

Time Series

Haeji Yun

2023-06-05

Dans cette étude, nous allons analyser l'évolution du prix en dollar d'un kilo de café et prédire son prix future par différentes modélisations de séries temporelles.

Le jeu de données contient les prix journaliers de café: le prix le plus élevé du jour, le prix le moins élevé du jour, les prix au début et la fin du jour, ainsi que le volume journalier de transactions sur la période du janvier 2005 au décembre 2020.

Par la suite, nous allons utiliser le prix le plus élevé pour l'étude.

##	Date	Open	High	Low	Close	Volume	Currency
## 1242	2005-01-03	103.25	105.10	102.5	103.10	9787	USD
## 1243	2005-01-04	101.50	102.50	99.8	101.45	12032	USD
## 1244	2005-01-05	100.85	104.75	98.0	98.50	14840	USD
## 1245	2005-01-06	97.70	99.00	95.6	96.80	16983	USD
## 1246	2005-01-07	97.00	98.65	96.1	96.25	8470	USD
## 1247	2005-01-10	95.25	97.00	95.1	95.60	8380	USD

0. Préparation de données

Nettoyage de données

Pour effectuer notre étude, les données doivent être sous forme d'une série temporelle où les valeurs étudiées forment une suite de données indexées par le temps.

Dans le jeu de données d'origine, le temps est exprimé en jour. Étant donné que seules les dates de la semaine (weekend exclu) sont disponibles, nous avons opté de travailler sur la fréquence du mois. Les valeurs mensuelles sont obtenues par la moyenne des prix journalières sur le mois correspondant.

```
## # A tibble: 1 x 192
##   '2005-01' '2005-02' '2005-03' '2005-04' '2005-05' '2005-06' '2005-07'
##   <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>     <dbl>
## 1    103.      116.      129.      123.      121.      114.      103.
## # i 185 more variables: '2005-08' <dbl>, '2005-09' <dbl>, '2005-10' <dbl>,
## #   '2005-11' <dbl>, '2005-12' <dbl>, '2006-01' <dbl>, '2006-02' <dbl>,
## #   '2006-03' <dbl>, '2006-04' <dbl>, '2006-05' <dbl>, '2006-06' <dbl>,
## #   '2006-07' <dbl>, '2006-08' <dbl>, '2006-09' <dbl>, '2006-10' <dbl>,
## #   '2006-11' <dbl>, '2006-12' <dbl>, '2007-01' <dbl>, '2007-02' <dbl>,
## #   '2007-03' <dbl>, '2007-04' <dbl>, '2007-05' <dbl>, '2007-06' <dbl>,
## #   '2007-07' <dbl>, '2007-08' <dbl>, '2007-09' <dbl>, '2007-10' <dbl>, ...
```

Conversion en Time Series

Par la conversion, nous obtenons un jeu de données en series temporelles :

```
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 2005 102.99750 115.54474 129.27045 122.90476 121.36190 114.27273 102.70250
## 2006 120.93000 113.67632 107.78261 110.02368 103.85227  97.66364  98.74500
##           Aug      Sep      Oct      Nov      Dec
## 2005 101.46304  92.82857 101.12381 101.97000  98.47143
## 2006 105.99348 104.03750 105.71818 116.71250 125.28000
```

Les prix varient entre 92,83\$ à 285,13\$, plutôt concentrés sur les valeurs élevées, et il ne semble pas y avoir de valeurs extrêmes.

```
##           V1
## Min.      : 92.83
## 1st Qu.:114.22
## Median :127.16
## Mean   :139.95
## 3rd Qu.:150.09
## Max.    :285.13
```

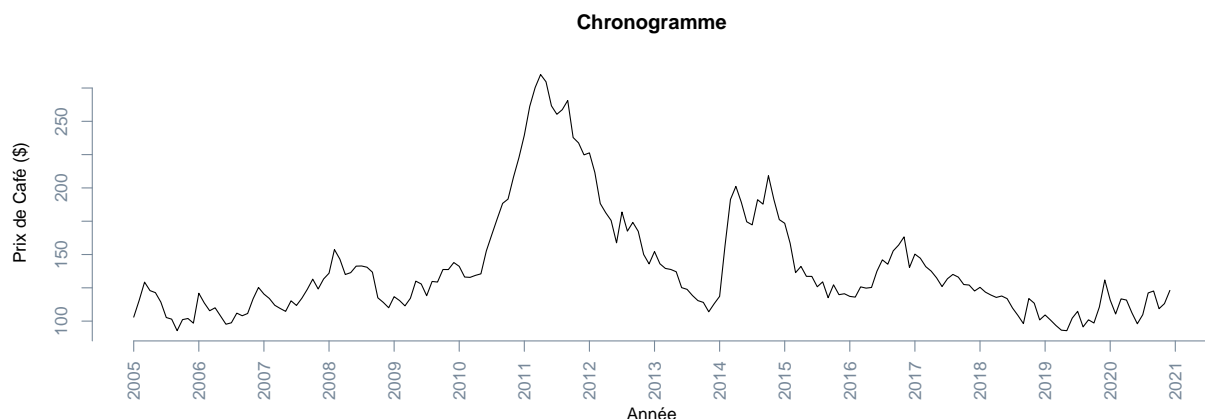
1. Étude préliminaire

Chronogramme

Nous pouvons visualiser le chronogramme pour avoir une d'ensemble de série.

- Nous avons une **tendance** non linéaire qui varie au cours du temps à la hausse et à la baisse.
- Il semble y avoir une **saisonnalité** avec une patterne qui se répète avec un effet plutôt multiplicatif mais sa variance reste floue à cause de la tendance qui varie beaucoup. La saisonnalité ne semble pas être régulière.
- Le prix a connu des **variations exceptionnelles** entre 2010 et 2012 ainsi qu'en 2014 où l'hausse et la baisse de prix sont particulièrement grandes par rapport aux autres années.

Dans la globalité, le prix de café varie en gardant un certain niveau entre 100\$ - 150\$ et il arrive des événements atypiques de temps en temps mais retrouve le niveau initial par la suite.

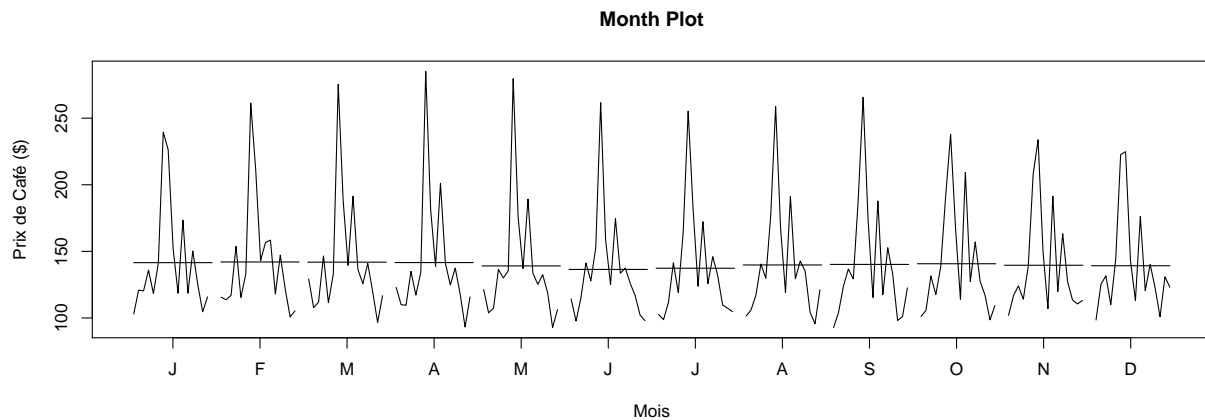


Monthplot

À travers un month plot, nous pouvons visualiser simultanément les chronogrammes associés à chaque mois pour détecter la saisonnalité.

Nous voyons que le prix moyen reste presque similaire sur les différents mois. Le prix baisse légèrement entre les mois de mai et de juillet et reste même sur le reste de mois.

Malgré une patterne apparente sur le chronogramme, la saisonnalité semble être très faible, voire presque inexistante.

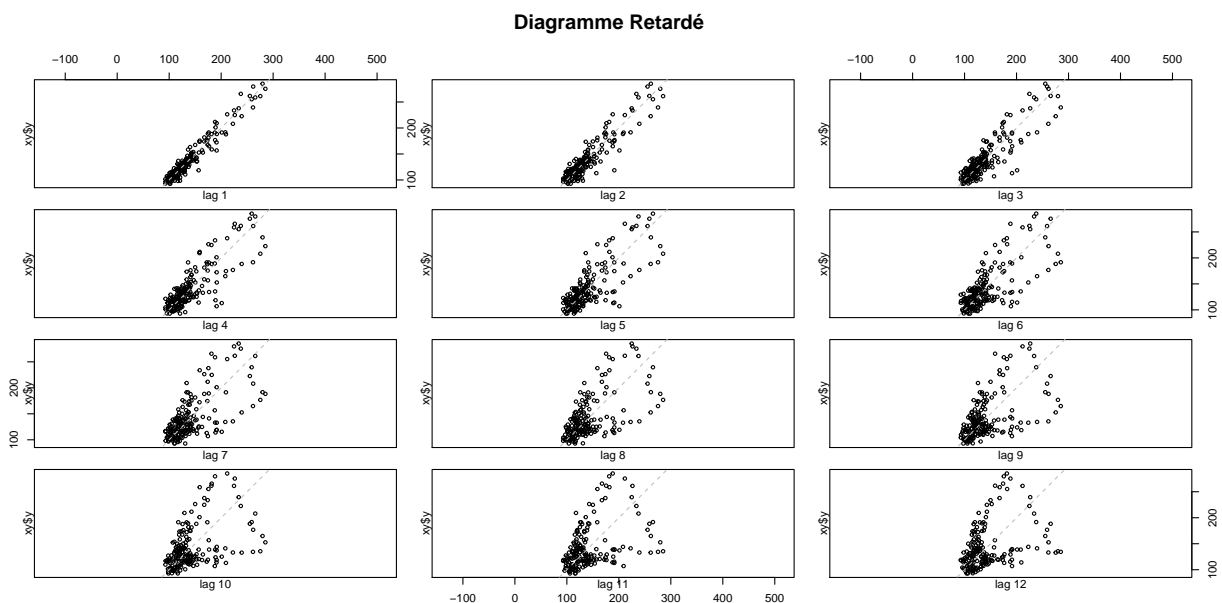


Lagplot

Nous pouvons voir l'autocorrélation de différents ordres avec un diagramme retardé.

Nous observons un lien linéaire de la série avec la série du lag 1. La série observée dépend de sa série décalé d'un mois. La série peut être la mieux expliquée par ce qui s'est passé le mois précédant. Nous remarquons également que le lien diminue avec le temps. C'est le passé proche qui explique le mieux la série.

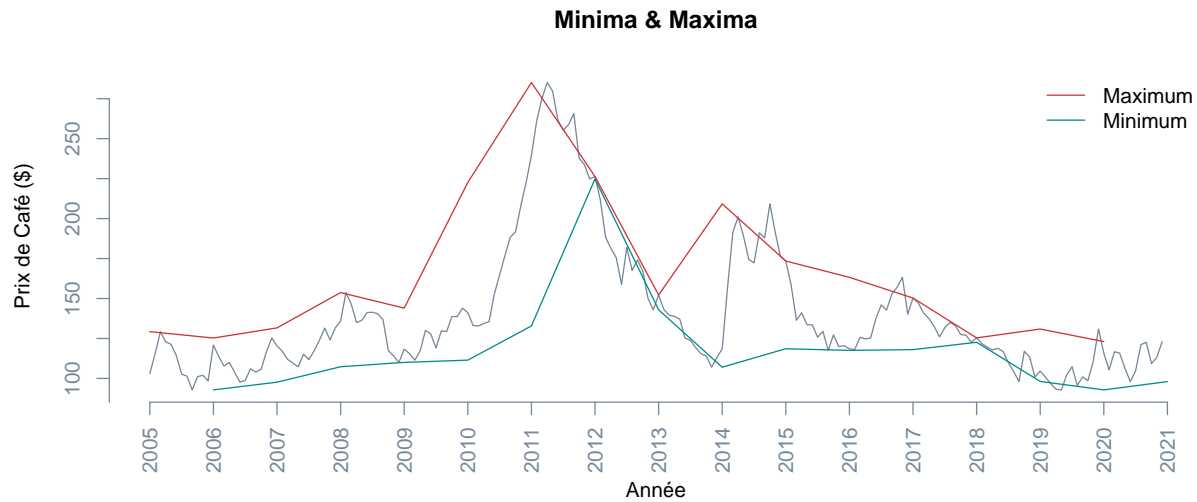
Avec le lien qui diminue de façon progressive, la saisonnalité semble être absente.



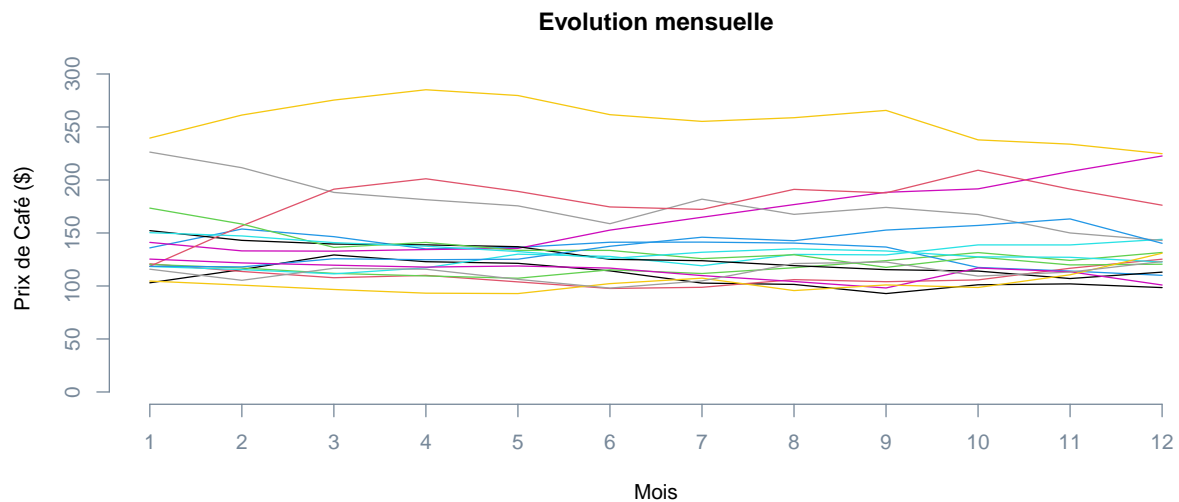
Identification du modèle

Sur le chronogramme, le modèle semble être multiplicatif. Nous pouvons le vérifier avec les méthodes de la bande et du profil.

Avec les courbes non-parallèles qui passe par les minima et les maxima du chronogramme, nous pouvons supposer que le modèle est multiplicatif.



Nous pouvons également superposer les courbes de chaque années sur les 12 mois. Les courbes n'étant pas parallèles entre elles, le modèle est multiplicatif.

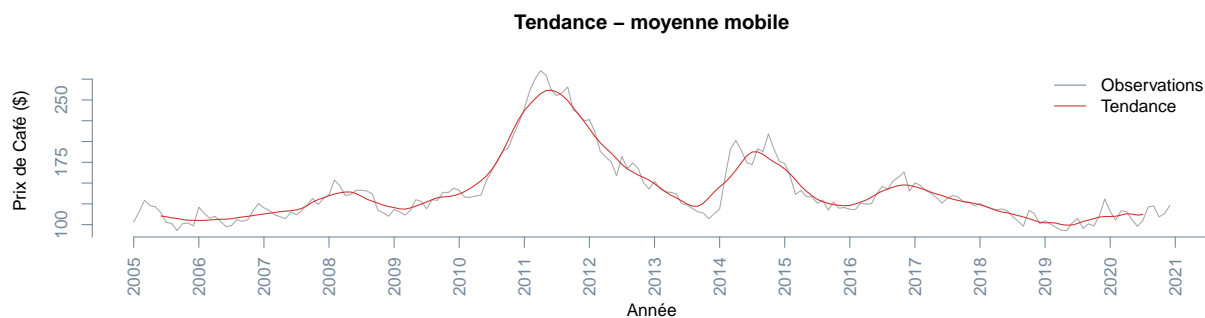


2. Décomposition

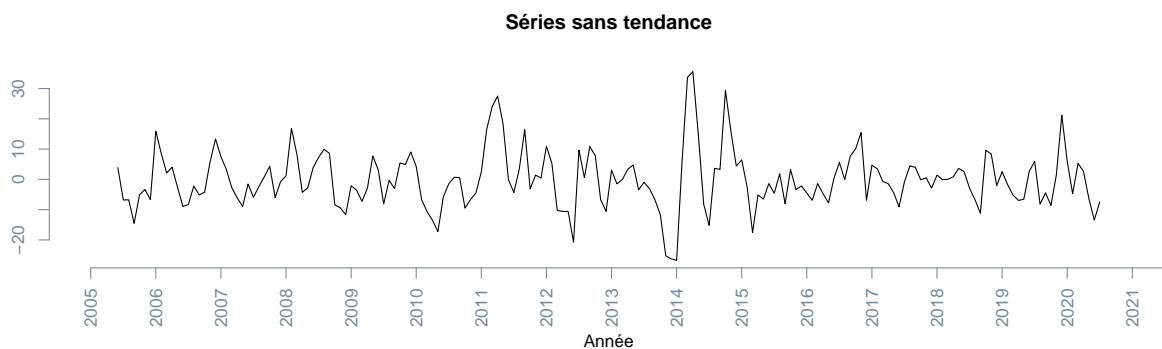
Maintenant que nous avons analysé la vue d'ensemble de la serie, nous pouvons la décomposer en composantes principales : la tendance, la saisonnalité, et le bruit.

Tendance

La tendance représente l'aspect général de la série. La tendance est estimée sans hypothèse à priori par la méthode de moyenne mobile. Nous observons la variation de la tendance au cours du temps, plutôt à la hausse jusqu'en 2011 puis à la baisse par la suite avec des variations entre temps.

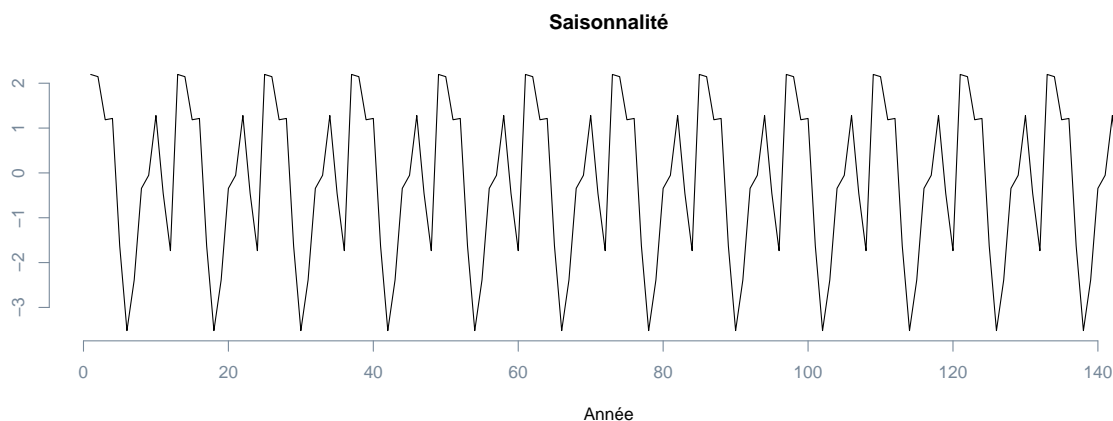


Lorsque nous enlevons la tendance de la série, nous retrouvons avec la saisonnalité et le bruit seuls.



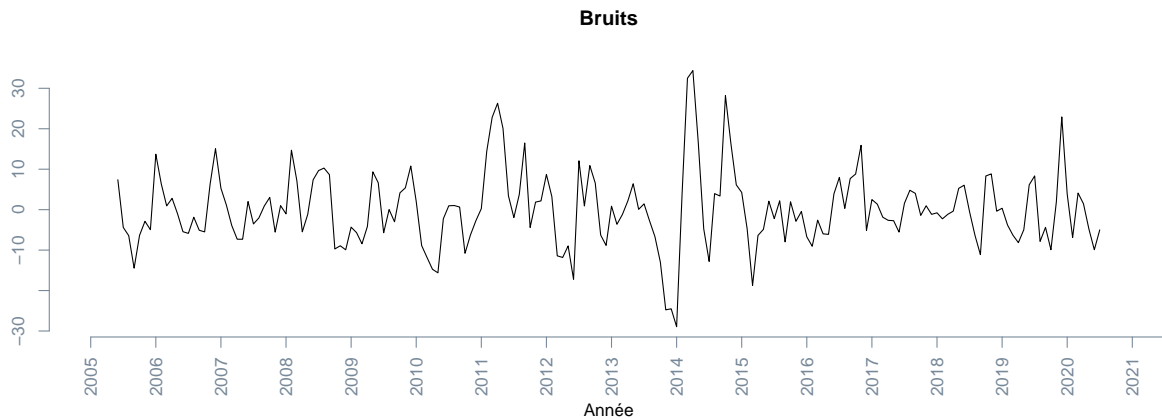
Saisonnalité

À partir de la série sans tendance, nous pouvons modéliser la saisonnalité en faisant la moyenne de chaque mois de toutes les années. Ce schéma obtenu est ensuite répété dans le temps pour obtenir la saisonnalité.



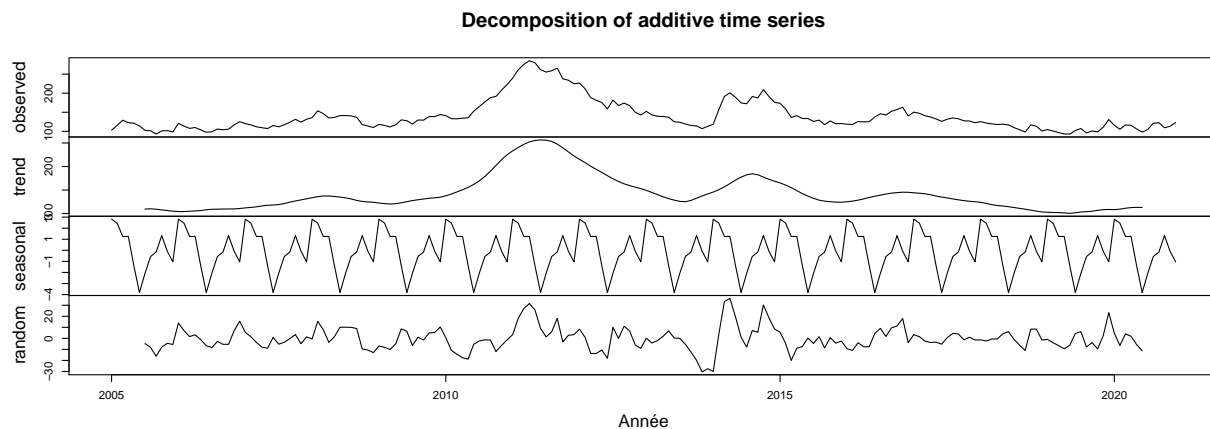
Bruit

Le bruit correspond aux variations résiduelles que nous ne pouvons pas expliquer. Nous pouvons le modéliser en enlevant la tendance et la saisonnalité de la série. Le bruit est centré à 0 avec une variance plus ou moins constante quimontre que la tendance et la saisonnalité ont été bien captées.



Décompose

Les différentes composantes peuvent être récapitulées avec la fonction *decompose*.



3. Lissage Exponentiel

Le lissage exponentiel consiste à prédire la série par la mise à jour de la tendance et de la saisonnalité avec des nouvelles observations disponibles.

Séparation de jeu de données

Nous séparons la série en deux pour pouvoir comparer la prédiction avec les vraies valeurs observées. Nous allons entraîner sur la série du 2005 à 2018 et prédire les valeurs du 2019 à 2020.

Comme nous avons vu précédemment que notre série est multiplicative, nous allons effectuer une modélisation multiplicative par la suite.

```
coffee_train = window(coffee, start = 2005, end = c(2018, 12))
coffee_test = window(coffee, start = 2019, end = c(2020, 12))
```

Lissage Exponentiel simple

Le lissage exponentiel simple prédit la série sans saisonnalité avec une tendance localement constante. C'est une moyenne pondérée qui attribue plus de poids au passé proche par le paramètre α . Il donne plus d'importance aux valeurs récentes.

Dans notre modèle, $\alpha = 0,9999$ très proche de 1. Il donne plus de poids au passé très proche.

```
## ETS(M,N,N)
##
## Call:
## ets(y = coffee_train, model = "MNN")
##
## Smoothing parameters:
##   alpha = 0.9999
##
## Initial states:
##   l = 102.5478
##
## sigma: 0.071
##
##      AIC      AICc      BIC
## 1635.874 1636.021 1645.246
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.009712432 10.26649 7.791354 -0.2413607 5.302575 0.2014017
##              ACF1
## Training set 0.1952324
```

Lissage Exponentiel double

Le lissage exponentiel double modélise la série sans saisonnalité mais avec une tendance localement linéaire. Ici, nous avons deux paramètres α et β où α consiste à mettre à jour le niveau et β la pente.

Dans notre modélisation α est proche de 1 et β est proche de 0. Le modèle provoque de fortes corrections de niveau et de faibles variations de la pente.

Nous observons que les AIC, AICc, et BIC sont faiblement plus élevés qu'avec le lissage exponentiel simple. En terme de ces trois mesures de qualité, le lissage exponentiel double n'est pas préférable au lissage exponentiel simple.

```
## ETS(M,M,N)
##
## Call:
## ets(y = coffee_train, model = "MMN", damped = F)
##
```

```
## Smoothing parameters:
##   alpha = 0.9999
##   beta  = 1e-04
##
## Initial states:
##   l = 123.4065
##   b = 1.0013
##
## sigma: 0.0725
##
##      AIC      AICc      BIC
## 1645.603 1645.974 1661.223
##
## Training set error measures:
##               ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
## Training set -0.3247696 10.3977 7.943766 -0.4936384 5.447539 0.2053415 0.176425
```

Lissage Exponentiel Triple

Le lissage exponentiel triple modélise la série à la fois avec la saisonnalité et la tendance. Ici, nous avons un autre paramètre γ qui consiste à mettre à jour la saisonnalité.

Dans notre modélisation, α est proche de 1, β et γ sont proches de 0. Le modèle donne plus d'importance au passé très récent pour la mise à jour du niveau mais le passé récent n'a pas autant d'impact pour les mises à jour de la pente et de la saisonnalité.

Les AIC, AICc, et BIC sont faiblement plus élevés que le lissage exponentiel double. Donc le lissage exponentiel triple n'est pas meilleur que les deux précédents.

```
## ETS(M,M,M)
##
## Call:
## ets(y = coffee_train, model = "MMM", damped = F)
##
## Smoothing parameters:
##   alpha = 0.9999
##   beta  = 0.0866
##   gamma = 1e-04
##
## Initial states:
##   l = 111.9594
##   b = 1.0106
##   s = 0.9654 0.982 1.0066 0.9845 1.0016 0.9882
##       0.992 1.0035 1.0206 1.0258 1.0253 1.0045
##
## sigma: 0.0734
##
##      AIC      AICc      BIC
## 1661.101 1665.181 1714.209
##
##Training set error measures:
##               ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.9307888 10.41572 7.729437 -0.5943357 5.282681 0.1998012
##               ACF1
## Training set 0.1357523
```


Lissage Exponentiel Automatique

Nous avons également la possibilité de laisser R choisir le meilleur modèle en terme de AICc. Le meilleur modèle retenu est le modèle multiplicatif avec une tendance localement constante et sans tendance qui correspond à notre premier modèle de lissage exponentiel simple.

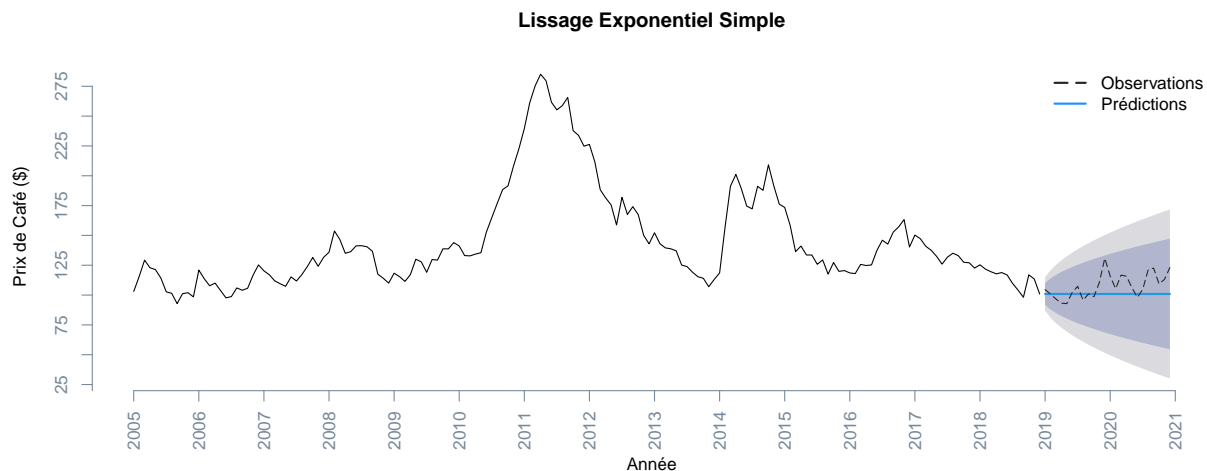
```
## ETS(M,N,N)
##
## Call:
## ets(y = coffee_train)
##
## Smoothing parameters:
##   alpha = 0.9999
##
## Initial states:
##   l = 102.5478
##
## sigma: 0.071
##
##      AIC      AICc      BIC
## 1635.874 1636.021 1645.246
##
## Training set error measures:
##               ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.009712432 10.26649  7.791354 -0.2413607  5.302575  0.2014017
##               ACF1
## Training set 0.1952324
```

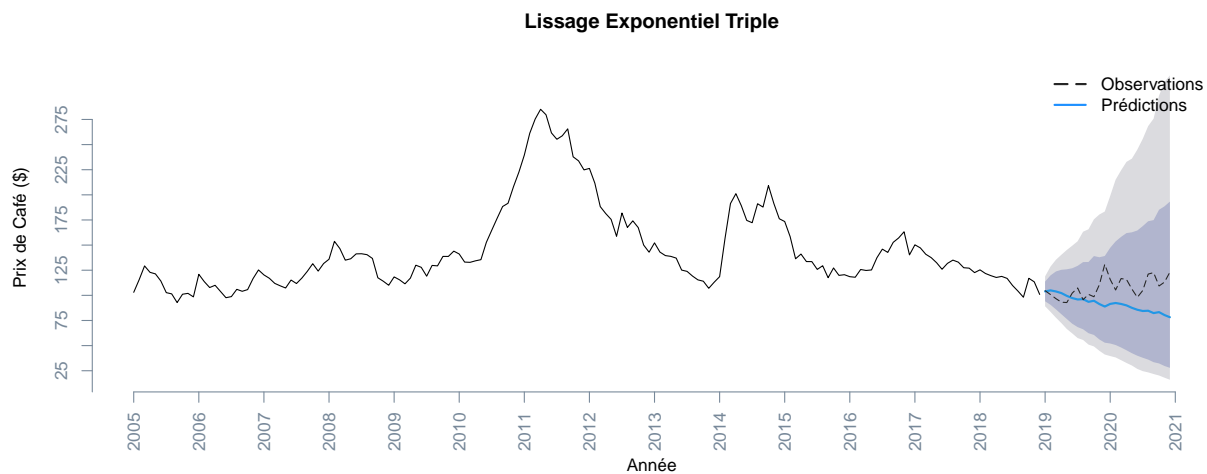
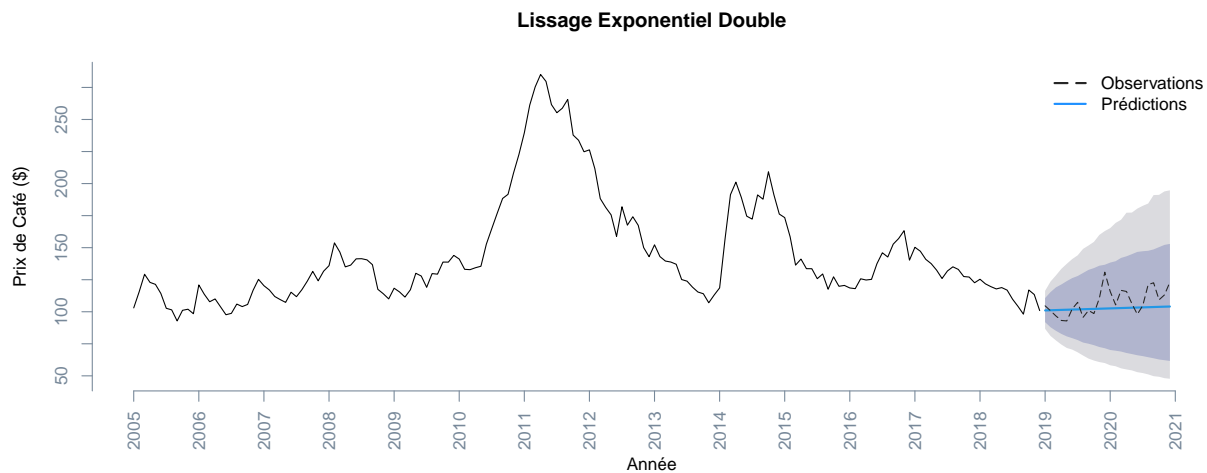
Prédiction

Nous pouvons maintenant prédire avec les trois modèles.

Les prédictions de deux premiers modèles sans saisonnalité sont plus proches aux valeurs réelles observées que le lissage triple.

Au lissage triple, les prédictions ont tendance à décroître alors que les vraies valeurs ont une tendance presque constante. Il présente également des intervalles de prédictions plus larges, donc plus d'incertitude.





Le fait que la prédiction constante du lissage exponentiel simple prédit le mieux la série laisse la possibilité que la tendance et la saisonnalité observées sur le chronogramme sont plutôt liés à des éléments autres que le temps et que la tendance de la série est en fait constante sans saisonnalité régulière.

Ou peut-être qu'autres modèles plus complexes de lissage exponentiel seraient plus adaptés à la série.

4. SARIMA

Maintenant nous allons effectuer la prédiction avec le modèle SARIMA.

Dans un premier temps avec le modèle ARMA après avoir différencié la série, puis avec le modèle SARIMA sur la série d'origine.

Transformation

Tout d'abord, nous commençons par transformer la série pour résoudre le problème d'hétéroscédasticité et séparer à nouveau la série obtenue en séries d'entraînement et de test.

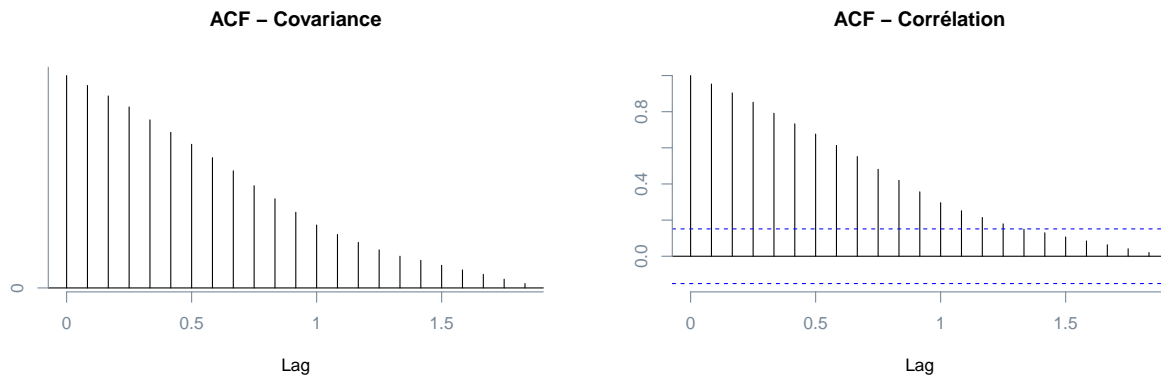
```
log_coffee = (log(coffee))

log_coffee_train = window(log_coffee, start = 2005, end = c(2018, 12))
log_coffee_test = window(log_coffee, start = 2019, end = c(2020, 12))
```

Stationnarité

Pour appliquer ARMA(p,q), la série doit être stationnaire. Nous pouvons examiner la stationnarité de série avec un corrélogramme.

Sur les corrélogrammes, nous observons que la corrélation est toujours positive et diminue lentement avec le lag. La série est fortement corrélée à ses séries retardées dont l'intensité diminue avec le temps d'écart. Sa moyenne et sa variance ne sont pas constantes donc la série n'est pas stationnaire. De plus la corrélation diminue de façon linéaire et nous n'observons pas de saisonnalité apparente.

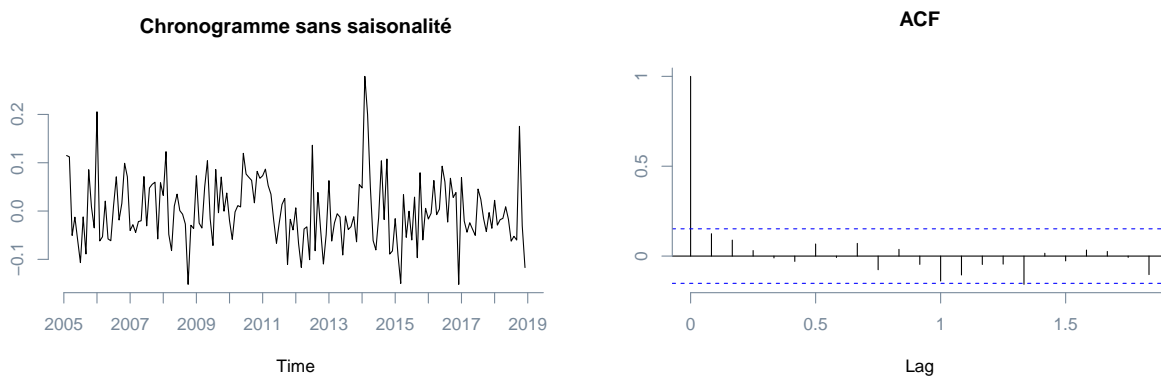


Pour rendre une série stationnaire, nous devons enlever la saisonnalité et la tendance.

Nous pouvons enlever la saisonnalité par l'opérateur différence. Étant donnée que la série présente un lien linéaire avec sa série retardée d'un mois, la différenciation est effectuée sur le lag 1.

Après la différenciation, nous observons que l'autocorrélation s'annule à partir du lag 1. La série n'est plus corrélée avec la série retardée.

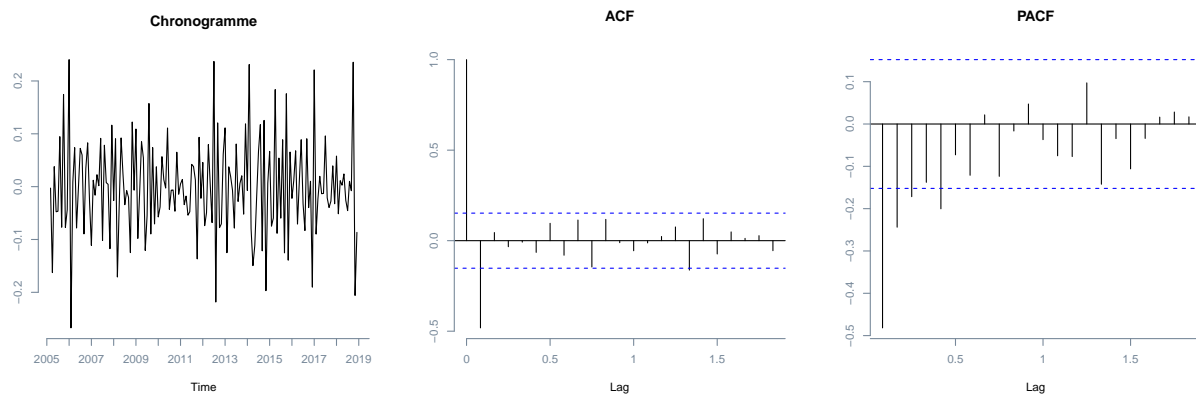
Sur le chronogramme, nous ne voyons plus de saisonnalité. Il reste à enlever la tendance.



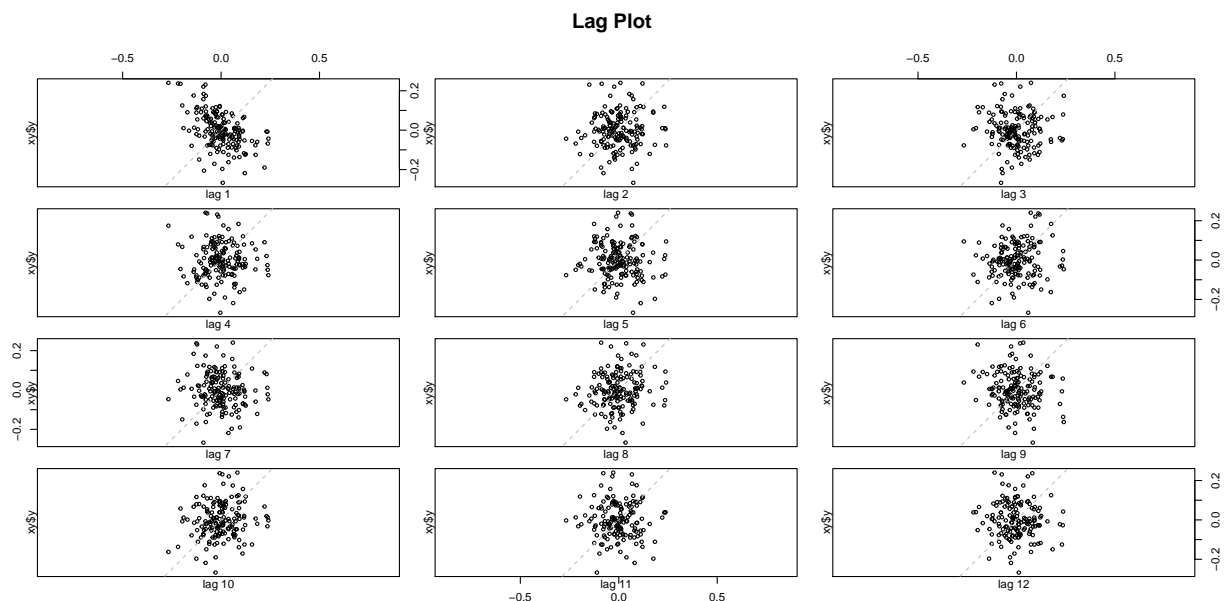
Pour éliminer la tendance, nous pouvons utiliser l'opérateur différence. En générale, une différenciation de degré 2 est suffisante pour obtenir une série stationnaire. Comme nous avons déjà effectuée une différenciation

de degré 1 pour éliminer la saisonnalité, nous allons nous contenter de faire une différenciation de degré 1 pour éliminer la tendance.

Après la différenciation, nous obtenons un chronogramme centré à 0 avec une variance constante. Il ne révèle plus de tendance ni d'aspect saisonnier. Nous pouvons vérifier avec l'ACF et la PACF que la série n'est plus corrélée avec les séries retardées.



Nous pouvons également vérifier avec le lag plot l'absence de lien linéaire entre la série et les séries retardées. Les variations d'un instant à l'autre sont aléatoires et le passé n'explique plus le présent.



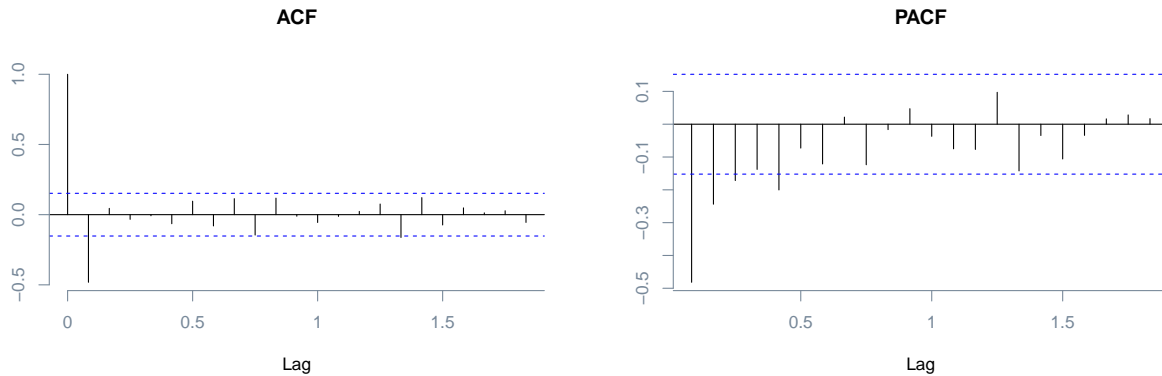
Modèle ARMA(p,q)

Maintenant que nous avons une série stationnaire, nous pouvons identifier le modèle ARMA qui est la combinaison du processus autorégressifs et moyennes mobiles. Pour cela, nous allons nous aider de l'ACF et la PACF de la série ci-dessous.

Nous avons l'ACF qui décroît exponentiellement avec oscillations vers 0 à partir de l'ordre 3, et la PACF qui décroît exponentiellement avec oscillations vers 0 à partir de l'ordre 6. Nous détectons donc un AR(2) et MA(5).

De plus nous observons que l'ACF s'annule dès le lag 2 alors que la PACF s'annule au bout du lag 6. Donc

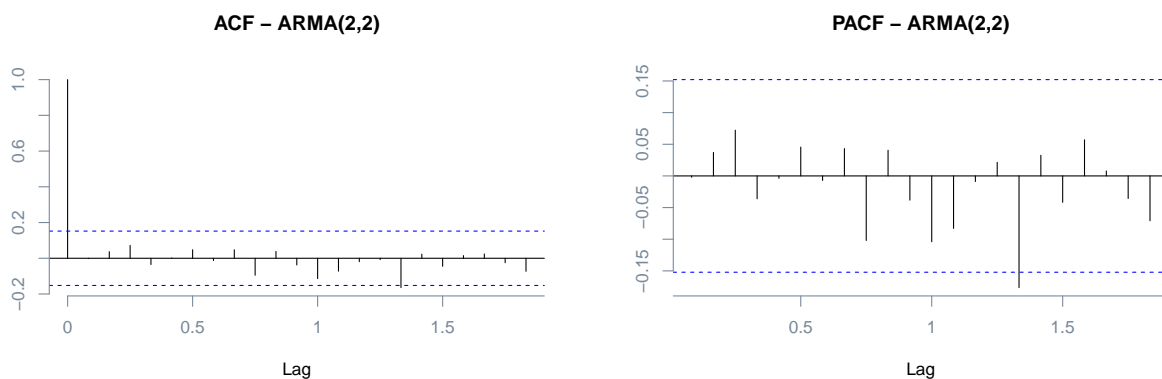
c'est l'AR qui est dominant. Nous allons donc garder l'ordre de AR détecté et donner un ordre assez bas à MA pour avoir un modèle ARMA(2,2)



ARMA(2,2)

Avec le modèle ARMA(2,2), nous pouvons vérifier avec l'ACF et la PACF des résidus du modèle, que les coefficients de l'autocorrelation des résidus sont nuls.

```
## Series: sans_saison_sans_tendance
## ARIMA(2,0,2) with non-zero mean
##
## Coefficients:
##          ar1      ar2      ma1      ma2      mean
##      -0.6606  0.1564 -0.2085 -0.7915 -1e-04
## s.e.   0.1960  0.0787  0.1834  0.1832  2e-04
##
## sigma^2 = 0.004756: log likelihood = 208.5
## AIC=-404.99  AICc=-404.46  BIC=-386.32
```



Nous pouvons tester si ces coefficients sont significativement nuls avec le test de blancheur. Le test confirme bien que la nullité. Les résidus restant sont bien des bruits blancs. Nous avons capté tout ce que nous pouvons modéliser.

```
##
## Box-Pierce test
```

```
##
## data: model$residuals
## X-squared = 12.202, df = 20, p-value = 0.9089
```

Nous allons examiner s'il existe la colinéarité entre les paramètres avec la fonction *cor.arma*. Nous voyons qu'il y a plusieurs paramètres qui présentent la colinéarité. Le modèle n'est pas optimal.

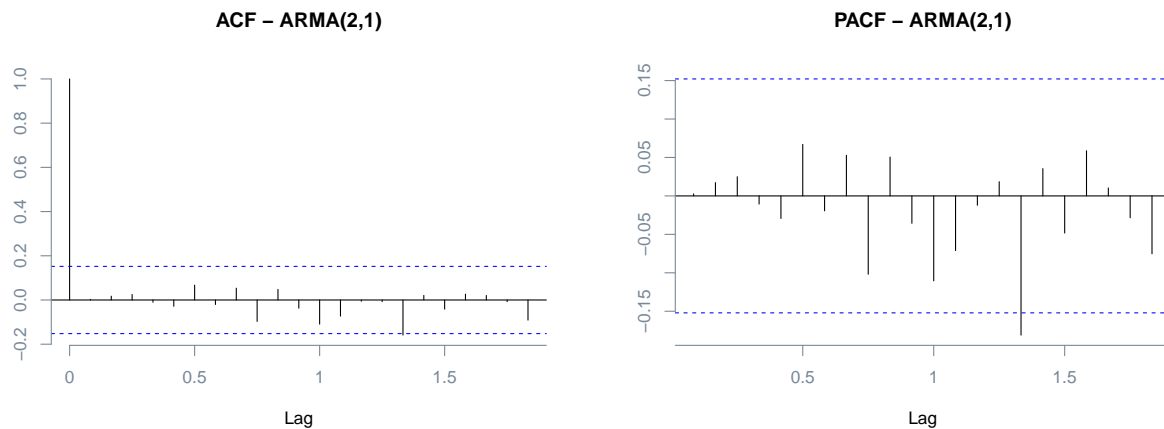
```
##           ar1           ar2           ma1           ma2  intercept
## ar1      1.00000000  0.29030451 -0.91372790  0.91511871 -0.03166079
## ar2      0.29030451  1.00000000  0.02894395 -0.02899218 -0.04981241
## ma1     -0.91372790  0.02894395  1.00000000 -0.98691710  0.01669732
## ma2      0.91511871 -0.02899218 -0.98691710  1.00000000 -0.01672204
## intercept -0.03166079 -0.04981241  0.01669732 -0.01672204  1.00000000
```

De plus, le test de Student montre que les paramètres ar1 et ar2 ne sont pas significatifs. L'ordre de notre modèle ne semble pas être adapté, nous allons modéliser à nouveau avec d'ordres plus bas.

```
##           ar1           ar2           ma1           ma2 intercept
## t.stat -3.369508  1.987788 -1.136623 -4.321081 -0.973900
## p.val   0.000753  0.046835  0.255696  0.000016  0.330106
```

ARMA(2,1)

En baissant l'ordre de MA à 1, nous avons modélisé ARMA(2,1) et il y a toujours les problèmes de significativité et de colinéarité des paramètres.



```
## [1] "ARMA(2,1)"

##
## Box-Pierce test
##
## data: model_2$residuals
## X-squared = 11.386, df = 20, p-value = 0.9356

## [1] "Test Student - ARMA(2,1)"
```

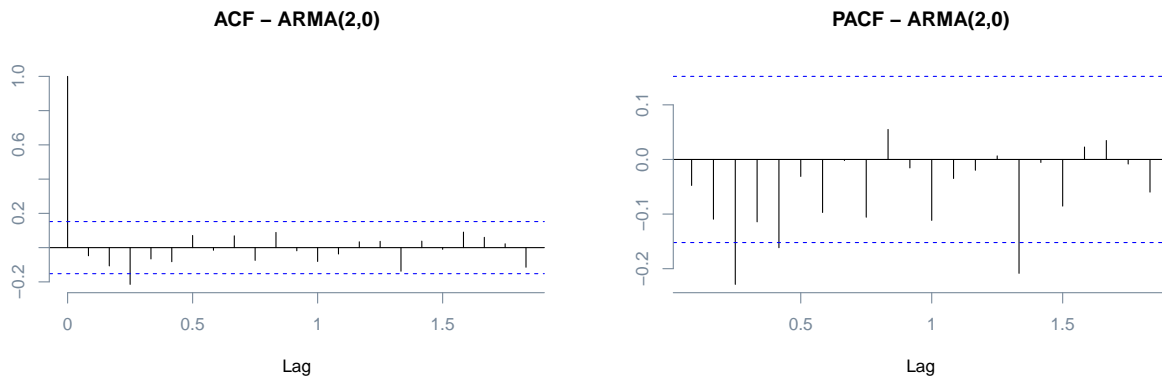
```
##          ar1      ar2      ma1 intercept
## t.stat 1.545174 1.055436 -61.1777 -0.957884
## p.val  0.122304 0.291226  0.0000  0.338121
```

```
## [1] "Test de Colinéarité - ARMA(2,1)"
```

```
##          ar1      ar2      ma1      intercept
## ar1      1.000000e+00 -9.825713e-02 -2.612127e-06 -3.668791e-02
## ar2      -9.825713e-02  1.000000e+00 -2.614638e-06 -5.785009e-02
## ma1      -2.612127e-06 -2.614638e-06  1.000000e+00  1.801035e-07
## intercept -3.668791e-02 -5.785009e-02  1.801035e-07  1.000000e+00
```

ARMA(2,0)

Avec un ordre MA plus bas, ARMA(2,0), le modèle semble être bon. L'ACF, la PACF et le test du blanchement vérifient que les résidus obtenus sont du bruit blanc et les paramètres sont significatifs sans problème de colinéarité.



```
## [1] "ARMA(2,0)"
```

```
##
## Box-Pierce test
##
## data: model_3$residuals
## X-squared = 22.819, df = 20, p-value = 0.2978
```

```
## [1] "Test Student - ARMA(2,0)"
```

```
##          ar1      ar2 intercept
## t.stat -8.12984 -3.540912 -0.263402
## p.val  0.00000  0.000399  0.792240
```

```
## [1] "Test de Colinéarité - ARMA(2,0)"
```

```
##          ar1      ar2      intercept
## ar1      1.00000000  0.51017900 -0.01403019
## ar2      0.51017900  1.00000000 -0.02830942
## intercept -0.01403019 -0.02830942  1.00000000
```

ARMASELECT

Nous pouvons également laisser R choisir le modèle.

La fonction *armaselect* nous sélectionne les modèles dans l'ordre de performance en terme de SBC. Néanmoins, les cinq premiers modèles présentent des problèmes de colinéarité donc nous n'allons pas utiliser les modèles proposés par la fonction pour notre série.

```
##      p q      sbc
## [1,] 0 1 -865.8463
## [2,] 1 2 -864.1460
## [3,] 1 3 -862.2816
## [4,] 1 1 -861.7682
## [5,] 2 2 -861.6684
## [6,] 0 2 -860.1013
## [7,] 0 3 -858.5876
## [8,] 2 3 -857.4190
## [9,] 5 1 -857.1320
## [10,] 2 1 -856.9376
```

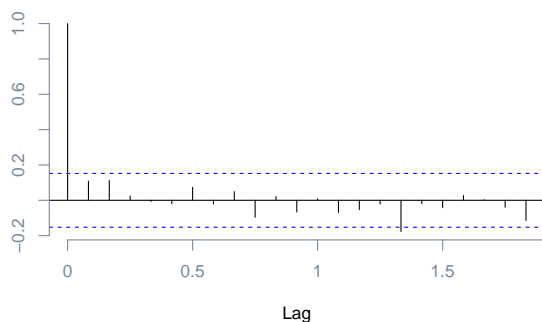
AUTOARIMA

Quant à la fonction *auto.arima*, elle nous donne le meilleur modèle en terme de AICc. Pour notre série, elle propose le modèle SARIMA(0,0,1)(1,0,0)[12].

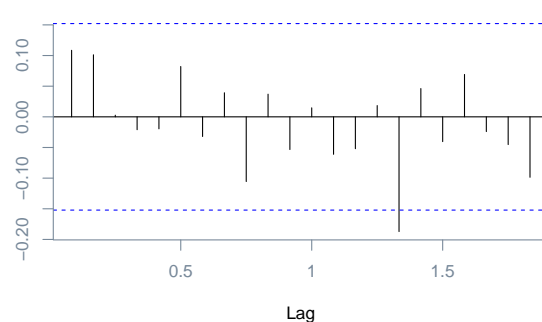
L'ACF, la PACF et le test du blancher du modèle vérifient que les résidus obtenus sont du bruit blanc et que les paramètres sont significatifs sans problème de colinéarité.

```
## Series: sans_saison_sans_tendance
## ARIMA(0,0,1)(1,0,0)[12] with zero mean
##
## Coefficients:
##      ma1      sar1
##    -0.9847 -0.1579
## s.e.  0.0242  0.0825
##
## sigma^2 = 0.004743: log likelihood = 207.58
## AIC=-409.17  AICc=-409.02  BIC=-399.83
```

ACF – ARIMA(0,0,1)(1,0,0)[12]



PACF – ARIMA(0,0,1)(1,0,0)[12]



```
## [1] "ARIMA(0,0,1)(1,0,0)[12]"
```



```
##
## Box-Pierce test
##
## data: model_auto$residuals
## X-squared = 14.9, df = 20, p-value = 0.7821

## [1] "Test Student - ARIMA(0,0,1)(1,0,0)[12]"

##          ma1          sar1
## t.stat -40.76653 -1.913110
## p.val   0.00000  0.055734

## [1] "Test de Colinéarité - ARIMA(0,0,1)(1,0,0)[12]"

##          ma1          sar1
## ma1   1.00000000 -0.09760075
## sar1 -0.09760075  1.00000000
```

Nous pouvons comparer ce modèle avec notre modèle ARMA(2,0) retenu.

Le modèle SARIMA(0,0,1)(1,0,0)[12] présente AIC, AICc, BIC plus bas que le modèle ARMA(2,2). Donc c'est le modèle SARIMA que nous allons garder pour effectuer la prédiction.

```
## [1] "ARMA(2,0)"

## [1] -374.8583

## [1] -374.6099

## [1] -362.4104

## [1] "SARIMA(0,0,1)(1,0,0)[12]"

## [1] -409.1674

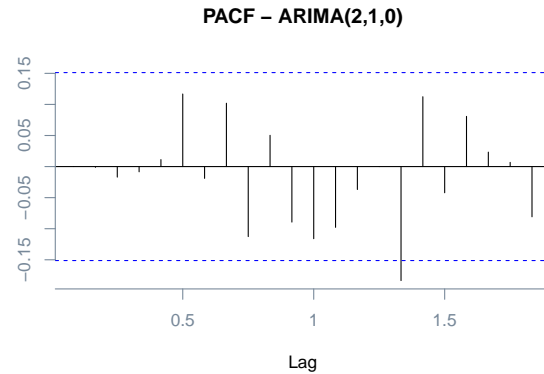
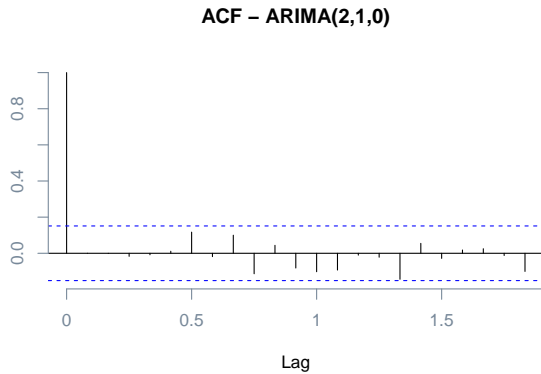
## [1] -409.0192

## [1] -399.8314
```

Nous pouvons également appliquer la fonction *auto.arima* directement sur la série origine sans différenciation.

Elle propose le modèle ARIMA(2,1,0). Néanmoins le modèle présente un problème de significativité des paramètres. Nous n'allons pas garder ce modèle.

```
## Series: coffee_train
## ARIMA(2,1,0)
##
## Coefficients:
##          ar1          ar2
##          0.1765  0.1102
## s.e.   0.0771  0.0774
##
## sigma^2 = 101.9: log likelihood = -622.1
## AIC=1250.21 AICc=1250.36 BIC=1259.56
```



```
## [1] "ARIMA(2,1,0)"

##
## Box-Pierce test
##
## data: sarima$residuals
## X-squared = 15.31, df = 20, p-value = 0.7584

## [1] "ARIMA(2,1,0)"

##          ar1      ar2
## t.stat 2.287899 1.423818
## p.val  0.022143 0.154499

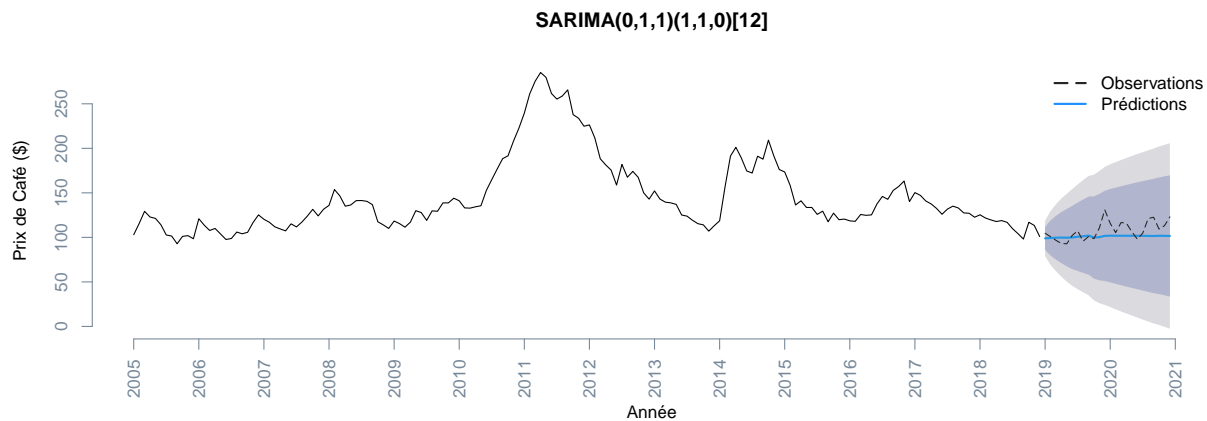
## [1] "ARIMA(2,1,0)"

##          ar1      ar2
## ar1  1.0000000 -0.1881458
## ar2 -0.1881458  1.0000000
```

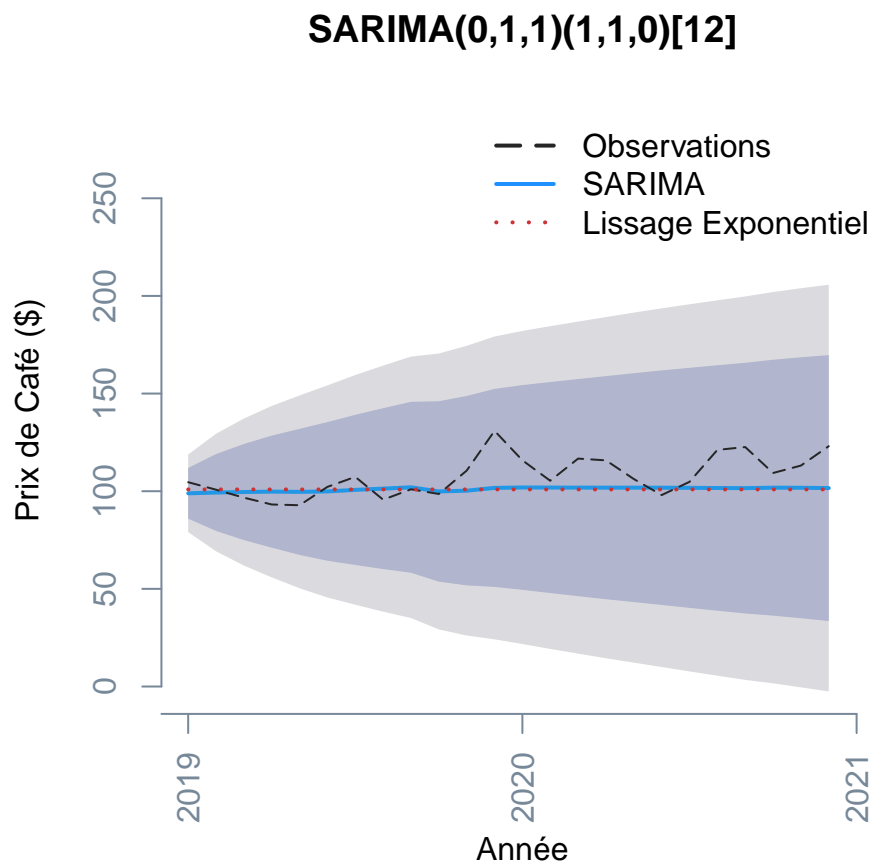
5. Prédiction

Nous pouvons maintenant effectuer la prédiction avec le modèle SARIMA(0,0,1)(1,0,0)[12] retenu. Le modèle a été obtenu en entraînant sur la série différenciée de 2 degrés, nous pouvons donc modéliser SARIMA(0,1,1)(1,1,0)[12] sur la série d'origine pour avoir le même modèle.

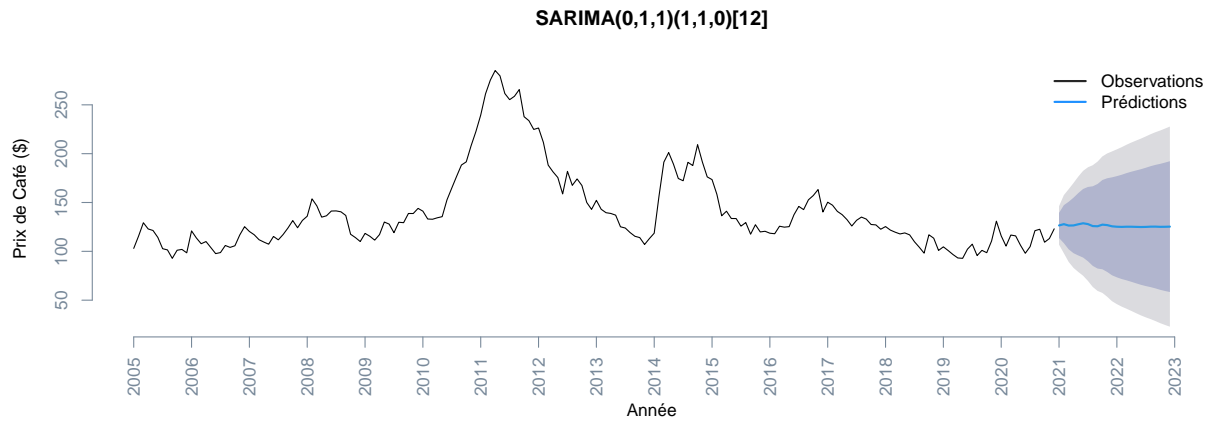
Nous observons que les prédictions sont au même niveau que les valeurs réelles observées mais elles restent constante alors que les vraies observations connaissent des variations.



Nous remarquons également que les prédictions obtenues sont presque mêmes que celles obtenues par le lissage exponentiel simple. Les variations qui existent sur les valeurs réelles ne sont pas captées par les deux modèles.



Nous pouvons quand même effectuer la prédiction sur les deux années suivantes qui ne sont pas observées. Pour les années 2022 et 2023, le modèle prédit une tendance constante avec des variations légères à la deuxième moitié de l'année 2022.



6. Conclusion

Malgré la tendance non-linéaire et la saisonnalité visibles sur le chronogramme, les différents modèles nous indiquent que la série a une tendance constante sans saisonnalité apparente au cours du temps.

Nous pouvons supposer que les variations observées sur la série sont des variations aléatoires et des variations liées aux événements externes au temps.

En effet, en 2010, le prix de café a connu une hausse brutale à cause des aléas climatiques subis par les principaux pays producteurs de café. Cette hausse avait enregistré le prix le plus haut de café depuis 13 ans, donc c'était un événement atypique qui ne se produit pas régulièrement.

Avec beaucoup plus d'observations ou en éliminant les périodes atypiques de la série, nous pourrions probablement trouver un modèle qui explique mieux la série et prédit les variations futures.