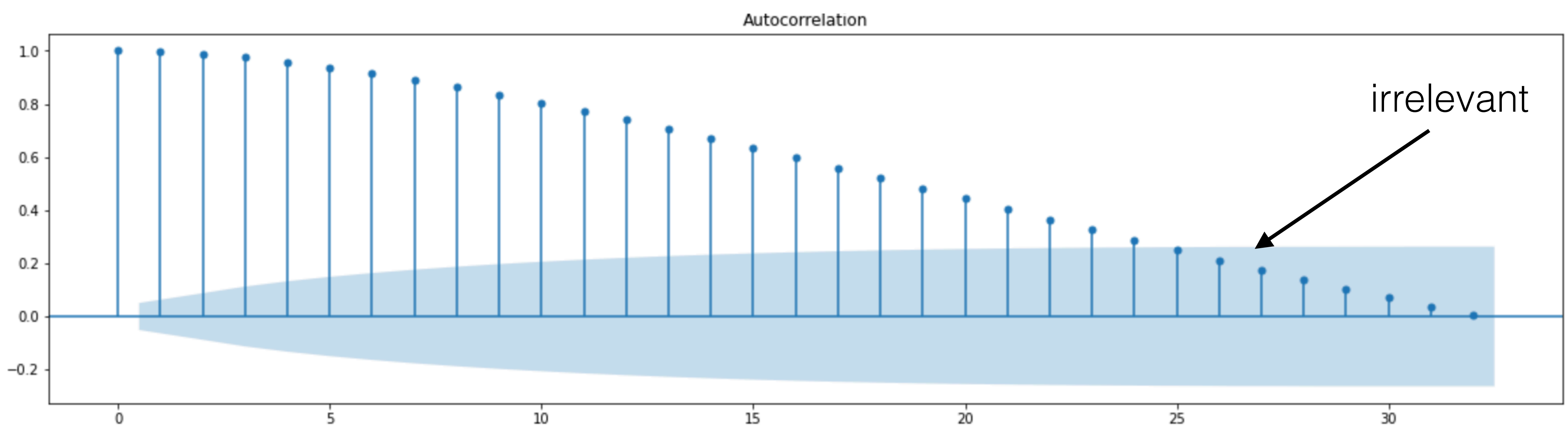
Explore the notebook *TS_components.ipynb*

AutoCorrelation Function (ACF)

The Pearson's correlation coefficient is a number between -1 and 1 that describes a negative or positive correlation respectively. A value of zero indicates no correlation.

We can calculate the correlation for time series observations with observations with previous time steps, called lags. Because the correlation of the time series observations is calculated with values of the same series at previous times, this is called a serial correlation, or an autocorrelation.

A plot of the autocorrelation of a time series by lag is called the **A**uto**C**orrelation **F**unction, or the acronym ACF. This plot is sometimes called a correlogram or an autocorrelation plot.

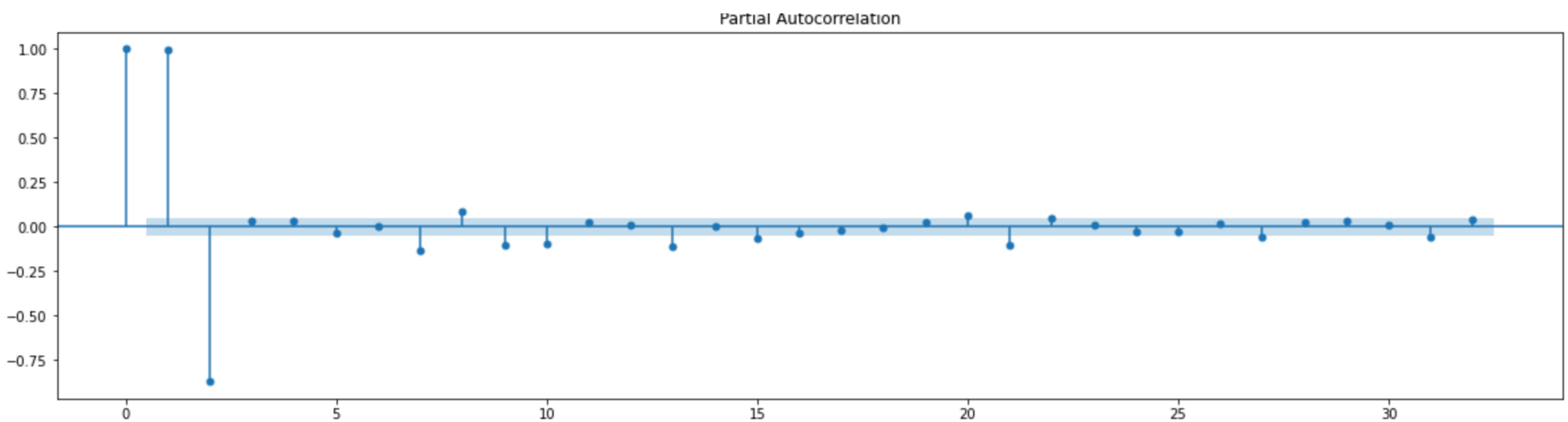


Partial AutoCorrelation Function (PACF)

A partial autocorrelation is a summary of the relationship between an observation in a time series with observations at prior time steps with the relationships of intervening observations removed.

> *The partial autocorrelation at lag k is the correlation that results after removing the effect of any correlations due to the terms at shorter lags.*

— Page 81, Section 4.5.6 Partial Autocorrelations, Introductory Time Series with R.



SARIMAX Variant Models

Autoregressive (AR) Models

Suppose we have a time series given by y_t. An $AR(p)$ model can be specified by

$$y_t = \beta + \epsilon_t + \sum_{i=1}^{p} \theta_i y_{t-i}$$

Where p is the number of time lags to regress on, ϵ_t is the noise at time t and β is a constant.

This equation can be made more concise through the use of the lag operator, L.

$$L^n y_t = y_{t-n}$$

Taking $\Theta(L)^p$ to be an order p polynomial function of L, we can instead define an autoregressive model by

$$y_t = \Theta(L)^p y_t + \epsilon_t$$

Taking note that the constant has been absorbed into the polynomial Θ.

SARIMAX Variant Models

Moving average (MA) Models.

Whereas autoregressive models regress on prior values of y_t, moving average models regress on prior values of error. An $MA(q)$ model can be specified by

$$y_t = \Phi(L)^q \epsilon_t + \epsilon_t$$

Where q is the number of time lags of the error term to regress on and Φ is defined analagously to Θ.

Autoregressive Moving Average (ARMA) Models

$ARMA(p, q)$ models are simply a sum of $AR(p)$ and $MA(q)$ models.

$$y_t = \Theta(L)^p y_t + \Phi(L)^q \epsilon_t + \epsilon_t$$

SARIMAX Variant Models

Autoregressive Integrated Moving Average (ARIMA) Models

To help tackle non-stationary data, we introduce an integration operator Δ^d, defined as follows

$$y_t^{[d]} = \Delta^d y_t = y_t^{[d-1]} - y_{t-1}^{[d-1]}$$

where $y_t^{[0]} = y_t$ and d is the order of differencing used.

We can now fit an $ARMA(p, q)$ model to $y_t^{[d]}$ rather than y_t.

$$y_t^{[d]} = \Theta(L)^p y_t^{[d]} + \Phi(L)^q \epsilon_t^{[d]} + \epsilon_t^{[d]}$$

This is equivalent to an $ARIMA(p, d, q)$ model on y_t

$$\Delta^d y_t = \Theta(L)^p \Delta^d y_t + \Phi(L)^q \Delta^d \epsilon_t + \Delta^d \epsilon_t$$

With some algebra, we can re-arrange the equation and absorb constants into the polynomials Θ and Φ.

$$\Theta(L)^p \Delta^d y_t = \Phi(L)^q \Delta^d \epsilon_t$$

SARIMAX Variant Models

SARIMA

SARIMA models take seasonality into account by essentially applying an ARIMA model to lags that are integer multiples of seasonality. Once the seasonality is modelled, an ARIMA model is applied to the leftover to capture non-seasonal structure.

To see this more clearly, suppose we have a time series y_t with seasonality s. We can try to eliminate the seasonality with differencing, by applying the differencing operator Δ_s^D to take the seasonal differences of the time series. Here s is the number of time lags comprising one full period of seasonality. D takes on a similar meaning to d in ARIMA models, but instead applies to *seasonal* lags.

We can then capture any remaining structure by applying an $ARMA(P, Q)$ model, to the differenced values, but using seasonal lags. i.e. instead of using a regular lag operator L, we use L^s. P and Q are again seasonal time lags

$$\Delta_s^D y_t = \theta(L^s)^P \Delta_s^D y_t + \phi(L^s)^Q \Delta_s^D \epsilon_t + \Delta_s^D \epsilon_t$$

SARIMAX Variant Models

As with ARIMA, massaging the equation and absorbing constants into polynomials yields the following concise form

$$\theta(L^s)^P \Delta_s^D y_t = \phi(L^s)^Q \Delta_s^D \epsilon_t$$

With any seasonality now removed, we can apply another $ARIMA(p, d, q)$ model to $\Delta_s^D y_t$ by multiplying the seasonal model by the new ARIMA model.

$$\Theta(L)^p \theta(L^s)^P \Delta^d \Delta_s^D y_t = \Phi(L)^q \phi(L^s)^Q \Delta^d \Delta_s^D \epsilon_t$$

This is the general form of a $SARIMA(p, d, q)(P, D, Q, s)$ model.

ARIMAX and SARIMAX

ARIMAX and SARIMAX models simply take exogenous variables into account - ie variables measured at time t that influences the value of our time series at time t, but that are not autoregressed on. To do this, we simply add the terms in on the right hand side of our ARIMA and SARIMA equations.

SARIMAX Variant Models

For n exogenous variables defined at each time step t, denoted by x_t^i for $i \leq n$, with coefficients β_i, the $ARIMAX(p, d, q)$ model is defined by

$$\Theta(L)^p \Delta^d y_t = \Phi(L)^q \Delta^d \epsilon_t + \sum_{i=1}^{n} \beta_i x_t^i$$

and the $SARIMAX(p, d, q)(P, D, Q, s)$ model by

$$\Theta(L)^p \theta(L^s)^P \Delta^d \Delta_s^D y_t = \Phi(L)^q \phi(L^s)^Q \Delta^d \Delta_s^D \epsilon_t + \sum_{i=1}^{n} \beta_i x_t^i$$

Further References

Chatfield, C 2004, *The Analysis of Time Series : An Introduction*, 6th ed., Chapman & Hall/CRC, Boca Raton, Fla.