Time Series Management

Michele Linardi Ph.D.

michele.linardi@orange.fr
https://github.com/mlinardiCYU/GestionEtAnalyseTS_25-26

Required skills

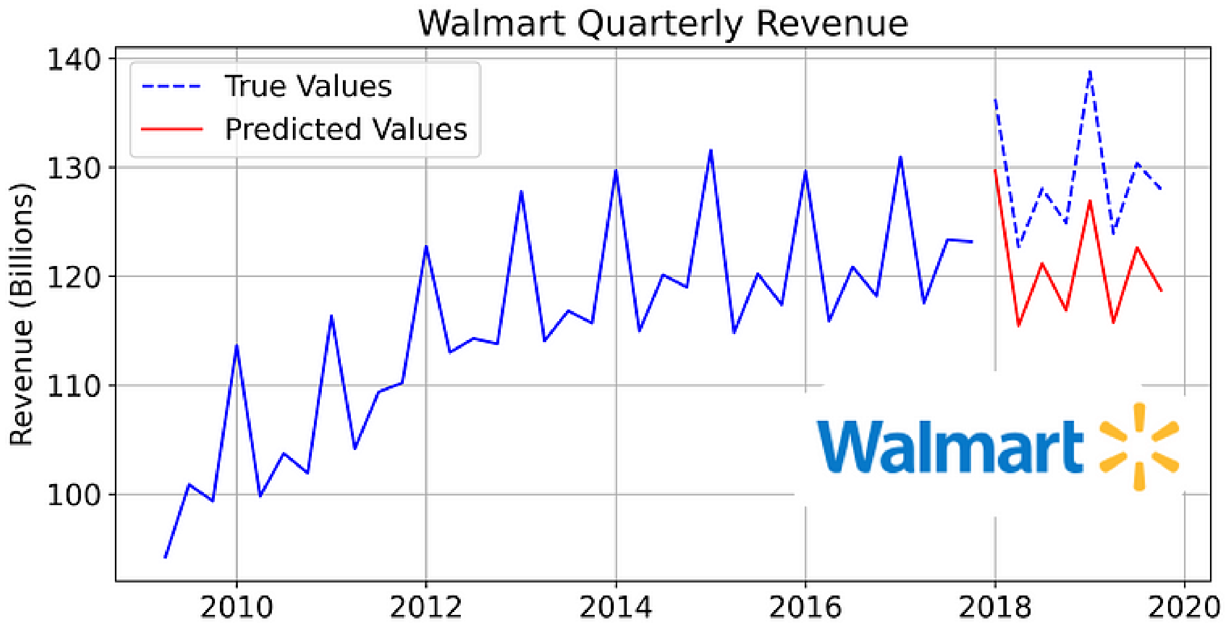
- Calculus and ML
- Excel
- Programming (Python)
- Attending classes (participation)

Syllabus

- Time series data
- Trend, seasonality, cycles and residuals
- Stationary processes
- Autoregressive processes.
- Moving average processes.
- ACF & PACF
- Fitting AR(p) MA(q) models
- ARIMA model
- Backshift Notation
- Model selection and forecasting
- MLE (Maximum likelihood estimation)
- AIC Akaike's Information Criterion

Time series data

- A **univariate time series** is a sequence of measurements of the same variable collected over time. Most often, the measurements are made at regular time intervals.



https://www.kaggle.com/code/anushkaml/walmart-time-series-sales-forecasting

TS Trend


divyant published on TradingView.com, Apr 10, 2022 21:19 UTC+5:30
BAJAJ AUTO LTD, 1W, NSE 3776.20 +70.45 (+1.88%)
Higher tops
Higher lows
Uptrend
INR
3000.00
2830.55
2600.00
2400.00
2200.00
2000.00
1800.00
1650.00
1530.00
1410.00
1290.00
1190.00
1090.00
1010.00
935.00
Jun
2011
Jun
2012
Jun
2013
Jul
2014
Jul
2015
Jul
2016
Jun
2017
TradingView

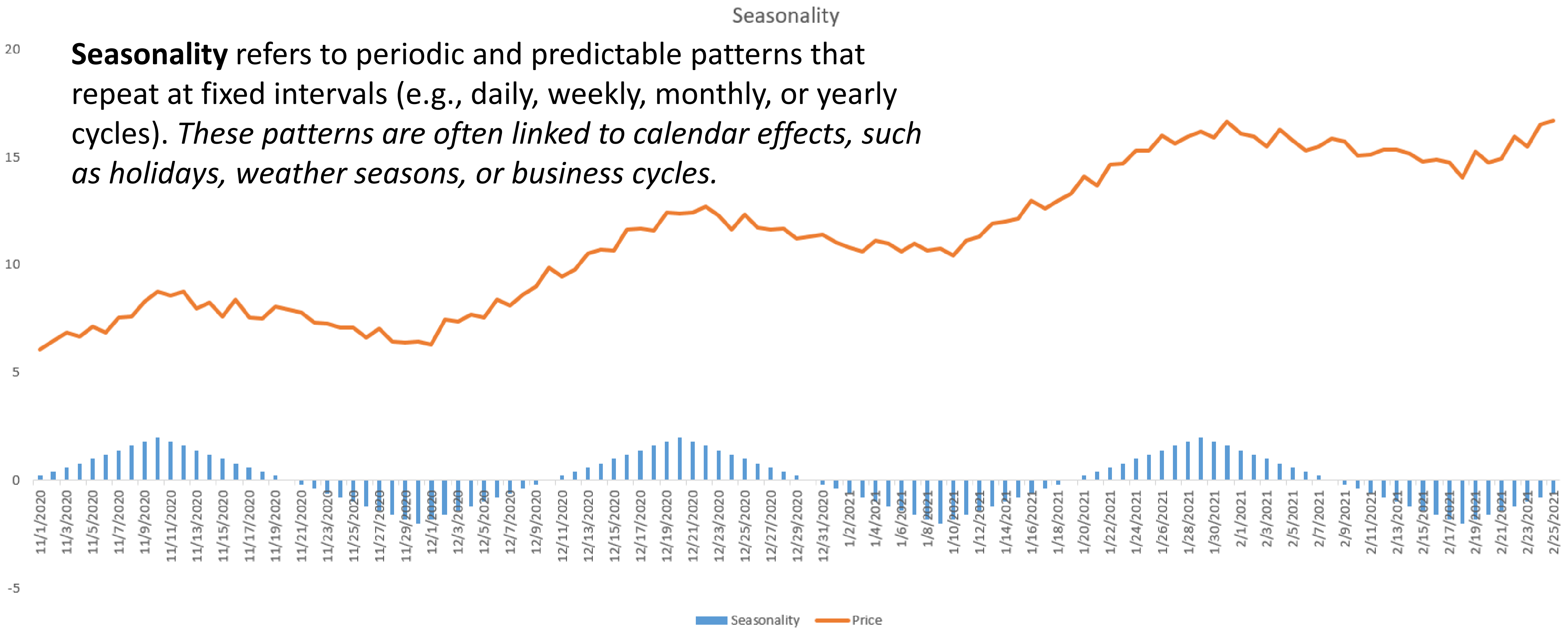
TS Trend

The trend is the long-term movement or direction in the data, which may be upward, downward, or even stable over time. It reflects changes due to underlying factors, such as economic growth, technological changes, or other gradual effects.

Example: Monthly sales increasing steadily over several years.

TS Seasonality and Ciclicality

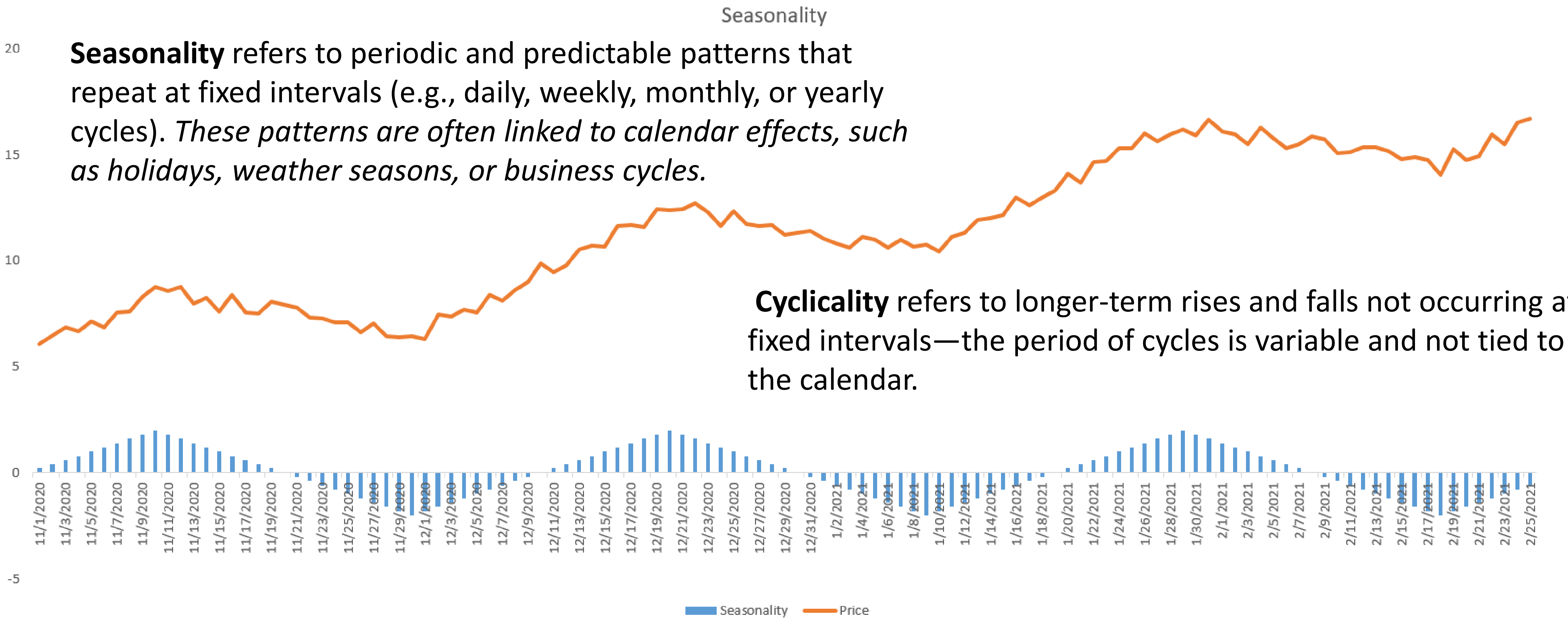
Seasonality refers to periodic and predictable patterns that repeat at fixed intervals (e.g., daily, weekly, monthly, or yearly cycles). *These patterns are often linked to calendar effects, such as holidays, weather seasons, or business cycles.*



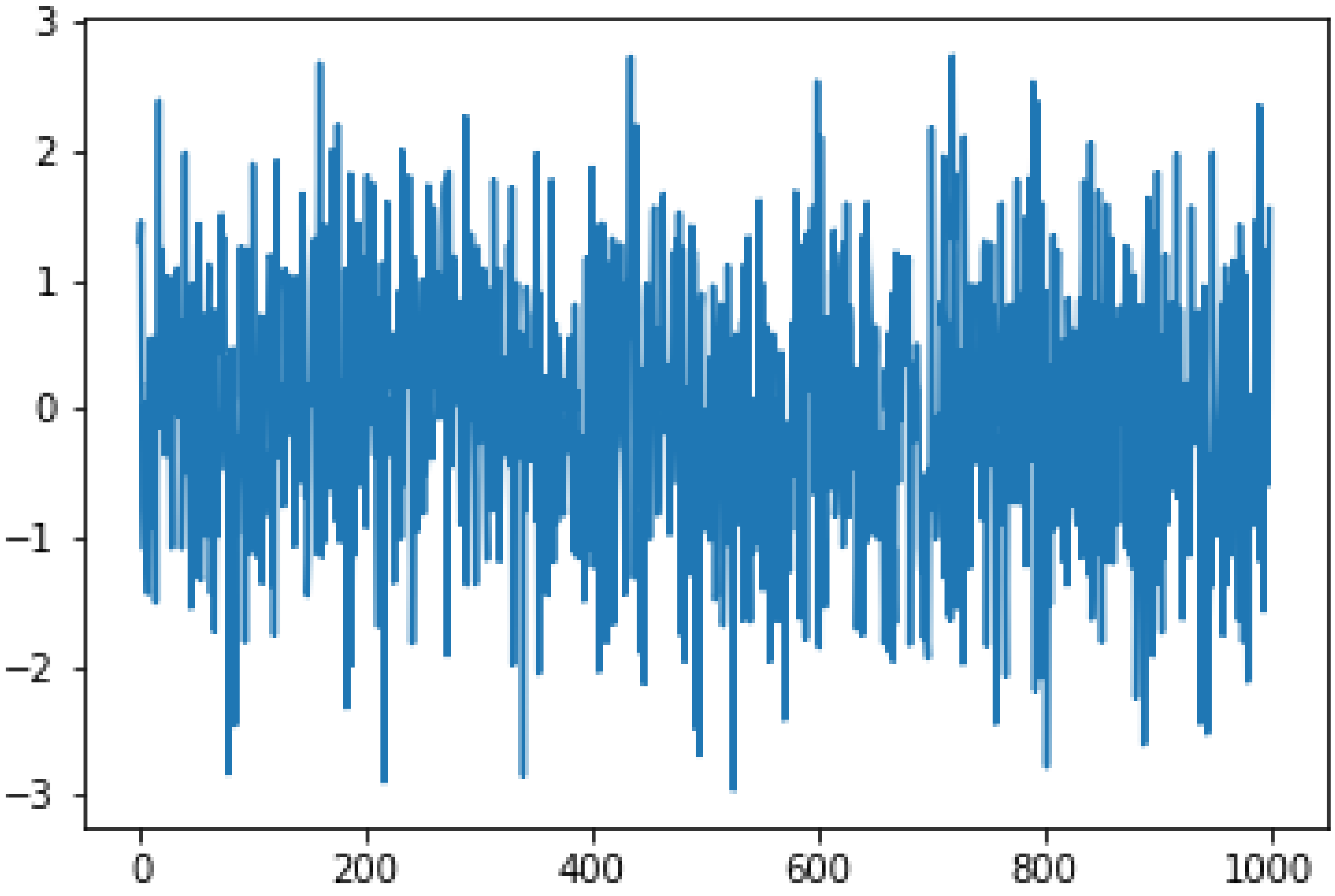
TS Seasonality and Ciclicality

Seasonality refers to periodic and predictable patterns that repeat at fixed intervals (e.g., daily, weekly, monthly, or yearly cycles). *These patterns are often linked to calendar effects, such as holidays, weather seasons, or business cycles.*

Cyclicality refers to longer-term rises and falls not occurring at fixed intervals—the period of cycles is variable and not tied to the calendar.

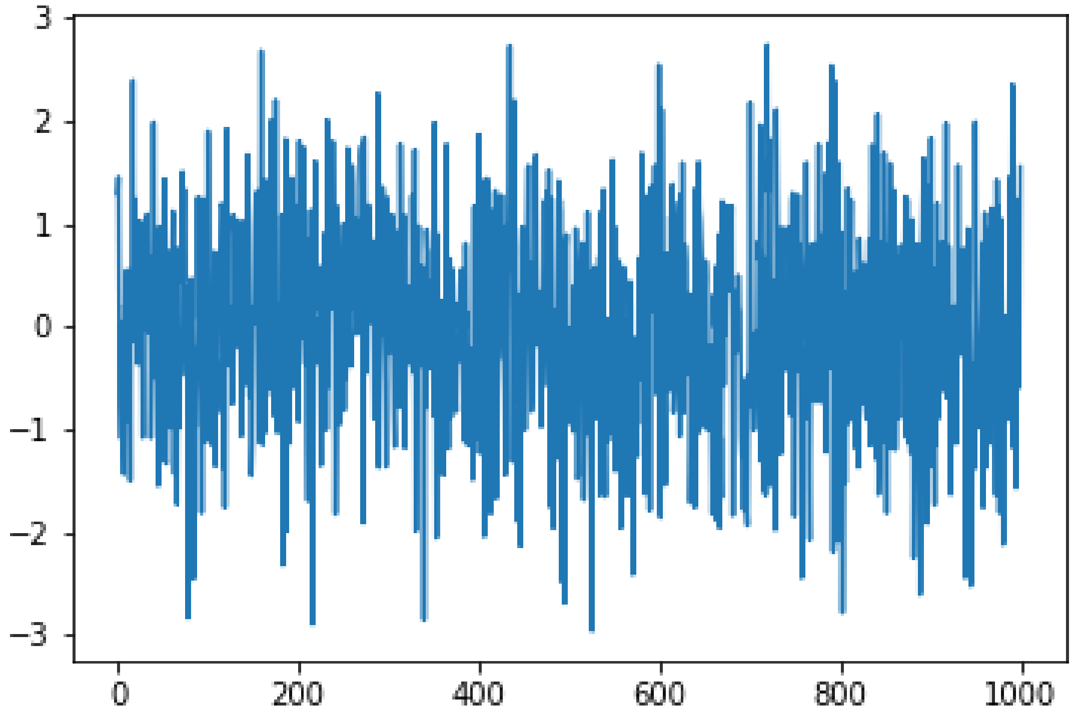


TS Noise



TS Noise

Noise is the irregular, random fluctuation or variation remaining after removing trend and seasonality from the data. It arises from unpredictable events, measurement errors, or other unexplained sources.



Notation

We represent a time series whose values are recorded (at regular and unitary interval) until time t with $\boldsymbol{X_t}$ **(equivalently $\boldsymbol{Y_t}$).**

We represent a lagged series (starting after h intervals) with X_{t-h}, where h is called lag.

Example: $X_t = \boldsymbol{x_1}, \boldsymbol{x_2}, \boldsymbol{x_3}, ..$; $X_{t-1} = \boldsymbol{x_2}, \boldsymbol{x_3}, ..$; $X_{t-2} = \boldsymbol{x_3}$,...

Stationary (weakly a.k.a second order) TS

- A time series is second-order stationary if the mean is constant and the covariance between any two values only depends on the time difference between those two values (and not on the value of t itself):

A time series X_t is second-order stationary if:

- Mean: $E(X_t) = \mu$, the mean does not depend on time t
- Variance: $Var(X_t) = \mu$, the variance is constant for all t
- Autocovariance: $Cov(X_t, X_{t-h}) = \gamma(h)$, the covariance depends only on the lag h, not on the actual time points.

ACF – Autocorrelation function

- **The sample autocorrelation function (ACF)** measures the correlation between observations in a time series separated by a certain number of time steps (lags). In simple terms, it quantifies how similar a time series is to a delayed (lagged) version of itself :

$$\text{ACF (lag h)} = \rho_h = \frac{Covariance(X_t, X_{t-h})}{Std.Dev.(X_t)Std.Dev.(X_{t-h})}$$

PACF – Partial Autocorrelation function

The Partial Autocorrelation Function (PACF) in time series analysis measures the correlation between observations at a certain lag, after removing the effects of correlations at all shorter lags. Essentially, it captures the direct relationship between a time series and its lagged values while controlling for intermediate time points.

If we assume that $\boldsymbol{x_t = C + \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \ldots}$

We may want to consider the (partial) dependency between $\boldsymbol{X_{t-3}}$ to $\boldsymbol{X_t}$. Hence, PACF of order 3:

PACF (lag h) = $\rho_{h,h}$= $\dfrac{Covariance(X_t, X_{t-3} | X_{t-2}, X_{t-1})}{\sqrt{Variance(X_t | X_{t-1}, X_{t-2}) Variance(X_{t-3} | X_{t-1}, X_{t-2})}}$

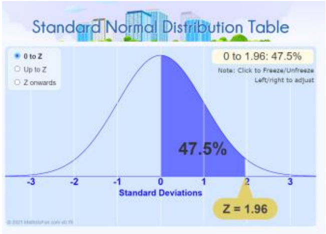
ACF and PACF

ACF (lag h) = $\rho_h = \dfrac{Covariance(X_t, X_{t-h})}{Std.Dev.(X_t)Std.Dev.(X_{t-h})}$

Significance limits ACF et PACF

$$S^+ = \frac{+Z}{N-h} \qquad S^- = \frac{-Z}{N-h}$$

N $\coloneqq$ $time\ series\ values$ $\qquad h \coloneqq lag$



Pick Z such that the area of N(0,1) between (– Z) and (Z) is equal to the desired amount of confidence
(in general for 95% c.i. , Z = 1.96)

https://www.mathsisfun.com/data/standard-normal-distribution-table.html

Direct (partial) effect of X_{t-h} on X_t , removing X_{t-h+1} , … , X_{t-2}

PACF (lag 3) = $\rho_{3,3} \dfrac{Covariance(X_{[t]}, X_{[t-3]} | X_{[t-2]}, X_{[t-1]})}{\sqrt{Variance(X_{[t]} | X_{[t-1]}, X_{[t-2]}) Variance(X_{[t-3]} | X_{[t-1]}, X_{[t-2]})}}$

Levinson and Durbin recursive algorithm (for PACF computation)

$$\text{PACF (lag h)} = \rho_{h,h} = \frac{\rho_h - \sum_{j=1}^{h-1} \rho_{h-1,j}\, \rho_{h-j}}{1 - \sum_{j=1}^{h-1} \rho_{h-1,j}\, \rho_j}$$

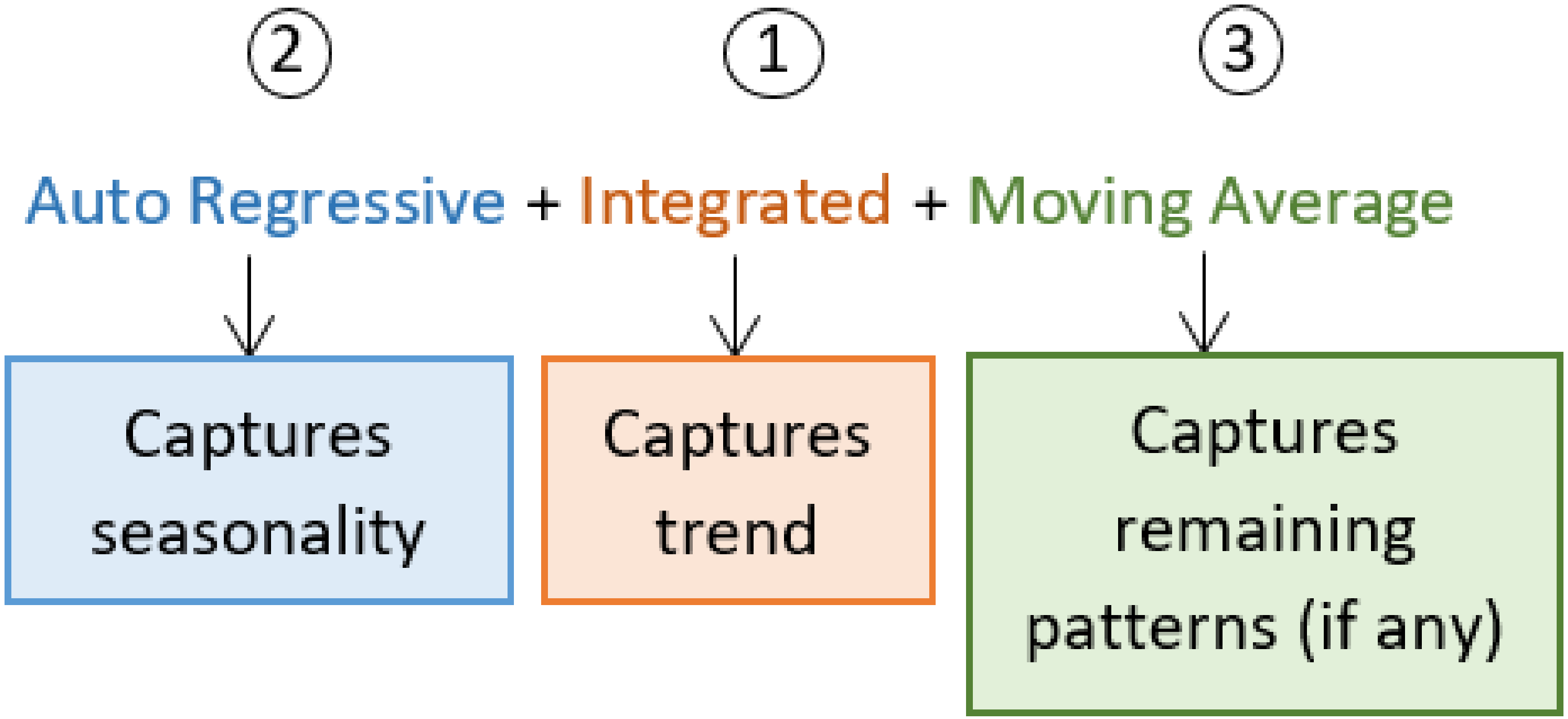
where :

$$\rho_{h,j} = \rho_{h-1,j} - \rho_{h,h}\, \rho_{h-1,h-j}$$

for :

$$j = 1,2, \dots, h-1$$

Time series modelling : AR.I.MA



https://www.linkedin.com/pulse/time-series-part-2-introduction-arima-models-using-excel-agarwal

Autoregression (AR) models

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + w_t$$

Where c is a constant, ϕ_i are parameters to estimate and w_t the error term (iid ~N(0,σ_{ω_t}))

Constraints:

- For p = 1, -1 < ϕ_1 < 1.
- For p = 2, -1 < ϕ_2 < 1, ϕ_2 + ϕ_1 < 1 and ϕ_2 - ϕ_1< 1.

- p := order of the Autoregression

Integrated model (d)

It is important to remove trends, or non-stationarity, from time series data prior to model building, since such autocorrelations dominate the ACF. One way of removing non-stationarity is through the method of differencing. :

$$Y'_t = Y_t - Y_{t-1}$$

Occasionally, such taking of first differences is insufficient to remove non-stationarity. In that case, second-order differences usually produce the desired effect:

$$Y''_t = Y'_t - Y'_{t-1}$$

d := order of the integration

Moving Average (MA) models

$$Y_t = c + w_t - \theta_1 w_{t-1} - \theta_2 w_{t-2} + \cdots$$

Where c is a constant, θ_i are parameters to estimate and w_i the error terms (iid ~N(0,1))

Constraints:

- For q = 1, -1 < θ_1 < 1.
- For q = 2, -1 < θ_2< 1, θ_2 + θ_1 < 1.
- q := order of the Moving Average

Backshift notation

The backshift notation is commonly used to represent ARIMA models. It uses the operator B, which shifts data back one period:

$$BY_t = Y_{t-1}$$

Two applications of B shift the data back two periods:

$$\mathrm{B}(BY_t) = B^2Y_t = Y_{t-2}$$

In general, Bs represents “shift back s time periods”. Note that a first difference is represented by $1 - B$:

$$\mathrm{Y}_{\mathrm{t}}' = Y_t - Y_{t-1} = Y_t - BY_t = (1\text{-B})\, Y_t$$

ARIMA(p,d,q) = AR(p) , I(d) , MA(q)

ARIMA(1,1,1) ➔ Y_t = c + $\phi_1 Y_{t-1} + \theta_1 w_{t-1} + w_t$

➔ (1 – ϕ_1B)(1 – B)$Y_t = c + (1 + \theta_1 B)\, w_t$

The general expression of the ARIMA(p, d, q) in backshift notation :

ARIMA(p,d,q) ➔ (1 – ϕ_1B – ⋯ – $\phi_p B^p$)(1– B)$Y_t = c +$ (1 + θ_1B + ⋯ +$\theta_q B^q$)w_t

Model selection and forecasting

Phase 1 (Identification):

Preliminary analysis and data preparation:

- Difference data to obtain a stationary series.

Model selection:

- Analyse time plots, ACF, PACF to identify potential models.

Model selection and forecasting

Phase 2 (Estimation and testing (1/2)):

1. Estimate parameters in potential models
2. Select best model using suitable criterion…

3. If satisfying model is not found go back to Phase 1
4. Otherwise…

Model selection and forecasting

Phase 2 (Estimation and testing (1/2)):

1. **Estimate parameters in potential models**
2. Select best model using suitable criterion…

3. If satisfying model is not found go back to Phase 1
4. Otherwise…

Likelihood

Likelihood function := L(Model with parameters θ | observed data)

:= Prob(observed data | Model with parameters θ)

The likelihood of a fully-specified model with a set of parameters θ, given some observed data, is equal to the probability of observing these data, given the defined model with those specific parameter values.

Likelihood is a quantitative measure of model fit. Higher likelihoods correspond to a higher probability of the model producing the observed data (the data "fit" the model well).

Maximum Likelihood estimation (MLE)

$$\widehat{\Theta}_{MLE} = \underset{\Theta}{\arg max}\, L(\Theta)$$

The goal of MLE is to find the parameter Θ that maximizes the likelihood function (or equivalently, the log-likelihood).

Likelihood of the Normal Distribution

Consider you have a sample of data $x_1, x_2, \dots, x_n$ drawn from a normal distribution with unknown mean μ and variance σ^2.

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The log-likelihood function is:

$$\log L(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

Maximizing this with respect to μ and σ^2 gives the MLE estimates:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

Sum of squares error (SSE):

$$\boldsymbol{SSE(\theta)} = \sum_{i=1}^{T} (x_i - f(x_{i-1}, \dots; \theta))^2 = \sum_{i=1}^{T} r_i^2(\theta)$$

Parameters

Model prediction

Find the values of the parameters which maximise the probability of obtaining the data that we have observed (minimize $\boldsymbol{SSE(\theta)}$)

AIC (Akaike's Information Criterion)

$$AIC = -2\ln(L) + 2m$$

L:= likelyhood

$m = p + q$.

By varying the choices of p, q, we want to pick the lowest AIC.

AIC penalizes overfitting

BIC (Bayesian Information Criterion)

$$\text{BIC} = -2\ \ln(L) + k(ln(\text{n}))$$

L:= likelyhood

K := number of parameters to estimate

By varying the choices of p, q, we want to pick the lowest BIC.

The BIC introduces a penalty term for the number of parameters in the model. The penalty term is larger in BIC than in AIC.

Model quality - R^2

- The R^2 quantifies the degree of any linear correlation between *the observation* and *the prediction*.

$$\boldsymbol{S(\theta) = \sum_{i=1}^{T} (x_i - f(x_{i-1}, \ldots; \theta))^2 = \sum_{i=1}^{T} r_i^2(\theta)}$$

$$\boldsymbol{S_{tot}(\theta) = \sum_{i=1}^{T} (x_i - \mu_X)^2}$$

$$R^2 = 1 - \frac{\boldsymbol{S(\theta)}}{\boldsymbol{S_{tot}(\theta)}}$$

Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{T}(x_i - \hat{x})^2}{N}}$$

Prediction

The RMSE (or RMSD) of a sample is the quadratic mean of the differences between the observed values and predicted ones. These deviations are called **residuals** when the calculations are performed over the data sample that was used for estimation and are called **errors** (or prediction errors) when computed out-of-sample.

Error

$$\boldsymbol{MAPE} = \sum_{i=1}^{N} \sqrt{\frac{(x_i - \widehat{x_i})^2}{x_i}}$$

Prediction

The **mean absolute percentage error** (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of prediction accuracy of a forecasting method in statistics.

Model selection and forecasting

Phase 2 (Estimation and testing (2/2)):

Testing:

1. Check ACF/PACF of residuals (errors)
2. Are residuals iid ?

References

- Andrew V. Metcalfe, **Paul S.P. Cowpertwait, Introductory Time Series with R** (2009)
- Aileen Nielsen, **Practical Time Series Analysis: Prediction with Statistics and Machine Learning** (2019)
- Changquan Huang , Alla Petukhina, **Applied Time Series Analysis and Forecasting with Python** (2022)
- Peter J. Brockwell, Richard A. Davis, **Introduction to Time Series and Forecasting** (2022)