Lecture 8:
Time Series Analysis

Olexandr Isayev

Department of Chemistry, CMU

olexandr@cmu.edu

Lab 3: Wide data and linear models

Disease type (phenotype) prediction from transcriptomics data

Dataset: differential expression from The Cancer Genome Atlas

~550 patients

UCEC (uterine corpus endometrial carcinoma)

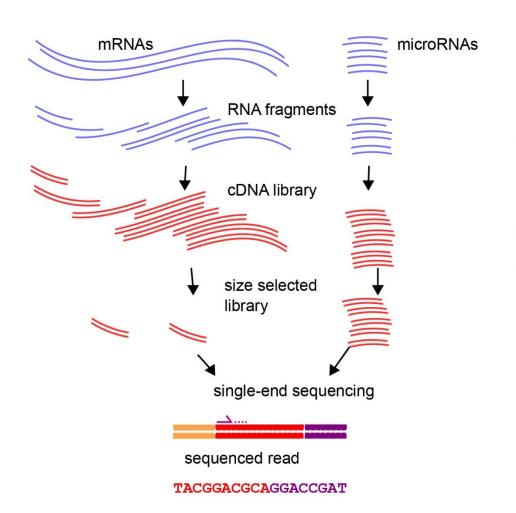Predicting label – tumor grade
"II-" vs. "III+"

RNA Seq data for ~17,000 coding genes

Deadlines

Recital on Thursday and Some Q&A Tuesday after Fall break

Lab 3 Due: October 24th

RNA Sequencing

Features

Expression levels for each coding human gene (~17k in total) for each sample.
Expression levels represented in form of normalized signal: TPM and CLR.

TPM: The counts are transformed into transcripts per million (tpm), which account for gene length to normalize reads. The TPM value is determined in terms of the counts as,

$$\text{tpm}_i = \frac{10^6 \cdot (\text{counts}_i/\text{length}_i)}{\sum_j (\text{counts}_j/\text{length}_j)},$$

in which i and j index genes.

CLR: A centered-log-ratio transform is carried out on the TPM vectors. The CLR value is determined in terms of the TPM values as,

$$\text{clr}_i = \log(\text{tpm}_i) - \frac{1}{N} \sum_j \log(\text{tpm}_j),$$

in which N is the number of genes.

	ENSG00000000003	ENSG00000000005	ENSG00000000419	ENSG00000000457	ENSG00000000938	ENSG00000000971	ENSG00000001036
46C422D1-E1A1-42E5-8FB1-918C0DAE533D	6.610916	0.299408	5.20356	4.022508	5.245761	6.57066	6.57554
FCD43986-D877-4856-A4E0-F35AF0A5117D	7.390681	-5.74411	6.75426	4.888669	3.679566	5.895638	6.882263
122D1D3C-8C13-4BFF-A135-2A308C6A9483	7.887847	-4.31143	7.276095	5.395927	4.09808	5.904496	7.730568
BACA9F84-971B-43B5-834F-8C524D378998	6.186265	0.915815	6.548932	5.30702	2.816588	6.035357	6.513611
38F5957E-0A99-432D-8F0B-B48E03F17668	6.447163	-0.47291	6.18441	4.102008	4.039482	2.318558	6.681119
48A2AAEB-7F60-4284-808E-2058A531B17C	7.024267	5.067618	8.07971	5.446709	4.217084	5.035486	7.447464
0E4F9C8A-3DE5-4893-8ED9-BBF12169F199	7.152781	-5.06512	6.750094	4.881583	4.099117	4.344411	7.517147
D50BAB19-5050-4C27-9F6B-6CB1C865552D	6.945063	-6.69913	7.54104	5.325336	4.034267	5.132126	7.106275
ABCF168A-484E-4AF7-8E2C-918F8244765F	5.93018	-6.03718	6.682619	4.242678	6.054994	4.708649	7.435929
D4E58BFB-9859-4D77-A581-1D713AA54B0B	6.932277	1.543796	7.022409	5.085165	4.761884	5.242993	7.517907
278DBFBE-6831-4335-8750-0B76406779FB	6.451089	-5.08311	6.631651	5.432744	4.443306	5.197162	7.055614
2A02A359-28A1-41B9-8EDD-2477C95CD00A	7.40088	1.529062	6.753776	5.40047	4.833505	6.668592	7.349799
3DA9560E-8181-4BAD-A90E-E4EE558E6EAF	6.997104	1.148314	6.64975	4.890514	4.283081	3.978705	7.006432
A8D65DDB-516A-4F8C-8038-483A254D41B1	7.09325	0.812931	7.145081	5.000056	3.849058	6.076678	7.206288

Labels

	xml_neoplasm_histologic_grade
46C422D1-E1A1-42E5-8FB1-918C0DAE533D	0
FCD43986-D877-4856-A4E0-F35AF0A5117D	1
122D1D3C-8C13-4BFF-A135-2A308C6A9483	0
BACA9F84-971B-43B5-834F-8C524D378998	1
38F5957E-0A99-432D-8F0B-B48E03F17668	0
48A2AAEB-7F60-4284-808E-2058A531B17C	1
0E4F9C8A-3DE5-4893-8ED9-BBF12169F199	0
D50BAB19-5050-4C27-9F6B-6CB1C865552D	1
ABCF168A-484E-4AF7-8E2C-918F8244765F	1
D4E58BFB-9859-4D77-A581-1D713AA54B0B	0
278DBFBE-6831-4335-8750-0B76406779FB	0
2A02A359-28A1-41B9-8EDD-2477C95CD00A	0
3DA9560E-8181-4BAD-A90E-E4EE558E6EAF	0
A8D65DDB-516A-4F8C-8038-483A254D41B1	0
1D34604D-44C1-431F-B6E5-8C78B11F5A9E	0
F3C17A4E-3940-4DAB-8D00-FA6A16118B75	1
2AE27B42-030A-4801-8B03-0AFF4F1AEA3E	1
6A37EAD0-5083-426C-A6D9-274EB0F8D42C	0
EF5618C6-6AFB-4FB5-A32B-7946F18E9736	1
B3029936-1EEF-4378-AA28-D5B2A32CDC86	0
39A9B95F-186B-4550-9271-001151B1A626	1
536FEED2-6251-41BC-AA61-EA46744EAB05	1
AA26E72A-924D-423E-90F6-96C67A7DFBBE	0

Tasks:

- Please use binary classification (0/1) using a linear models. Measure the classification performance using accuracy and F1-score on the Kaggle test set.

- Develop a pipeline to try different linear models (linear regression, logistic regression, Ridge regression, LASSO etc.)

- Study effect of regularization parameters and type on model performance. What model is the best?

- Compare your best model accuracy with random guessing (Hint: scramble labels aka Y-randomization)

- What are the top important genes for the model decision?

- Try to visualize dataset and see if you could visually separate two groups of patients

Bonus Qs

- You could use https://www.uniprot.org/ and search for Gene IDs.

- See if there is a meaningful connection between top 10 most important genes and disease.

- Did your model recapitulate of known association between Genes and Disease?

Linear Regression

Linear models can be used to model the dependence of a regression target y on some features x. The learned relationships are linear and can be written for a single instance i as follows:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p + \epsilon$$

The predicted outcome of an instance is a weighted sum of its p features. The betas (β_j) represent the learned feature weights or coefficients. The first weight in the sum (β_0) is called the intercept and is not multiplied with a feature. The epsilon (ϵ) is the error we still make, i.e. the difference between the prediction and the actual outcome.

These errors are assumed to follow a Gaussian distribution, which means that we make errors in both negative and positive directions and make many small errors and few large errors.

Logistic Regression

$$\hat{y}^{(i)} = \beta_0 + \beta_1 x_1^{(i)} + \ldots + \beta_p x_p^{(i)}$$

$$P(y^{(i)} = 1) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 x_1^{(i)} + \ldots + \beta_p x_p^{(i)}))}$$

GLM

The GLM mathematically links the weighted sum of the features with the mean value of the assumed distribution using the link function g, which can be chosen flexibly depending on the type of outcome

$$g(E_Y(y|x)) = \beta_0 + \beta_1 x_1 + \ldots \beta_p x_p$$

The link function g, the weighted sum $X^T\beta$ (sometimes called linear predictor) and a probability distribution from the exponential family that defines E_Y.

Generalized Additive Models (GAMs)

Why not 'simply' allow the (generalized) linear model to learn nonlinear relationships? That is the motivation behind GAMs.

GAMs relax the restriction that the relationship must be a simple weighted sum, and instead assume that the outcome can be modeled by a sum of arbitrary functions of each feature.

$$g(E_Y(y|x)) = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_p(x_p)$$

The core of a GAM is still a sum of feature effects, but you have the option to allow nonlinear relationships between some features and the output.

The Ridge regression

The solution to the ordinary least squares fitting procedure is the vector $\hat{\beta} = (\hat{\beta}_0, ..., \hat{\beta}_p)$ that minimizes the Residual Sum of Squares
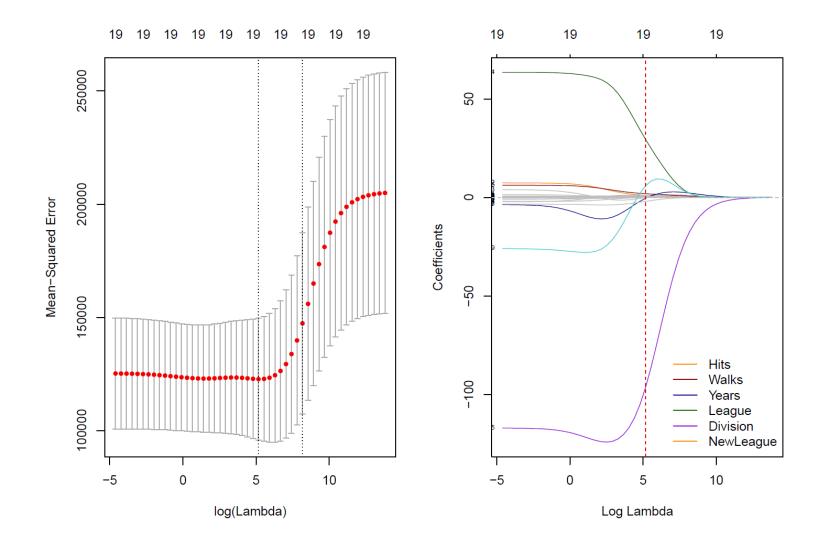
$$RSS = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2.$$

Ridge regression, similarly, seeks the vector $\hat{\beta}^{\text{Ridge}}$ that minimizes the *penalized* or *regularized* RSS

$$\sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 = RSS + \lambda \sum_{j=1}^{p} \beta_j^2$$

where $\lambda \geq 0$ is a **complexity parameter**.

Choosing λ: k-fold cross-validation

The LASSO Regression
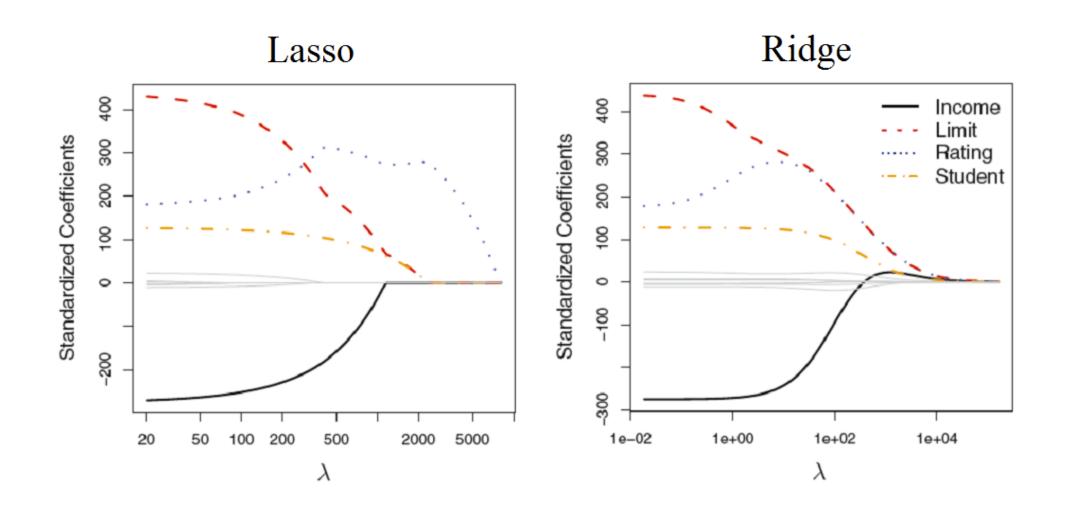
"LASSO" stands for **L**east **A**bsolute **S**hrinkage and **S**election **O**perator

The idea of constraining the size of the OLS estimates can be extended to consider different kinds of penalizations.

While the Ridge penalty encompasses the *L*2 norm of the estimates vector, the **LASSO** makes use of the *L*1 norm:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = RSS + \lambda\sum_{j=1}^{p}|\beta_j|$$

Comparing LASSO and Ridge

Comparing LASSO and Ridge

A hybrid penalty: the Elastic Nets

The LASSO sometimes does not perform well with highly correlated variables, and often performs worse than Ridge in prediction.

To overcome this limitations, a penalty that combines the $L1$ and $L2$ constraints has been developed.

An **elastic net** is a regularization and variable selection procedure that makes use of the penalty
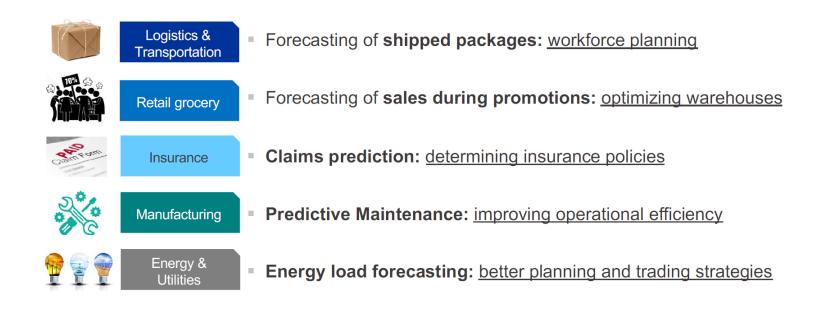
$$\lambda \left[\frac{1}{2}(1-\alpha) \sum_{j=1}^{p} \beta_j^2 + \alpha \sum_{j=1}^{p} |\beta_j| \right]$$

where $\alpha \in [0, 1]$ is called the **mixing** parameter and λ has the usual interpretation. LASSO and Ridge are special cases, respectively for $\alpha = 1$ and $\alpha = 0$.

Applications of Time series Analysis

The fields of application of Time series Analysis are numerous:

Demand Planning is one of the most common application, however, from industry to industry there are other possible uses. For instance:

Logistics & Transportation	▪	Forecasting of **shipped packages:** <u>workforce planning</u>
Retail grocery	▪	Forecasting of **sales during promotions:** <u>optimizing warehouses</u>
Insurance	▪	**Claims prediction:** <u>determining insurance policies</u>
Manufacturing	▪	**Predictive Maintenance:** <u>improving operational efficiency</u>
Energy & Utilities	▪	**Energy load forecasting:** <u>better planning and trading strategies</u>

Time series

A time series is a series of data points indexed in time order.

Most commonly, a time series is a sequence taken at successive equally spaced points in time.

In a time series, time is often the independent variable and the goal is usually to make a forecast for the future.

Components of Time-Series Data

Typical questions

Is it **stationary**?

Is there a **seasonality**?

Is the target variable **autocorrelated**?
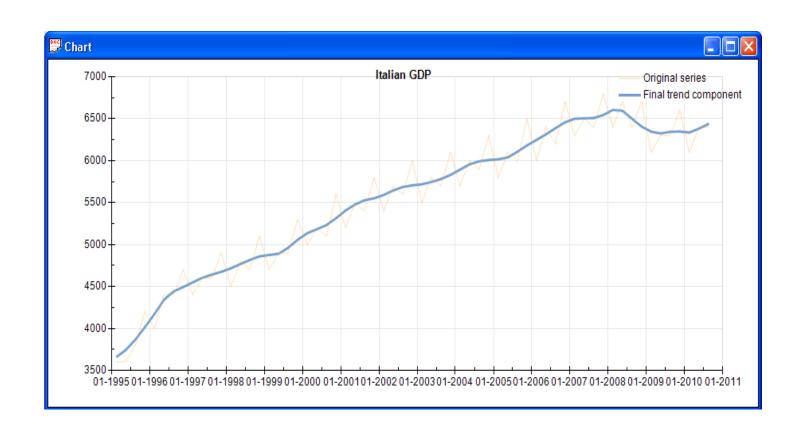
Usual Components

- **The Trend Component**
  - The Trend is the long term evolution of the series that can be observed on several decades

- **The Cycle Component**
  - The Cycle is the smooth and quasi-periodic movement of the series that can usually be observed around the long term trend

- **The Seasonal Component (Seasonality)**
  - Fluctuations observed during the year (each month, each quarter) and which appear to repeat themselves on a more or less regular basis from one year to other
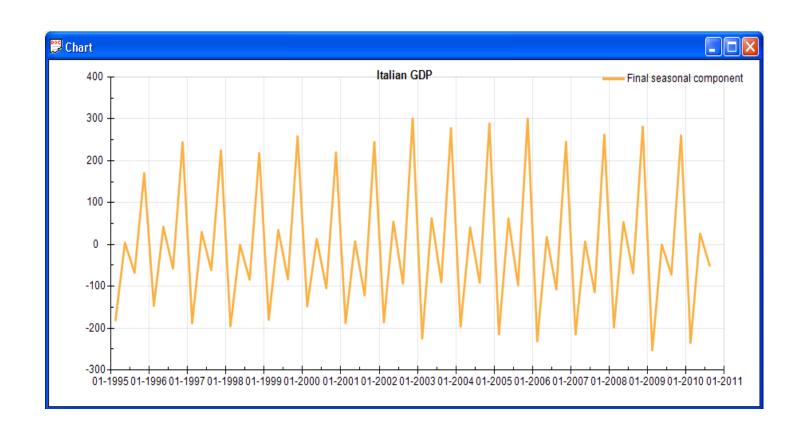
Usual Components

- The Calendar Effect
  - Any economic effect which appears to be related to the calendar (e.g. one more Sunday in the month can affect the production)

- The Irregular Component
  - The Irregular Component is composed of residual and random fluctuations that cannot be attributed to the other "systematic" components

- Outliers
  - Different kinds of Outliers can be defined
  - year to other
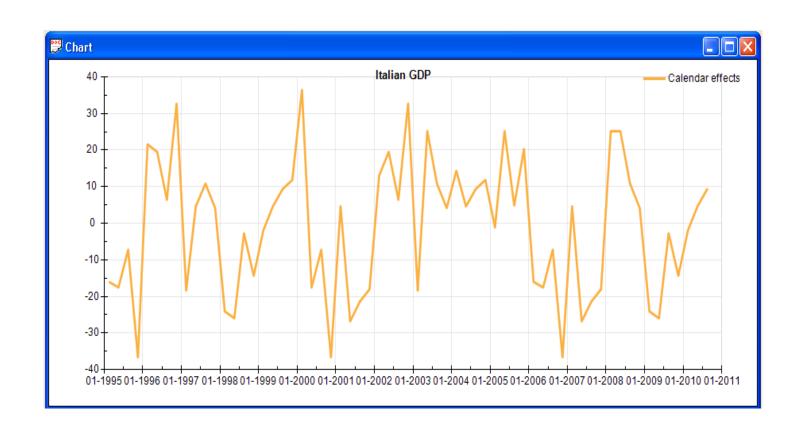
Italian GDP – Trend component
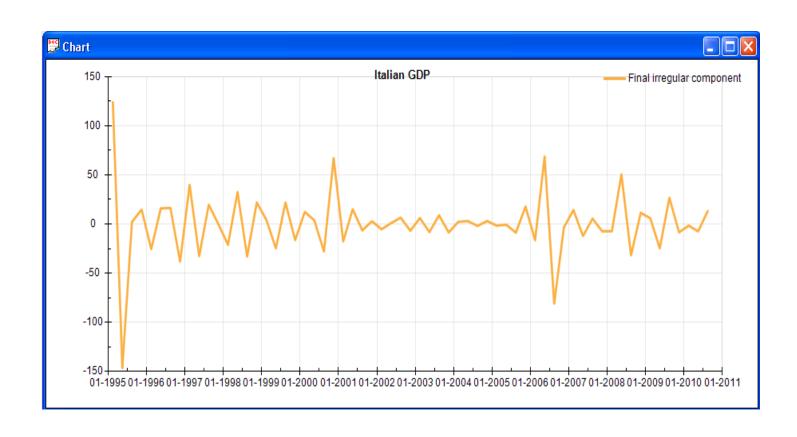
Italian GDP – Seasonal component
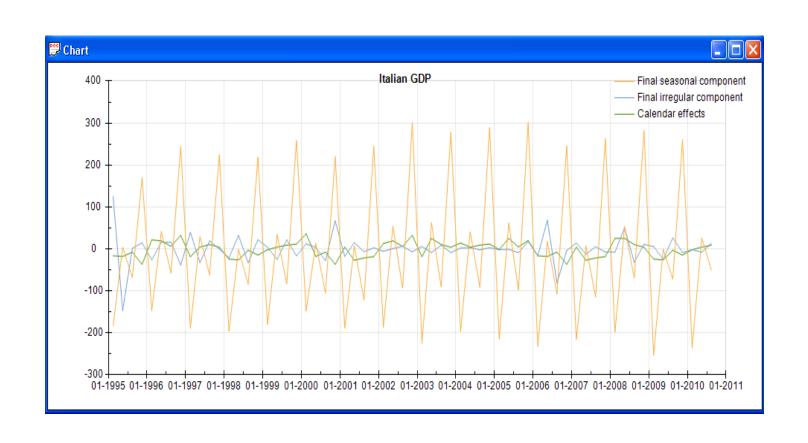
Italian GDP – Calendar component

Italian GDP – Irregular component

Italian GDP – Irregular, Seasonal and Calendar

Trend

The Trend Component is defined as the **long-term movement** in a series

The Trend is a reflection of the underlying level of the series. This is typically due to influences such as population growth, price inflation and general economic development

The Trend Component is sometimes referred to as the Trend-Cycle (see Cycle Component)

Cause of Seasonality

Seasonality and Climate: due to the variations of the weather and of the climate (seasons!)
- Examples: agriculture, consumption of electricity (heating)

Seasonality and Institutions: due to the social habits and practices or to the administrative rules
- Examples: effect of Christmas on the retail trade, of the fiscal year on some financial variables, of the academic calendar

Indirect Seasonality: due to the Seasonality that affects other sectors
- Examples: toy industry is affected a long time before Christmas. A Seasonal increase in the retail trade has an impact on manufacturing, deliveries, etc..

Seasonal Adjustment

Seasonal Adjustment is the process of **estimating and removing the Seasonal Effects** from a Time Series, and by Seasonal we mean an effect that happens at the same time and with the same magnitude and direction every year

The basic goal of Seasonal Adjustment is to decompose a Time Series into several different components, including a Seasonal Component and an Irregular Component

Because the Seasonal effects are an unwanted feature of the Time Series, Seasonal Adjustment can be thought of as focused noise reduction

Seasonal Adjustment

Since Seasonal effects are annual effects, <u>the data must be collected at a frequency less than annually</u>, usually monthly or quarterly

For the data to be useful for Time Series analysis, the data should be comparable over time. This means:

- The measurements should be taken over discrete (nonoverlapping) consecutive periods, i.e., every month or every quarter

- The <u>definition of the concept and the way it is measured should be consistent</u> over time

Seasonal Adjustment
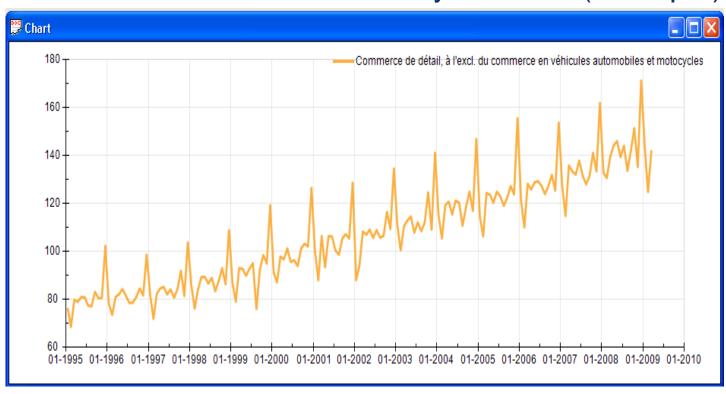
- Keep in mind that <u>longer series are NOT necessarily better</u>. If the series has changed the way the data is measured or defined, it might be better to cut off the early part of the series to keep the series as homogeneous as possible

- The best way <u>to decide if your series needs to be shortened</u> is to investigate the data collection methods and the economic factors associated with your series and choose a length that gives you the most homogeneous series possible

Types of Stationarity

- **Strict stationarity** - This means that the unconditional joint distribution of any moments (e.g. expected values, variances, third-order and higher moments) remains constant over time. This type of series is rarely seen in real-life practice.

- **First-order stationarity** - These series have a mean constant over time. Any other statistics (like variance) can change at the different points in time.

- **Second-order stationarity (also called weak stationarity)** - These time series have a constant mean and variance over time. Other statistics in the system are free to change over time. This is one of the most commonly observed series in real-life practice.

- **Trend stationarity** - These series are the series with a trend. This trend when removed from the series leaves a stationary series.

- **Difference stationarity** - These series are the series that need one or more differencing to become stationary.

Seasonal Adjustment
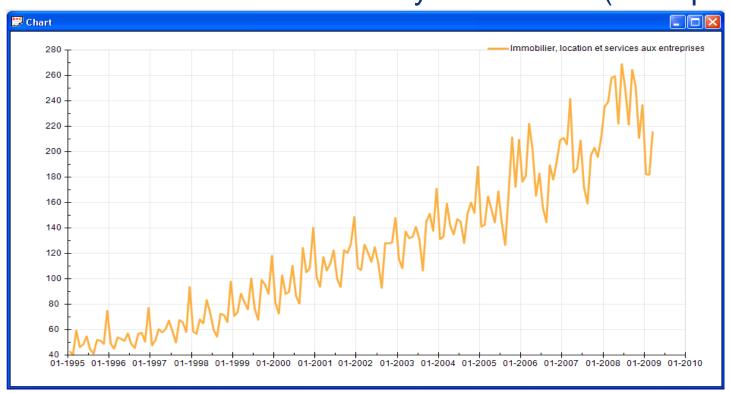
A first overview – No Stationary in mean (example)

Seasonal Adjustment

A first overview – No Stationary in variance (example)
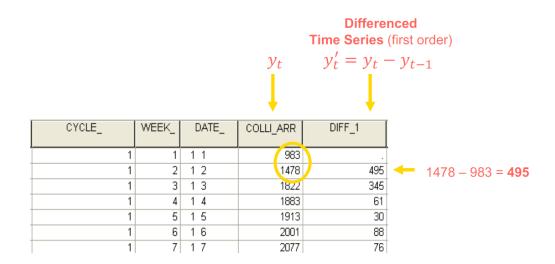
Differencing

One way to make a time series stationary is to compute the differences between consecutive observations.
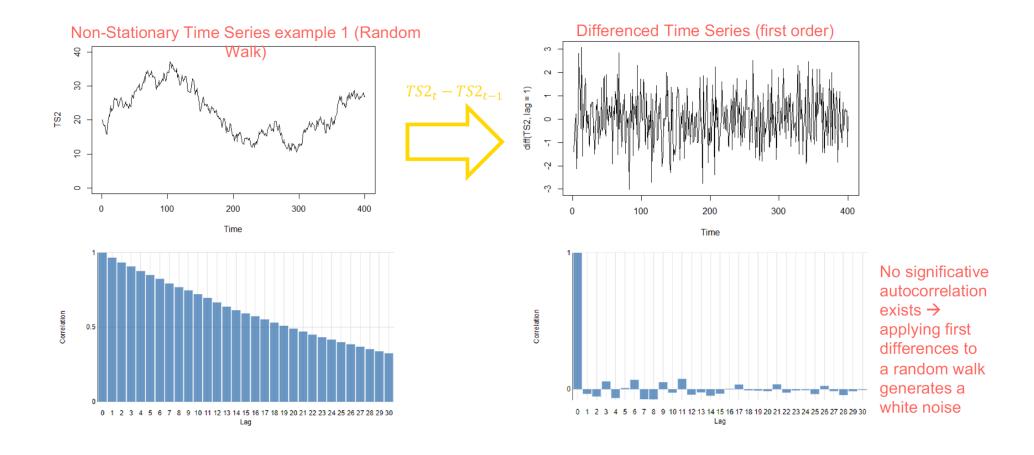
This Is known as DIFFERENCING

Differencing can help **stabilize the mean** of a time series by removing changes in the level of a time series, and so eliminating trend (and also seasonality, using a specific differencing order)

The **Order of Integration** for a Time Series, denoted **I(d)**, reports the minimum number of differences (d) required to obtain a stationary series (*note: I(0)* → it means the series is stationary!)

Differenced
Time Series (first order)

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

CYCLE_	WEEK_	DATE_	COLLI_ARR	DIFF_1
1	1	1 1	983	.
1	2	1 2	1478	495
1	3	1 3	1822	345
1	4	1 4	1883	61
1	5	1 5	1913	30
1	6	1 6	2001	88
1	7	1 7	2077	76

← 1478 – 983 = **495**

Differencing

Non-Stationary Time Series example 1 (Random Walk)

$TS2_t - TS2_{t-1}$

Differenced Time Series (first order)

No significative autocorrelation exists → applying first differences to a random walk generates a white noise

Example: use differencing to make stationary a non-stationary series

Differencing

Occasionally the differenced data will not appear stationary and it may be necessary to difference the data a second time to obtain a stationary series

Decomposition Models

Usual Additive and Multiplicative Models

$$X_t = T_t + C_t + S_t + I_t$$

$$X_t = T_t * C_t * S_t * I_t$$
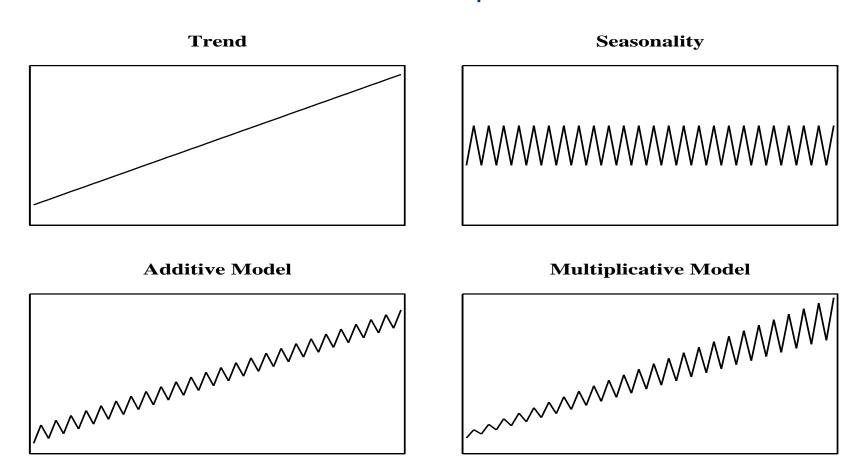
$$X_t = T_t * (1 + C_t) * (1 + S_t) * (1 + I_t)$$

More components: Outliers, Calendar Effects
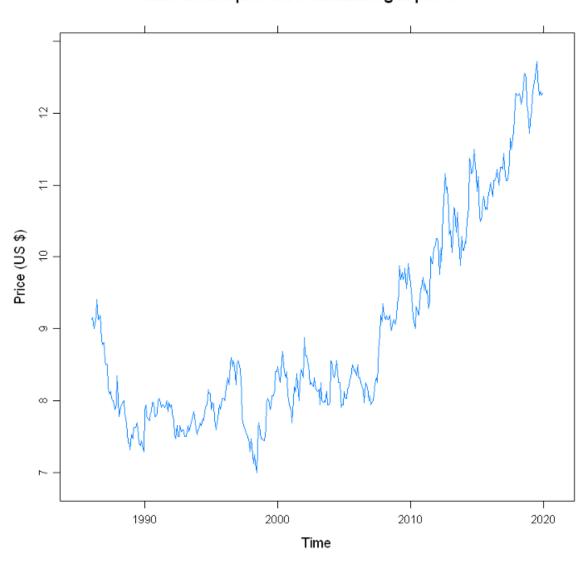
$$X_t = T_t + C_t + S_t + O_t + TD_t + MH_t + I_t$$

Decomposition Models
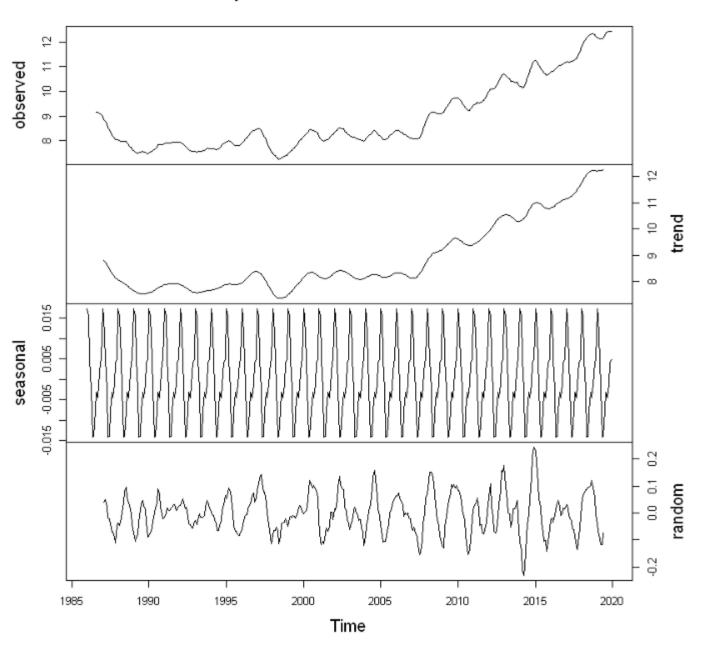
Some usual shapes

Time series plot for Schlumberger price

Decomposition of additive time series

Calendar Adjustment

Calendar Effects typically include:

- Different number of Working Days in a specific period

- Composition of Working Days

- Leap Year effect

- Moving Holidays (Easter, Ramadan, etc.)

Calendar Adjustment - Trading Day Effect
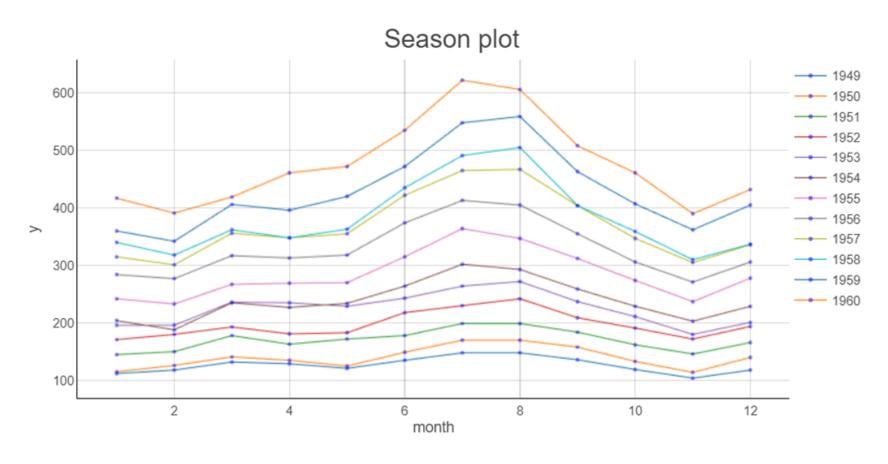
- Recurring effects associated with individual days of the week. This occurs because only non-leap-year Februarys have four of each type of day: four Mondays, four Tuesdays, etc.

- All other **months have an excess of some types of days**. If an activity is higher on some days compared to others, then the series can have a Trading Day effect. For example, building permit offices are usually closed on Saturday and Sunday

- Thus, the number of building permits issued in a given month is likely to be higher if the month contains a surplus of weekdays and lower if the month contains a surplus of weekend days

Calendar Adjustment - Moving Holiday Effect

Effects from **holidays that are not always on the same day of a month**, such as Labor Day or Thanksgiving.
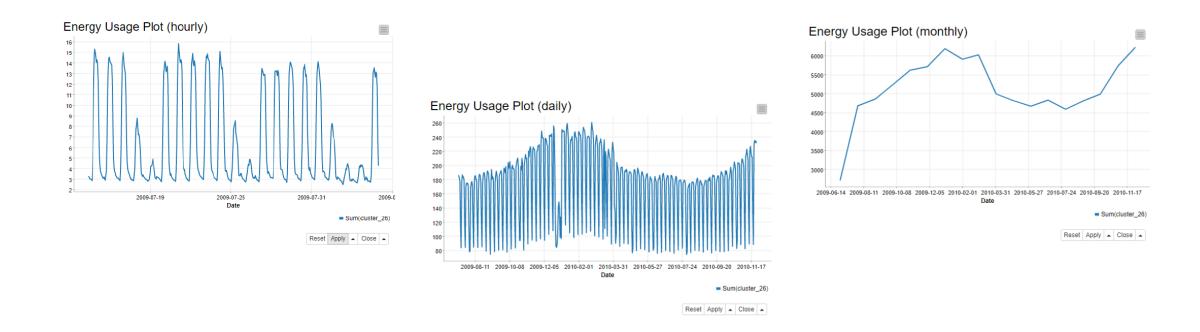
The most important Moving Holiday in the US and European countries is Easter, not only because it moves between days, but also because it moves between months since it can occur in March or April
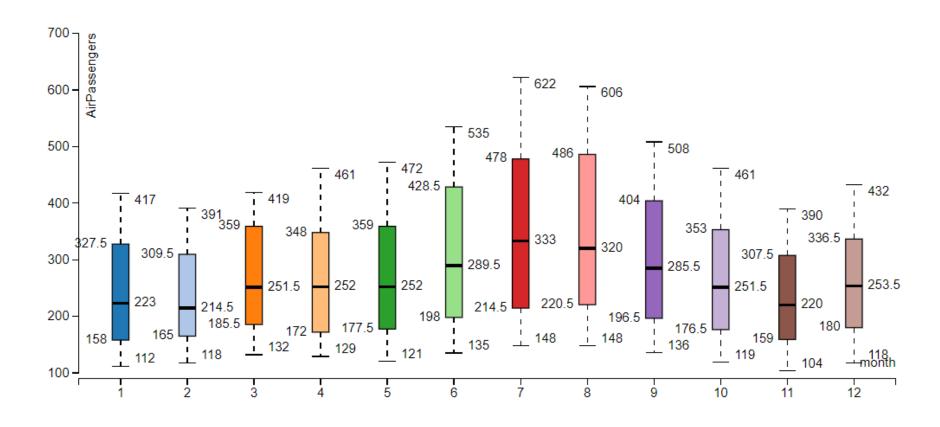
Graphical Analysis: Seasonal Plot

Produce the **Seasonal plot** of the Time series in order to analyze more in detail the seasonal component (and possible changes in seasonality over time)

Granularity and Line Plots

Show time series by hour, day, and month in line plots
Identify daily, weekly, and yearly seasonality

Graphical Analysis: Box Plot

Create the **conditional Box plot** of the Time series in order to deeply understand the distribution of data in the same period of each seasons and focusing on specific aspects such as outliers, skewness, variability,…

Irregular Component

The Irregular Component is the remaining component of the series **after the Seasonal and Trend Components have been removed** from the original data

For this reason, it is also sometimes referred to as the Residual Component. It attempts to capture the remaining short term fluctuations in the series which are neither systematic nor predictable

Exploration – Basic tools

Exploration is a very essential step when analyzing a Time Series

Looking for "structures" in the series
- Trend, Seasonality, "strange" points or behavior etc.
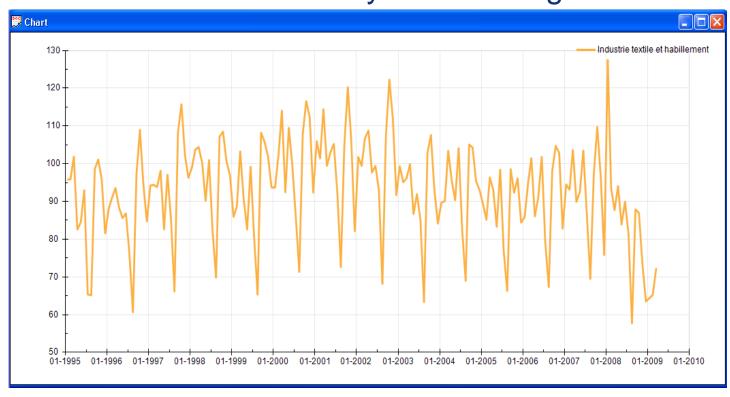
Helps to formulate a global or decomposition model for the series

Graphics are a key player in this exploration
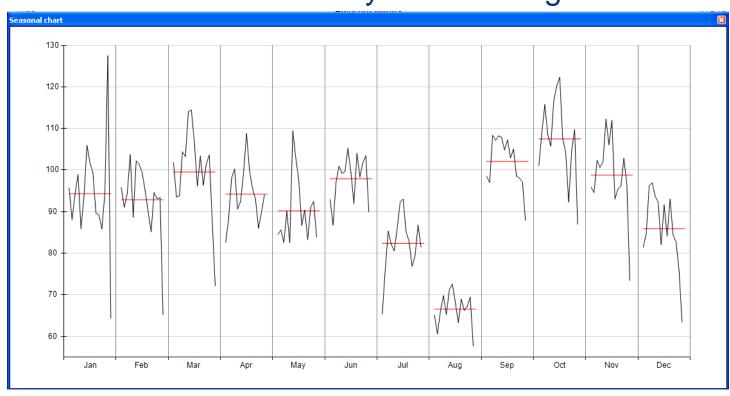
Exploration – Usual representation

Textile industry and clothing

Exploration – Seasonal Chart

Textile industry and clothing

Why Seasonal Adjustment?

Business cycle analysis

To improve comparability:

Over time:
- o Example: how to compare the first quarter (with February) to the fourth quarter (with Christmas)?

Across space:
- o Never forget that while we are freezing at work, Australians are burning on the beach!
- o Very important to compare national economies (convergence of business cycles) or sectors

Why Seasonal Adjustment?

The aim of Seasonal Adjustment is to eliminate Seasonal and Calendar Effects. Hence **there are no Seasonal and Calendar Effects in a perfectly Seasonally Adjusted series**
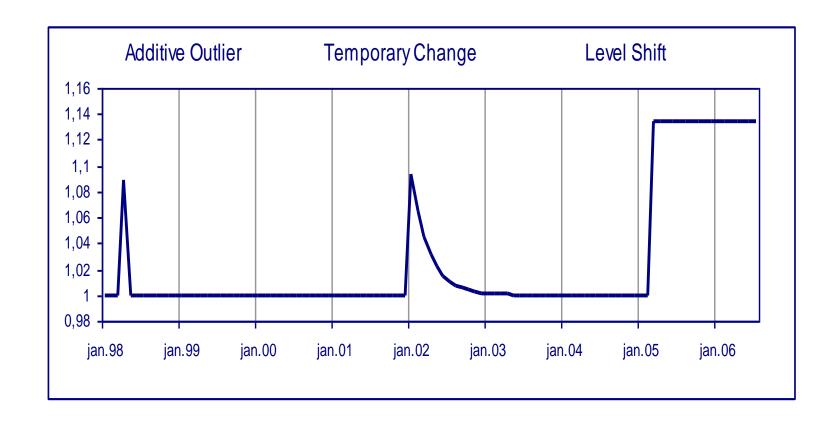
In other words: Seasonal Adjustment transforms the world we live in into a world where no Seasonal and Calendar Effects occur. In a Seasonally Adjusted world the temperature is exactly the same in winter as in the summer, there are no holidays, Christmas is abolished, people work every day in the week with the same intensity (no break over the weekend), etc.

Outliers

Outliers are data which do not fit in the tendency of the Time Series observed, which fall outside the range expected on the basis of the typical pattern of the Trend and Seasonal Components
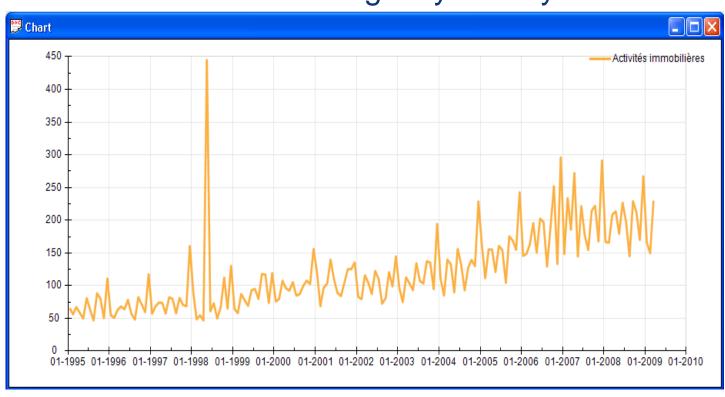
Additive Outlier (AO): the value of only one observation is affected. AO may either be caused by random effects or due to an identifiable cause as a strike, bad weather or war

Temporary Change (TC): the value of one observation is extremely high or low, then the size of the deviation reduces gradually (exponentially) in the course of the subsequent observations until the Time Series returns to the initial level
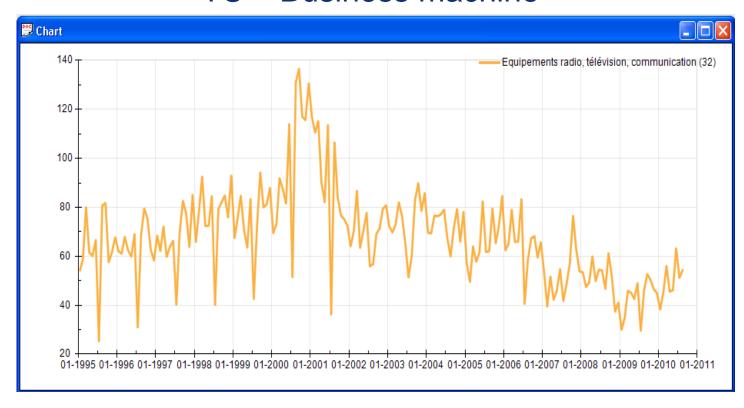
Level Shift (LS): starting from a given time period, the level of the Time Series undergoes a permanent change. Causes could include: change in concepts and definitions of the survey population, in the collection method, in the economic behavior, in the legislation or in the social traditions
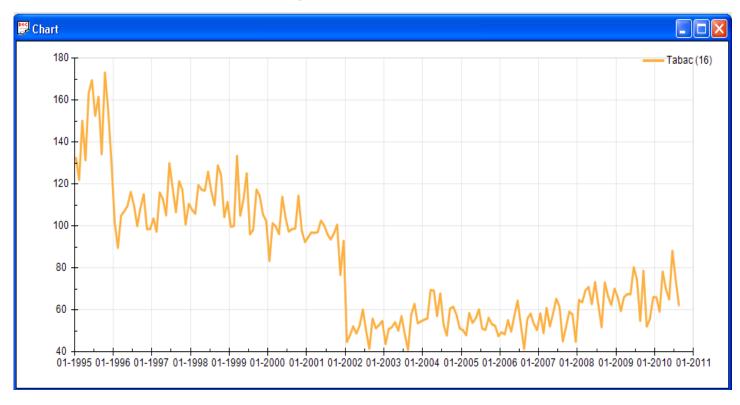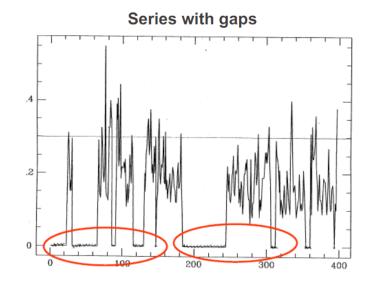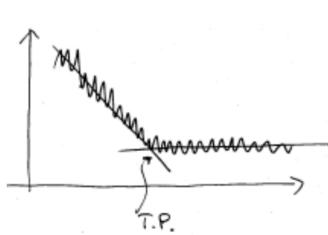
Outliers

Outliers

AO - Estate agency activity

Outliers

TC – Business machine

Outliers
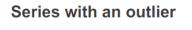
LS – Tobacco

Series with gaps

Series with a turning point

Series with an outlier

Outliers

Additive Outliers:
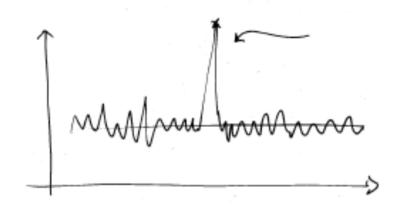Unusual high or low singular values in the data series

Transitory Changes:
Transitory changes in the trend, followed by slow comebacks to the initial tendency

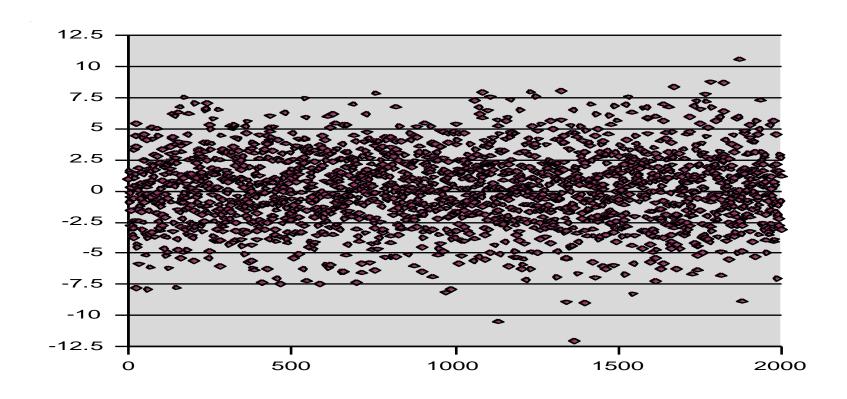**Level Shift:**
Clear changes of the trend

Assimilated to the Irregular Component

Assimilated to the Trend Component

Common Models

- White Noise
- AR
- MA
- ARMA
- ARIMA
- SARIMA
- ARMAX
- Kalman Filter
- Exponential Smoothing, trend, seasons

White Noise

mean=zero, Finite std.dev., often unknown
Often, but not always, Gaussian

Autoregressive Models (AR)

An **autoregressive model** is when a value from a time series is regressed on previous values from that same time series. For example, y_t on y_{t-1}:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t.$$

In this regression model, the response variable in the previous time period has become the predictor and the errors have our usual assumptions about errors in a simple linear regression model. The **order** of an autoregression is the number of immediately preceding values in the series that are used to predict the value at the present time. So, the preceding model is a first-order autoregression, written as AR(1).
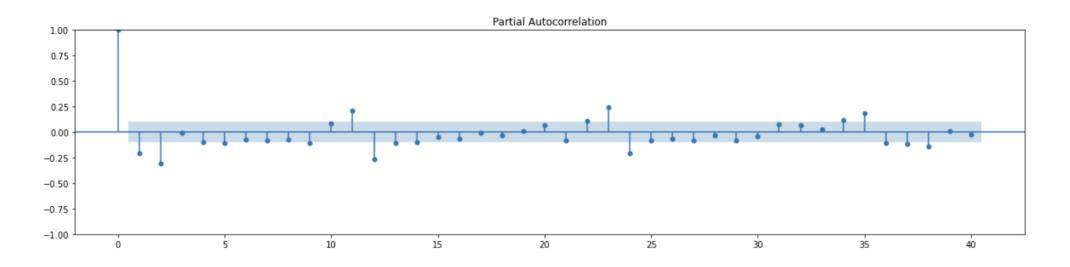
If we want to predict y using measurements of the previous two observations (y_{t-1}, y_{t-2}), then the autoregressive model for doing so would be:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \epsilon_t.$$

This model is a second-order autoregression, written as AR(2), since the value at time t is predicted from the values at times t−1 and t−2.

How to Determine AR Order?

PACF (Partial Autocorrelation Function) is typically used for determining p values. For a given observation in a time series X_t, it may be correlated with a lagged observation X_{t-3} which is also impacted by its lagged values (e.g. X_{t-2}, X_{t-1}). PACF visualizes the **direct** contribution of the past observation to the current observations. For example, the PACF below when lag = 3 the PACF is roughly -0.60, which reflects the impact of lag 3 on the original data point, while the compound factor of lag 1 and lag 2 on lag 3 are not explained in the PACF value. The p values for the AR(p) model is then determined by when the PACF drops to below significant threshold (blue area) for the first time, i.e. p = 4 in this example below.

Moving Average (MA)

Moving average model (MA) model generates the current values based on the **ERRORS** from the past forecasts instead of using the past values like AR. Past errors are analyzed to produce the current value. Perfecting a baking recipe will be like a moving average model. You will do adjustments for needed sugars or butter for today's baking depending on the previous days' amount to perfect the recipe.
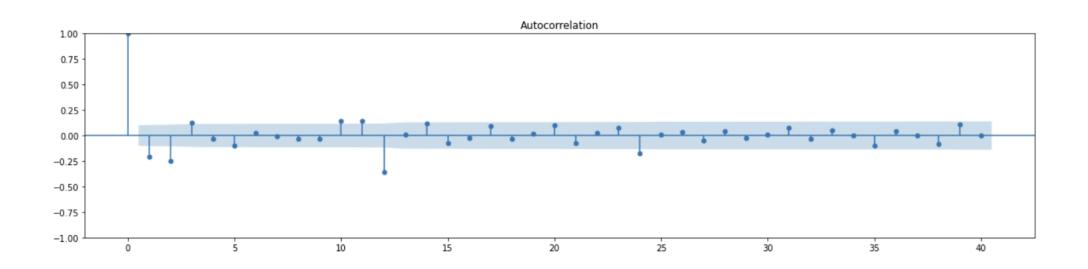
$$MR(X_t, q) = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q}$$

MR(q)
Moving average model, MR(q) adjusts the model based on the average predictions errors from previous q observations, which can be stated as below, where e represents the error terms and ϑ represents the weights. q value determines the number of error terms to include in the moving average window.

How to determine the q value in MA?

ACF can be used for determining the q value. It is typically selected as the first lagged value of which the ACF drops to nearly 0 for the first time. For example, we would choose q=4 based on the ACF plot below.
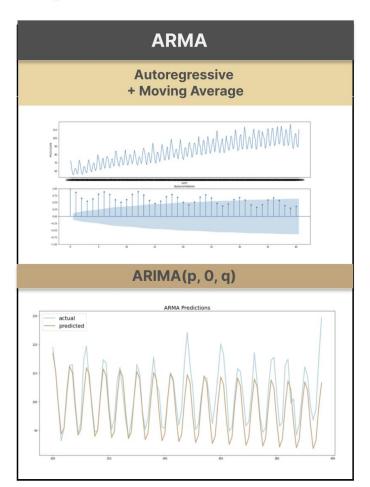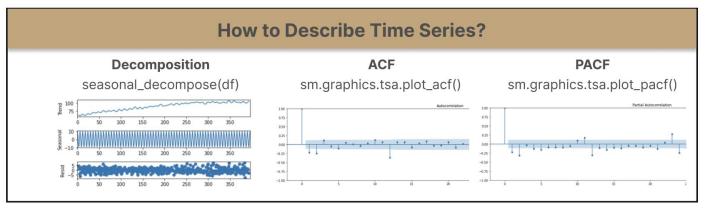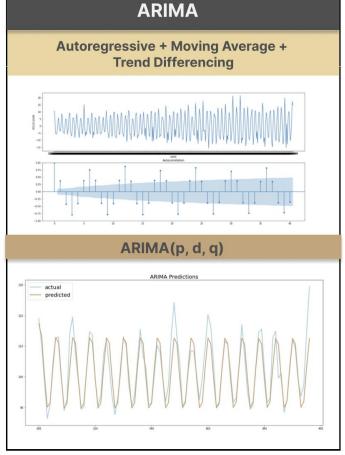
Time Series Analysis

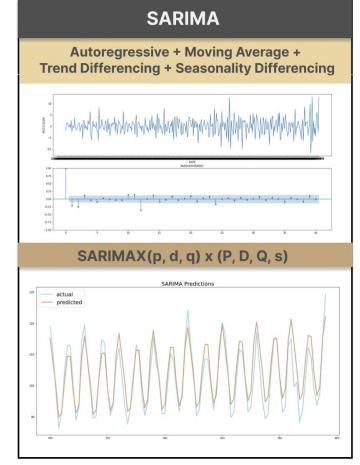Time Series is a unique type of machine learning where time plays a critical role in model predictions.

Time Series Components:
1. Trend
2. Seasonality
3. Residual

How to Describe Time Series?

Decomposition
seasonal_decompose(df)

ACF
sm.graphics.tsa.plot_acf()

PACF
sm.graphics.tsa.plot_pacf()

ARMA

Autoregressive + Moving Average

ARIMA(p, 0, q)

ARIMA

Autoregressive + Moving Average + Trend Differencing

ARIMA(p, d, q)

SARIMA

Autoregressive + Moving Average + Trend Differencing + Seasonality Differencing
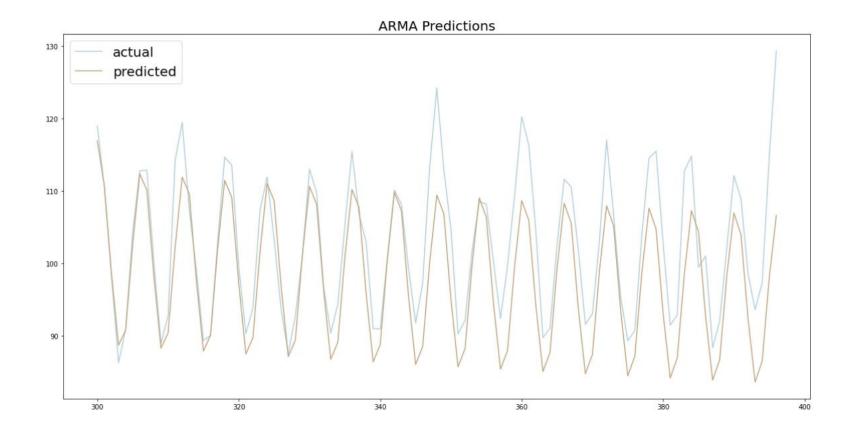

SARIMAX(p, d, q) x (P, D, Q, s)

Autoregressive Integrated Moving Average

- ARIMA stands for **Autoregressive Integrated Moving Average**, which extends from ARMA model and incorporates the integrated component (inverse of differencing).

- ARIMA builds upon autoregressive model (AR) and moving average model (MA) by introducing degree of differencing components (specified as the parameter d) - ARIMA (p, d, q).
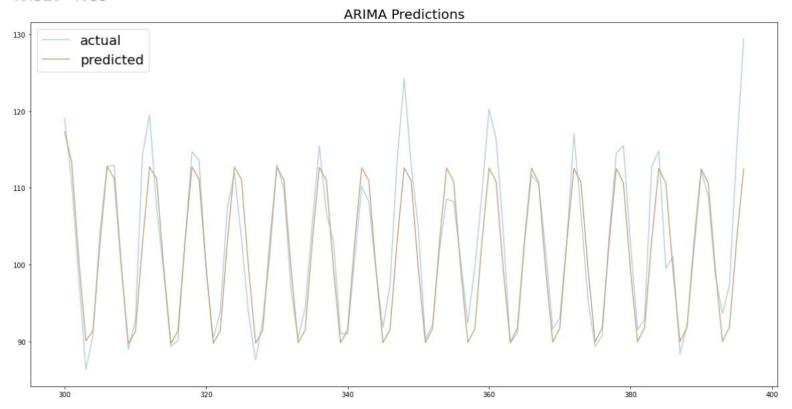
Seasonal ARIMA

SARIMA stands for Seasonal ARIMA which addresses the periodic pattern observed in the time series. Previously we have introduced how to use seasonal differencing to remove seasonal effects. SARIMA incorporates this functionality to predict seasonally changing time series and we can implement it using SARIMAX(p, d, q) x (P, D, Q, s).
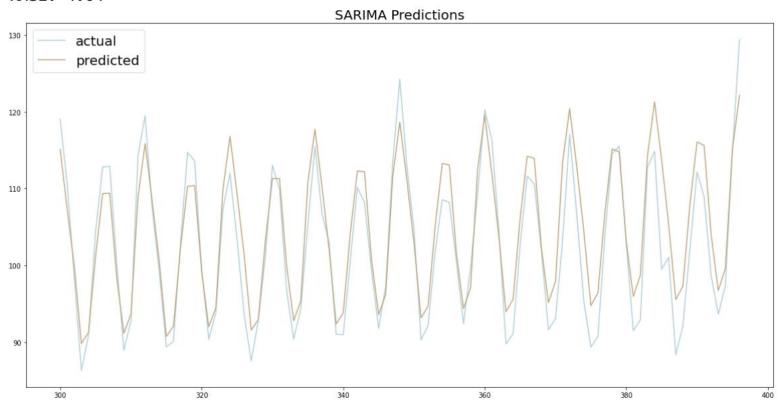
The first term (p, d, q) represents the order of the ARIMA model and (P, D, Q, s) represents the seasonal components. P, D, Q are the autoregressive, differencing and moving average terms of the seasonal order respectively. s is the number of observations in each period.

ARMA Predictions

RMSE: 4.35

ARIMA Predictions

RMSE: 4.04

SARIMA Predictions

Reading

1. Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods.* Oxford university press.

2. Giovanni Petris & Sonia Petrone (2011), *State Space Models in R, Journal of Statistical Software*

3. G Petris, S Petrone, and P Campagnoli (2009). *Dynamic Linear Models with R. Springer*

4. Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice. OTexts.*

-