

Dossier de recherche en économétrie des séries temporelles :

Prévision du taux d'occupation des hôtels à Los Angeles : saisonnalité et implications pandémiques sur le tourisme

Par

Barthel Valentin

Coroama Georgiana-Ioana

M1 Analyse et politique économique

Parcours DS2E

Année universitaire 2023/2024

Enseignant : M. Jamel Trabelsi

SOMMAIRE

I. INTRODUCTION	1
II. CADRE STATISTIQUE	2
II.1. Données et statistiques descriptives.....	2
.....	2
III. MODELISATION, INFERENCE STATISTIQUE ET ANALYSE DES RESULTATS.....	5
III.1. Modélisation SARIMAx (0,1,2) (1,0,0) [12]	5
III.2 Inférence statistique : tests de stationnarité (DF, Phillips-Perron, KPSS)	8
Test ADF	8
Phillips-Perron	10
KPSS	11
IV. PREVISIONS	14
Cas 1. Persistance de la pandémie sur un horizon de 12 mois, prévision sur 2 ans	14
Cas 2. Persistance de la pandémie sur un horizon de 6 mois, prévisions sur 2 ans	15
.....	17
Cas 3. Persistance de la pandémie sur un horizon de 3 mois, prévision sur 2 ans	17
IV.2 Modèle SARIMA (0,1,1) (0,1,1) [12]	19
V. CONCLUSION.....	23
V.1 Synthèse, implications et recommandations	23
V.2 Limites et extensions	23
REFERENCES BIBLIOGRAPHIQUES	24

I. INTRODUCTION

Au cœur de l'économie nationale, le secteur du tourisme contribue de manière significative à la production économique des Etats-Unis, avec un total estimé à 1500 milliards USD en 2016 selon l'OCDE. Au-delà de cela, il est bien connu que ce secteur fait partie des ceux qui créent le plus d'emplois au niveau national, avec environ 8 millions d'emplois enregistrés en 2016. De plus, les statistiques de l'OCDE révèlent également que les dépenses collectives des touristes internationaux à destination des Etats-Unis pour la même année ont atteint 245 milliards USD, le principal marché d'exportation en termes de contribution étant la Chine.

Dans le contexte de la pandémie de Covid-19, qui a engendré des conséquences catastrophiques sur l'économie au niveau global, les restrictions sanitaires ont fortement déstabilisé l'industrie du tourisme (Škare et al. (2021)), ayant provoqué une chute considérable dans le taux d'occupation des établissements hôteliers à partir de 2020.

En nous appuyant sur cela, notre objectif va être d'estimer un modèle de prévision du taux d'occupation des hôtels pour la ville de Los Angeles, en tenant compte à la fois de la saisonnalité dans les cycles touristiques, ainsi que du changement de régime induit par la crise sanitaire. Plus précisément, nous allons utiliser des données journalières sur la période début janvier 2013- fin octobre 2020 issues d'une analyse comparative plus large sur l'effet de la pandémie sur les revenus de l'industrie hôtelière dans huit grandes villes des Etats-Unis, basée sur l'emploi des techniques complexes de Deep Learning.¹

Nous chercherons à répondre à plusieurs questions telles que *La pandémie induira-t-elle un changement permanent dans le taux d'occupation des hôtels sur un horizon temporel de deux ans ? Pourrait-on s'attendre à un retour à la configuration touristique prépandémique si la pandémie s'était prolongée encore 3, 6 ou bien 12 mois après octobre 2020 ?* Nous allons ainsi pouvoir comparer nos prévisions pour la période novembre 2020- octobre 2022 avec les statistiques réelles enregistrées pendant cette période.

En ce sens, la section suivante comportera un modèle ARIMAx (p, d, q) (P, D, Q) saisonnier, avec l'inclusion d'une variable binaire qui prend la valeur 1 si l'observation est issue de la période pandémique dont le début aux Etats-Unis est officiellement enregistré en janvier 2020 et 0 sinon.

Quant à la littérature existante à ce sujet, il convient de mentionner que les modèles de variation ont été largement employés durant la crise sanitaire pour des prévisions concernant de différents indicateurs pandémiques, tels que le nombre de nouveaux cas de Covid-19 ou le nombre de décès causés par le virus. Fanelli et Piazza (2020) ont développé un modèle ARIMA (p,d,q) pour étudier la tendance épidémiologique dans trois des pays européens les plus affectés (Espagne, Italie et France), en concluant que celui-ci permettait

¹ Données disponibles à l'adresse suivante : <https://paperswithcode.com/dataset/hotel-sales>

de réaliser des prévisions fiables pour la prévalence du virus sur un horizon de quelques jours. Chintalapudi et al. (2020) ont adopté un modèle SARIMA pour faire des prévisions du nombre de cas de Covid-19 en Italie en utilisant des données jusqu'en mars 2020. Les auteurs ont intégré dans leur analyse l'effet des deux mois de confinement en Italie, les prévisions leur ayant permis d'observer une diminution des cas confirmés pendant le confinement.

Nous tenons à mentionner qu'à la différence de ces analyses, notre étude cherchera à prédire le taux d'occupation des hôtels et non l'évolution pandémique en soi, l'inclusion de la variable binaire Pandémie nous permettant de saisir le changement de régime touristique lors des prévisions et de comparer ces dernières en fonction de la durée anticipée de la pandémie en partant de notre échantillon sur la période 2013-2020.

II. CADRE STATISTIQUE

II.1. Données et statistiques descriptives

Liste des packages à installer et à charger

```
packages <- c("readxl", "dplyr", "ggplot2", "lubridate", "fmsb", "quantmod",
"AER",
               "xts", "astsa", "urca", "forecast", "tseries")
```

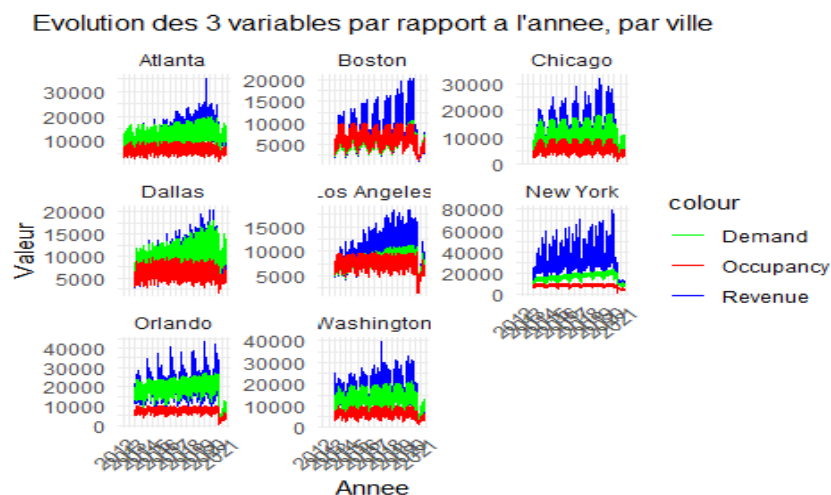
Boucle pour installer les packages s'ils ne sont pas déjà installés et chargement

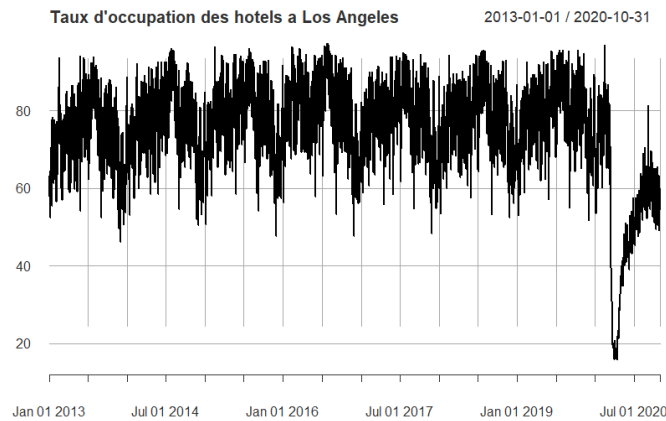
```
for (pkg in packages) {
  if (!require(pkg, character.only = TRUE)) {
    install.packages(pkg)
    library(pkg, character.only = TRUE)
  }
}
```

Message de fin d'installation et de chargement

```
cat("Tous les packages requis sont installés et chargés.\n")
```

```
## Tous les packages requis sont installés et chargés.
```





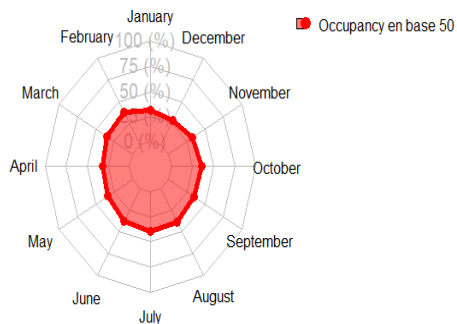
En procédant à l'examen comparatif des 8 graphiques ci-dessus, ainsi que de l'évolution du taux d'occupation des hôtels pour le cas précis de la ville de Los Angeles, plusieurs aspects notables peuvent être mis en avant.

Dans un premier temps, il convient de mentionner que l'ensemble des 8 séries représentées comportent des similitudes en termes d'évolution du taux d'occupation des hôtels sur la période étudiée. Plus précisément, nous pouvons remarquer que ce dernier semble atteindre son pic début juillet toutes les années, en diminuant graduellement vers la fin du mois. Autrement dit, toutes les séries considérées comportent une composante saisonnière annuelle aisément identifiable jusqu'à 2019, ainsi qu'une marche aléatoire correspondant notamment à la période 2019-2020.

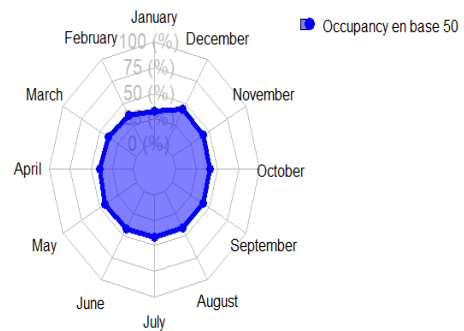
Cela s'avère assez intuitif si on s'appuie sur les cycles touristiques des grandes villes aux Etats-Unis, qui sont fortement influencés par les périodes des vacances et de congés, ainsi que par la fête nationale, célébrée le 4 juillet.

Par ailleurs, nous pouvons saisir que pour chaque ville, une rupture dans la saisonnalité intervient en 2020. En effet, le taux d'occupation des hôtels atteint sa valeur minimale (18%-20%) en juillet 2020, à la suite d'une chute considérable à partir de la fin de 2019. Sans surprise, cette diminution drastique peut s'expliquer par les effets néfastes de la première vague de la pandémie de Covid-19, marquée par un double choc d'offre et de demande. Par conséquent, le confinement, ainsi que la fermeture de nombreux établissements touristiques en raison des difficultés financières ont vraisemblablement mis les bases d'un climat touristique très précaire en 2020.

Comparaisons saisonnières pour Los Angeles



Comparaisons saisonnières pour New York



Ces graphiques radars nous permettent d'observer les différences saisonnières du taux d'occupation des chambres d'hôtels pour l'ensemble des mois de l'année pour les villes de New York et de Los Angeles.

Il n'est pas surprenant de constater que les périodes pour lesquelles le taux d'occupation des hôtels à New York atteint ses plus hauts points est en période d'été et en période de fin d'année, particulièrement durant les mois de juillet et d'août, mais également avec Thanksgiving ou encore les fêtes de Noël et de la Saint Sylvestre. Le taux d'occupation semble être moins élevé durant les périodes hivernales et automnales du fait qu'il ne s'agisse pas de saison touristique.

Contrairement à la ville de New York, la ville de Los Angeles ne semble pas atteindre son pic d'attractivité touristique en période de fin d'année (novembre et décembre) bien que la ville soit tout de même très fréquentée sur la période prolongée de février à novembre, s'expliquant par son climat doux et l'attraction d'événements majeurs d'envergure internationale tels que le « Grammy Awards » en février, l'« Academy Awards » en mars ou encore le « Los Angeles Film Festival » durant l'été.

Ces résultats prennent donc en compte les taux d'occupation moyens pour chaque mois et nous permettent d'observer les tendances saisonnières pendant la période étudiée. Cependant, ces moyennes tendancielle seront tirées vers le bas en raison de la crise Covid qui a fortement impacté le tourisme et la production d'événements.

Dans la suite de notre étude, nous procéderons à l'estimation d'un modèle Arima saisonnier (p, d, q) (P, D, Q) à variable exogène nous permettant de réaliser des prévisions sur le taux d'occupation des hôtels pour la ville de Los Angeles en tenant compte également de la présence de la pandémie.

III. MODELISATION, INFERENCE STATISTIQUE ET ANALYSE DES RESULTATS

III.1. Modélisation SARIMAx (0,1,2) (1,0,0) [12]

Dans l'optique d'assurer la validité de notre modèle, une première étape indispensable consiste à identifier les ordres correspondant au processus autorégressif et au processus de la moyenne mobile de la composante non saisonnière, ainsi que de la composante saisonnière.

Il est également essentiel de vérifier la stationnarité de la série, car modéliser une série non stationnaire conduit à des résultats non fiables. Pour ce faire, nous allons d'abord nous appuyer sur une analyse visuelle des fonctions d'ACP et de PACF, ainsi que de la série considérée, pour procéder dans la section suivante aux tests formels de stationnarité.

#Mensualisation des données journalières pour la ville de Los Angeles, en prenant la valeur moyenne pour chaque mois :

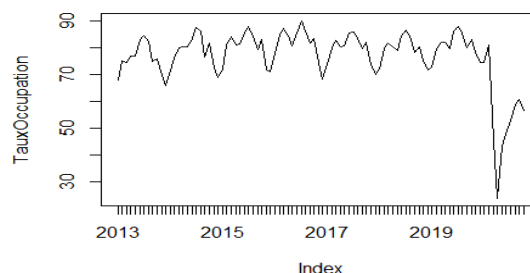
```
library(dplyr)
LosAng<- LA %>% group_by(Year, Month) %>% summarise(TauxOccupation=(mean(Occupancy)))
```

#On transforme Les données en série temporelle :

```
LosAng <- LosAng %>%
  mutate(Date = as.yearmon(paste(Year, Month, sep = "-")))
TauxOccupation<- xts(LosAng$TauxOccupation, order.by = LosAng$Date)

TauxOccupation <- as.zoo(TauxOccupation)
plot.zoo(TauxOccupation)
```

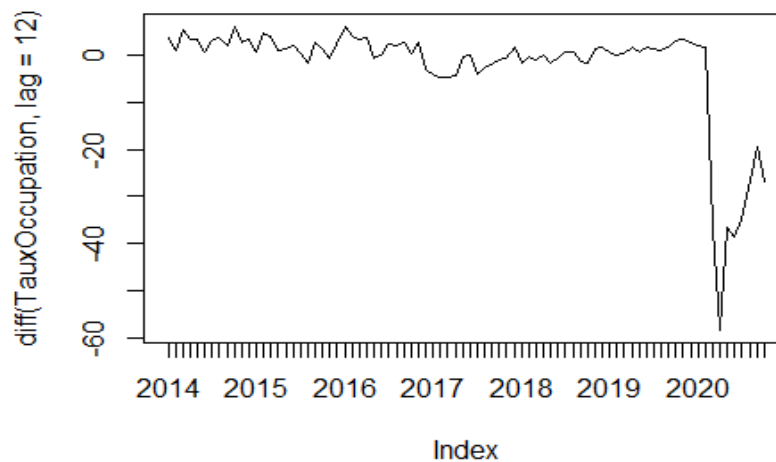
Après avoir transformé nos données pour qu'elles comportent une fréquence mensuelle, l'analyse graphique de l'évolution du taux d'occupation des hôtels à LA rend encore plus claire l'existence d'une saisonnalité à période annuelle dans les données avant le début de la pandémie.



Par conséquent, il s'avère nécessaire de réaliser une première différenciation saisonnière dans le but d'enlever cette composante.

En analysant la série après l'extraction de la composante saisonnière, nous pouvons observer qu'elle n'est pas devenue stationnaire dans son ensemble et que la marche aléatoire persiste notamment sur la période correspondant à la pandémie. Cela n'est pas surprenant, car la différenciation saisonnière ne permet pas d'enlever la tendance stochastique qui rend la série non stationnaire spécifique à la période fin 2019-octobre 2020.

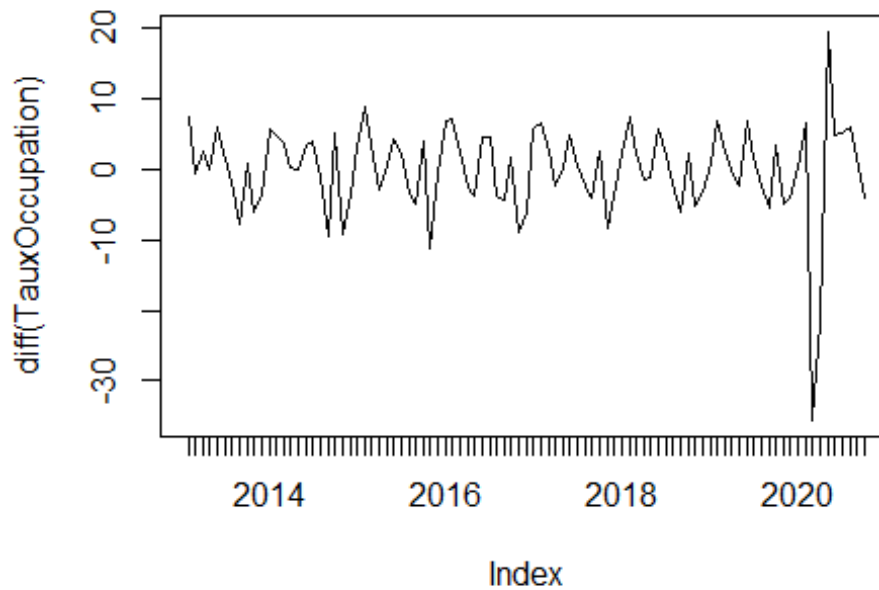
```
plot.zoo(diff(TauxOccupation, lag=12))
```



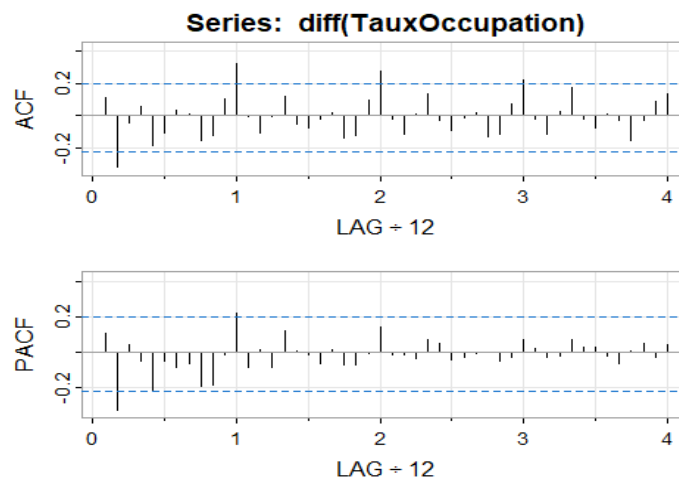
Ainsi, il serait judicieux de procéder à une double différenciation (saisonnière et ordinaire). Néanmoins, étant donné le cadre de notre modélisation, qui comporte une saisonnalité non uniforme sur l'ensemble de la période considérée (il n'y a pas de saisonnalité de la série pour la période pandémique), la double différenciation présente le risque de rendre la série trop volatile, ce qui pourrait conduire à des prévisions non fiables par la suite. De ce fait, nous allons procéder dans la section suivante à la mise en place des tests de stationnarité permettant de vérifier l'existence d'une racine unitaire et de tester également si seule la différenciation ordinaire est suffisante pour rendre la série stationnaire ou si une double différenciation s'avère impérative. Pour cela, nous allons nous baser sur le test de DF augmenté, ainsi que sur les tests de Phillips-Perron et KPSS.

En nous appuyant pour l'instant uniquement sur le graphique de la série différencié avec un filtre (1-L), il apparaît notamment que la série est devenue plutôt stationnaire dans son ensemble.


```
plot.zoo(diff(TauxOccupation))
```



```
acf2(diff(TauxOccupation))
```



De plus, en regardant également la PACF après une différenciation ordinaire avec un filtre $(1-L)$, elle indique bien l'existence d'une corrélation significative avec la variable retardée à la période $t-12$, ainsi qu'une corrélation significative avec le premier lag. Quant à l'ACF, elle

révèle aussi l'existence d'une composante de moyenne mobile saisonnière, ainsi qu'une composante non-saisonnière. Cela semble être en concordance avec la représentation observée au niveau graphique, qui indique que la variation du taux d'occupation des hôtels à LA pour la période prépandémique à l'instant t dépend de la variation de ce même taux au même mois de l'année précédente. Cela nous conduit au choix d'un modèle ARIMA saisonnier avec une intégration ordinaire d'ordre 1. Nous allons nous servir aussi de la fonction `auto.arima` pour l'identification précise des ordres de la composante saisonnière et non saisonnière du processus AR et du processus MA respectivement. Avant de faire cela, nous allons vérifier si les résultats concernant l'ordre d'intégration retenu peuvent être validés à travers les trois tests de stationnarité mentionnés précédemment.

III.2 Inférence statistique : tests de stationnarité (DF, Phillips-Perron, KPSS)

Test ADF

```
library(urca)

#La série non différenciée :

adf<- ur.df(TauxOccupation, type="none", selectlags = "AIC") #La série non di
fférenciée n'est pas stationnaire
summary(adf)

##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -35.797  -2.833   1.128   4.158  21.989
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -0.006075   0.008812  -0.689   0.492
## z.diff.lag    0.110924   0.104078   1.066   0.289
##
## Residual standard error: 6.582 on 90 degrees of freedom
## Multiple R-squared:  0.01709,    Adjusted R-squared:  -0.004747
```

```
## F-statistic: 0.7826 on 2 and 90 DF,  p-value: 0.4603
##
##
## Value of test-statistic is: -0.6894
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

#La série différenciée :

```
adf_test <- ur.df(diff(TauxOccupation), type = "none", selectlags = "AIC")
summary(adf_test)
```

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -36.470  -2.288   0.651   3.368  11.030
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -1.17883    0.13336  -8.840 8.07e-14 ***
## z.diff.lag   0.32672    0.09971   3.277  0.0015 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.267 on 89 degrees of freedom
## Multiple R-squared:  0.5032, Adjusted R-squared:  0.4921
## F-statistic: 45.08 on 2 and 89 DF,  p-value: 3.01e-14
##
##
## Value of test-statistic is: -8.8397
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.6 -1.95 -1.61
```

Les résultats du test ADF ne permettent pas de rejeter l'hypothèse nulle de présence de la racine unitaire dans la série, car la valeur statistique -0.6894 est supérieure à la valeur critique à 1, 5 et 10%. Par conséquent, nous concluons à la non-stationnarité de la série retenue et procédons à sa différenciation ordinaire. Les résultats du test ADF appliqué à la série différenciée permettent de rejeter l'hypothèse de non-stationnarité aux seuils de 1%, 5% et 10%, car -8.84 est inférieur aux valeurs critiques à tous ces seuils et de conclure, par conséquent, à la stationnarité de la série différenciée une fois.

Phillips-Perron

```
library(tseries)

La série non différenciée :

pp.test(TauxOccupation)

##
##  Phillips-Perron Unit Root Test
##
## data:  TauxOccupation
## Dickey-Fuller Z(alpha) = -21.242, Truncation lag parameter = 3, p-value
## = 0.04415
## alternative hypothesis: stationary

La série différenciée :

pp.test(diff(TauxOccupation))

## Warning in pp.test(diff(TauxOccupation)): p-value smaller than printed p-value

##
##  Phillips-Perron Unit Root Test
##
## data:  diff(TauxOccupation)
## Dickey-Fuller Z(alpha) = -69.307, Truncation lag parameter = 3, p-value
## = 0.01
## alternative hypothesis: stationary
```

En regardant la p-value associée au test de Phillips-Perron sur la série non différenciée, il est à noter qu'elle indique une non-stationnarité de la série uniquement au seuil de 1%. Après différenciation, la p-value est inférieure aux seuils de 1, 5 et 10%, ce qui nous permet de rejeter l'hypothèse nulle de présence de racine unitaire dans le cas de la série différenciée avec un filtre (1-L).

KPSS

```

kpss_test <- kpss.test(TauxOccupation)

#La série non différenciée :

print(kpss_test)

##
## KPSS Test for Level Stationarity
##
## data: TauxOccupation
## KPSS Level = 0.46026, Truncation lag parameter = 3, p-value = 0.05118

#La série différenciée :

serie_diff <- diff(TauxOccupation)

kpss_test_diff <- kpss.test(serie_diff)

## Warning in kpss.test(serie_diff): p-value greater than printed p-value

print(kpss_test_diff)

##
## KPSS Test for Level Stationarity
##
## data: serie_diff
## KPSS Level = 0.096014, Truncation lag parameter = 3, p-value = 0.1

```

Dernièrement, en s'appuyant sur le test KPSS, qui à la différence des deux précédents, teste l'hypothèse nulle de stationnarité, il apparaît notamment que nous pouvons rejeter l'hypothèse nulle d'absence de racine unitaire uniquement au seuil de 10%. Quant à la série différenciée, le test permet de ne pas rejeter l'hypothèse nulle de stationnarité aux seuils de 1%, 5% et 10%.

Ainsi, les trois tests présentent des résultats sensiblement différents au niveau de seuils de rejet de l'hypothèse nulle de présence de la racine unitaire (ou de son absence pour le test KPSS), mais ils permettent tous de conclure que la série différenciée est bien stationnaire.

En nous basant sur ce résultat, nous allons retenir comme ordre de différenciation ordinaire $d=1$, sans procéder à une double différenciation avec différenciation saisonnière. ($D=0$)

Par ailleurs, une variété de méthodes qui s'offre à nous peut être mise en place pour l'identifier les ordres p, q et P, Q de la composante non saisonnière et de la composante saisonnière respectivement.

Nous avons identifié une potentielle modélisation en nous basant sur l'analyse graphique des fonctions ACF et PACF de la série différenciée. Pour un aperçu plus précis, nous pouvons également employer un critère d'information tel que le critère d'Akaike ou bien le critère BIC. Comme indiqué précédemment, nous allons privilégier dans notre analyse l'utilisation de la fonction `auto.arima` qui propose le modèle optimal en intégrant les critères mentionnés (minimisation de la valeur du BIC /AIC et ordre de différenciation nécessaire à la stationnarisation)

```
library(forecast)

sarima<- auto.arima(TauxOccupation)
summary(sarima)

## Series: TauxOccupation
## ARIMA(0,1,2)(1,0,0)[12]
##
## Coefficients:
##          ma1      ma2      sar1
##          0.1319 -0.3464  0.6199
## s.e.  0.1068   0.1154  0.1059
##
## sigma^2 = 29.3:  log likelihood = -290.54
## AIC=589.09   AICc=589.54   BIC=599.22
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.3067688 5.296206 2.91814 -1.213321 4.902596 0.5549843
##              ACF1
## Training set 0.01305676

residuals <- residuals(sarima_model)
SSE <- sum(residuals^2)
SST <- sum ((TauxOccupation - mean (TauxOccupation))^2)
R_squared <- 1 - (SSE / SST)
R_squared

## 0.7499256
```

Il est à noter que l'utilisation de la fonction `auto.arima` nous a permis de retrouver le modèle **ARIMA(0,1,2)(1,0,0)[12]**, qui comporte un processus autorégressif saisonnier d'ordre 1, une différenciation ordinaire et une composante non saisonnière de moyenne mobile d'ordre 2.

En regardant les coefficients estimés, nous pouvons remarquer que la variation du terme saisonnier annuel (t-12) induit un effet positif significatif statistiquement sur le taux d'occupation des hôtels à LA à l'instant t. Celui-ci ne peut pas être perçu comme un multiplicateur, mais il peut s'interpréter comme l'effet d'une impulsion saisonnière appliquée 12 mois avant sur le taux d'occupation des hôtels à l'instant t. Autrement dit, une impulsion saisonnière de 1 point de % en t-12 induirait une hausse de 0.62 point de % du taux d'occupation des hôtels à l'instant t, toutes choses égales par ailleurs. De plus, nous pouvons observer que la qualité d'ajustement de notre modèle est élevée (75%), ce qui implique que le terme saisonnier contribue à bien prédire le taux d'occupation des hôtels en t.

Dans l'optique de tenir compte également de l'effet de la pandémie, nous allons procéder par la suite à l'inclusion d'une variable binaire prenant la valeur 1 pour toutes les observations issues à partir de janvier 2020. Notre spécification ajustée prend donc la forme suivante :

$$\Delta TauxOccupation_t = \varphi_1 \Delta TauxOccupation_{t-12} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \beta Pandemie_t$$

Le tableau ci-dessous présente les résultats des estimations obtenus :

```
LosAng <- LosAng %>%
  mutate(Pandemie = ifelse(Year == 2020, 1, 0))

library(astsa)
sarima_model <- arima(TauxOccupation, order=c(0,1,2), seasonal=list(order=c(1
,0,0), period=12), xreg=LosAng$Pandemie)
summary(sarima_model)

##
## Call:
## arima(x = TauxOccupation, order = c(0, 1, 2), seasonal = list(order = c(1,
##      0), period = 12), xreg = LosAng$Pandemie)
##
## Coefficients:
##          ma1          ma2          sar1  LosAng$Pandemie
##      -0.2928   -0.7072    0.6044         -24.2290
## s.e.    0.0726    0.0682    0.1069          2.5614
##
## sigma^2 estimated as 23.78:  log likelihood = -283.36,  aic = 576.72
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
ACF1
## Training set 0.596411 4.850885 2.96799 -0.2445276 4.790805 0.6488437 0.102
2506
```

L'analyse des estimations après avoir contrôlé de l'effet de la pandémie met en avant un aspect important. Si l'effet d'une impulsion saisonnière appliquée en t-12 ne change pas significativement, l'effet négatif de la pandémie lui est très fort. Plus précisément, la pandémie induit une diminution de 24 points de pourcentage du taux moyen d'occupation

des hôtels à LA, ce qui traduit une forte perturbation de la saisonnalité manifestée en absence de la pandémie. Il est à mentionner aussi que tous les coefficients sont statistiquement significatifs.

IV. PREVISIONS

Dans cette section de notre étude, nous allons procéder à l'utilisation de ce modèle pour la réalisation des prévisions sur de différents horizons temporels, tout en intégrant la présence de la pandémie avec plusieurs scénarios possibles.

Pour ce faire, nous avons créé un vecteur d'une longueur égale aux nombres de périodes futures pour lesquelles nous anticipons que l'effet de la pandémie persisterait à partir de la dernière observation (octobre 2020) et nous avons ajusté les prévisions correspondant à ces périodes en soustrayant de leur valeur moyenne l'effet négatif lié à la pandémie. Les graphiques suivants illustrent les prévisions selon les différents scénarios possibles.

Cas 1. Persistance de la pandémie sur un horizon de 12 mois, prévision sur 2 ans

```
library(forecast)

forecast_result <- forecast(sarima_model, h = 24)

effet_pandemie <- rep(0, length(forecast_result$mean))
effet_pandemie[1:12] <- -24.23

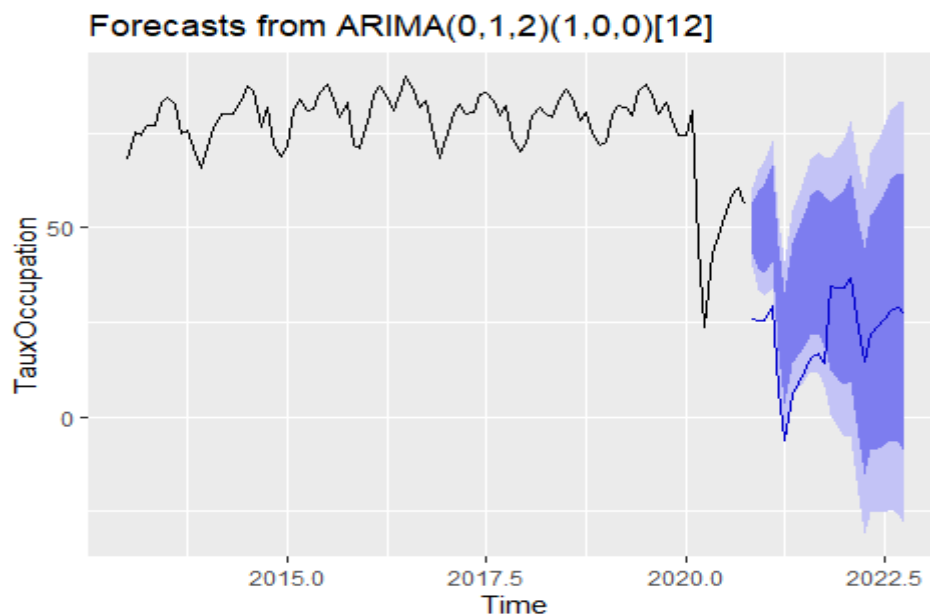
forecast_result_with_pandemic <- forecast_result
forecast_result_with_pandemic$mean <- forecast_result_with_pandemic$mean + ef
fet_pandemie

print(forecast_result_with_pandemic)
```

##		Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
##	Nov 2020	25.833063	43.239315	56.88681	39.6270410	60.49909
##	Dec 2020	25.361969	39.285468	59.89847	33.8295366	65.35440
##	Jan 2021	25.475691	38.088584	61.32280	31.9388595	67.47252
##	Feb 2021	29.507636	40.943481	66.53179	34.1706649	73.30461
##	Mar 2021	7.458261	17.816578	45.55994	10.4733519	52.90317
##	Apr 2021	-6.060139	3.298518	33.04120	-4.5738948	40.91362
##	May 2021	5.982487	14.404576	46.02040	6.0363739	54.38860
##	Jun 2021	8.881892	16.419880	49.80390	7.5836628	58.64012
##	Jul 2021	12.014258	18.712673	53.77584	9.4320130	63.05650
##	Aug 2021	15.688612	21.585863	58.25136	11.8810921	67.95613
##	Sep 2021	16.785549	21.915211	60.11589	11.8041030	70.22699
##	Oct 2021	14.239855	18.631606	58.30810	8.1298717	68.80984
##	Nov 2021	34.515090	12.480263	56.54992	0.8157307	68.21445
##	Dec 2021	34.223051	9.963471	58.48263	-2.8787752	71.32488


```
## Jan 2022      34.293549    8.526771 60.06033   -5.1133377 73.70044
## Feb 2022      36.793016    9.602457 63.98358   -4.7913554 78.37739
## Mar 2022      23.124255   -5.419153 51.66766  -20.5291205 66.77763
## Apr 2022      14.743982  -15.090994 44.57896  -30.8846766 60.37264
## May 2022      22.209399   -8.863506 53.28230  -25.3125083 69.73131
## Jun 2022      24.006787   -8.256584 56.27016  -25.3357803 73.34935
## Jul 2022      25.948591   -7.462855 59.36004  -25.1498057 77.04699
## Aug 2022      28.226382   -6.294979 62.74774  -24.5694832 81.02225
## Sep 2022      28.906391   -6.690294 64.50308  -25.5340412 83.34682
## Oct 2022      27.328274   -9.312190 63.96874  -28.7084802 83.36503
```

```
autoplot(forecast_result_with_pandemic)
```



Cas 2. Persistance de la pandémie sur un horizon de 6 mois, prévisions sur 2 ans

```
forecast_result <- forecast(sarima_model, h = 24)

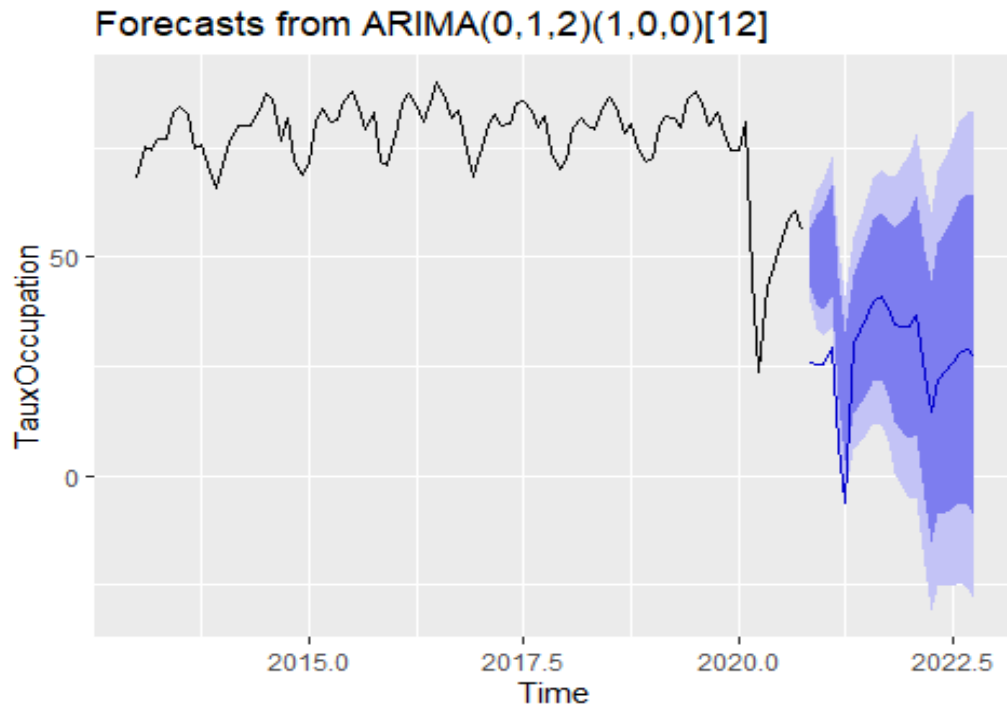
effet_pandemie <- rep(0, length(forecast_result$mean))
effet_pandemie[1:6] <- -24.23

forecast_result_with_pandemic <- forecast_result
forecast_result_with_pandemic$mean <- forecast_result$mean + effet_pandemie

print(forecast_result_with_pandemic)
```

##		Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## Nov 2020		25.833063	43.239315	56.88681	39.6270410	60.49909
## Dec 2020		25.361969	39.285468	59.89847	33.8295366	65.35440
## Jan 2021		25.475691	38.088584	61.32280	31.9388595	67.47252
## Feb 2021		29.507636	40.943481	66.53179	34.1706649	73.30461
## Mar 2021		7.458261	17.816578	45.55994	10.4733519	52.90317
## Apr 2021		-6.060139	3.298518	33.04120	-4.5738948	40.91362
## May 2021		30.212487	14.404576	46.02040	6.0363739	54.38860
## Jun 2021		33.111892	16.419880	49.80390	7.5836628	58.64012
## Jul 2021		36.244258	18.712673	53.77584	9.4320130	63.05650
## Aug 2021		39.918612	21.585863	58.25136	11.8810921	67.95613
## Sep 2021		41.015549	21.915211	60.11589	11.8041030	70.22699
## Oct 2021		38.469855	18.631606	58.30810	8.1298717	68.80984
## Nov 2021		34.515090	12.480263	56.54992	0.8157307	68.21445
## Dec 2021		34.223051	9.963471	58.48263	-2.8787752	71.32488
## Jan 2022		34.293549	8.526771	60.06033	-5.1133377	73.70044
## Feb 2022		36.793016	9.602457	63.98358	-4.7913554	78.37739
## Mar 2022		23.124255	-5.419153	51.66766	-20.5291205	66.77763
## Apr 2022		14.743982	-15.090994	44.57896	-30.8846766	60.37264
## May 2022		22.209399	-8.863506	53.28230	-25.3125083	69.73131
## Jun 2022		24.006787	-8.256584	56.27016	-25.3357803	73.34935
## Jul 2022		25.948591	-7.462855	59.36004	-25.1498057	77.04699
## Aug 2022		28.226382	-6.294979	62.74774	-24.5694832	81.02225
## Sep 2022		28.906391	-6.690294	64.50308	-25.5340412	83.34682
## Oct 2022		27.328274	-9.312190	63.96874	-28.7084802	83.36503

```
autoplot(forecast_result_with_pandemic)
```



Cas 3. Persistance de la pandémie sur un horizon de 3 mois, prévision sur 2 ans

```
forecast_result <- forecast(sarima_model, h = 24)

effet_pandemie <- rep(0, length(forecast_result$mean))
effet_pandemie[1:3] <- -24.23

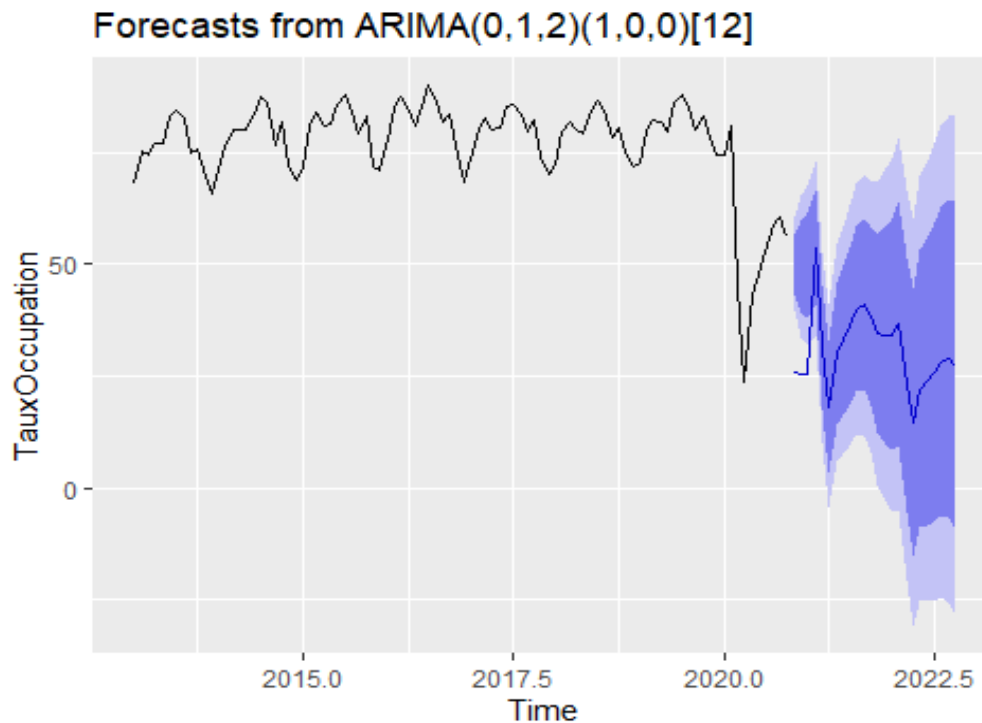
forecast_result_with_pandemic <- forecast_result
forecast_result_with_pandemic$mean <- forecast_result_with_pandemic$mean + ef
fet_pandemie

print(forecast_result_with_pandemic)
```

##	Point	Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## Nov 2020		25.83306	43.239315	56.88681	39.6270410	60.49909
## Dec 2020		25.36197	39.285468	59.89847	33.8295366	65.35440
## Jan 2021		25.47569	38.088584	61.32280	31.9388595	67.47252
## Feb 2021		53.73764	40.943481	66.53179	34.1706649	73.30461
## Mar 2021		31.68826	17.816578	45.55994	10.4733519	52.90317
## Apr 2021		18.16986	3.298518	33.04120	-4.5738948	40.91362
## May 2021		30.21249	14.404576	46.02040	6.0363739	54.38860
## Jun 2021		33.11189	16.419880	49.80390	7.5836628	58.64012
## Jul 2021		36.24426	18.712673	53.77584	9.4320130	63.05650
## Aug 2021		39.91861	21.585863	58.25136	11.8810921	67.95613

## Sep 2021	41.01555	21.915211	60.11589	11.8041030	70.22699
## Oct 2021	38.46986	18.631606	58.30810	8.1298717	68.80984
## Nov 2021	34.51509	12.480263	56.54992	0.8157307	68.21445
## Dec 2021	34.22305	9.963471	58.48263	-2.8787752	71.32488
## Jan 2022	34.29355	8.526771	60.06033	-5.1133377	73.70044
## Feb 2022	36.79302	9.602457	63.98358	-4.7913554	78.37739
## Mar 2022	23.12426	-5.419153	51.66766	-20.5291205	66.77763
## Apr 2022	14.74398	-15.090994	44.57896	-30.8846766	60.37264
## May 2022	22.20940	-8.863506	53.28230	-25.3125083	69.73131
## Jun 2022	24.00679	-8.256584	56.27016	-25.3357803	73.34935
## Jul 2022	25.94859	-7.462855	59.36004	-25.1498057	77.04699
## Aug 2022	28.22638	-6.294979	62.74774	-24.5694832	81.02225
## Sep 2022	28.90639	-6.690294	64.50308	-25.5340412	83.34682
## Oct 2022	27.32827	-9.312190	63.96874	-28.7084802	83.36503

`autoplot(forecast_result_with_pandemic)`



A la lumière de ces trois graphiques, plusieurs aspects notables peuvent être mis en avant. Premièrement, nous pouvons observer que le caractère plus ou moins optimiste des prévisions sur le taux mensuel d'occupation des hôtels à LA est fortement influencé par la durée anticipée du prolongement de la pandémie après octobre 2020. En procédant à l'analyse comparative des 3 graphiques, il apparaît notamment que le scénario le plus optimiste est sans surprise celui qui limite la durée anticipée du prolongement de la crise sanitaire à uniquement trois mois. Dans ce cas, les prévisions montrent une diminution du

taux d'occupation des hôtels à LA vers 25%, avec une reprise importante en février 2021 (53.73%) suivie par une nouvelle baisse, suite à laquelle le niveau prévu jusqu'en octobre 2022 fluctue entre 30 et 40% sans retrouver ses niveaux pré pandémiques.

Quant au deuxième scénario, envisageant une durée anticipée de la pandémie de 6 mois après octobre 2020, il est à noter que les prévisions comportent un caractère beaucoup plus pessimiste. En effet, dans ce cas l'industrie hôtelière serait la plus affectée en mars et avril 2021, avec des niveaux prévus en dessous de 0% et une reprise prévue à partir de mai 2021.

Le scénario le plus néfaste est sans doute le premier, qui fait l'hypothèse que la durée de la pandémie serait prolongée pour encore 12 mois après octobre 2020. Dans ce cas, nous pouvons observer que les prévisions révèlent une dégradation importante du cycle touristique traditionnel, avec de très faibles niveaux prévus et une reprise beaucoup plus lente que dans les deux autres scénarios précédemment mis en avant.

Deuxièmement, un aspect important à mettre en évidence est que pour tout scénario considéré, la pandémie semble opérer une reconfiguration complète de la demande adressée à l'industrie hôtelière sur un horizon de deux ans, les prévisions indiquant une persistance des taux d'occupation réduits même en octobre 2022.

Il serait également intéressant de comparer ces 3 scénarios avec l'évolution prévue du taux mensuel d'occupation des hôtels à LA sous l'hypothèse que la pandémie de Covid-19 n'aurait jamais émergé. Pour ce faire, nous allons soustraire de notre échantillon initial toutes les observations correspondant à la période 2013-2019, et nous allons nous servir à nouveau de la fonction `auto.arima` pour choisir la spécification optimale, comme suit:

IV.2 Modèle SARIMA (0,1,1) (0,1,1) [12]

```
LosAng<- LA %>% group_by(Year, Month) %>% summarise(Tauxoccupation=(mean(Occu
pancy)))

## `summarise()` has grouped output by 'Year'. You can override using the
## `.groups` argument.

LosAng<- LosAng%>% filter(Year!=2020)
LosAng <- LosAng %>%
  mutate(Date = as.yearmon(paste(Year, Month, sep = "-")))
Tauxoccupation<- xts(LosAng$Tauxoccupation, order.by = LosAng$Date)

serie<- auto.arima(Tauxoccupation)
summary(serie)

## Series: Tauxoccupation
## ARIMA(0,1,1)(0,1,1)[12]
##
## Coefficients:
```

```
##          ma1      sma1
##      -0.5340  -0.4689
## s.e.   0.1338   0.1777
##
## sigma^2 = 3.156: log likelihood = -142.18
## AIC=290.36  AICc=290.72  BIC=297.15
##
## Training set error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.06401351 1.610191 1.234627 -0.0954055 1.563588 0.5914087
##              ACF1
## Training set 0.08558148
```

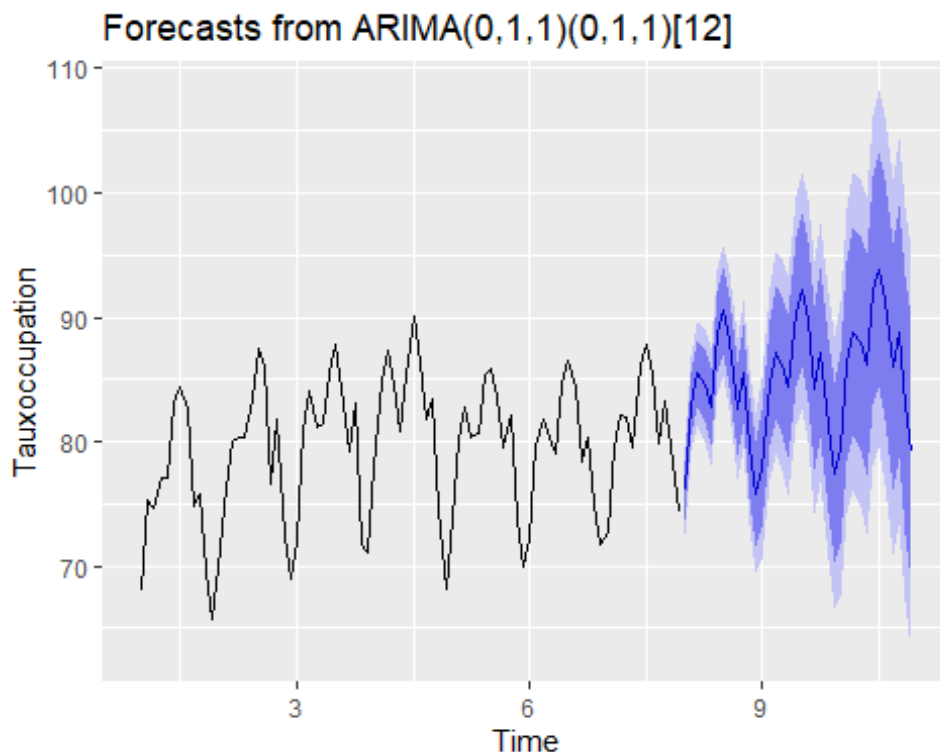
Nous pouvons observer que dans le cas de la série comportant uniquement des observations prépandémiques, une double différenciation (saisonnière et ordinaire) a été nécessaire pour la rendre stationnaire. De plus, il apparaît notamment que pour ce sous-échantillon, le taux d'occupation mensuel des hôtels à LA suit uniquement un processus non-saisonnier de moyenne mobile d'ordre $q=1$ et un processus saisonnier de moyenne mobile d'ordre $Q=1$, et pas de processus autorégressif, comme c'est le cas pour la série allant jusqu'en octobre 2020. Nous allons procéder maintenant à la réalisation des prévisions concernant l'évolution du taux d'occupation des hôtels à LA sur un horizon de 36 mois en utilisant ce modèle.

```
library(forecast)
forecast_result<-forecast(serie, h=36)
summary(forecast_result)

##
## Forecast method: ARIMA(0,1,1)(0,1,1)[12]
##
## Model Information:
## Series: Tauxoccupation
## ARIMA(0,1,1)(0,1,1)[12]
##
## Coefficients:
##          ma1      sma1
##      -0.5340  -0.4689
## s.e.   0.1338   0.1777
##
## sigma^2 = 3.156: log likelihood = -142.18
## AIC=290.36  AICc=290.72  BIC=297.15
##
## Error measures:
##              ME      RMSE      MAE      MPE      MAPE      MASE
## Training set -0.06401351 1.610191 1.234627 -0.0954055 1.563588 0.5914087
##              ACF1
## Training set 0.08558148
##
## Forecasts:
##      Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan  8          76.03863 73.76173 78.31552 72.55641 79.52084
```

## Feb 8	83.01670	80.50476	85.52865	79.17501	86.85839
## Mar 8	85.47240	82.74559	88.19921	81.30211	89.64269
## Apr 8	84.56202	81.63609	87.48796	80.08720	89.03685
## May 8	82.78685	79.67451	85.89920	78.02694	87.54677
## Jun 8	88.76234	85.47413	92.05054	83.73346	93.79121
## Jul 8	90.50414	87.04902	93.95927	85.21999	95.78830
## Aug 8	88.07018	84.45584	91.68453	82.54252	93.59785
## Sep 8	82.61747	78.85063	86.38431	76.85658	88.37836
## Oct 8	85.45687	81.54346	89.37027	79.47184	91.44189
## Nov 8	79.59782	75.54315	83.65248	73.39674	85.79889
## Dec 8	75.80197	71.61080	79.99314	69.39213	82.21181
## Jan 9	77.72558	72.95902	82.49215	70.43575	85.01541
## Feb 9	84.70366	79.66791	89.73941	77.00214	92.40517
## Mar 9	87.15936	81.86810	92.45062	79.06707	95.25164
## Apr 9	86.24898	80.71399	91.78397	77.78395	94.71401
## May 9	84.47381	78.70538	90.24224	75.65176	93.29586
## Jun 9	90.44929	84.45652	96.44207	81.28413	99.61446
## Jul 9	92.19110	85.98207	98.40013	82.69521	101.68699
## Aug 9	89.75714	83.33914	96.17513	79.94166	99.57262
## Sep 9	84.30443	77.68405	90.92480	74.17944	94.42941
## Oct 9	87.14382	80.32708	93.96056	76.71851	97.56913
## Nov 9	81.28477	74.27716	88.29239	70.56756	92.00199
## Dec 9	77.48893	70.29551	84.68234	66.48754	88.49031
## Jan 10	79.41254	71.68107	87.14401	67.58827	91.23680
## Feb 10	86.39061	78.35555	94.42568	74.10204	98.67918
## Mar 10	88.84631	80.51872	97.17391	76.11035	101.58227
## Apr 10	87.93594	79.32574	96.54613	74.76778	101.10409
## May 10	86.16077	77.27695	95.04458	72.57415	99.74738
## Jun 10	92.13625	82.98700	101.28550	78.14368	106.12881
## Jul 10	93.87805	84.47086	103.28525	79.49099	108.26512
## Aug 10	91.44409	81.78583	101.10235	76.67306	106.21512
## Sep 10	85.99138	76.08843	95.89434	70.84612	101.13665
## Oct 10	88.83078	78.68902	98.97253	73.32030	104.34125
## Nov 10	82.97173	72.59667	93.34679	67.10445	98.83901
## Dec 10	79.17588	68.57265	89.77911	62.95965	95.39212

```
autoplot(forecast_result)
```



Comme nous pouvons l'observer en regardant le graphique ci-dessus, en absence de la pandémie, selon les prévisions la demande adressée à l'industrie hôtelière à LA suivrait ses cycles traditionnels, avec des périodes où le taux d'occupation approcherait 95%, correspondant à la forte demande touristique dans les mois de juin et juillet et des fluctuations cycliques entre 75% et 90% les autres mois, en fonction de la saisonnalité. Il est également à noter que les prévisions semblent indiquer une hausse graduelle de ces valeurs d'un cycle à un autre sur un horizon de 36 mois.

En comparant cette situation avec les scénarios intégrant la présence de la pandémie, il devient donc évident que le changement de régime, qui se traduit par le passage à la période pandémique a généré une perturbation notable du cycle touristique enregistré auparavant et de celui prévu en absence de cette dernière. Si sous l'hypothèse que la crise sanitaire n'avait jamais émergé le taux d'occupation prévu pour juillet 2021 aurait été aux alentours de 92%, en présence de celle-ci les prévisions s'avèrent très pessimistes, comme nous avons pu le voir avec les trois scénarios. En effet, selon que le prolongement anticipé de la pandémie après octobre 2020 soit de 3, 6 ou 12 mois, le taux prévu varierait entre 12 et 36% et il ne toucherait pas les niveaux prépandémiques même vers la fin du 2022.

Toutefois, il convient de mentionner que notre modélisation comporte certaines caractéristiques qui nécessitent une approche prudente dans l'interprétation des résultats. Les limites y liées seront présentées dans la section V.2 de la Conclusion.

V. CONCLUSION

V.1 Synthèse, implications et recommandations

Au terme de notre étude, plusieurs aspects essentiels sont à souligner. En effet, nous pouvons constater que faire des prévisions du taux d'occupation des hôtels sur un horizon de deux ans en présence de la pandémie renvoie à une analyse complexe. Dans un premier temps, nous avons estimé un modèle SARIMA $(0,1,2)(1,0,0)$ [12] en nous appuyant sur l'analyse de l'ACF et de la PACF et à l'aide de la fonction `auto.arima`. Les tests de stationnarité de ADF, Phillips-Perron et KPSS nous ont permis de conclure à la non-stationnarité de la série non différenciée et à l'absence de racine unitaire après différenciation, raison pour laquelle nous avons retenu l'ordre d'intégration $d=1$.

Dans le but de tenir compte du changement de régime caractérisant les observations de la période janvier-octobre 2020, nous avons par la suite intégré la variable binaire *Pandémie* et nous avons procédé aux estimations d'un modèle SARIMAX.

La dernière étape de notre analyse a été consacrée à la réalisation des prévisions en suivant trois scénarios possibles portant sur le prolongement anticipé de la crise sanitaire. Cela nous a donné un aperçu plus précis de la façon dont la pandémie produirait une rupture dans la cyclicité touristique annuelle, ainsi que de l'intensité de cette rupture. Plus précisément, nos prévisions pour la période novembre 2020-octobre 2022 ont bien révélé qu'une persistance de la pandémie pour un an après octobre 2020 conduirait à une diminution drastique du taux d'occupation des hôtels avec une reprise très lente. Si nous comparons ces prévisions avec la situation réellement enregistrée durant la pandémie selon les statistiques officielles, nos prévisions comportent quand même un caractère plus pessimiste.

Si nous nous plaçons dans une perspective de recommandations à faire pour amortir l'effet de la crise sanitaire sur le tourisme à LA, en supposant que nos prévisions seraient intervenues en pleine pandémie, en octobre 2020, il serait essentiel que les établissements touristiques adaptent leur offre en procédant à une diversification des marchés (voyages d'affaires essentiels), ainsi qu'à une amélioration de l'expérience client par l'adaptation aux nouvelles normes de santé et sécurité.

V.2 Limites et extensions

Bien que l'utilisation d'une variable binaire indiquant la présence de la pandémie dans notre modèle SARIMA ait l'avantage de pouvoir capturer le changement de régime, ce qui est essentiel pour expliquer la perturbation de la cyclicité enregistrée en absence de la pandémie, il est essentiel de mentionner que ce modèle ne prend pas en compte la dynamique d'ajustement vers le long terme. Comme vu précédemment, cela nous a permis de réaliser des prévisions sur un horizon de court terme à partir de différents scénarios portant sur la durée anticipée de la pandémie. Autrement dit, l'emploi de la variable binaire nous a permis uniquement d'identifier la perturbation de la saisonnalité induite par la

pandémie à chaque instant de prévision, sans pour autant capturer un potentiel échelonnage de l'effet de la pandémie sur le tourisme. Etant donné que l'objectif de notre étude s'articule autour de la prévision du taux mensuel d'occupation des hôtels à LA en prenant en compte à la fois la saisonnalité et la manière dont la présence de la pandémie perturbe le cycle touristique à court terme, le modèle ARIMAx saisonnier employé répond de manière satisfaisante à cet objectif.

Néanmoins, dans l'optique de tenir compte de l'effet graduel de la pandémie sur le tourisme, le passage de notre modèle ARIMAx saisonnier à un modèle à retards échelonnés incluant comme variable explicative le nombre hebdomadaire de décès causés par le Covid-19 pourrait être envisagé. Dans ce cas, l'accent serait mis plutôt sur l'effet total d'une variation de cette variable, et non plus sur la façon dont un changement de régime (présence vs absence de la pandémie) perturbe le processus autorégressif saisonnier, comme vu dans notre modélisation. Cela pourrait faire l'objet d'une étude ultérieure.

REFERENCES BIBLIOGRAPHIQUES

1. Nalini Chintalapudi, Gopi Battineni and Francesco Amenta, "Covid-19 disease outbreak forecasting of registered and recovered cases after sixty day lockdown in Italy: A data driven model approach", *Journal of Microbiology Immunology and Infection*, 2020.
2. Duccio Fanelli and Francesco Piazza, "Analysis and forecast of covid-19 spreading in China, Italy and France", *Chaos Solitons & Fractals*, vol. 134, pp. 109761, 2020
3. Marinko Škare and Domingo Riberio Soriano and Małgorzata Porada-Rochoń "Impact of COVID-19 on the travel and tourism industry", *Technological Forecasting and Social Change*, 2021.
4. Farhangi, A., Huang, A. and Guo, Z., A Novel Deep Learning Model For Hotel Demand and Revenue Prediction amid COVID-19. Proceedings of the 55th Hawaii International Conference on System Sciences (HICSS), 2022