Chapter 3
Time series

UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

Outline of the chapter

Chapter on time series analysis

1. Decomposition of a time series
2. Determination of the trend
3. Identification of the seasonal component
4. Seasonally adjusted series (SA series)
5. Forecasting

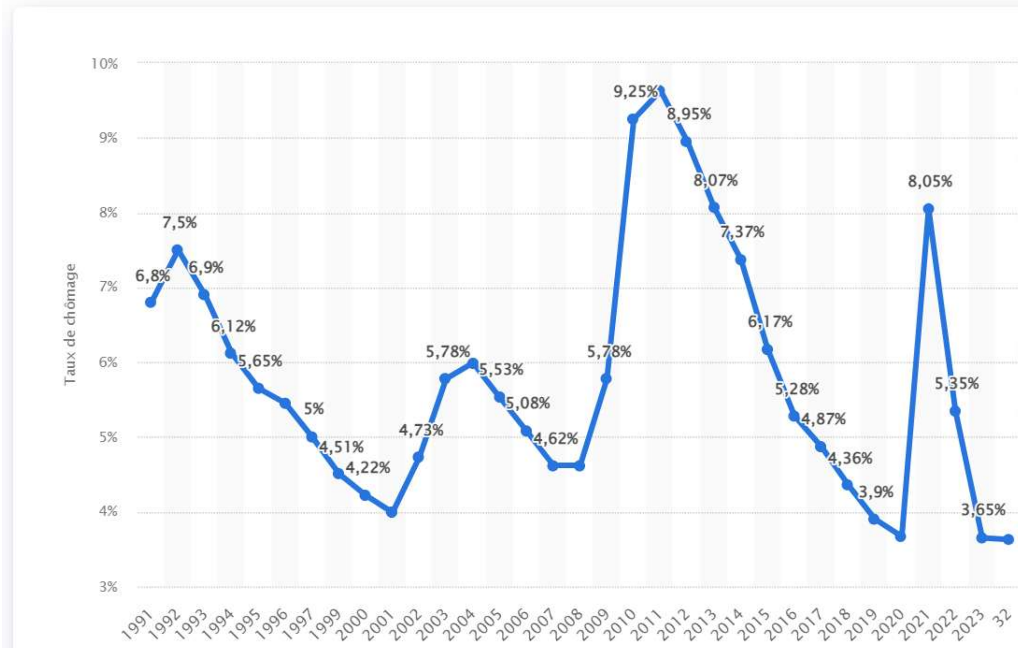
Introduction

- Evolution of a variable **over time**: time series. Series of observations, each observation is measured at a point in time.
- The order is important: specificity of time series.
- How do we measure time?
  - Calendar time: days, months, quarters, years...

Example 1

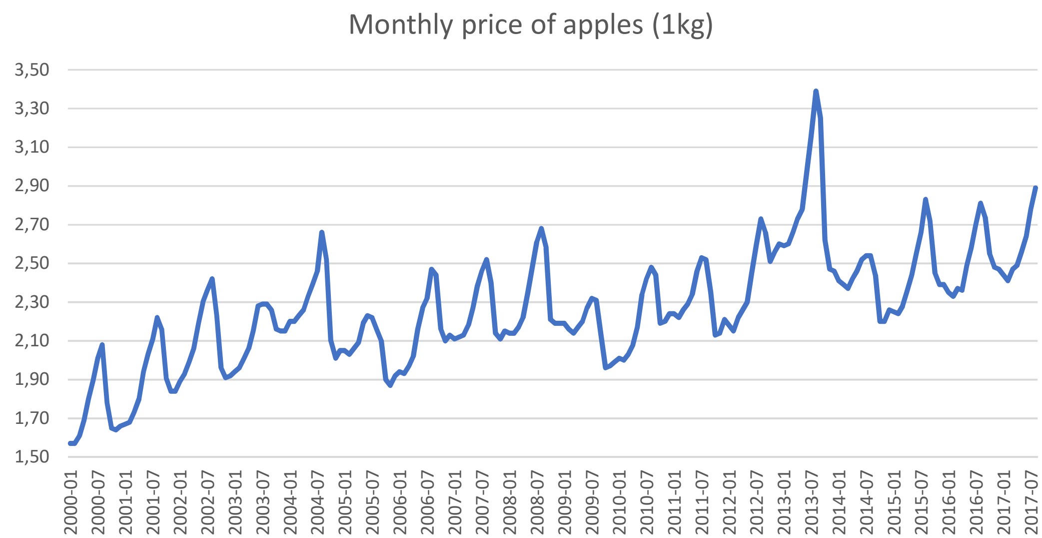
- Example 1: american unemployment rate (% of the active popolation), measured at the end of each year.
- Bounded (between 0 and 100%)
- Variations due to the economic cycle

Unemployment rate in the US between 1991 and 2023



Example 2

- Example 2: monthly price of apples (1kg), in euros.
- Variations due to seasons: high price in summer, low in winter.
- Presence of a trend.



What we learned from these 2 examples

- Time series are functions of time: bivariate statistical series (t, x_{t})
- **Example 1:** A cyclical variable fluctuating around a stable level → stable central tendency → **stationary variable**.
- **Example 2:** The central tendency appears to be increasing → **non-stationary variable**.
- It is not always easy to clearly distinguish the nature and actual evolution of the variable.

Objectives

- **Analysing a time series**: we have different objectives:
  - Describing the evolution of a variable over time
  - Interpreting this evolution
  - Predicting its future evolution
- **Decomposition** of the global evolution

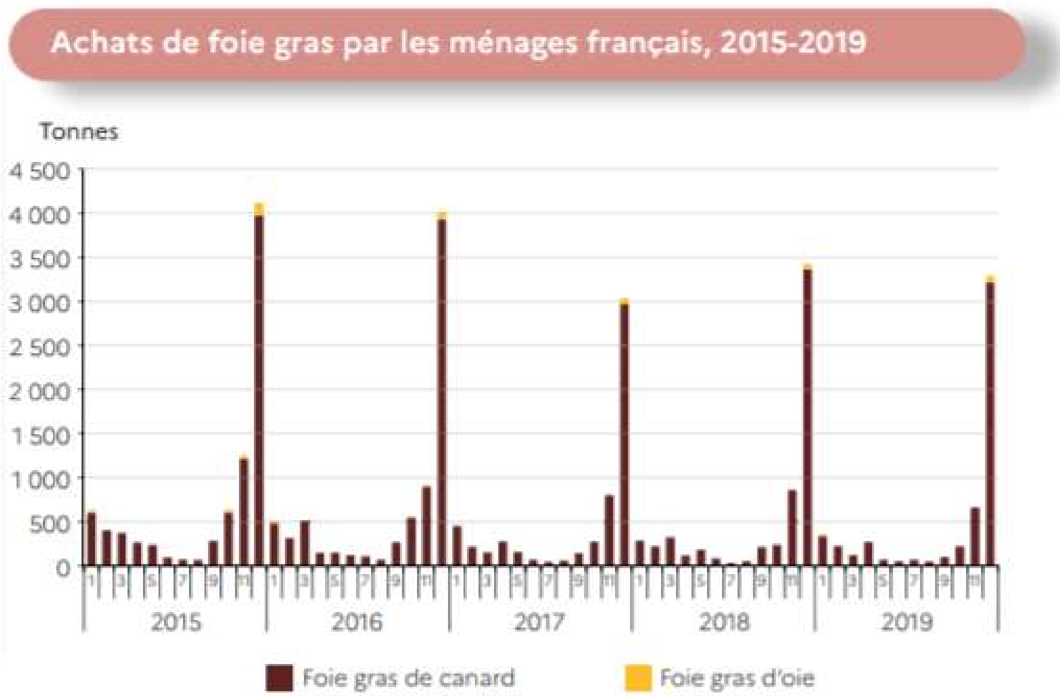
1.Decomposition of a time series

a) Different elements

b) Additive and multiplicative models

1- a) Different elements

- Ex: ”Foie gras” sales in a supermarket

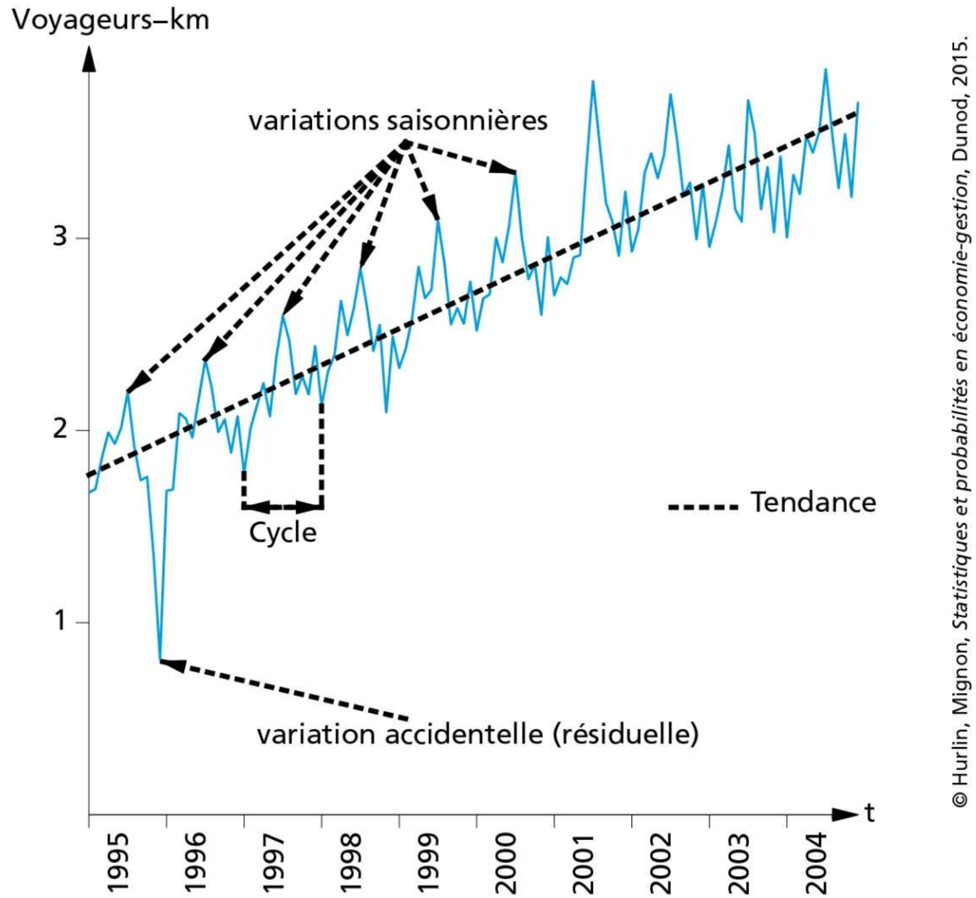


Source : FranceAgriMer d'après Kantar Worldpanel

1- a) Different elements

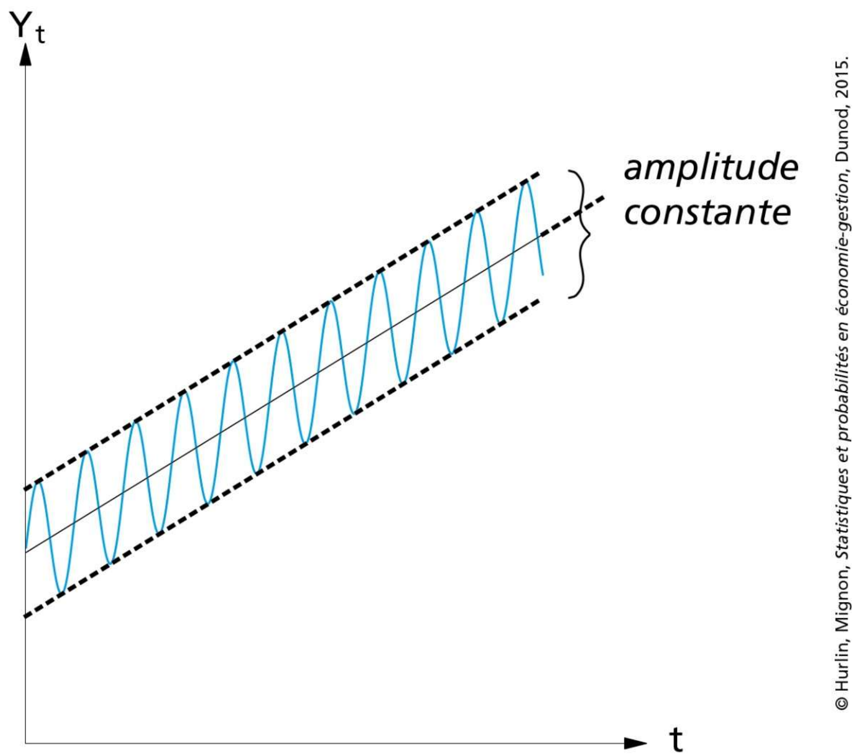
- 3 elements:
  - The trend: g_t
  - The seasonal component: s_t
  - The error (noise): e_t
  - Eventually, we also have a cyclic component
- **Two models**:
  - Additive model
  - Multiplicative model

1- a) Different elements

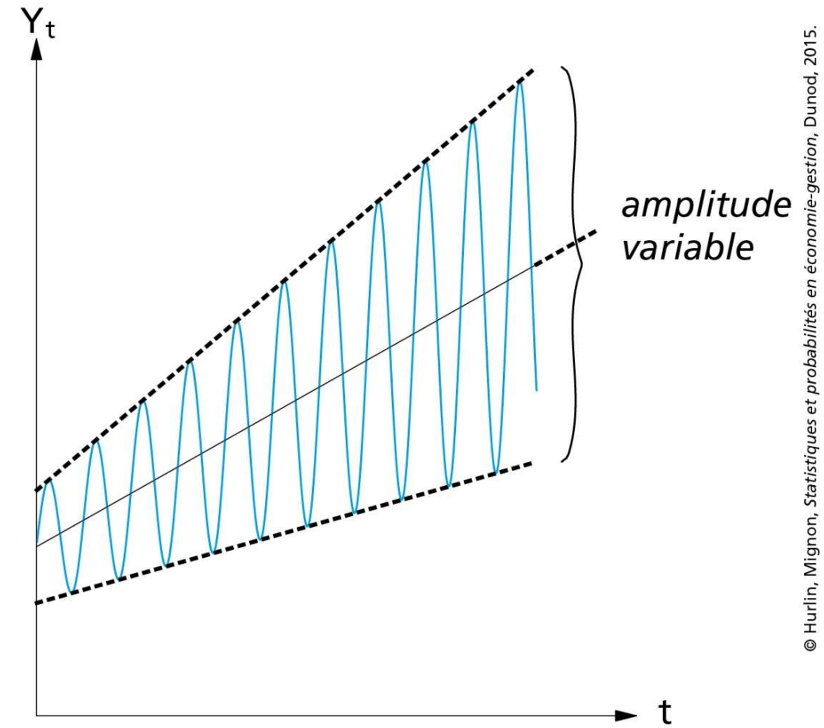


Number of travellers on TGV trains between january 1995 and december 2005, monthly data

1- b) Additive and multiplicative models



Additive model: Constant amplitude of variations



Multiplicative model: variable amplitude of variations

b) Additive and multiplicative models

- Additive model:

$$x_t = g_t + s_t + e_t$$

Hypothesis: the elements are independent from each other.

Corresponds to an arithmetic sequence:

$$x_t = x_{t-1} + r$$

- Multiplicative model:

$$x_t = g_t \times s_t \times e_t$$

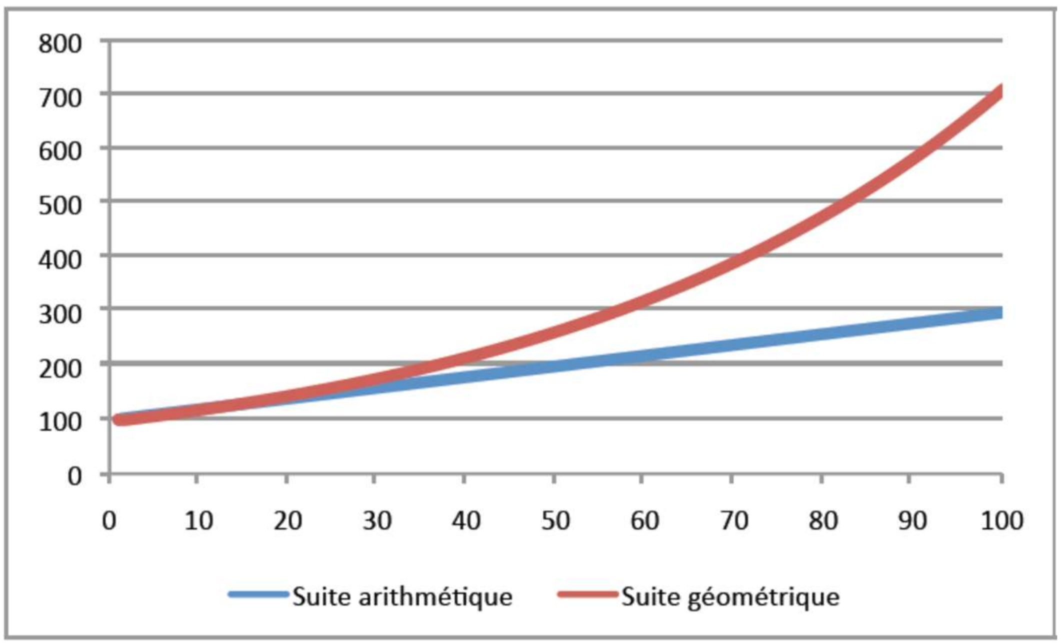
Hypothesis: the elements depend from each other.

Corresponds to a geometric sequence:

$$x_t = x_{t-1} \times q$$

1- b) Additive and multiplicative models

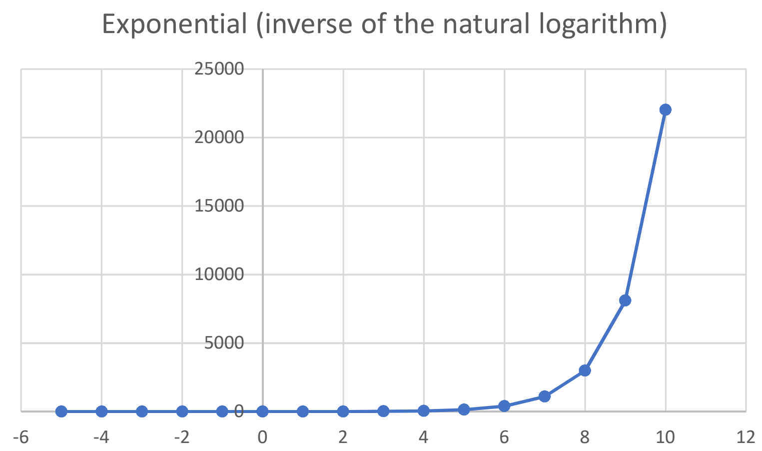
- Arithmetic sequence: here, $y_t = y_{t-1} + 2$
- Geometric sequence: here, $y_t = y_{t-1} \times 1{,}02$

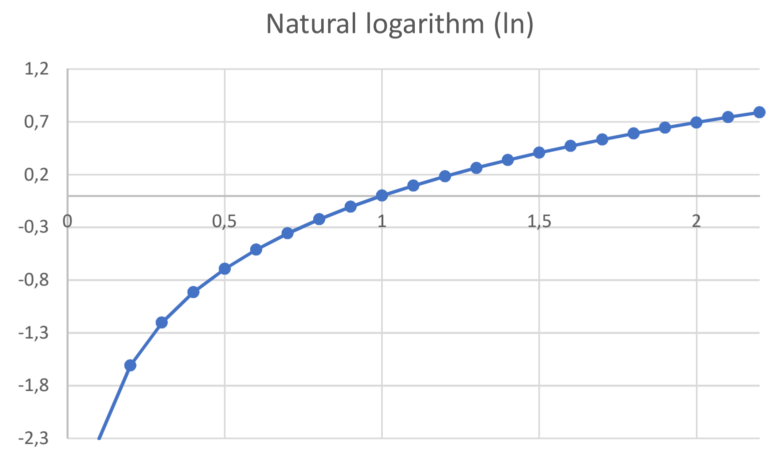


Additive and multiplicative models

- The multiplicative model can be converted into an additive model by applying a natural logarithmic transformation:

$$x_t = g_t \times s_t \times e_t \text{ becomes } \ln x_t = \ln g_t + \ln s_t + \ln e_t$$





Additive and multiplicative models

- How to choose between the additive and the multiplicative model:
- **Graphically**: we look whether the variable has constant amplitude of variations (additive) or an increasing/decreasing one (multiplicative)
- **Algebrically**: Buys-Ballot model. We look at the differences between means (per year for instance) and the standard deviations (each year also), which represent the deviations with the mean. Then we compute the coefficients of variations (standard deviation / means).
  - If the coefficients of variation are stable over time, the additive model is preferred, other multiplicative model.
- We will mainly study the additive model.

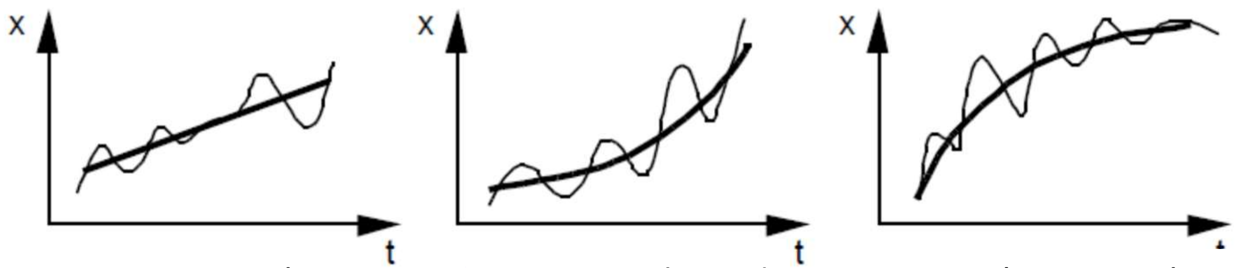
2- Determination of the trend

a) Ordinary Least Squared

b) Moving Averages

2- a) Ordinary Least Squared (OLS)

- **Parametric Trend Fitting:** Based on a visual inspection of the time series, we choose a curve form (e.g. linear, quadratic, exponential) and then fit this curve to the data.



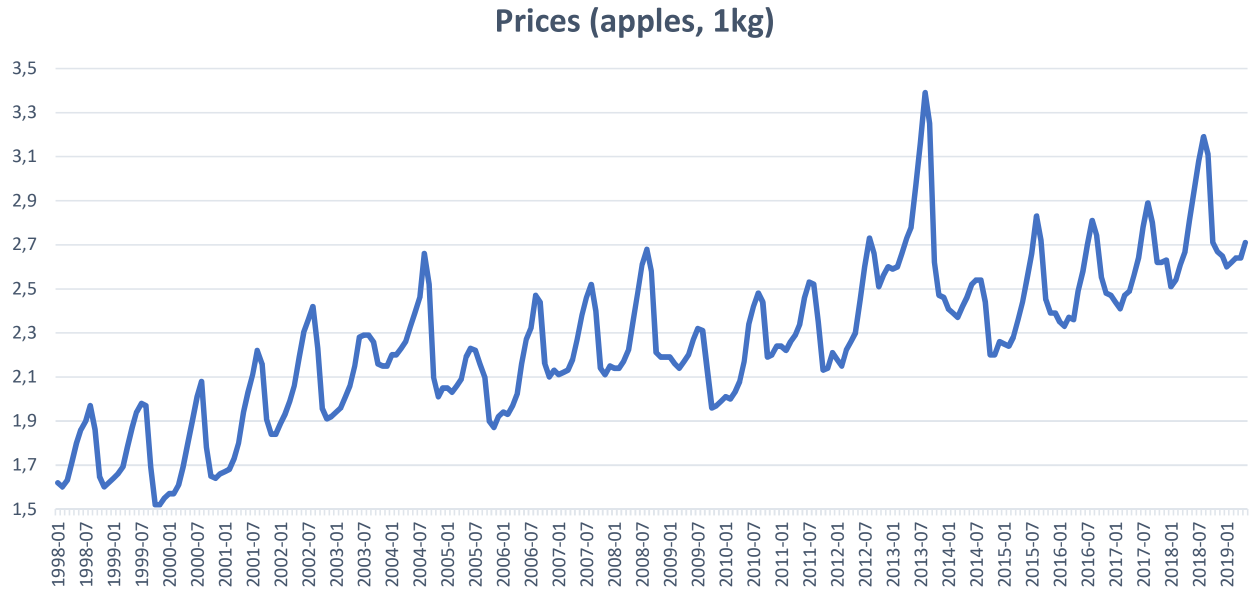
Linear trend $g_t = a \cdot t + b$

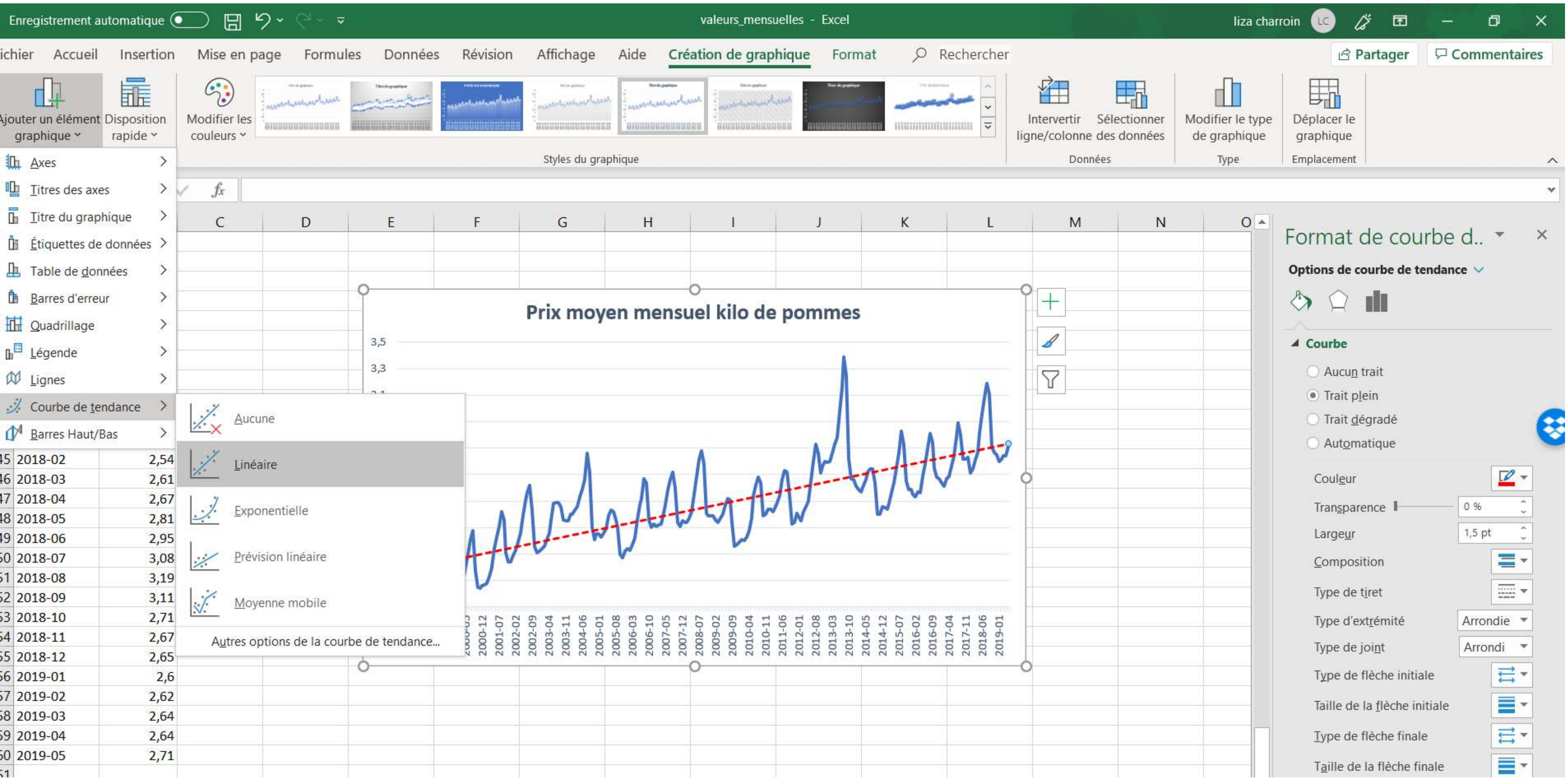
Exponential trend $g_t = b \cdot a^t$

Logarithmic trend $g_t = \ln(a \cdot t + b)$

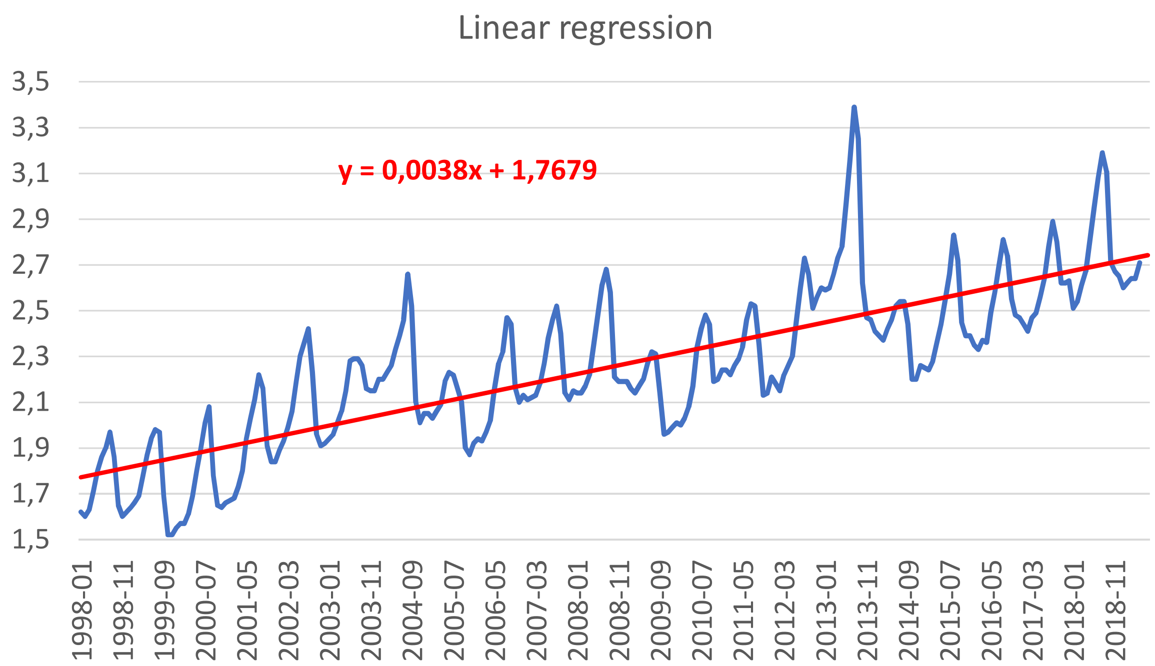
2- a) Ordinary Least Squared (OLS)

- https://www.insee.fr/fr/statistiques/serie/000641367#Telechargement




Enregistrement automatique
valeurs_mensuelles - Excel
liza charroin
ichier
Accueil
Insertion
Mise en page
Formules
Données
Révision
Affichage
Aide
Création de graphique
Format
Rechercher
Partager
Commentaires
Ajouter un élément graphique
Disposition rapide
Modifier les couleurs
Styles du graphique
Intervertir ligne/colonne
Sélectionner des données
Données
Modifier le type de graphique
Type
Déplacer le graphique
Emplacement
Axes
Titres des axes
Titre du graphique
Étiquettes de données
Table de données
Barres d'erreur
Quadrillage
Légende
Lignes
Courbe de tendance
Barres Haut/Bas
Aucune
Linéaire
Exponentielle
Prévision linéaire
Moyenne mobile
Autres options de la courbe de tendance...
Prix moyen mensuel kilo de pommes
3,5
3,3
2018-02 2,54
2018-03 2,61
2018-04 2,67
2018-05 2,81
2018-06 2,95
2018-07 3,08
2018-08 3,19
2018-09 3,11
2018-10 2,71
2018-11 2,67
2018-12 2,65
2019-01 2,6
2019-02 2,62
2019-03 2,64
2019-04 2,64
2019-05 2,71
Format de courbe d..
Options de courbe de tendance
Courbe
Aucun trait
Trait plein
Trait dégradé
Automatique
Couleur
Transparence
0 %
Largeur
1,5 pt
Composition
Type de tiret
Type d'extrémité
Arrondie
Type de joint
Arrondi
Type de flèche initiale
Taille de la flèche initiale
Type de flèche finale
Taille de la flèche finale

2- a) Ordinary Least Squared (OLS)



Interpretation?

2- a) Ordinary Least Squared (OLS)

The linear trend g_t is given by the equation of the linear regression

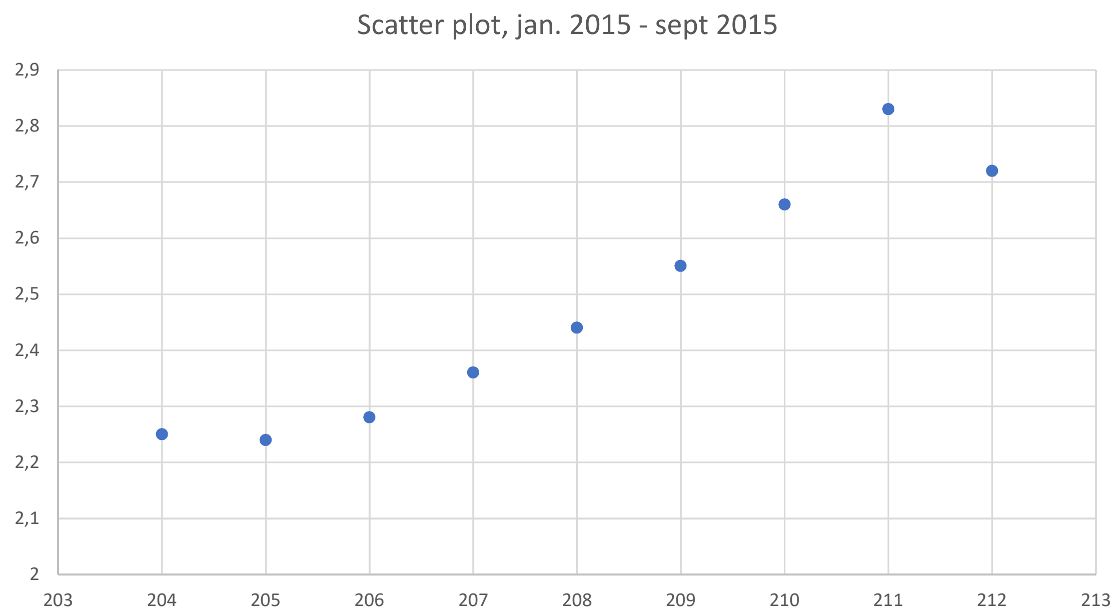
SOMME	=0.0038*E4+1.7679

	A	B	C	D	E	F	G
1						Average monthly prices in France -	
2							
3	Date	Year	Month	Month#	t	Price	Linear trend (g_t)
4	1998-01	1998	january	1	0	1.62	=0.0038*E4+1.7679
5	1998-02	1998	february	2	1	1.6	1.7717
6	1998-03	1998	march	3	2	1.63	1.7755
7	1998-04	1998	april	4	3	1.71	1.7793
8	1998-05	1998	may	5	4	1.8	1.7831

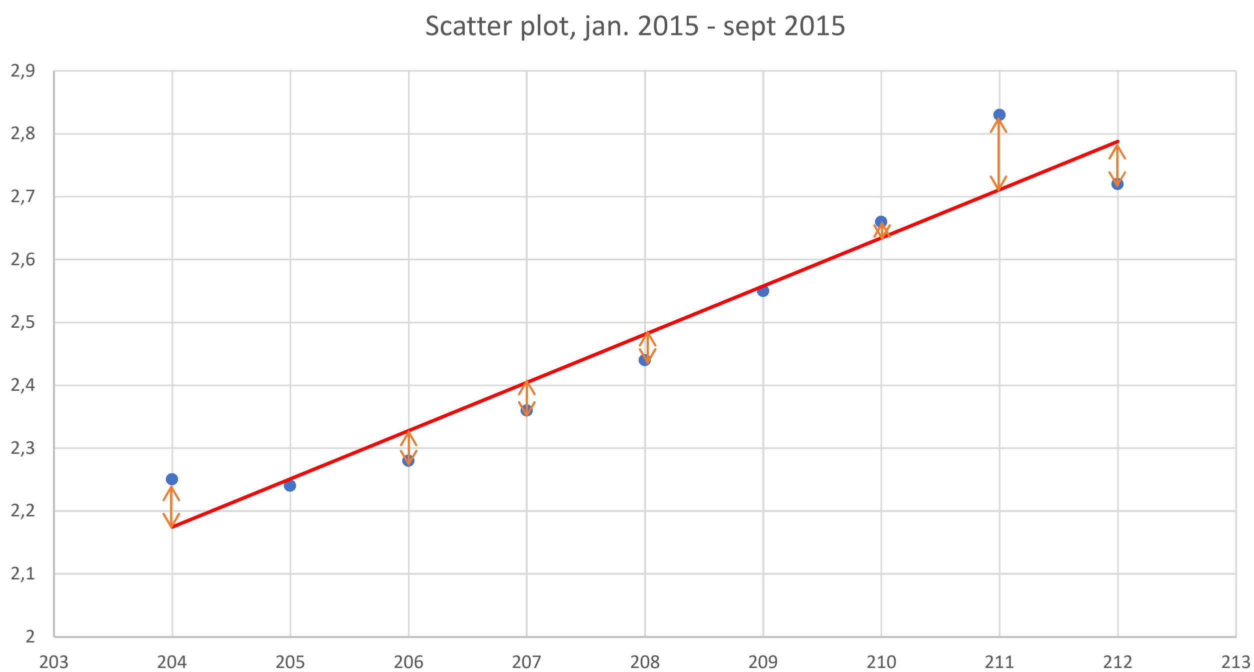
2- a) Ordinary Least Squared (OLS)

- How do we determine **the OLS function**?
  - Here, we focus on linear regressions: $y = ax + b$ (here x represents time, so we usually write it $y = at + b$)
  - a represents the **slope of the curve** (if >0, positive relationship, if <0 negative)
  - b represents the **intercept**
- The goal is to draw the line that minimize the gaps between each point and this line.

2- a) Ordinary Least Squared (OLS)



2- a) Ordinary Least Squared (OLS)



We want to draw the trend (line) while **minimizing** the deviations between our data and the line. We want to **minimize the size of the arrows.**

2- a) Ordinary Least Squared (OLS)

- a is the slope (coefficient) and b is the intercept.
- We choose a and b such that:

$$\min_{\mathrm{a,b}} \sum_{t=1}^{T} \{y_t - (a \times t + b)\}^2$$

We want to minimize the distance between each point and the trend line $y_t = a \times t + b$.

We will go further in Chapter 5.

2- b) Moving averages

- Other method: **moving averages.** Objective: obtain the trend by smoothing. Non-parametric method.
- **No assumption** about the underlying model (linear, exponential, etc.)
- Principle of the calculation: we define a series of averages that are
  - Calculated over a certain number of consecutive observations, called the order.
  - Shifted by one time unit at each new calculation (“moving”)

Moving averages: order 3

x_t		$\overline{x_t}^{III}$
x_1	…	
x_2		$\overline{x_2}^{III} = \frac{x_1 + x_2 + x_3}{3}$
x_3		$\overline{x_3}^{III} = \frac{x_2 + x_3 + x_4}{3}$
x_4		$\overline{x_4}^{III} = \frac{x_3 + x_4 + x_5}{3}$
x_5		
…		
…		
x_{T-2}		
x_{T-1}		$\overline{x_{T-1}}^{III} = \frac{x_{T-2} + x_{T-1} + x_T}{3}$
x_T		

General formula: $\overline{x_t}^{III} = \frac{x_{t-1}+x_t+x_{t+1}}{3}$

Moving averages

- Moving averages of order 5

$\overline{\boldsymbol{x_3}}^V = \frac{x_1+x_2+x_3+x_4+x_5}{5}$ estimation of the trend at $t = 3$

$\overline{\boldsymbol{x}_4}^V = \frac{x_2+x_3+x_4+x_5+x_6}{5}$ estimation of the trend at $t = 4$

etc.

Genral formula: $\boxed{\overline{\boldsymbol{x}_t}^V = \frac{x_{t-2}+x_{\mathrm{t}-1}+x_{\mathrm{t}}+x_{\mathrm{t}+1}+x_{t+2}}{5}}$

Moving averages with an **odd order** are easy to calculate.

Moving averages of even order (e.g., order 4):

- **Problem**: which date (t) to associate with an average of 4 values?
- **Solutions**: we average two consecutive averages of order 4:

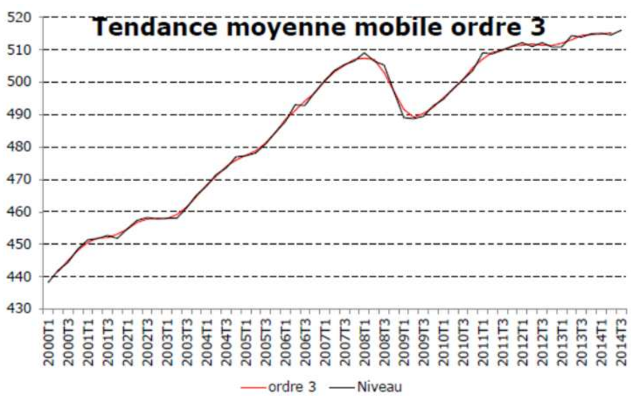
$$\frac{1}{2}\left[\frac{x_1 + x_2 + x_3 + x_4}{4} + \frac{x_2 + x_3 + x_4 + x_5}{4}\right]$$

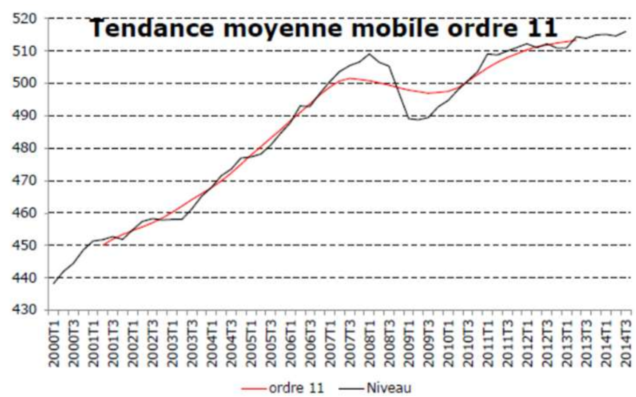
$$= \frac{\frac{1}{2}x_1 + x_2 + x_3 + x_4 + \frac{1}{2}x_5}{4}$$

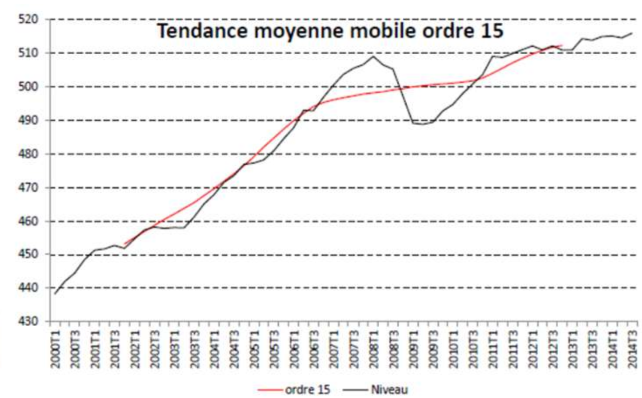
General formula: $\boxed{\overline{x_t}^{IV} = \frac{\frac{1}{2}x_{t-2} + x_{t-1} + x_t + x_{t+1} + \frac{1}{2}x_{t+2}}{4}}$

Moving averages

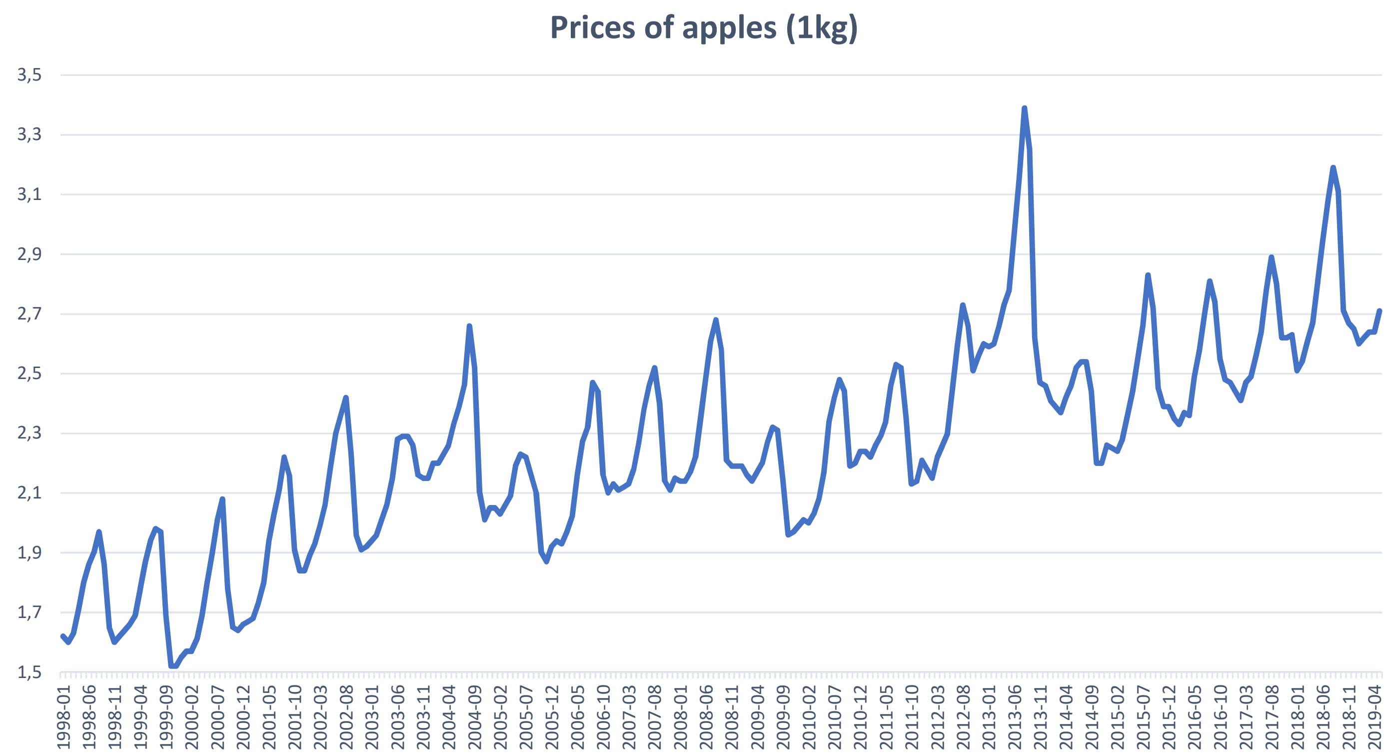
- Which order should we choose?
- The order must be adapted to the periodicityof seasonal variations.
- However, the bigger the order,
  - The stronger the smoothing
  - the more values are “lost”







Example



Prices of apples (1kg)
3,5
3,3
3,1
2,9
2,7
2,5
2,3
2,1
1,9
1,7
1,5
1998-01
1998-06
1998-11
1999-04
1999-09
2000-02
2000-07
2000-12
2001-05
2001-10
2002-03
2002-08
2003-01
2003-06
2003-11
2004-04
2004-09
2005-02
2005-07
2005-12
2006-05
2006-10
2007-03
2007-08
2008-01
2008-06
2008-11
2009-04
2009-09
2010-02
2010-07
2010-12
2011-05
2011-10
2012-03
2012-08
2013-01
2013-06
2013-11
2014-04
2014-09
2015-02
2015-07
2015-12
2016-05
2016-10
2017-03
2017-08
2018-01
2018-06
2018-11
2019-04

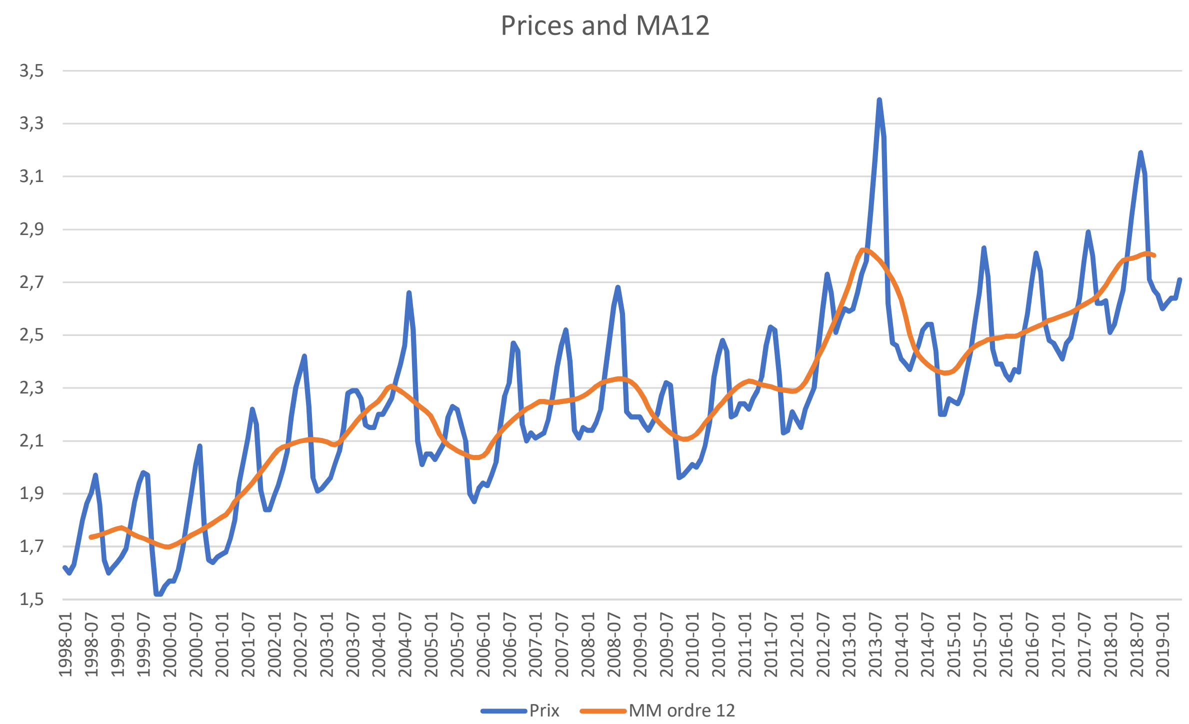
2- b) Moving averages

- Moving averages of order 12

	Average monthly prices in France - Apples (1kg)			
t	Price	Linear trend (g_t)	Moving averages (order 12)	Deviations g_t)
0	1.62	1.7679		
1	1.6	1.7717		
2	1.63	1.7755		
3	1.71	1.7793		
4	1.8	1.7831		
5	1.86	1.7869		
6	1.9	1.7907	=(1/2*F4+SOMME(F5:F15)+1/2*F16)/12	
7	1.97	1.7945	SOMME(**nombre1**; [nombre2]; ...)	
8	1.86	1.7983	1.744166667	
9	1.65	1.8021	1.749583333	
10	1.6	1.8059	1.755416667	
11	1.62	1.8097	1.761666667	
12	1.64	1.8135	1.768333333	

Excel formula: SOMME in French and SUM in English

2- b) Moving averages

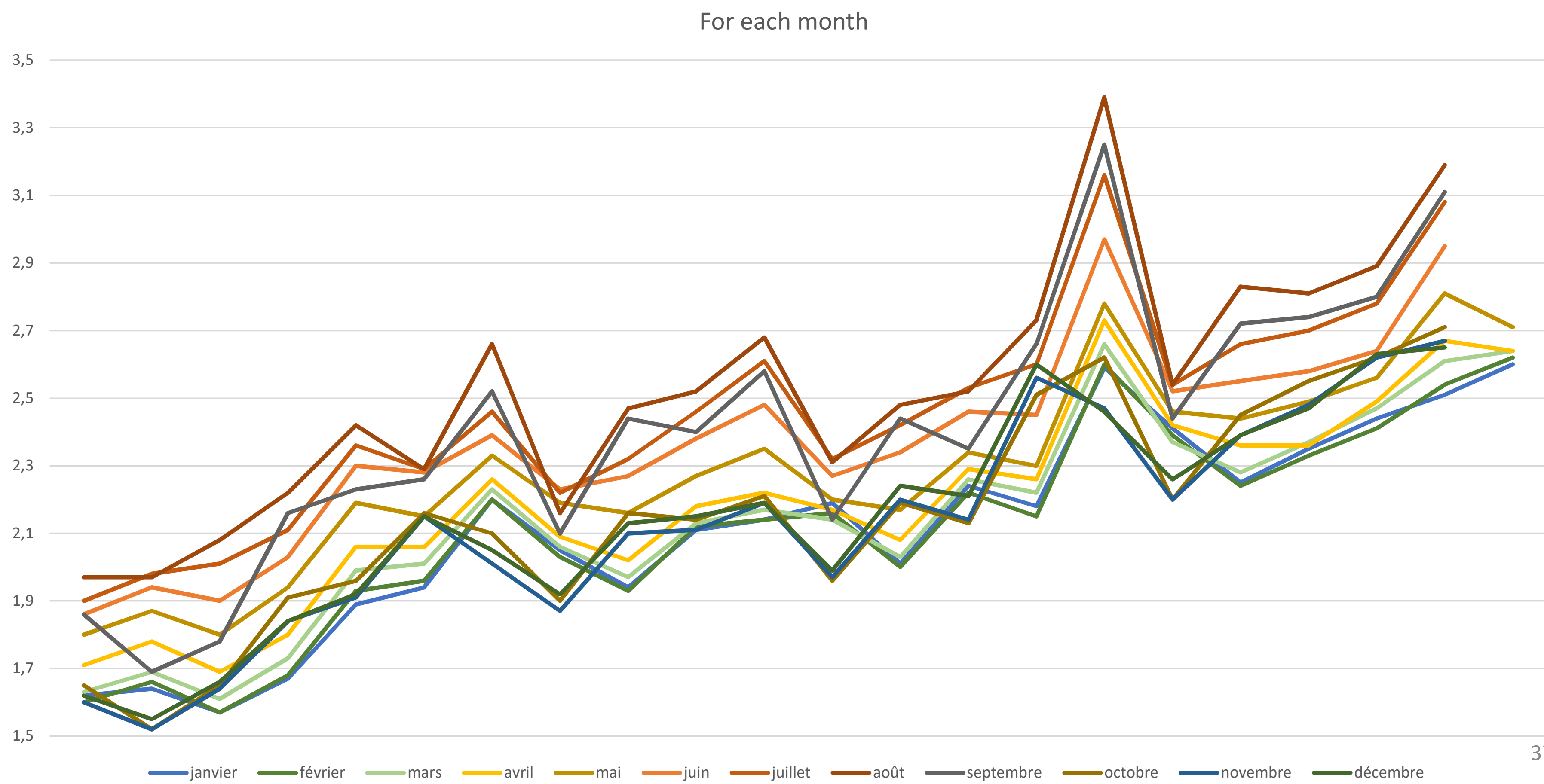


Comparison between OLS and moving averages

- **OLS**:
  - Simple to calculate on Excel
  - Can be computed for the whole period
  - Easy to interpret
  - Enables the trend to be extrapolated
  - But: not flexible AND which trend to choose (linear, exponential, etc.)?
- **Moving averages**:
  - Flexible (can adapt to big variations across time)
  - But: difficult to choose the order, cannot be computed for the whole period and does not enable the trend to be extrapolated (so, no forecast possible)

3. Identification of the seasonal component

Seasons



For each month
3,5
3,3
3,1
2,9
2,7
2,5
2,3
2,1
1,9
1,7
1,5
janvier
février
mars
avril
mai
juin
juillet
août
septembre
octobre
novembre
décembre

3. Identification of the seasonal component

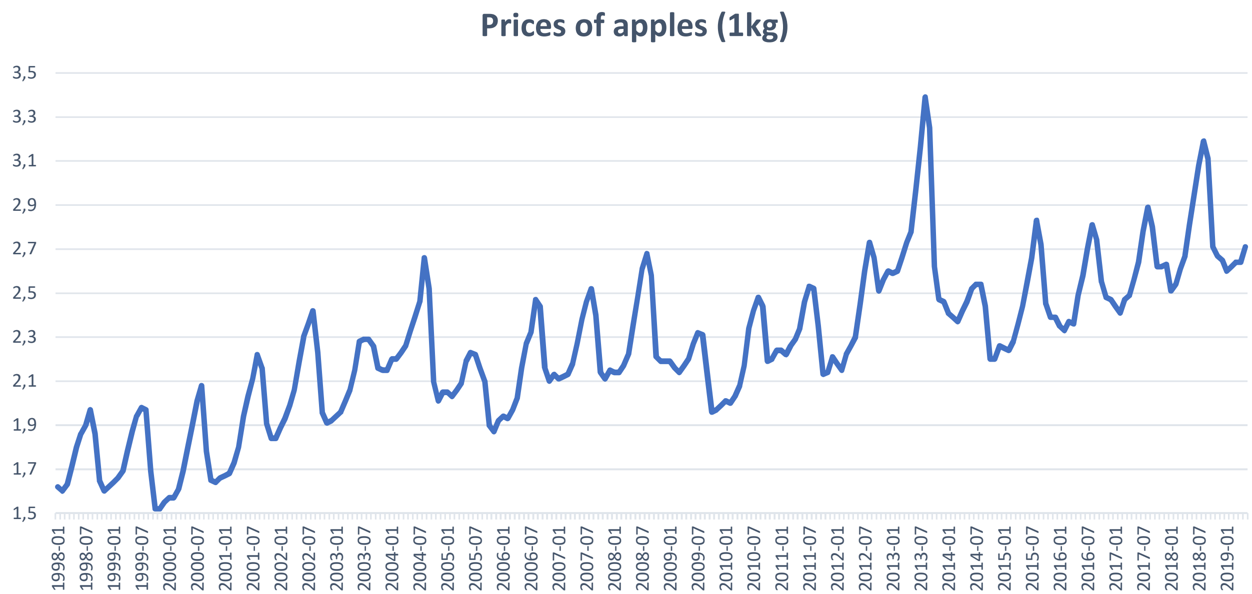
- For each date t, we have
  - One observation x_t
  - One estimation of the linear trend g_t
- The **seasonal component** is determined from these two elements. We compute the deviations from trend. For the:
  - Additive model: $s_t = x_t - g_t$
  - Multiplicative model: $s_t = \frac{x_t}{g_t}$

3. Identification of the seasonal component

- Then, we **compute the means** of $\boldsymbol{s_t}$ for each season of the year (each month, each quarter… depending on the data)
- This corresponds to the seasonal coefficient of each month (or quarter…):
  - 12 seasonal coefficients (SC) with monthly data
  - 4 SC with quarterly data

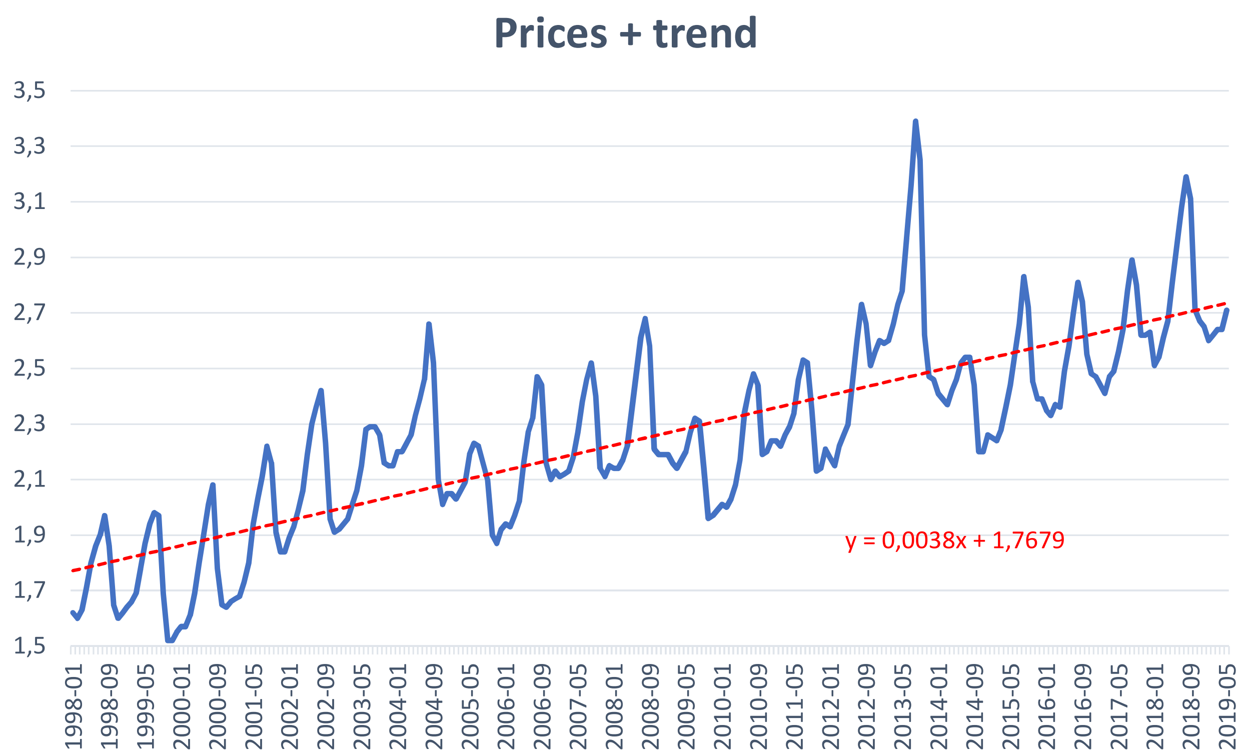
Example

- Which model do we choose?



Raw data

Example

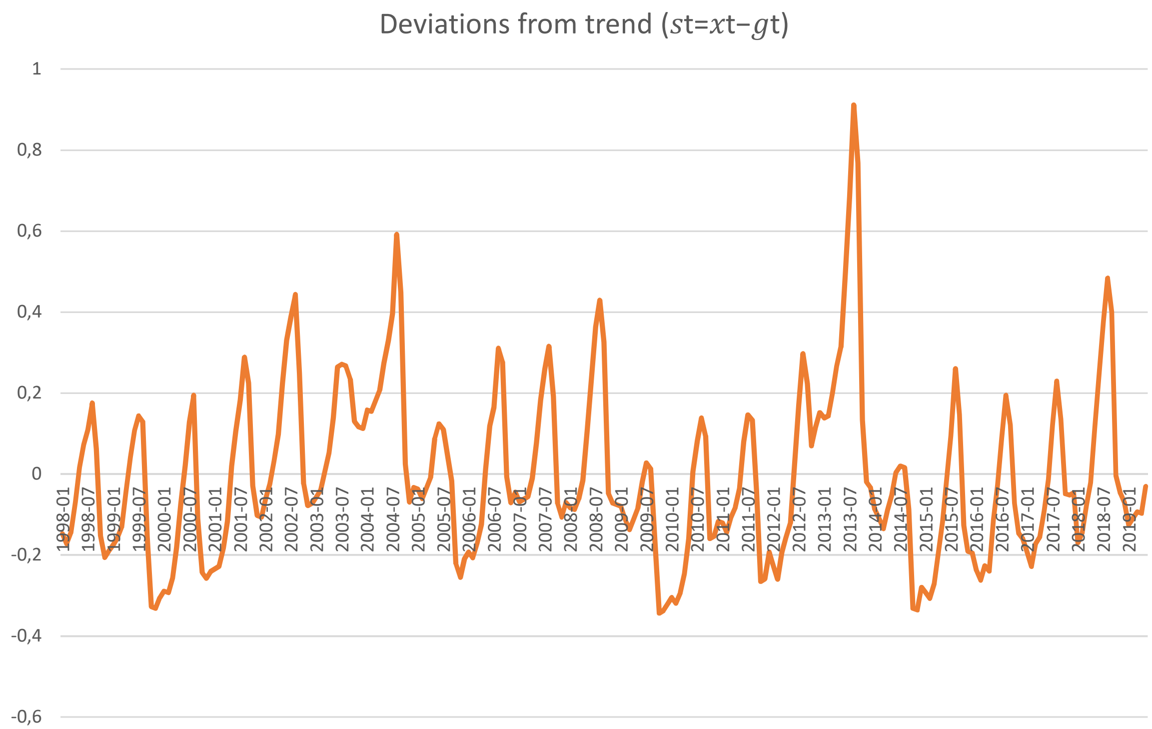


Linear trend

Computation of the deviations from trend

t	Price	Linear trend (g_t)	Moving averages (order 12)	Deviations from trend ($s_t = x_t - g_t$)
0	1.62	1.7679		-0.1479
1	1.6	1.7717		-0.1717
2	1.63	1.7755		-0.1455
3	1.71	1.7793		-0.0693
4	1.8	1.7831		0.0169
5	1.86	1.7869		0.0731
6	1.9	1.7907	1.735833333	0.1093
7	1.97	1.7945	1.739166667	0.1755
8	1.86	1.7983	1.744166667	0.0617
9	1.65	1.8021	1.749583333	-0.1521
10	1.6	1.8059	1.755416667	-0.2059

Deviations from trend



3. Identification of the seasonal component

- **Monthly data from** jan. 1998 to may 2019
  - We have **22 observations each month** for january, february… till may
  - **21 observations each month** for june, july, … till december
- **SC in January**: mean of 22 deviations from the trend observed in January (same in february, etc. until may).
- **SC in June:** mean of 21 deviations from the trend (same in July, … december).

3- a) Seasonal coefficients

3- b) Eventual correction of the seasonal coefficients

- Assumption of **neutrality of seasonal variations** over one period
- Mean of the seasonal coefficients (SCs):
  - Zero in the additive model
  - Equal to 1 in the multiplicative model
- If this assumption is not satisfied, it is necessary to **adjust** the seasonal coefficients:
  - Additive model → subtract the mean of the SCs from each seasonal coefficient
  - Multiplicative model → divide each seasonal coefficient by the mean of the SCs

3- b) Eventual correction of the seasonal coefficients

- If we have quarterly data:
  - Means of SC: $M_{SC} = \frac{SC_I + SC_{II} + SC_{III} + SC_{IV}}{4}$
  - Additive model:
    - M_{SC} is significantly different from 0?
    - If yes, adjusted SC: $SC_I^* = SC_I - M_{SC}$, $SC_{II}^* = SC_{II} - M_{SC}$, etc.
  - Multiplicative model:
    - M_{SC} is significantly different from 1?
    - If yes, adjusted SC: $SC_I^* = SC_I / M_{SC}$, $SC_{II}^* = SC_{II} / M_{SC}$, etc.

Example

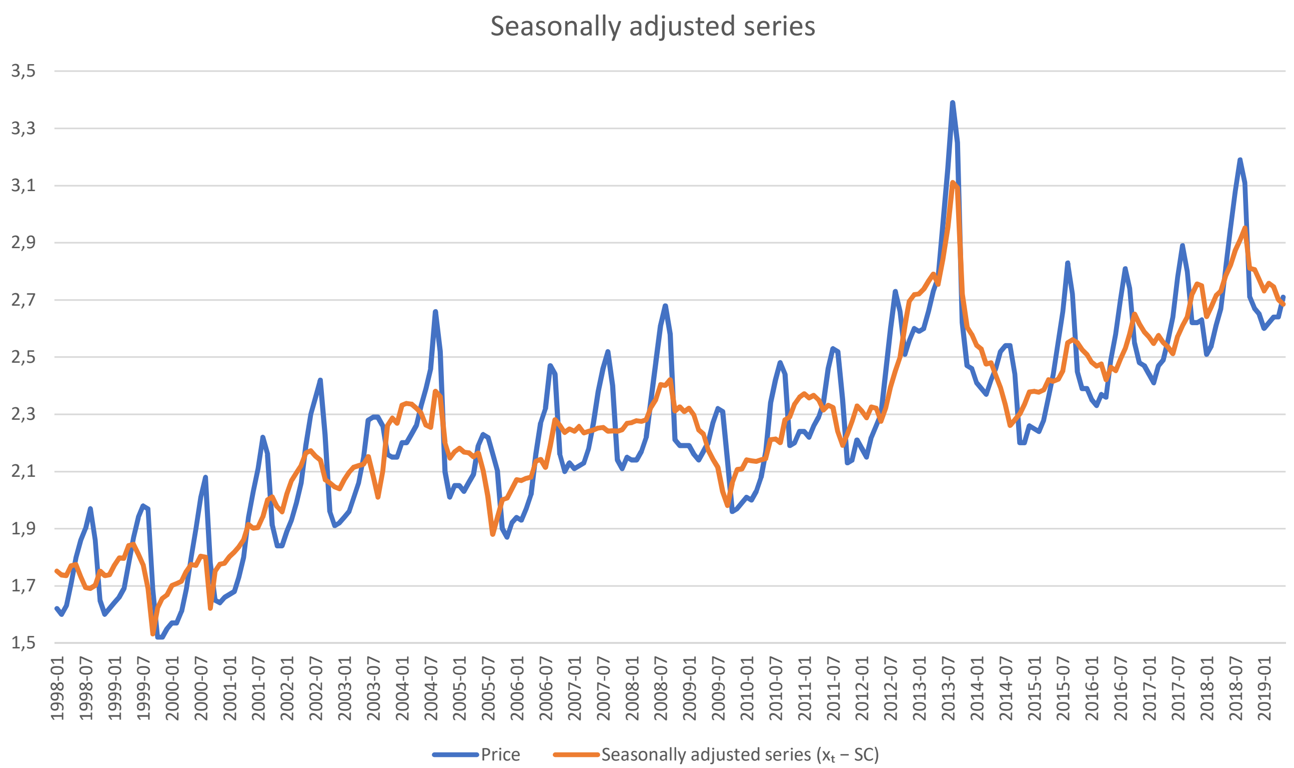
Seasonal coefficients
-0,131245455
-0,137772727
-0,105663636
-0,060827273
0,024918182
0,128052381
0,206157143
0,2795
0,158557143
-0,100957143
-0,136185714
-0,119033333
0,000458297

Mean close enough to 0 here

4- Seasonally adjusted series (SA series)

Seasonal adjustment

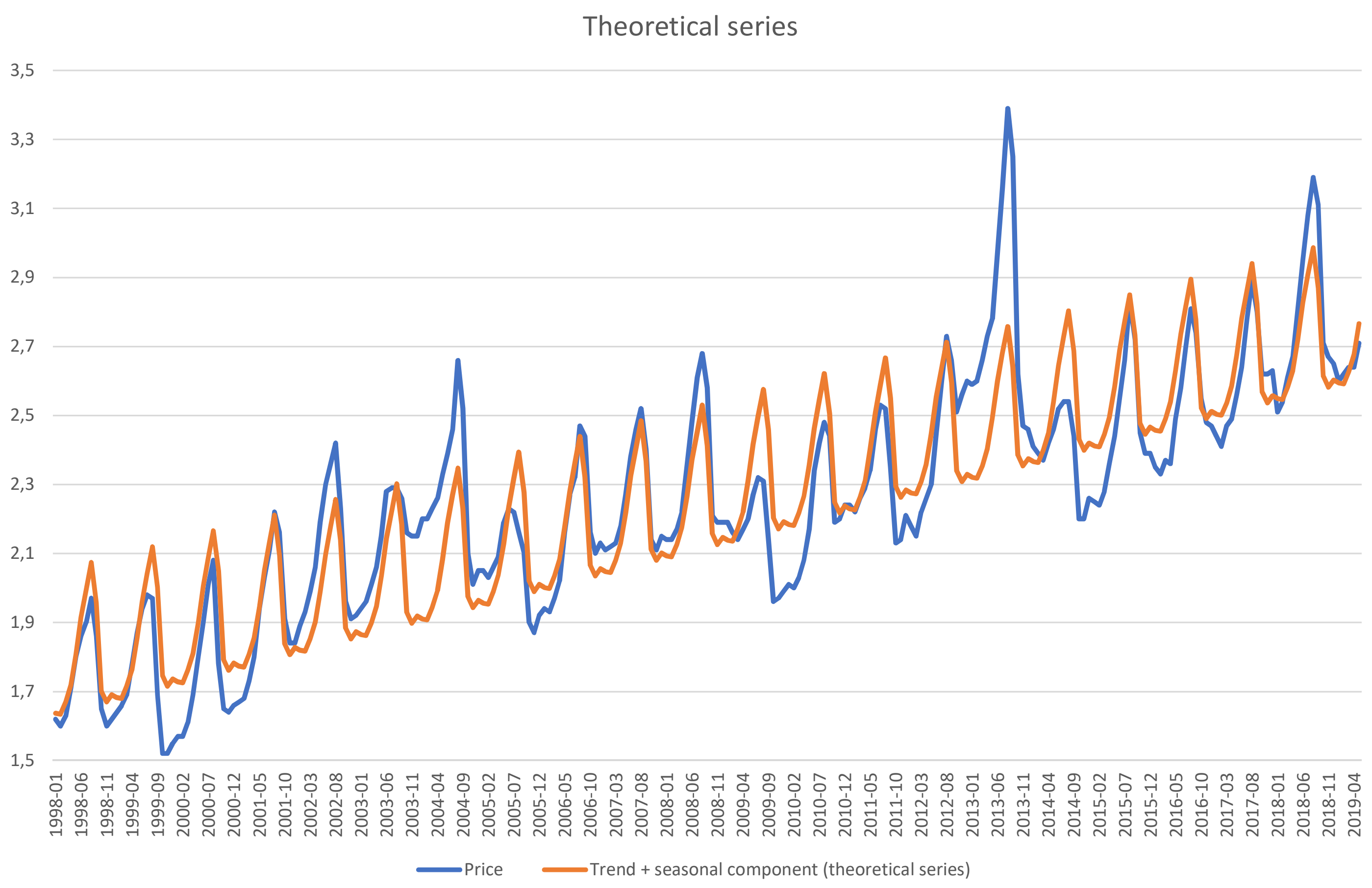
- Why **seasonally adjust**? Eliminate the effect of seasonal fluctuations to better observe underlying trends.
- How to seasonally adjust?
  - **Additive model**: subtract from the value of the variable in each month (or quarter) the seasonal coefficient associated with that month (or quarter).
    - Seasonally adjusted series (SA series) consists of $x_t - SC_{mt}$ where m represents the month and t the date.
  - **Multiplicative model**: divide the value of the variable in each month (or quarter) by the seasonal coefficient associated with that month (or quarter).
    - SA series consists of x_t/SC_{mt}


Seasonally adjusted series
3,5
3,3
3,1
2,9
2,7
2,5
2,3
2,1
1,9
1,7
1,5
1998-01
1998-07
1999-01
1999-07
2000-01
2000-07
2001-01
2001-07
2002-01
2002-07
2003-01
2003-07
2004-01
2004-07
2005-01
2005-07
2006-01
2006-07
2007-01
2007-07
2008-01
2008-07
2009-01
2009-07
2010-01
2010-07
2011-01
2011-07
2012-01
2012-07
2013-01
2013-07
2014-01
2014-07
2015-01
2015-07
2016-01
2016-07
2017-01
2017-07
2018-01
2018-07
2019-01
Price
Seasonally adjusted series (x_t – SC)

5- Forecasting

Forecasting

- We reconstruct a **theoretical series** from the trend and the seasonal coefficients
- With:
  - The **additive** model: $x_t^f = g_t + SC_{mt}$
  - The **multiplicative** model: $x_t^f = g_t \times SC_{mt}$
- To make predictions, we **extend the theoretical series**
- **It's only possible to extend the trend g_t with the OLS method**


Theoretical series
3,5
3,3
3,1
2,9
2,7
2,5
2,3
2,1
1,9
1,7
1,5
Price
Trend + seasonal component (theoretical series)

Forecasting

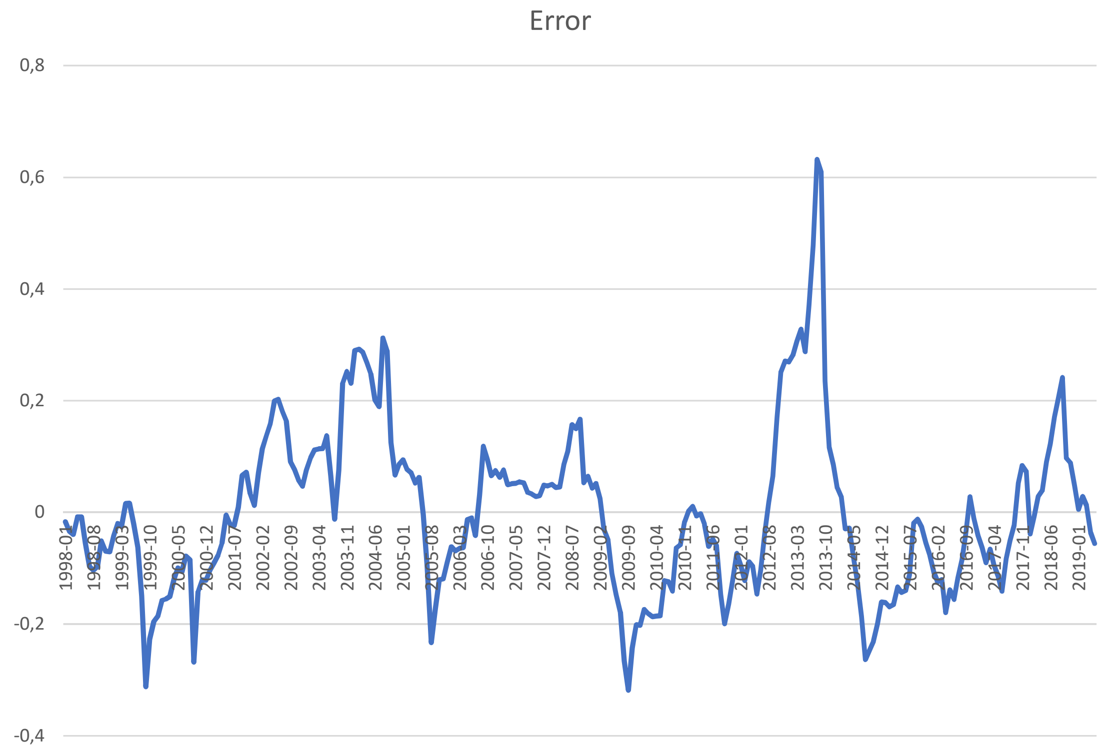
- What should be the price of 1kg of apples in june 2019?
  - Trend (OLS): $\mathrm{g_t} = 0.0038t + 1.7679$
  - June 2019 corresponds to $t = 257$
  - $\mathrm{g}_{257} = 0.0038 \times 257 + 1.7679 = 2.7445$ ➔ linear trend
  - We add the corresponding seasonal coefficient: $SC_{june} = 0.1280$
- Price forecast in june 2019: $x^{f}_{june19} = 2.7445 + 0.1280 = 2.8725.$
- Price observed in june 2019? 2.79. We are quite close. How can we explain the gap between both values?

To improve our forecasts

- To be more precise, one can take into account the cyclical component (for example, every 5 years the apple harvest may be better).
- Choose another functional form to estimate the trend (nonlinear, exponential, etc.).
- Focus on the determinants of the price of apples (climate, etc.).

Error and/or cycle

- Error (and/or cycle) = price observed – theoretical series



Summary to analyse a time series

1. We choose the model (additive or multiplicative) thanks to the variations (constant or variable)
2. We determine the trend (OLS or Moving averages) to see the evolution of the time series on the long term
3. We compute the seasonal coefficients (thanks to the deviations to the trend)
4. We can compute the seasonally adjusted series by removing the SC from the observations.
5. We determine the theoretical series (trend + SC) and can forecast future values (only with OLS trend).

Summary

Additive or mult model depending on the graph or Buys Ballot

Unpredicted: $x_t - \mathit{theoretical\ series}$

Theoretical series

Data (x_t) = Trend (g_t) + or x Seasonal component + or x Error

Two methods:

- OLS: computed on Excel (linear, exponential, etc) = essential to forecast
- Moving averages (order, even or odd): cannot forecast with it

Compute seasonal coefficients:

1. Compute deviations from trend ($x_t - g_t$ or x_t / g_t)
2. Compute the means of deviations for each month (or quarter) ➔ 12 or 4 SC
3. Check if the means of SC = 0 (add) or 1 (mult)
4. If yes, SC-mean of SC = SCc or SC/mean of SC = SCc

Seasonally adjusted series = $x_t - SC$