INSTITUT NATIONAL
DES SCIENCES
APPLIQUÉES
RENNES

GUILLAUD Baptiste

ROUX Matthieu

Projet de séries temporelles

4MA

Modélisation du trafic aérien Mondial

Table des matières

I. Introduction

L'objectif de ce travail est de modéliser le trafic aérien mondial puis de prédire ce trafic sur les prochaines années. On s'intéresse plus précisément à la capacité mensuelle mondiale qui est exprimée en ASK (Available Seat Kilometers). Pour un vol donné, il s'agit du produit du nombre de sièges avec la distance parcourue pendant le vol. On dispose de la valeur de cet indicateur de 2011 à 2020. On veut utiliser ces données au mieux afin de prédire la capacité mensuelle mondiale pour les années 2021 à 2025.

On trace d'abord l'allure de la série :

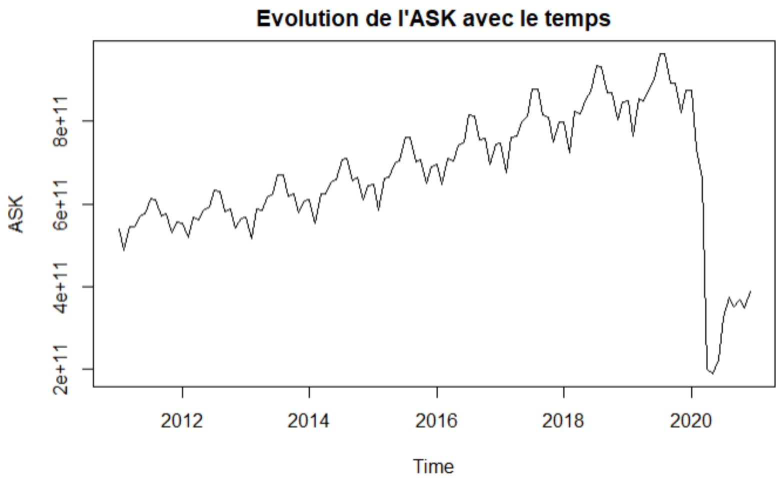


Figure 1 : Evolution de l'ASK entre 2011 et 2020

De janvier 2011 à mars 2020, la courbe est très régulière : on observe une tendance à la hausse, linéaire ou quadratique ainsi qu'une saisonnalité marquée, de période 1 année. Les mois où la capacité mondiale est la plus faible sont les mois d'hiver : janvier, février et mars. Les mois où cette quantité est la plus élevée sont juillet et août. La raison à cela est probablement que les trajets liés au départ en vacances sont plus nombreux l'été.

Cependant, il apparaît une brusque diminution de la capacité mensuelle mondiale en avril 2020, diminution qui invalide totalement la tendance haussière linéaire observée jusqu'à cette période. De plus, cette diminution perturbe la saisonnalité. Cette diminution est la conséquence de la pandémie de Covid et des mesures de limitation de déplacements qui ont suivi. Les valeurs de capacité mensuelle mondiale d'avril 2020 à décembre 2020 vont sans doute être des points influents ou aberrants.

Cette analyse peut être complétée en utilisant la fonction decompose de R sur cette série :

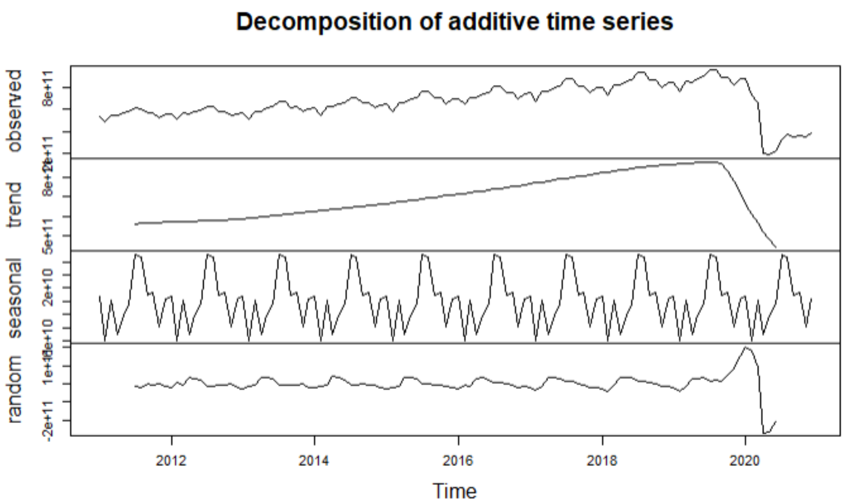


Figure 2 : décomposition de la série avec R

Ici, la tendance est estimée par la méthode de lissage avec moyenne mobile comme suit :

$$trend = \frac{1}{12}(\frac{X_{t-6}}{2} + X_{t-5} + \cdots + X_{t-1} + X_t + X_{t+1} + \cdots + X_{t+5} + \frac{X_{t+6}}{2})$$

On observe bien une tendance linéaire jusqu'à début 2020, puis une chute de la tendance sur 2020. On observe également la saisonnalité évoquée précédemment.

Ainsi, il est judicieux de modéliser le trafic avec la série avant 2020 puis avec la série contenant l'année 2020. Par ailleurs, cela nous permettra d'étudier les effets de la crise sur les prédictions du trafic aérien mondial.

C'est pourquoi dans un premier temps, nous réaliserons des prévisions grâce aux données d'avant 2020. Puis, nous ferons des prévisions en prenant en compte l'ensemble des données. Enfin, nous comparerons nos 2 méthodes et nous analyserons l'impact de la crise sanitaire sur le trafic aérien.

Remarque sur la méthodologie : nous avons divisé les données par 10^{11} afin d'éviter de manipuler des valeurs trop grandes, et ainsi, d'avoir des variances trop importantes.

Pour chaque approche, nous modéliserons l'évolution du ASK de 3 manières différentes : avec un modèle linéaire, un lissage de Holt-Winters et un modèle ARMA. Chaque prédiction aura un intervalle de confiance de niveau 95%.

II. Prévisions en utilisant les données d'avant 2020

On trace l'évolution du ASK avant 2020 :

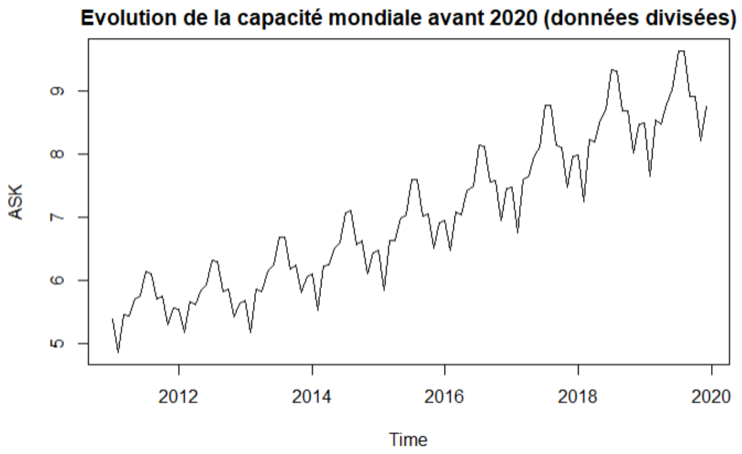


Figure 3 : Série ASK avant 2020

L'amplitude des variations annuelles semble augmenter au cours du temps. Nous choisissons donc de modéliser cette évolution de manière multiplicative. On passe donc les données divisées au logarithme.

Nous définissons également une série d'entrainement, les données de 2011 à 2018, ainsi qu'une série de validation, 2018 et 2019. Cela nous permettra de vérifier la qualité de notre modèle.

A. Modèle linéaire

Nous modélisons l'évolution du ASK de la manière suivante :

$$log(X_t) = \beta_1 t + \gamma_1 \ 1(t = janvier) + \cdots + \gamma_{12} \ 1(t = décembre) + \varepsilon_t$$

Avec pour tout t, ε_t identiquement distribué suivant une loi normale centrée.

Puis on réalise un premier modèle linéaire sur la série d'entrainement. Comme on peut le voir ci-dessous, la tendance et toutes les saisons sont significatives.

```
Coefficients:
         Estimate Std. Error t value Pr(>|t|)    
trend    4.894e-03  8.003e-05   61.15   <2e-16 ***
season1  1.642e+00  7.283e-03  225.53   <2e-16 ***
season2  1.547e+00  7.315e-03  211.44   <2e-16 ***
season3  1.653e+00  7.349e-03  224.99   <2e-16 ***
season4  1.647e+00  7.383e-03  223.04   <2e-16 ***
season5  1.688e+00  7.418e-03  227.58   <2e-16 ***
season6  1.696e+00  7.454e-03  227.56   <2e-16 ***
season7  1.763e+00  7.491e-03  235.40   <2e-16 ***
season8  1.757e+00  7.528e-03  233.38   <2e-16 ***
season9  1.677e+00  7.565e-03  221.61   <2e-16 ***
season10 1.677e+00  7.604e-03  220.59   <2e-16 ***
season11 1.592e+00  7.643e-03  208.28   <2e-16 ***
season12 1.640e+00  7.683e-03  213.53   <2e-16 ***
---
```

Figure 4 : coefficients du modèle linéaire sur les données divisées et logarithmisées

On vérifie les hypothèses sur les résidus avec les graphiques suivants

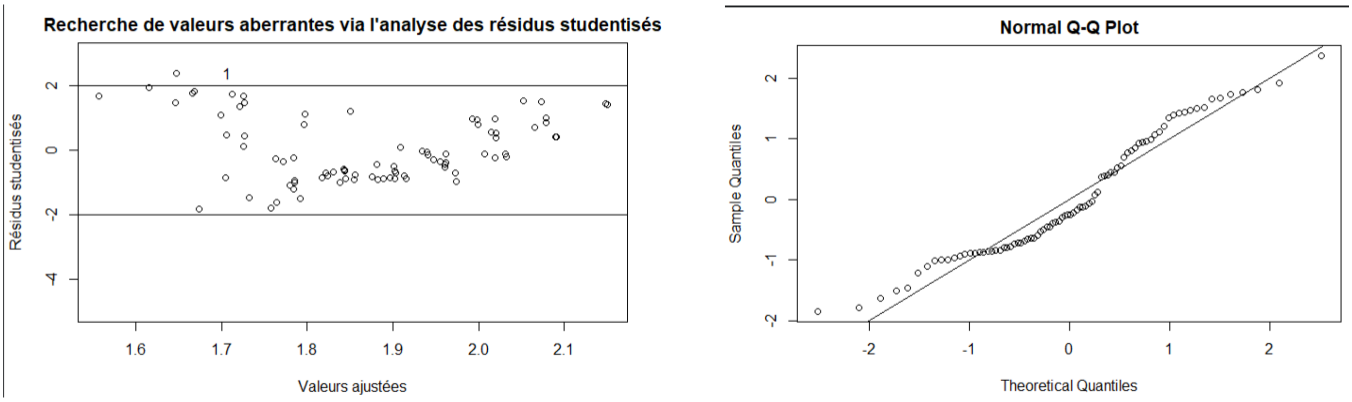


Figure 5 : Analyse des résidus du modèle linéaire

On n'observe pas de forme particulière sur le graphique des résidus studentisés ce qui confirme l'homoscédasticité. Une seule valeur paraît aberrante mais elle ne paraît pas trop importante. La moyenne des résidus est très proche de 0 et le diagramme quantile-quantile confirme que les résidus suivent bien une loi normale de variance 1.

En passant les données issues de la modélisation à l'exponentielle et en multipliant par 10^{11}, on évalue notre modèle sur les années de validation :

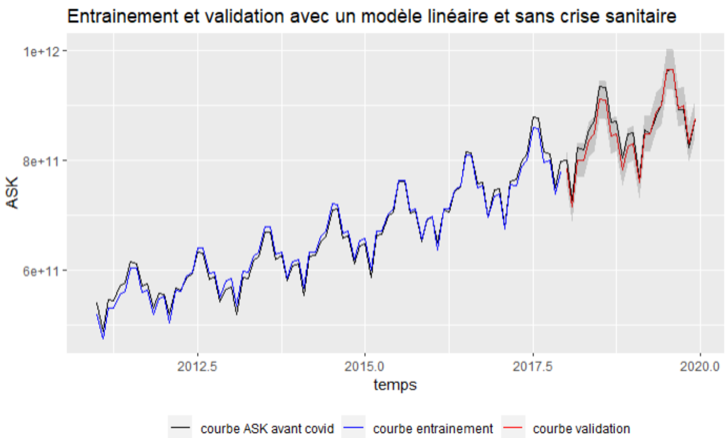


Figure 6 : Validation des prédictions avec le modèle linéaire

Notre modèle s'ajuste bien. De plus, la prédiction est très proche des données de validation. On regarde donc la prédiction jusqu'à 2025 avec ce modèle :

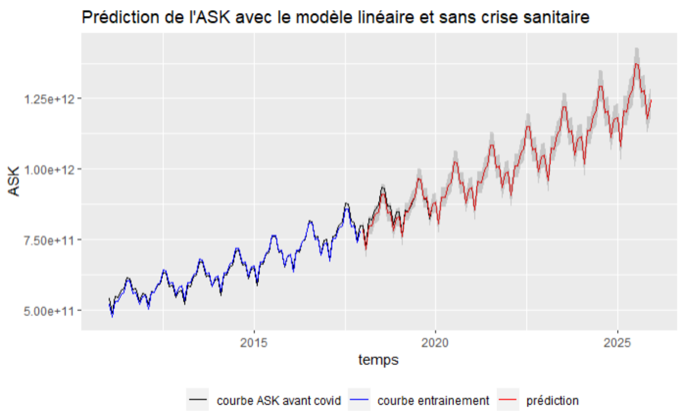


Figure 7 : Prédictions du ASK avec le modèle linéaire

B. Modèle de Holt-Winters

Le lissage de Holt-Winters pour les séries avec saisonnalité est une méthode de prédiction pouvant donner de bons résultats. Comme nous travaillons sur les données logarithmisées, nous gardons le modèle de Holt-Winters additif.
Pour rappel, cette méthode se base sur les 3 équations suivantes :

$$\begin{cases} \hat{a}_{T,1} = \alpha\left(X_T - \hat{S}_{T-d}\right) + (1-\alpha)(\hat{a}_{T-1,1} + \hat{a}_{T-1,2}) \\ \hat{a}_{T,2} = (1-\beta)\hat{a}_{T-1,2} + \beta(\hat{a}_{T,1} + \hat{a}_{T-1,1}) \\ \hat{S}_T = \gamma\left(X_t - \hat{a}_{T,1}\right) + (1-\gamma)\hat{S}_{T-d} \end{cases}$$

Les trois coefficients α, β, γ permettent de moduler le lissage. Ces coefficients peuvent être choisis manuellement ou déterminés via un critère tel que l'erreur de prédiction à 1 pas :

$$\sum_{t=1}^{T-1} (X_{t+1} - \widehat{X_t(1)})^{\,2}$$

La prédiction s'effectue ensuite de la manière suivante :

$$\begin{cases} \widehat{X_T}(h) = \hat{a}_{T,1} + \hat{a}_{T,2}h + \hat{S}_{T+h-d} \quad si\ 1 \le h \le d \\ \widehat{X_T}(h) = \hat{a}_{T,1} + \hat{a}_{T,2}h + \hat{S}_{T+h-2d} \quad si\ d+1 \le h \le 2d \\ \quad \dots \qquad\qquad\qquad\qquad \dots \end{cases}$$

Les différents paramètres de ce lissage sont estimés directement grâce à la fonction Holtwinters de R. On applique cette fonction à nos données divisées et logarithmisées.
On vérifie les hypothèses sur les résidus

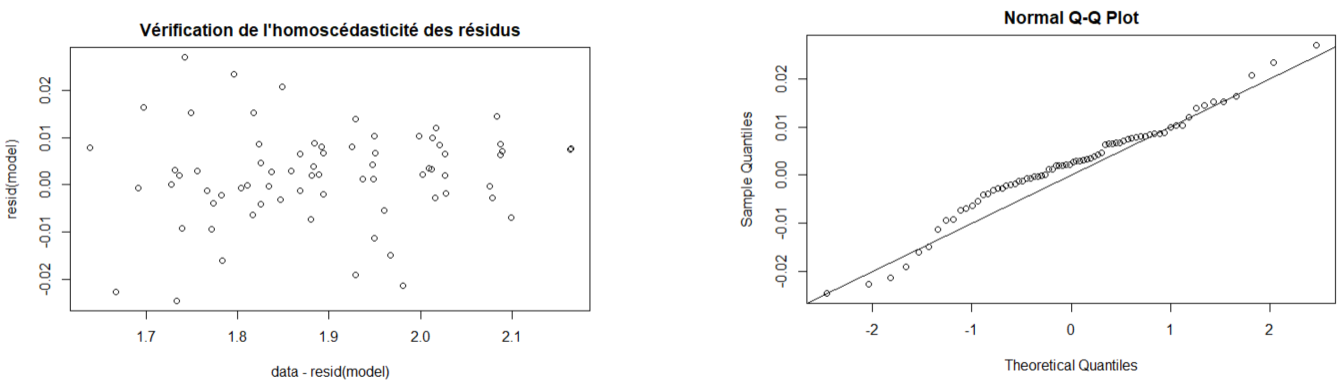


Figure 8 : analyse des résidus du lissage de Holtwinters

L'homoscédasticité, tout comme le centrage, sont vérifiés. Les résidus semblent suivre une loi normale mais la variance très petite. C'est pourquoi le diagramme quantile-quantile ne suit pas la première bissectrice.

On valide maintenant les prédictions

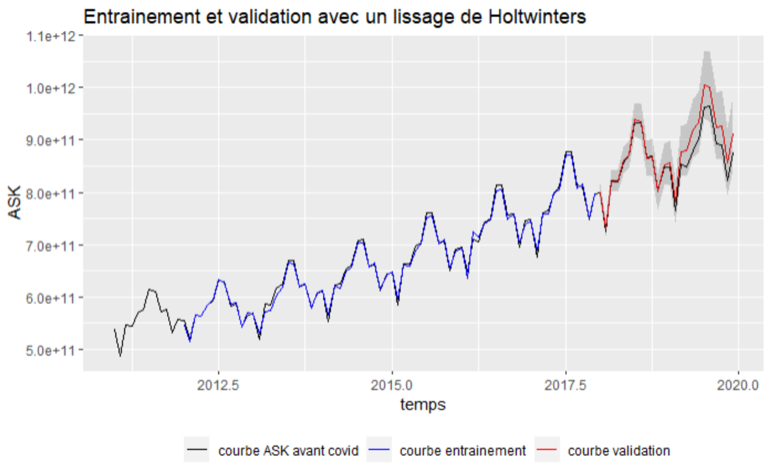


Figure 9 : Validation des prédictions avec le modèle de Holtwinters

Le modèle s'ajuste bien et les prédictions sont proche de la série de validation.

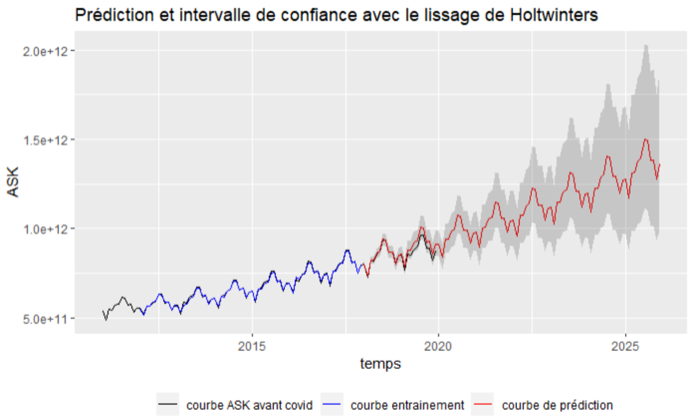


Figure 10 : prédictions avec le modèle de Holt-Winters

Comme pour le modèle linéaire, les prédictions nous montrent une augmentation importante du ASK. Ici, l'intervalle de confiance est plus large.

C. Modèle ARMA(p,q)

Les modèles Arma sont utilisés sur des processus stationnaires, c'est-à-dire, des processus de carré intégrable, de moyenne constante et dont la fonction d'autocovariance $\gamma(t+h,t)$ est indépendante de t pour tout h.

On sait d'après le modèle linéaire avec saisonnalité, que notre série possède une tendance et une saisonnalité. Elle ne peut donc être stationnaire. On enlève donc la tendance et la saisonnalité issue du modèle linéaire.

On réalise ensuite le test de Dickey-Fuller sur ces données afin de vérifier que le processus est stationnaire.

Pour ce faire, on définit ASK', le processus sans tendance et sans saisonnalité.

Le test s'effectue sur $\Delta ASK'_t = c + bt + (\phi - 1)ASK'_{t-1} + \varepsilon_t$

Les étapes du test sont les suivantes :

- On teste b=0
- Si oui, on teste c=0
- On teste $(\phi - 1) = 0$
- Si $\phi - 1$ est différent de 0, le processus est stationnaire

Ce test nous confirme que le processus sans tendance et sans saisonnalité est bien stationnaire. On peut donc le modéliser à l'aide d'un modèle ARMA.

Les graphiques d'autocorrélations et d'autocorrélations partiels ci-dessous suggère un modèle Arma(1,11).

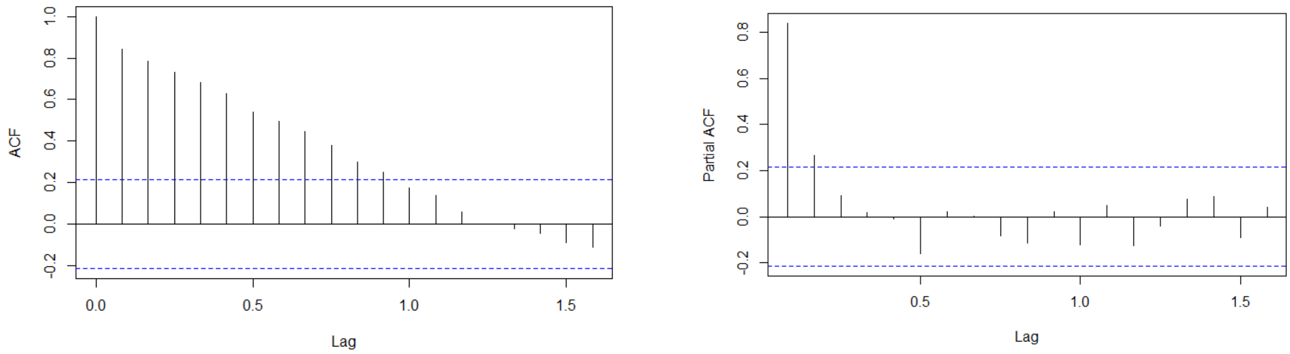


Figure 11 : ACF (à gauche) et PACF (à droite) du processus stationnaire

Cependant, on cherche à avoir le modèle le plus parcimonieux. On applique la fonction auto.arima sur ASK' et on obtient un modèle ARMA(1,1). Cette fonction permet de trouver le modèle ARIMA qui minimise l'AIC avec des coefficients significatifs.

L'AIC est défini par $AIC = -2ll + 2(p + q + 1)$ avec ll, la log vraisemblance du modèle ajusté.

```
Coefficients:
         ar1      ma1
      0.9770  -0.4000
s.e.  0.0251   0.1043

sigma^2 estimated as 5.317e-05:  log likelihood=294.05
AIC=-582.09   AICc=-581.79   BIC=-574.8
```

Le modèle obtenu est donc $ASk'_t = 0.977ASK_{t-1} - 0.4\varepsilon_{t-1} + \varepsilon_t$

La fonction t_stat nous confirme la significativité des coefficients. Cette fonction réalise le rapport entre le coefficient et son écart-type. Ce rapport doit suivre une loi normale centrée et réduite sous H0. Ici, les p-value de cette statistique de test nous confirment que ces 2 coefficients sont différents de 0.

On vérifie les hypothèses sur les résidus

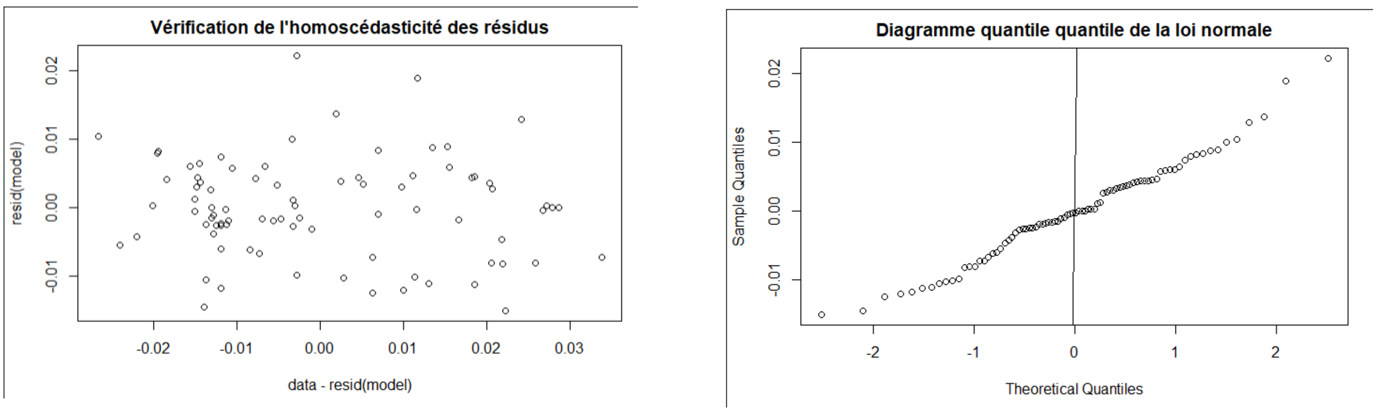


Figure 12 : Analyse des résidus du modèle ARMA (1,1)

Tout comme avec le lissage de Holtwinters, l'homoscédasticité, tout comme le centrage, sont vérifiés. Les résidus semblent suivre une loi normale mais on une variance très petite. C'est pourquoi le diagramme quantile-quantile ne suit pas la première bissectrice.

On regarde la prédiction sur le processus stationnarisé :

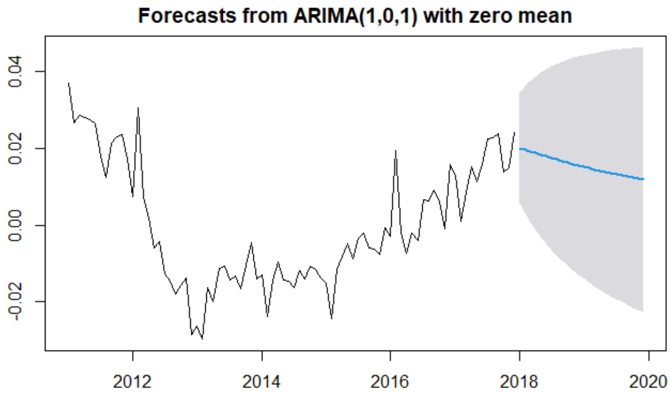


Figure 13 : prédiction du modèle arma sur le processus stationnarisé

La prédiction diminue nous laisse voir une diminution. On ajoute la tendance et la saisonnalité du modèle linéaire. Puis passe cette prédiction à l'exponentielle et on multiplie par 10^{11} ce qui nous donne le graphique suivant.

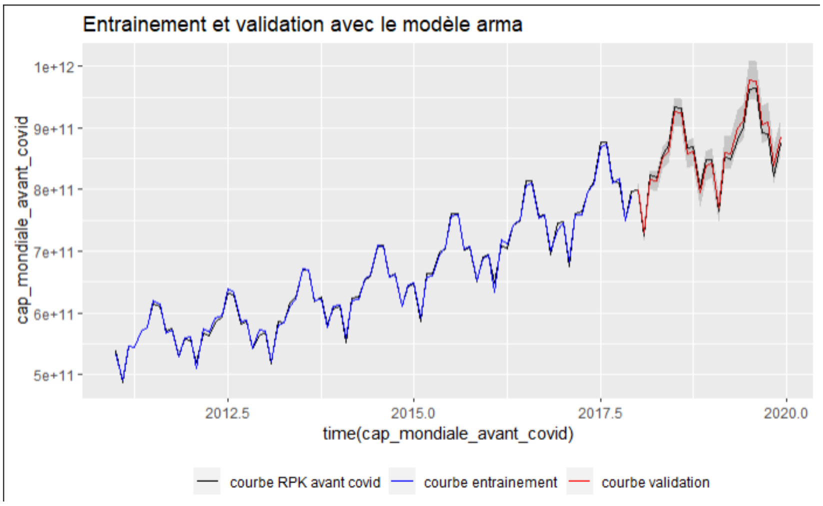


Figure 14 : Validation des prédictions du modèle arma

Les prédictions sont proches des données de validation, le modèle est donc performant pour les données d'avant covid. On réalise notre prédiction avec celui-ci.

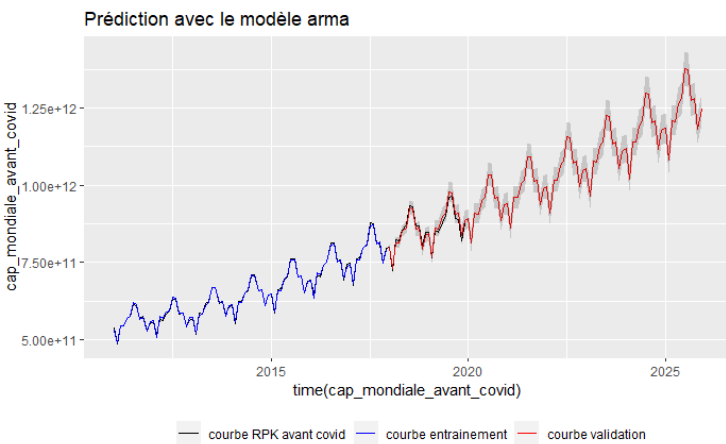


Figure 15 : Prédiction avec le modèle arma sur les données d'avant covid

Comme les 2 modèles précédents, la prédiction nous donne une augmentation du ASK.

D. Comparaison des modèles et des prédictions

On peut faire l'hypothèse que la crise du covid n'était qu'une parenthèse et que l'évolution du ASK reprendra comme avant la crise, avec la même tendance et la même saisonnalité. On met donc nos prédictions précédentes au niveau de décembre 2020 et on regarde la prédiction de 2021 à 2025.

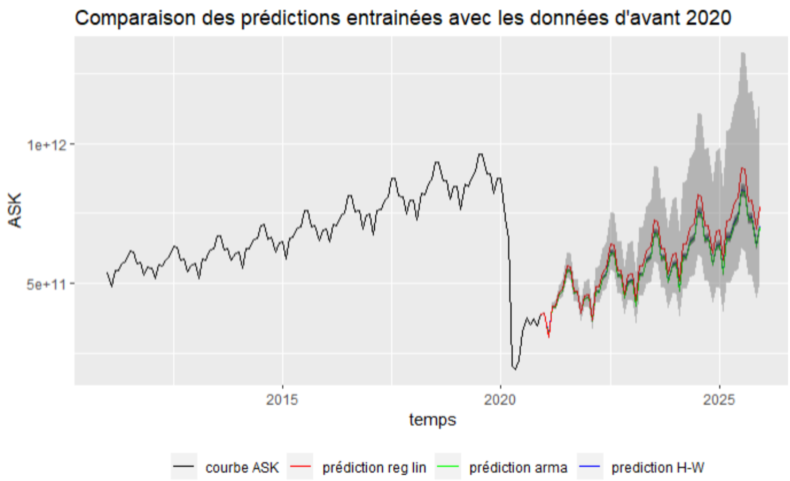


Figure 16 : Comparaison des modèles entrainés avec les données d'avant crise

Les 3 prédictions sont à peu près au même niveau. La prédiction réalisée avec le modèle linéaire est cependant légèrement plus haute que les 2 autres. La prédiction réalisée avec la méthode de Holt-Winters a un intervalle de confiance plus large que les 2 autres. Cette prédiction est donc moins précise mais elle nous permet de prendre moins de risque si un événement exceptionnel, telle qu'une crise sanitaire, venait à faire chuter ou à faire augmenter rapidement l'ASK.

On regarde également la somme des écarts entre les données de validation prédites et réelles, ainsi que la distribution des résidus :

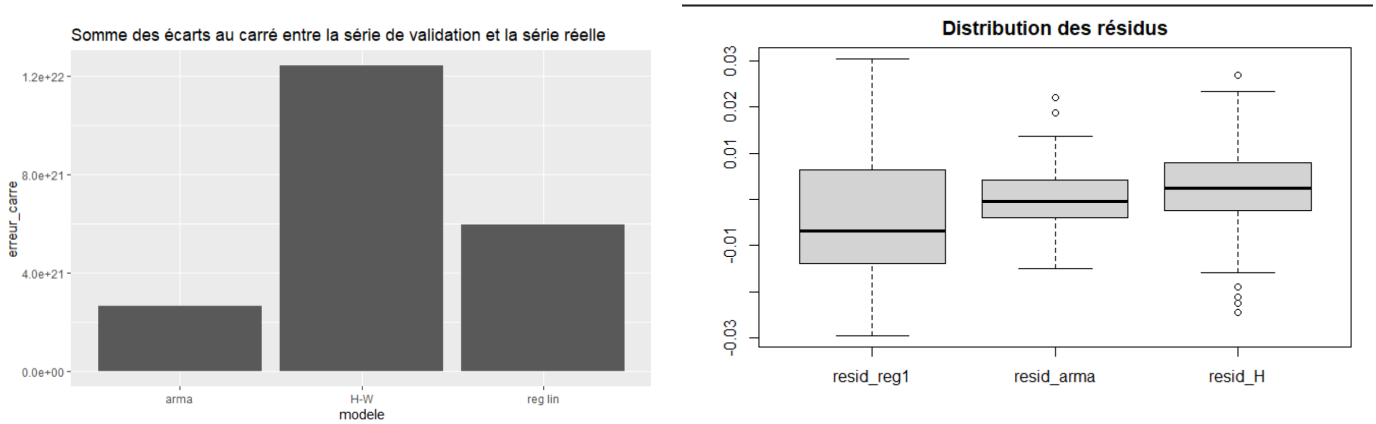


Figure 17 : Comparaison de la performance de nos modèles

Le modèle arma(1,0,1) a une meilleure performance sur le jeu de validation et a la distribution de résidus la plus étroite autour de zéro. C'est donc le modèle à priori, le plus fiable. Cependant, comme évoqué précédemment, il est intéressant de garder la prédiction réalisée avec Holt-Winters car l'intervalle de confiance est large.

III. Prédiction avec les données de 2020

Ici, on fait le choix de ne pas séparer la série en une série d'entrainement et de validation. En effet, si l'on veut que le modèle prenne en compte la crise du covid, notre série de validation sera trop petite ce qui risque d'empêcher une bonne comparaison.

La crise du covid nous donnera des points influents et/ou aberrants. De plus, cette chute soudaine risque d'invalider nos hypothèses sur les résidus.

A. Modèle linéaire

Comme en II.A, on réalise un modèle linéaire sur les données logarithmisées et divisées. La seule différence est que nous prenons l'ensemble de la série pour entrainer le modèle. On obtient les coefficients suivants :

```
Coefficients:
         Estimate Std. Error t value Pr(>|t|)
trend    0.0026668  0.0003712   7.184 9.47e-11 ***
season1  1.7198400  0.0487969  35.245  < 2e-16 ***
season2  1.6234566  0.0489534  33.163  < 2e-16 ***
season3  1.7340528  0.0491122  35.308  < 2e-16 ***
season4  1.7286128  0.0492732  35.082  < 2e-16 ***
season5  1.7710087  0.0494365  35.824  < 2e-16 ***
season6  1.7827631  0.0496021  35.941  < 2e-16 ***
season7  1.8503232  0.0497699  37.178  < 2e-16 ***
season8  1.8457750  0.0499399  36.960  < 2e-16 ***
season9  1.7689366  0.0501120  35.300  < 2e-16 ***
season10 1.7709677  0.0502863  35.218  < 2e-16 ***
season11 1.6878633  0.0504627  33.448  < 2e-16 ***
season12 1.7390350  0.0506413  34.340  < 2e-16 ***
---
```

Figure 18 : Coefficients obtenus avec le modèle linéaire sur l'ensemble de la série

Ces coefficients sont tous significatifs. Cependant, le graphique des valeurs ajustées donne :

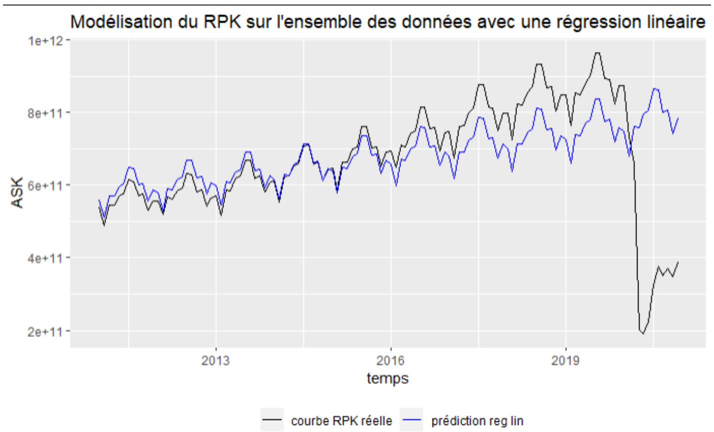


Figure 19 : comparaison de la série avec les valeurs ajustées du modèle linéaire

De plus, le graphique des résidus studentisés en fonction des valeurs ajustées montre une forme de parallélépipède, ce qui invalide l'hypothèse de l'homoscédasticité. Les points correspondant à 2020 sont tous considérés comme aberrants.

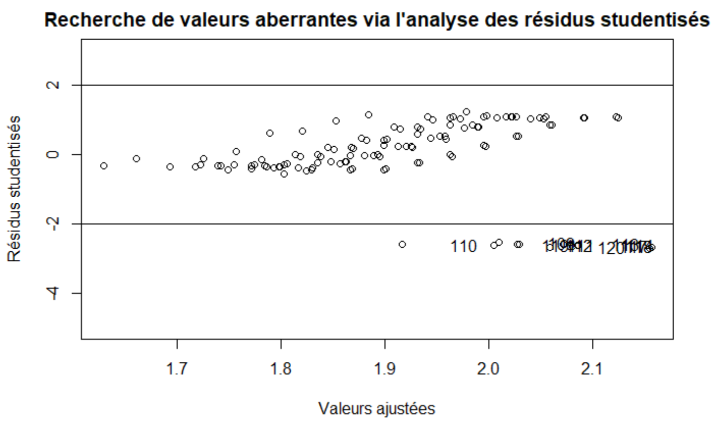


Figure 20 : Analyse des résidus studentisés du modèle linéaire

B. Modèle de Holt-Winters

1. Avec les données divisées et logarithmisées

On réalise un modèle de Holt-Winters avec les données initiales divisées et logarithmisées. Ce modèle est additif puisque les données ont été passées au logarithme. On obtient l'ajustement et la prédiction suivante.

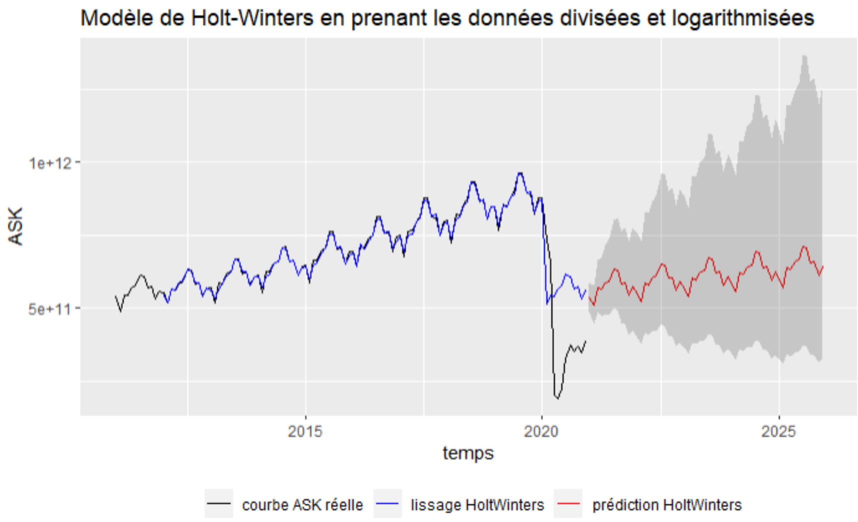


Figure 21 : Entrainement et prédiction sur le premier modèle de HoltWinters

Le modèle ne s'ajuste pas bien à toute la chute de 2020 et fait une prédiction qui reprend bien plus haut que le niveau de décembre 2020. C'est pourquoi on essaie un modèle en utilisant les données simplement logarithmisées.

2. Avec les données uniquement logarithmisées

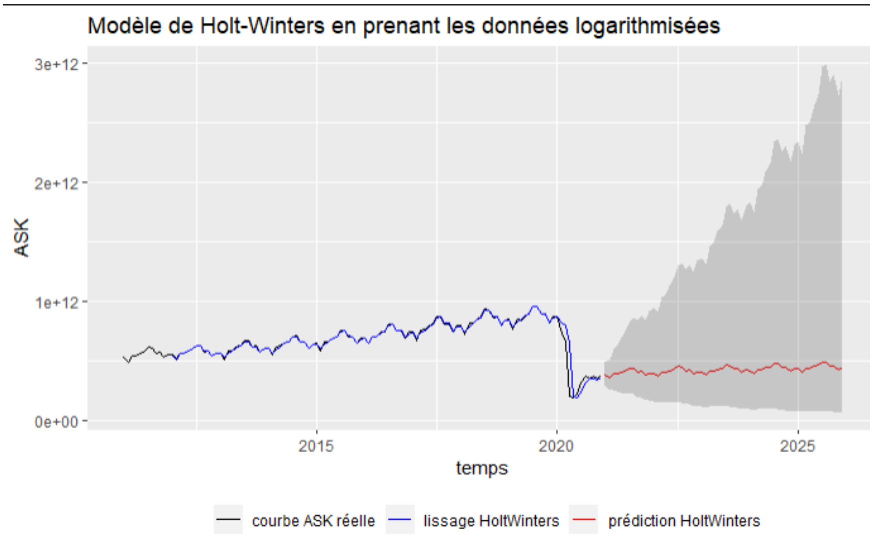


Figure 22 : Entrainement et prédiction avec le deuxième modèle de HoltWinters

Le modèle s'ajuste bien aux données. La prédiction a cependant un intervalle de confiance très large, la prédiction est donc difficilement interprétable. En effet, la borne supérieure indique un ASK à 3^{12} en 2025 tandis que la borne inférieure indique un ASK de l'ordre de 10^{10}. On a donc un facteur 100 entre la borne inférieure et la borne supérieur en 2025. Ce modèle peut donc être utile pour une prédiction dans l'année mais ne semble pas pertinent pour une prédiction à l'horizon 2025.

Une autre approche est également possible. En effet, il est difficile de faire des prévisions sur 5 ans, il est donc rassurant d'avoir un intervalle de confiance assez large pour prévoir les crises.

On valide les hypothèses sur les résidus du modèle :

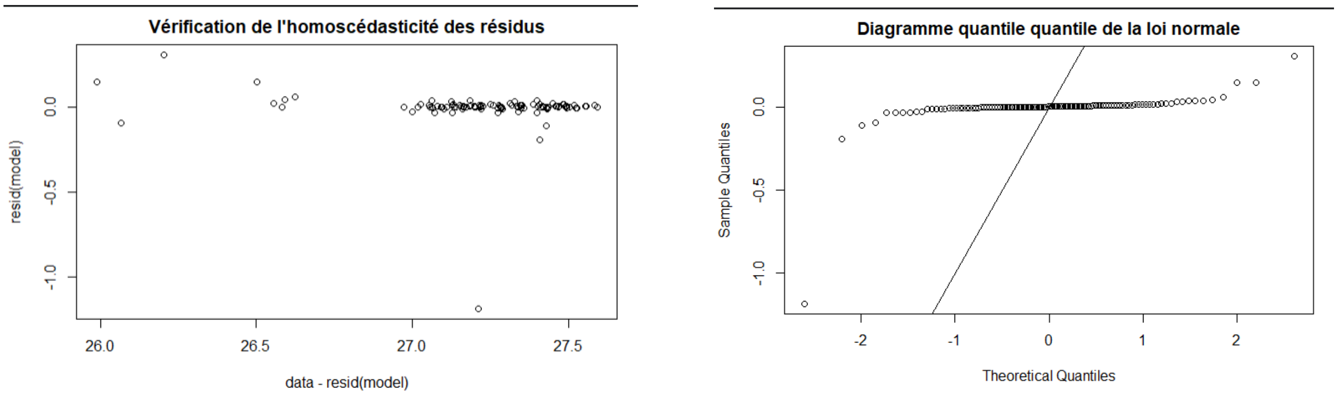


Figure 23 : Analyse des résidus du modèle de HoltWinters

L'hypothèse de normalité semble être vérifiée mais les variances sont très faibles. Le graphique des résidus en fonction des valeurs ajustées montre une dizaine de points aberrants, qui correspondent à l'année 2020. Les autres résidus semblent également regroupés dans un « paquet », mais sans forme particulière. On peut donc avoir quelques doutes sur l'homoscédasticité.

C. Modèle Arma(p,q)

On termine par modéliser le ASK avec un arma sur l'ensemble des données logarithmisées.

Afin d'enlever la saisonnalité, on différencie la série avec un lag de 12. Le test de Dickey-Fuller nous indique que le processus est DS (Differency Stationary). On travail donc sur la série logarithmisée, désaisonnalisée et différenciée.

En ajustant le modèle, on obtient la courbe suivante :

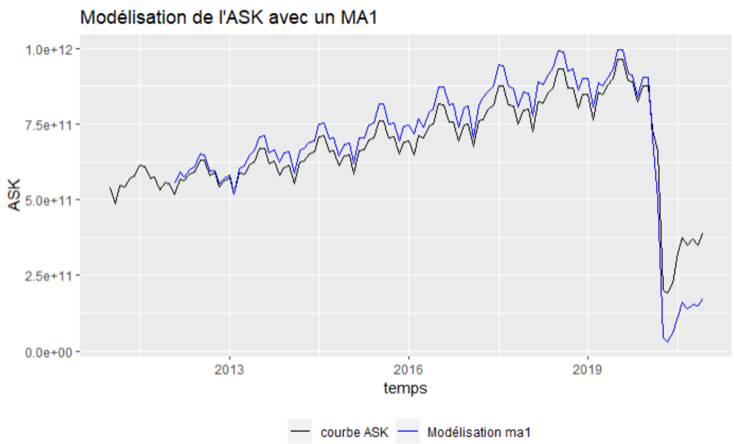
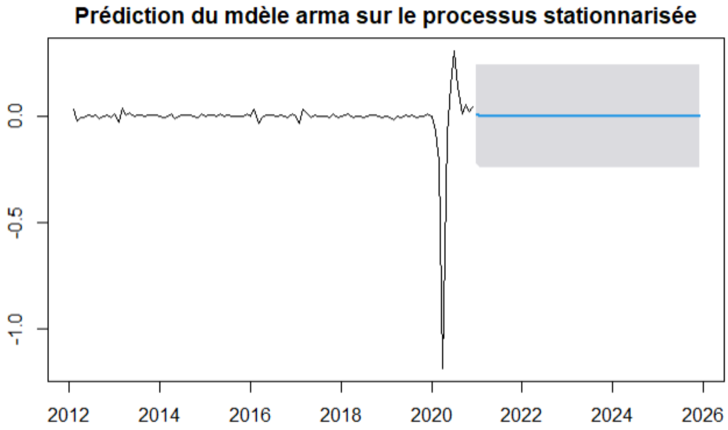


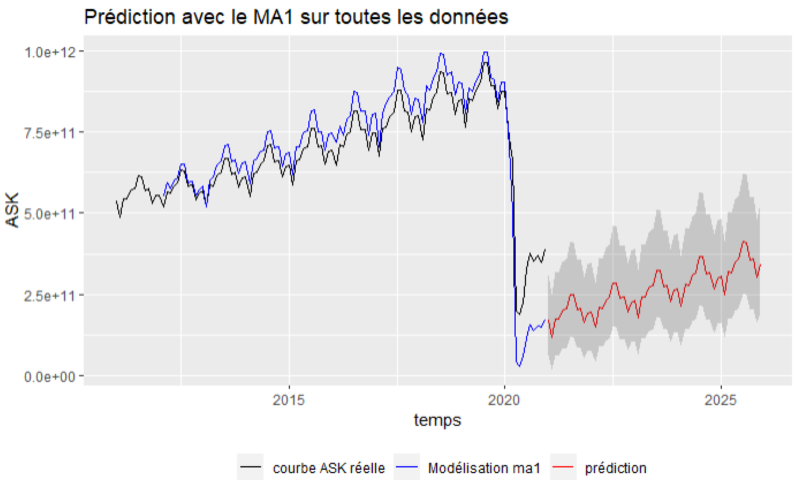
Figure 24 : Comparaison de la série réelle avec les valeurs ajustées du modèle arma

On réalise maintenant les prédictions sur le processus stationnaire :



Le modèle arma fait une prévision constante égale à 0.

On ajoute la tendance et la saisonnalité issue d'un modèle de régression sur les données logarithmisées avant covid. La seule chose qui changera par rapport à un modèle de régression classique réalisé sur les données avant covid est l'intervalle de confiance.



La prédiction semble en dessous de ce que pourrait donner une reprise du trafic aérien « normale » suite à la crise.

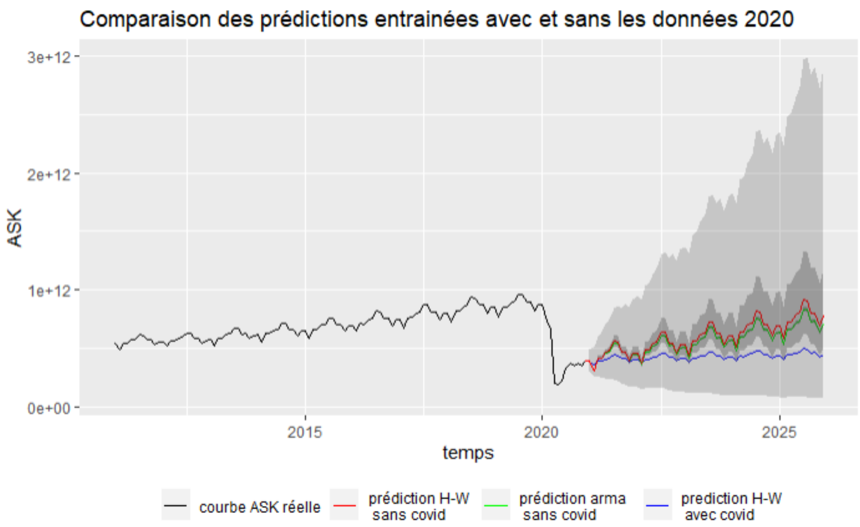
D. Choix du modèle

Ici, on voit très clairement que le modèle qui s'ajuste le mieux est notre 2e modèle de Holt-Winters. Le modèle linéaire n'est pas adapté car il ne prend pas en compte la chute de ASK de début 2020. Le modèle MA semble bien s'ajuster également mais la modélisation du processus stationnaire est une constante égale à 0, ce qui ne semble pas modéliser la réalité du processus.

On gardera donc le 2e modèle de Holt-Winters pour la comparaison de tous nos modèles.

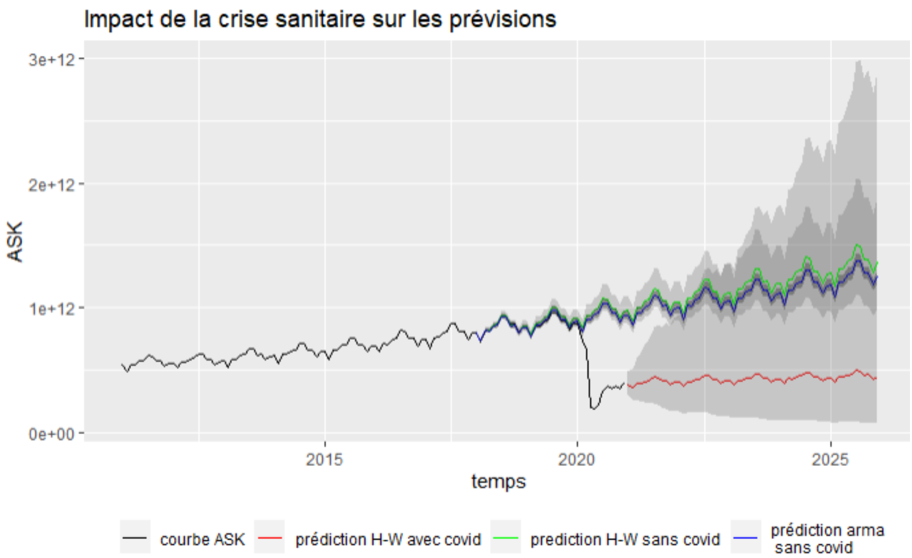
IV. Comparaison des différentes prédictions

On trace maintenant nos différentes prédictions sur le même graphique.



La prédiction de Holt-Winters est en dessous des 2 autres ce qui est logique vu qu'elle a été prédite grâce à un modèle entrainé sur toutes les données. De plus, l'intervalle de confiance de cette prédiction de H-W est large et comprend les 2 autres courbes, ainsi que leurs intervalles de confiance.

On regarde maintenant l'impact de la crise sur les prévisions :



Si l'on s'en tient aux prédictions sans intervalles de confiance. On remarque, que les modèles avec les données avant covid prédisent une augmentation continue avec une amplitude sur chaque année qui augmente. A l'inverse, la prédiction de Holt-Winters nous prédit un ASK relativement stable, restant au niveau de décembre 2021, avec une saisonnalité peu marquée.

Toutefois, l'intervalle de confiance de la prédiction de Holt-Winters est large (cf III.B) et, à partir de juillet 2022, il comprend les courbes de prédictions d'avant la crise. Ainsi, cet intervalle de confiance nous indique que l'ASK peut augmenter très rapidement et retrouver le niveau d'avant la crise dès 2022.

Enfin, l'imprécision de ce modèle de Holt-Winters nous confirme une chose : la reprise du trafic aérien mondiale est très incertaine.

V. Conclusion

Ce projet nous a permis de faire des prédictions sur l'ASK mondial avec différents modèles, et en prenant en compte 2 séries différentes : la série avant la crise et la série prenant compte de l'année 2020.

Ces prédictions ne sont pas aisées car le secteur du transport aérien a été durement touché par la crise. En effet, on peut voir l'écart important entre nos prédictions s'il n'y avait pas eu de crise et nos prédictions avec la crise.

De plus, selon Xerfi, leader des études sectorielles Français, ce secteur a une très mauvaise capacité d'adaptation à la crise. Cela nous confirme que la reprise du trafic aérien mondiale est incertaine et dépend de l'évolution de la pandémie.

Ainsi, la dernière prédiction réalisée avec le modèle de Holt-Winters semble le mieux représenter l'évolution de l'ASK : incertaine.

Annexe : Code R

Import des packages :

```{r}
library(tidyr)

library(forecast)

library(ggplot2)

library(caschrono)

#library(lubridate)
```

Chargement des données et allure de la série :

```{r}
data = read.csv('ST_4GM_Projet.csv', sep = ';', dec = ',', header = TRUE)

head(data)

summary(data)

cap_mondiale = ts(data$ASKs, start = c(2011, 1), end = c(2020, 12), frequency = 12)

plot(cap_mondiale, main="Evolution de l’ASK avec le temps", ylab="ASK")

plot(decompose(cap_mondiale))
```

Divisions des données et création d'une série "avant covid"

```{r}
div=10^11

cap_mondiale_avant_covid=window(cap_mondiale, start = c(2011, 1), end = c(2019,12))

plot(div_cap_mondiale_avant_covid, main="Evolution de la capacité mondiale avant 2020 (données divisées)", ylab="ASK")

div_cap_mondiale=ts(cap_mondiale_avant_covid/div, start=c(2011,1), end=c(2020,12), frequency=12)

div_cap_mondiale_avant_covid=window(div_cap_mondiale, start = c(2011, 1), end = c(2019,12))
```

Fonctions permettant de tracer les résidus :

Résidus studentisés pour la régression linéaire :

```{r}
PlotResStudent<-function(modele){

  plot(modele$fitted.values, rstudent(modele),ylim=c(-5,3), type="p", main="Recherche de valeurs aberrantes via l'analyse des résidus studentisés", xlab="Valeurs ajustées", ylab="Résidus studentisés", xy.labels=FALSE)

  abline(h=2)
```

````
  abline(h=-2)

  PtAb=which(rstudent(modele) < -2 | rstudent(modele) > 2)

  if(length(PtAb)!=0){

    text(modele$fitted.values[PtAb]+0.06, rstudent(modele)[PtAb],PtAb)

  }


  return(PtAb)

}


```
````

Pour les modèles de HoltWinters et Arima :

````
```{r}

plot_residuals=function(model,data){

 plot(resid(model), main="Vérification de la moyenne nulle des résidus")

 abline(h=0)

 hist(resid(model))

 qqnorm(resid(model), xy_labels=FALSE,main="Diagramme quantile quantile de la loi normale")

 abline(b=1, a=0)

 plot(data-resid(model), resid(model), xy.labels = FALSE, main="Vérification de l'homoscédasticité des résidus")

}

```
````

I. Prévisions avant covid

Pour tracer les courbes avec les données avant covid :

````
```{r}

plot_predict_av_cov <- function(train,predict, time_pred, title){

 ggplot()+

 geom_line(aes(x=time(cap_mondiale_avant_covid), y=cap_mondiale_avant_covid, color="black"))+

 geom_line(aes(x=time(cap_mondiale_avant_covid_train), y=train, color="blue"))+

 geom_line(aes(x=time_pred, y=predict$mean, color='red'))+

 geom_ribbon(aes(x=time_pred, ymin=predict$lower, ymax=predict$upper), alpha=0.2)+

 ggtitle(title)+

 scale_color_manual(name=NULL,

          values=c("black","blue", "red"),

          labels=c("courbe ASK avant covid","courbe entrainement", "courbe validation" ))+
````

theme(legend.position="bottom")

}

```

### Séparation du jeu de données en un jeu d'entrainement et un jeu de validation

```{r}

annee_separation=2018

len_train=7*12

div_cap_mondiale_avant_covid_train=ts(div_cap_mondiale_avant_covid[1:len_train], start = c(2011, 1), end = c(annee_separation-1,12), frequency = 12)

div_cap_mondiale_avant_covid_valid=ts(div_cap_mondiale_avant_covid[len_train:length(cap_mondiale_avant_covid)], start = c(annee_separation, 1), end = c(2019,12), frequency=12)

len_valid=length(cap_mondiale_avant_covid_valid)

```

Passage au logarithme

```{r}

log_div_cap_mondiale=log(div_cap_mondiale)

log_div_cap_mondiale_avant_covid = window(log_div_cap_mondiale, start = c(2011, 1), end = c(2019,12)) #On sélectionne les données avant février 2020

log_div_cap_mondiale_avant_covid_train=ts(log_div_cap_mondiale_avant_covid[1:len_train], start = c(2011, 1), end = c(annee_separation-1,12), frequency = 12)

log_div_cap_mondiale_avant_covid_valid=ts(log_div_cap_mondiale_avant_covid[len_train:length(log_div_cap_mondiale_avant_covid)], start = c(annee_separation, 1), end = c(2019,12), frequency=12)

```

Modèle linéaire :

```{r}

modele_reg1 = tslm(log_div_cap_mondiale_avant_covid_train ~ -1 +trend+season)

summary(modele_reg1)

```

```{r}

#Analyse des résidus studentisés

````
ptAb=PlotResStudent(modele_reg1)

qqnorm(rstudent(modele_reg1))

abline(a=0,b=1)

mean(modele_reg1$residuals)

```
````

Validation des prédictions

````
```{r}

validation=forecast(modele_reg1, h=len_valid, level=0.95)

validation$mean=exp(validation$mean)*div

validation$lower=exp(validation$lower)*div

validation$upper=exp(validation$upper)*div


recompose_train=exp(modele_reg1$fitted.values)*div


err_lin=sum((as.numeric(validation$mean)-cap_mondiale[seq(len_train+1, len_train+len_valid)])^2) #calcul de la somme des erreur s au carré.


ggplot()+

  geom_line(aes(x=time(log_div_cap_mondiale_avant_covid), y=exp(log_div_cap_mondiale_avant_covid)*div, color="black"))+

  geom_line(aes(x=time(log_div_cap_mondiale_avant_covid_train), y=recompose_train, color="blue"))+

  geom_line(aes(x=time(log_div_cap_mondiale_avant_covid_valid), y=validation$mean, color='red'))+

  geom_ribbon(aes(x=time(log_div_cap_mondiale_avant_covid_valid), ymin=validation$lower, ymax=validation$upper), alpha=0.2)+

  labs(x="temps", y="ASK")+

  ggtitle("Entrainement et validation avec un modèle linéaire et sans crise sanitaire")+

  scale_color_manual(name=NULL,

                     values=c("black","blue", "red"),

                     labels=c("courbe ASK avant covid","courbe entrainement", "courbe validation" ))+

  theme(legend.position="bottom")


```
````
````

Prédiction sur la régression linéaire

```{r}

nb_annee_pred=6

pred_reg1=forecast(modele_reg1, h=len_valid+nb_annee_pred*12, level=0.95)

pred_reg1$mean=exp(pred_reg1$mean)*div

pred_reg1$lower=exp(pred_reg1$lower)*div

pred_reg1$upper=exp(pred_reg1$upper)*div

time_pred_sans_covid=time(ts(pred_reg1$mean, start=c(annee_separation,1), end=c(2025,12), frequency=12))


ggplot()+

  geom_line(aes(x=time(cap_mondiale_avant_covid), y=cap_mondiale_avant_covid, color="black"))+

  geom_line(aes(x=time(cap_mondiale_avant_covid_train), y=recompose_train, color="blue"))+

  geom_line(aes(x=time_pred_sans_covid, y=pred_reg1$mean, color="red"))+

  geom_ribbon(aes(x=time_pred_sans_covid, ymin=pred_reg1$lower, ymax=pred_reg1$upper), alpha=0.2)+

  labs(x="temps", y="ASK")+

  ggtitle("Prédiction de l'ASK avec le modèle linéaire et sans crise sanitaire")+

  scale_color_manual(name=NULL,

                     values=c("black","blue", "red"),

                     labels=c("courbe ASK avant covid","courbe entrainement", "prédiction" ))+

  theme(legend.position="bottom")



```

Modèle de HoltWinters

```{r}


model_H=HoltWinters(log_div_cap_mondiale_avant_covid_train, seasonal = "additive")


validation_H=forecast(model_H, h=len_valid, level=0.95)

validation_H$mean=exp(validation_H$mean)*div

validation_H$lower=exp(validation_H$lower)*div

validation_H$upper=exp(validation_H$upper)*div
```

````
err_H=sum((as.numeric(validation_H$mean)-cap_mondiale[seq(len_train+1, len_train+len_valid)])^2)

recons_H=exp(model_H$fitted[,1])*div

```


```{r}

length(time(cap_mondiale_avant_covid_train)[-(1:12)])

ggplot()+

  geom_line(aes(x=time(cap_mondiale_avant_covid), y=cap_mondiale_avant_covid, color="black"))+

  geom_line(aes(x=time(cap_mondiale_avant_covid_train)[-(1:12)], y=recons_H,color="blue") )+

  geom_line(aes(x=time(cap_mondiale_avant_covid_valid), y=validation_H$mean,color="red"))+

  geom_ribbon(aes(x=time(cap_mondiale_avant_covid_valid), ymin=validation_H$lower,
ymax=validation_H$upper), alpha=0.2)+

  labs(x="temps", y="ASK", fill="")+

  ggtitle("Entrainement et validation avec un lissage de Holtwinters")+

  scale_color_manual(name=NULL,

                     values=c("black","blue", "red"),

                     labels=c("courbe ASK avant covid","courbe entrainement", "courbe validation" ))+

  theme(legend.position="bottom")


```
````

Etude des résidus du lissage de Holtwinters

````
```{r}

plot_residuals(model_H,log_div_cap_mondiale_avant_covid_train )

```
````

Prédiction avec le lissage de HoltWinters :

````
```{r}

nb_annee_pred=6

prediction_H=forecast(model_H, h=nb_annee_pred*12+len_valid, level=0.95)

prediction_H

prediction_H$mean=exp(prediction_H$mean)*div


prediction_H$lower=ts(exp(prediction_H$lower)*div, start=c(2018,01), end=c(2025,12), frequency=12)

prediction_H$upper=ts(exp(prediction_H$upper)*div, start=c(2018,01), end=c(2025,12), frequency=12)
````

```
length(prediction_H$mean)

prediction_H$upper

ggplot()+

  geom_line(aes(x=time(cap_mondiale_avant_covid), y=cap_mondiale_avant_covid, color="black"))+

  geom_line(aes(x=time(cap_mondiale_avant_covid_train)[-(1:12)], y=recons_H, color="blue"))+

  geom_line(aes(x=time_pred_sans_covid, y=prediction_H$mean, color="red"))+

  geom_ribbon(aes(x=time_pred_sans_covid, ymin=prediction_H$lower, ymax=prediction_H$upper),
alpha=0.2)+

  labs(x="temps", y="ASK", fill="")+

  ggtitle("Prédiction et intervalle de confiance avec le lissage de Holtwinters")+

  scale_color_manual(name=NULL,

                     values=c("black","blue", "red"),

                     labels=c("courbe ASK avant covid","courbe entrainement", "courbe de prédiction" ))+

  theme(legend.position="bottom")
```

Modèle ARIMA

Supression de la tendance et de la saisonnalité et test de dickey-Fuller :

```{r}
detrend_deseason=modele_reg1$residuals #désaisonnalisation

diff_detrend_deseason=diff(detrend_deseason)

n=length(detrend_deseason)

modele=lm(diff_detrend_deseason~-1+detrend_deseason[-n])

summary(modele)

plot(detrend_deseason)
```

La courbe désaisonnalisée et sans tendance est bien stationnaire

```{r}
plor(detrend_deseason)

acf(detrend_deseason)

pacf(detrend_deseason)

modele_arima=auto.arima(detrend_deseason, seasonal=F, ic="aic", d=0)

summary(modele_arima)

mod_arma=arima(detrend_deseason, order=c(1,0,1), include.mean=F)
```

t_stat(mod_arma)

```

On obtient un arma(1,1) avec le modèle désaisonnalisée et sans tendance

### Vérification des hypothèses sur les résidus

```{r}

plot_residuals(mod_arma, detrend_deseason)

```

Comme précédemment, la normalité, le centrage et l'homoscédasticité semble être respectée. Toutefois, la variance est très petite puisque le diagramme qqnorm suit une droite de pente très faible comparée à la première bissectrice

Entrainement et validation du modèle :

```{r}

recons=exp(detrend_deseason-resid(mod_arma)+modele_reg1$fitted.values)*div #reconstruction en ajoutant la tendance et la saisonnalité au modele arma

validation=forecast(modele_arima, h=len_valid, level=0.95) #on prédit le modèle arma

plot(validation)

trend_validation=seq(len_train+1,len_valid+len_train)*modele_reg1$coefficients[1] #On ajoutera la tendance

season_validation=rep(modele_reg1$coefficients[2:13],2) # On ajoutera la saisonnalité

validation$mean=exp(validation$mean+season_validation+trend_validation)*div

validation$lower=exp(validation$lower+season_validation+trend_validation)*div

validation$upper=exp(validation$upper+season_validation+trend_validation)*div


err_arma=sum((as.numeric(validation$mean)-cap_mondiale[seq(len_train+1, len_train+len_valid)])^2)


validation$mean

cap_mondiale[seq(len_train+1, len_train+len_valid)]

plot_predict_av_cov(recons, validation, time(cap_mondiale_avant_covid_valid), "Entrainement et validation avec le modèle arma")

```

Prediction

```{r}

nb_annee_pred=6

pred_arma=forecast(mod_arma, h=len_valid+nb_annee_pred*12, level=0.95)

trend_prediction=seq(len_train+1,len_valid+len_train+nb_annee_pred*12)*modele_reg1$coefficients[1]

season_prediction=rep(modele_reg1$coefficients[2:13],len_valid/12+nb_annee_pred)

```
pred_arma$mean=exp(pred_arma$mean+trend_prediction+season_validation)*div

pred_arma$lower=exp(pred_arma$lower+trend_prediction+season_validation)*div

pred_arma$upper=exp(pred_arma$upper+trend_prediction+season_validation)*div

plot_predict_av_cov(recons, pred_arma, time_pred1, "Prédiction avec le modèle arma")
```

Comparaison des prédictions avec les données sans covid

```{r}
rescale_2021=function(prediction){
  rescale=list()
  n=length(cap_mondiale)
  rescale$mean=window(prediction$mean, start=c(2020,12), end=c(2025,12))
  rescale$mean=rescale$mean-rescale$mean[1]+cap_mondiale[n]
  rescale$upper=window(prediction$upper, start=c(2020,12), end=c(2025,12))
  rescale$upper=rescale$upper-rescale$upper[1]+cap_mondiale[n]
  rescale$lower=window(prediction$lower, start=c(2020,12), end=c(2025,12))
  rescale$lower=rescale$lower-rescale$lower[1]+cap_mondiale[n]
  return(rescale)
}
pred_reg1_rescale=rescale_2021(pred_reg1)
pred_arma_rescale=rescale_2021(pred_arma)
pred_H_rescale=rescale_2021(prediction_H)
time_pred_rescale=time(pred_reg1_rescale$mean)
ggplot()+
  geom_line(aes(x=time(cap_mondiale), y=cap_mondiale, color="black"))+
  geom_line(aes(x=time_pred_rescale,y=pred_reg1_rescale$mean, color="red"))+
  geom_ribbon(aes(x=time_pred_rescale, ymax=pred_reg1_rescale$upper, ymin=pred_reg1_rescale$lower),
alpha=0.4)+


  geom_line(aes(x=time_pred_rescale, y=pred_arma_rescale$mean, color="green"))+
  geom_ribbon(aes(x=time_pred_rescale, ymax=pred_arma_rescale$upper, ymin=pred_arma_rescale$lower),
alpha=0.3)+


  geom_line(aes(x=time_pred_rescale, y=pred_H_rescale$mean, color="blue"))+
```

geom_ribbon(aes(x=time_pred_rescale, ymax=pred_H_rescale$upper, ymin=pred_H_rescale$lower), alpha=0.3)+

ggtitle("Comparaison des prédictions entrainées avec les données d'avant 2020")+

labs(x="temps", y="ASK", fill="")+

scale_color_manual(name=NULL,

values=c("black","red", "green", "blue"),

labels=c("courbe ASK","prédiction reg lin", "prédiction arma", "prediction H-W"))+

theme(legend.position="bottom")

```

Comparaison des erreurs de validation

```{r}

ggplot(data=data.frame(modele=c("reg lin", "H-W", "arma"), erreur_carre=c(err_lin, err_H, err_arma)), aes(x=modele, y=erreur_carre))+

geom_bar(stat="identity")+

ggtitle("Somme des écarts au carré entre la série de validation et la série réelle ")

```

Comparaison de la distribution des résidus

```{r}

resid=data.frame(resid_reg1=modele_reg1$residuals[-seq(1,12)], resid_arma=resid(mod_arma)[-seq(1,12)], resid_H=resid(model_H))

boxplot(resid, main="Distribution des résidus")

```

Prévisions avec les données de 2020

On ne définit pas de jeu de validation car la taille serait trop petite

Modèle linéaire sur les données divisées et logarithmisées

```{r}

modele_reg2 = tslm(log_div_cap_mondiale ~trend-1+season)

summary(modele_reg2)

````
```

```{r}
recompose=exp(modele_reg2$fitted.values)*div
ggplot()+
  geom_line(aes(x=time(cap_mondiale), y=cap_mondiale, color="black"))+
  geom_line(aes(x=time(cap_mondiale), y=recompose, color="blue"))+
  ggtitle("Modélisation de l’ASK sur l'ensemble des données avec une régression linéaire")+
  labs(x="temps", y="ASK", fill="")+
  scale_color_manual(name=NULL,
                     values=c("black", "blue"),
                     labels=c("courbe ASK","prédiction reg lin"))+
  theme(legend.position="bottom")

```
````

Aucun modèle linéaire ne pourra bien s'ajuster car la tendance linéaire ne permet pas de modéliser une chute si importante.

### Prédiction sur le modèle linéaire :

````
```{r}
nb_annee_pred=5
prediction_reg2=forecast(modele_reg2, h=nb_annee_pred*12, level=0.95)
prediction_reg2$mean=exp(prediction_reg2$mean)*div
prediction_reg2$lower=exp(prediction_reg2$lower)*div
prediction_reg2$upper=exp(prediction_reg2$upper)*div
time_pred=time(prediction_reg2$mean)
length(time_pred)
ggplot()+
  geom_line(aes(x=time(cap_mondiale), y=cap_mondiale))+
  geom_line(aes(x=time(cap_mondiale), y=recompose), color="blue")+
 geom_line(aes(x=time_pred, y=prediction_reg2$mean), color="red")+
  geom_ribbon(aes(x=time_pred, ymin=prediction_reg2$lower, ymax=prediction_reg2$upper), alpha=0.2)
ptAb=PlotResStudent(modele_reg2)
truc=qqnorm(rstudent(modele_reg2))
abline(a=0,b=1)
````

hist(rstudent(modele_reg2))

```

## Modèle de Holtwinters

```{r}

modele_H2=HoltWinters(log_div_cap_mondiale)


recons=exp(modele_H2$fitted[,1])*div


nb_annee_pred=5

prediction_H2=forecast(modele_H2, h=nb_annee_pred*12, level=0.95)

time_pred2=time(ts(prediction_H2$mean, start=c(2021,01), end=c(2025,12), frequency = 12))

prediction_H2$mean=exp(prediction_H2$mean)*div

prediction_H2$lower=exp(prediction_H2$lower)*div

prediction_H2$upper=exp(prediction_H2$upper)*div


ggplot()+

  geom_line(aes(x=time(cap_mondiale), y=cap_mondiale), label="courbe réelle")+

  geom_line(aes(x=time(cap_mondiale)[-seq(1,12)], y=recons), color="blue", label="fitted value")+

  geom_line(aes(x=time_pred2, y=prediction_H2$mean), color='red', label="validation predict")+

  geom_ribbon(aes(x=time_pred2, ymin=prediction_H2$lower, ymax=prediction_H2$upper), alpha=0.2)+

  ggtitle("Modèle de Holt-Winters en prenant les données divisées et logarithmisées")+

  labs(x="temps", y="ASK")


plot_residuals(modele_H2, log_div_cap_mondiale)


```

2e modèle de Holtwinters

```{r}

log_cap_mondiale=log(cap_mondiale)

modele_H3=HoltWinters(log_cap_mondiale)

recons=exp(modele_H3$fitted[,1])

nb_annee_pred=5

prediction_H3=forecast(modele_H3, h=nb_annee_pred*12, level=0.95)

```
prediction_H3$mean=exp(prediction_H3$mean)

prediction_H3$lower=exp(prediction_H3$lower)

prediction_H3$upper=exp(prediction_H3$upper)

ggplot()+

  geom_line(aes(x=time(cap_mondiale), y=cap_mondiale), label="courbe réelle")+

  geom_line(aes(x=time(cap_mondiale)[-seq(1,12)], y=recons), color="blue", label="fitted value")+

  geom_line(aes(x=time_pred2, y=prediction_H3$mean), color='red', label="validation predict")+

  geom_ribbon(aes(x=time_pred2, ymin=prediction_H3$lower, ymax=prediction_H3$upper), alpha=0.2)+

  ggtitle("Modèle de Holt-Winters en prenant les données logarithmisées")+

  labs(x="temps", y="ASK")

plot_residuals(modele_H3, log_cap_mondiale)
```

Modèle Arma

```{r}
deseason1=diff(log_div_cap_mondiale, lag=12) #supression de la saisonnalité


diff_deseason1=diff(deseason1)

trend=seq(1,length(diff_deseason1))

n=length(deseason1)

reg=lm(diff_deseason1~trend+deseason1[-n])

summary(reg)

reg2=lm(diff_deseason1~trend+deseason1[-n]-1)

summary(reg2)

reg3=lm(diff_deseason1~deseason1[-n]-1)

summary(reg3)

# Le processus est DS, on travaille sur la série différenciée et désaisonnalisée


modele_arima_tout1=auto.arima(diff_deseason1)

summary(modele_arima_tout1)


```

On obtient un arima(0,0,0) ce qui n'est pas utile puisque la modélisation corespond à une constante égale à 0.

Comme on peut le voir ci-après :

```{r}
pred_arma_tout1=forecast(modele_arima,h=5*12, level = 0.95)
plot(diff_deseason1-resid(modele_arima_tout1))
recons=exp(diff_deseason1-resid(modele_arima_tout1)+deseason1[-1]+log_div_cap_mondiale[-seq(1,13)])*div
ggplot()+
  geom_line(aes(x=time(cap_mondiale), y=cap_mondiale), label="courbe réelle")+
  geom_line(aes(x=time(cap_mondiale)[-seq(1,13)], y=recons), color="blue", label="fitted value")
```

On travail sur les données seulement logarithmisées

```{r}
deseason2=diff(log_cap_mondiale, lag=12)
plot(deseason)


diff_deseason2=diff(deseason2)
trend=seq(1,length(diff_deseason2))
n=length(deseason)
reg=lm(diff_deseason2~trend+deseason2[-n])
summary(reg)
reg2=lm(diff_deseason2~deseason2[-n])
summary(reg2)
reg3=lm(diff_deseason2~deseason2[-n]-1)
summary(reg3)
#Le processus est DS, on travaille sur la série différenciée.


modele_arima2=auto.arima(diff_deseason2)
summary(modele_arima2)
plot(diff_deseason2-resid(modele_arima2)+log_cap_mondiale[-seq(1,13)])


recons2=exp(modele_arima2$fitted+deseason2[-1]+log_cap_mondiale[-seq(1,13)])
ggplot()+
  geom_line(aes(x=time(cap_mondiale), y=cap_mondiale, color="black"))+
```

geom_line(aes(x=time(cap_mondiale)[-seq(1,13)], y=recons2, color="blue"))+

ggtitle("Modélisation du ASK avec un MA1")+

labs(x="temps", y="ASK")+

scale_color_manual(name=NULL,

values=c("black", "blue"),

labels=c("courbe ASK","Modélisation ma1"))+

theme(legend.position="bottom")

```

### Prédiction avec le modèle arima sur le processus stationnarisé

```{r}

nb_annee_pred=5

pred_arma2=forecast(modele_arima2,h=nb_annee_pred*12, level = 0.95)

plot(pred_arma2, main="Prédiction du mdèle arma sur le processus stationnarisée")

```

La prédiction est une constante de 0.

On estime la tendance et la saisonnalité sur les données logarithmisées avant covid.

```{r}

log_cap_mondiale_avant_covid=log(cap_mondiale_avant_covid)

plot(log_cap_mondiale_avant_covid)

reg=tslm(log_cap_mondiale_avant_covid~trend+season-1)

summary(reg)

trend_prediction=seq(1,length(pred_arma2$mean))*reg$coefficients[1]

seq(1,length(pred_arma2$mean))

seasonal_prediction=rep(reg$coefficients[2:13], 5)

```

On ajoute tendance et saisonnalité à la prédiction

```{r}

pred_arma2$mean=ts(exp(pred_arma2$mean+trend_prediction+seasonal_prediction), start=c(2021,01), end=c(2025,12), frequency=12)

time_pred3=time(pred_arma2$mean)

pred_arma2$lower=exp(pred_arma2$lower+trend_prediction+seasonal_prediction)

pred_arma2$upper=exp(pred_arma2$upper+trend_prediction+seasonal_prediction)

```
```

On remet les données au niveau du dernier ajustement du modèle.

```{r}
point=pred_arma2$mean[1]
pred_arma2$mean=pred_arma2$mean-pred_arma2$mean[1]+recons2[length(recons2)]
pred_arma2$lower=pred_arma2$lower-point+recons2[length(recons2)]
pred_arma2$upper=pred_arma2$upper-point+recons2[length(recons2)]
 pred_arma2$upper
ggplot()+
 geom_line(aes(x=time(cap_mondiale), y=cap_mondiale, color="black"))+
 geom_line(aes(x=time(cap_mondiale)[-seq(1,13)], y=recons2, color="blue"))+
 geom_line(aes(x=time_pred3, y=pred_arma2$mean, color="red"))+
 geom_ribbon(aes(x=time_pred3, ymin=pred_arma2$lower, ymax=pred_arma2$upper), alpha=0.2)+
 ggtitle("Prédiction avec le MA1 sur toutes les données")+
 labs(x="temps", y="ASK")+
 scale_color_manual(name=NULL,
          values=c("black", "blue", "red"),
          labels=c("courbe ASK","Modélisation ma1", "prédiction" ))+
 theme(legend.position="bottom")
```

III. Comparaison entre les 2 approches et études des effets de la crise

```{r}
length(pred_arma_rescale$mean)
ggplot()+
 geom_line(aes(x=time(cap_mondiale), y=cap_mondiale, color="black"))+
 geom_line(aes(x=time_pred2, y=prediction_H3$mean, color='red'))+
 geom_ribbon(aes(x=time_pred2, ymin=prediction_H3$lower, ymax=prediction_H3$upper), alpha=0.2)+


 geom_line(aes(x=time_pred_rescale, y=pred_arma_rescale$mean, color="green"))+
 geom_ribbon(aes(x=time_pred_rescale, ymax=pred_arma_rescale$upper, ymin=pred_arma_rescale$lower),
alpha=0.3)+


 geom_line(aes(x=time_pred_rescale, y=pred_H_rescale$mean, color="blue"))+
 geom_ribbon(aes(x=time_pred_rescale, ymax=pred_H_rescale$upper, ymin=pred_H_rescale$lower),
alpha=0.3)+
```

```
  ggtitle("Comparaison des prédictions entrainées avec et sans les données 2020")+
  labs(x="temps", y="ASK", fill="")+
  scale_color_manual(name=NULL,
                     values=c("black","red", "green", "blue"),
                     labels=c("courbe ASK","prédiction H-W \n sans covid", "prédiction arma  \n sans covid", "prediction H-W \n avec covid" ))+
  theme(legend.position="bottom")
```

Impact de la crise sur les prévisions :

```{r}
ggplot()+
  geom_line(aes(x=time(cap_mondiale), y=cap_mondiale, color="black"))+
  geom_line(aes(x=time_pred2, y=prediction_H3$mean, color='red'))+
  geom_ribbon(aes(x=time_pred2, ymin=prediction_H3$lower, ymax=prediction_H3$upper), alpha=0.2)+


  geom_line(aes(x=time_pred_sans_covid, y=prediction_H$mean, color="green"))+
  geom_ribbon(aes(x=time_pred_sans_covid, ymin=prediction_H$lower, ymax=prediction_H$upper), alpha=0.2)+


  geom_line(aes(x=time_pred_sans_covid, y=pred_arma$mean, color='blue'))+
  geom_ribbon(aes(x=time_pred_sans_covid, ymin=pred_arma$lower, ymax=pred_arma$upper), alpha=0.4)+
  labs(x="temps", y="ASK")+
  scale_color_manual(name=NULL,
                     breaks=c("black","red", "green", "blue"),
                     values=c("black","red", "green", "blue"),
                     labels=c("courbe ASK", "prédiction H-W avec covid","prediction H-W sans covid",  "prédiction arma \n sans covid" ))+
  theme(legend.position="bottom")+
  ggtitle("Impact de la crise sanitaire sur les prévisions")
```