

# Analyse de séries temporelles

Projet - Version 1.0

Année Universitaire 2024-2025

**Contexte.** Les séries temporelles, autrement dit "l'observation de variables quantitatives indicées par le temps", sont des objets omniprésents dans tous les domaines d'activité : domaine de la santé, domaine environnemental, domaine des transports ... mais aussi et surtout dans le domaine financier avec notamment la cotation des matières premières ou des métaux et l'évolution des indices boursiers.



Dans le cadre de ce projet, on se propose de travailler sur plusieurs cotations de matière première, cotations observées depuis le 1er janvier 2010 :

- Cotation du prix du café (Futures café US C - Devise USD)
- Cotation du prix du cacao (Futures cacao US - Devise USD)
- Cotation du prix du jus d'orange (Futures jus d'orange - Devise USD)
- Cotation du prix du sucre (Futures sucre Londres - Devise USD)
- Cotation du prix du baril de pétrole (Futures pétrole Brent - Devise USD)

Les données, contenues dans des fichiers PDF, ont été collectées depuis le site internet [Investing.com](https://www.investing.com) (cliquer sur l'illustration pour accéder au site)



Les fichiers disponibles sont les suivants :

- Data - Futures café US C
- Data - Futures cacao US
- Data - Futures jus d'orange
- Data - Futures sucre Londres
- Data - Futures pétrole Brent

Futures sucre Londres - Données Historiques					
Date	Dernier	Ouv.	Plus Haut	Plus Bas	
10/10/2024	564,30	562,30	566,50	559,10	
09/10/2024	563,90	575,00	575,00	554,60	
08/10/2024	575,10	567,50	582,40	565,00	
07/10/2024	570,20	572,50	585,50	565,40	
04/10/2024	576,90	585,00	590,30	569,00	
03/10/2024	584,00	573,90	586,30	566,10	
02/10/2024	574,00	585,80	586,50	570,00	

**Objectif.** Il va consister en plusieurs missions

- **Mission 1** - Elle a pour objectif d'importer les données de chacun des fichiers PDF, puis de les regrouper dans un même objet de classe `tibble`. On pourra faire usage en particulier de la bibliothèque de fonction `tabulapdf` disponible sur le CRAN.

Pour chacun des fichiers, il faudra définir 5 variables :

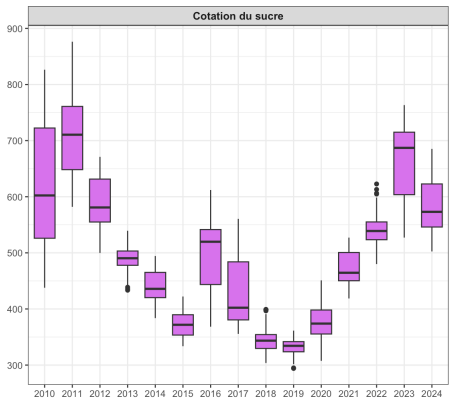
- Une variable nommée `Date` représentant la jour de cotation, avec le type adéquat
- Une variable nommée `Closed_Cotation` représentant la valeur de cotation à la fermeture des marchés
- Une variable nommée `Opened_Cotation` représentant la valeur de cotation à l'ouverture des marchés
- Une variable nommée `Highest_Cotation` représentant la valeur maximale de cotation de la journée
- Une variable nommée `Lowest_Cotation` représentant la valeur minimale de cotation de la journée

# A tibble: 15 x 5					
	Date	Closed_Cotation	Opened_Cotation	Highest_Cotation	Lowest_Cotation
	<date>	<dbl>	<dbl>	<dbl>	<dbl>
1	2024-10-10	459.03	476.3	476.3	458.05
2	2024-10-09	477.02	478.4	481.98	473.85
3	2024-10-08	477.55	489.25	491.18	473.18
4	2024-10-07	483.58	472.68	488.28	472.68
5	2024-10-04	470.75	470.02	470.75	464.25
6	2024-10-03	469.4	468.62	471.45	464.88
7	2024-10-02	468.2	468.53	469.55	459.08
8	2024-10-01	475.48	464.45	479.65	463.65
9	2024-09-30	461.7	473.12	479.63	454.48
10	2024-09-27	470.2	472	479.95	467
11	2024-09-26	475.3	489.25	490.95	471.5
12	2024-09-25	489.15	491.15	491.5	482.1
13	2024-09-24	489.05	481.35	491.5	474.55
14	2024-09-23	474.5	491.85	491.85	470.85
15	2024-09-20	490.85	478.75	492	476.7

**Remarque.** La lecture des données depuis les fichiers PDF peut amener à un comportement assez singulier selon le fichier étudié. Aussi, il peut s'avérer délicat d'automatiser cette tâche.

- **Mission 2** - Elle a plusieurs objectifs statistiques :
  - + Réaliser, uniquement pour la variable `Closed_Cotation`, des diagrammes proposant des boxplots en parallèle annuel représentant la valeur journalière de la cotation de chacune des matières premières. Commenter les graphiques individuellement et globalement.

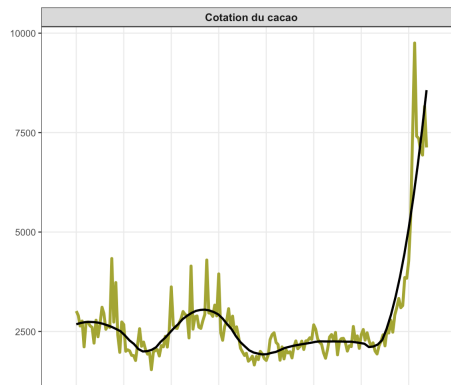
**Conseil.** Utiliser la fonction `facet_wrap()` pour afficher les diagrammes sur une seule représentation graphique.



Exemple d'un des diagrammes

- + Réaliser, uniquement pour la variable `Closed_Cotation`, un diagramme permettant de visualiser l'évolution moyenne mensuelle de la cotation journalière de chacune des matières premières au cours de la période d'étude, avec le cas échéant une courbe de régression lissée. Commenter les graphiques individuellement et globalement.

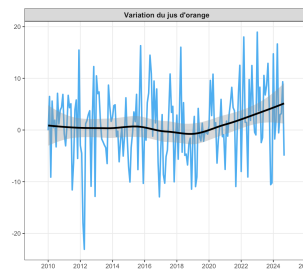
**Conseil.** Utiliser la fonction `facet_wrap()` pour afficher les diagrammes sur une seule représentation graphique.



Exemple d'un des diagrammes

- + Réaliser, uniquement pour la variable `Closed_Cotation`, le calcul du taux d'évolution de la moyenne mensuelle de la cotation journalière de chacune des matières premières au cours de la période d'étude, puis représenter cette évolution sous la forme d'un diagramme. Commenter les graphiques individuellement et globalement.

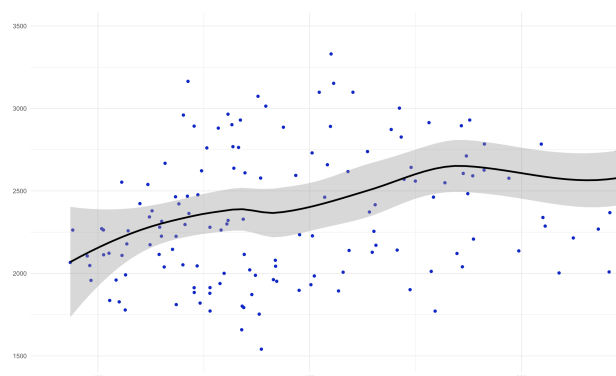
**Conseil.** Utiliser la fonction `facet_wrap()` pour afficher les diagrammes sur une seule représentation graphique.



Exemple d'un des diagrammes

- + Réaliser une étude à propos de l'éventuelle association pouvant exister entre les quantités suivantes, basées sur la variable `Closed_Cotation` :
  - La valeur moyenne mensuelle de la cotation journalière de la matière première "café"
  - La valeur moyenne mensuelle de la cotation journalière de la matière première "cacao"

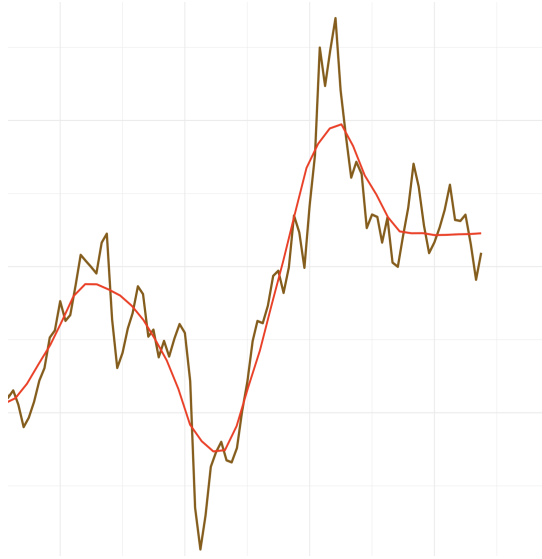
En particulier, on pourra étudier la présence de données atypiques, mettre en évidence la mise en évidence d'une association au travers d'une courbe de régression lissée, évaluer l'intensité linéaire de l'éventuelle association, et le cas échéant proposer une modélisation de l'association au travers d'un modèle de régression linéaire simple en présentant le coefficient de détermination, en réalisant un test d'hypothèses pour évaluer le caractère significatif de la pente de la droite de régression, en précisant l'équation des moindres carrés et en interprétant les indicateurs pertinents.



Extrait du graphique

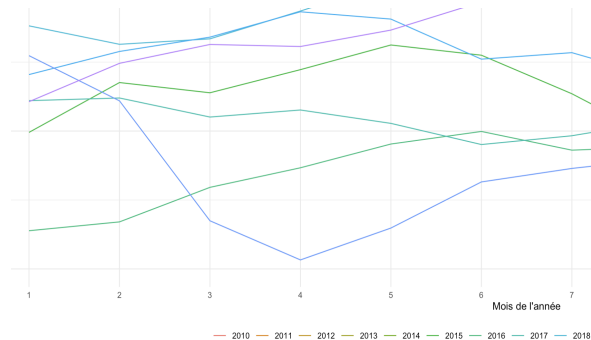
+ Réaliser une étude de la valeur moyenne mensuelle de la cotation journalière du **Brent** à la clôture depuis le 1er janvier 2010 en répondant à chacune des questions suivantes :

1. Visualiser le comportement de la valeur de cotation en y ajoutant une courbe de régression lissée. Commenter succinctement.



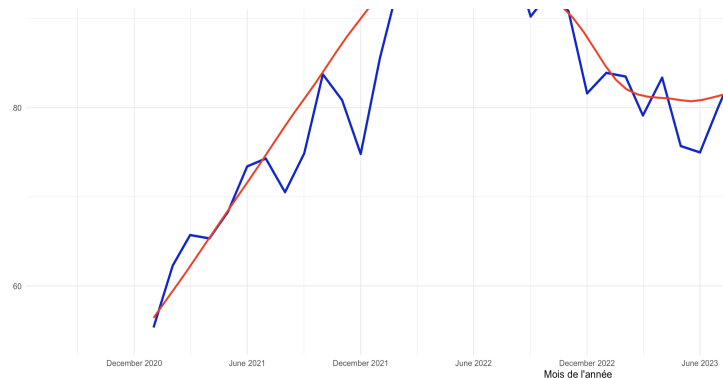
Extrait du graphique

2. Réaliser un diagramme des saisonnalités. Observe-t-on la présence d'une composante saisonnière ? Justifier votre réponse.



Extrait du graphique

3. Visualiser le comportement de la moyenne mensuelle de cotation journalière du **Brent** depuis le 1er janvier 2020 en y ajoutant une courbe de régression lissée. Proposer un modèle de régression linéaire par morceaux pour modéliser la tendance de cette série en précisant les instants de rupture de tendance.



Extrait du graphique

4. Réaliser la modélisation à l'aide de la fonction `lm()`, puis proposer une écriture explicite du modèle de régression selon la période de temps, puis vérifier la cohérence de la modélisation au travers d'un graphique.

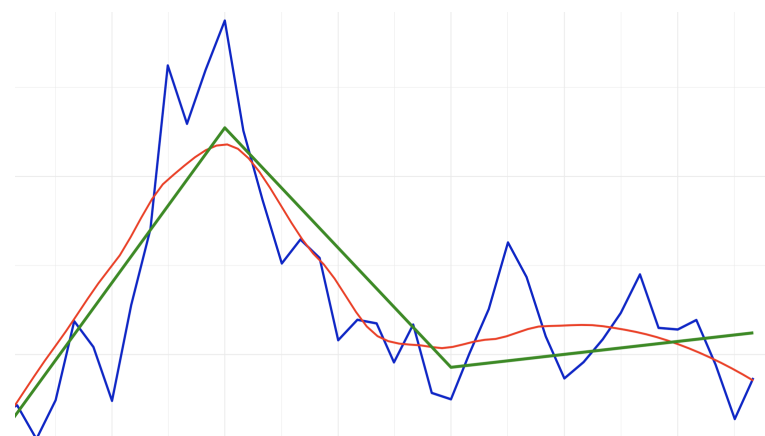
```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.723e+03  1.535e+02 -11.224 3.19e-14 ***
Year_Month    9.554e-02  8.114e-03  11.774 6.88e-15 ***
Var_t1       -1.692e-01  1.621e-02 -10.442 3.04e-13 ***
Var_t2         8.162e-02  1.678e-02   4.866 1.65e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.39 on 42 degrees of freedom
Multiple R-squared:  0.7699,    Adjusted R-squared:  0.7535
F-statistic: 46.85 on 3 and 42 DF,  p-value: 1.845e-13

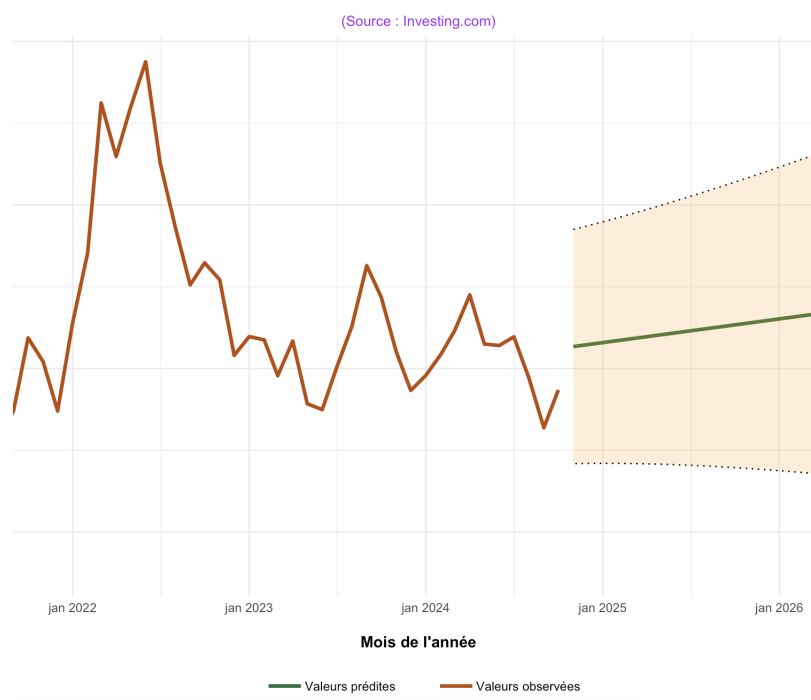
```

Extrait de la sortie textuelle



Extrait de la sortie graphique

5. Proposer une prévision de la valeur moyenne mensuelle de la cotation du **Brent** pour les 26 prochains mois en ajoutant une bande de confiance de 95% (utiliser la fonction `predict()` en y ajoutant l'argument `interval = "prediction"`).



Extrait de la sortie textuelle