

Résumé

L'objectif principal de ce dossier est d'étudier les prévisions touristiques et la relation potentielle avec la nouvelle variable Google Trends qui résume le comportement des internautes. Cette données mensuelles est une bonne opportunité de pouvoir prédire l'arrivée des tourismes sur le sol Polynésien. L'étude se potera en premier lieu sur la prévision de la fréquentation touristique mensuel entre 2007 et 2018. Dans un second lieu, elle se fera sur la relation de cointégration selon Engle-Granger entre le nombre de touriste arrivé dans le mois et la variable de Google Trends, mais aussi avec d'autre facteur pouvant expliquer les variations de cette fréquentation touristique. La variable google Trends pourrait être un nouvel outils pour les opérateurs du tourisme et optimiser les décisions politiques du gouvernement.

Abstract

The main goal of this paper is to study tourist forecasts and the potential relation with the Google Trends variable which summaries the behaviour of Internet Users. This monthly data is a good opportunity can be to predict arrival of tourism in French Polynesia. This study will focus first on the forecast of monthly tourist (2007-2018). In second time, It will be on the cointegration relation according to Engle-Granger between the number of tourists arriving in the month and the variable of Google Trends, but also with other factors that could explain the variations in the tourist frequentation. Google Trend could be a new tool for business operators in the sector of tourism or for policy makers.

Mots clés : Prévision, ARIMA, FAC, FACP, série temporelle, tourisme, Cointégration, Google Trends, Hôtel.

Key-words: Forecasting, times series, tourism, cointegration, Google Trends, ARIMA, FAC, FACP, hotel.

Sommaire

1. Introduction	4
2. Analyse économique du sujet	5
3. Analyse des variables	9
4. Présentation de la méthodologie ARIMA	14
5. Modèle ARIMA pour	22
6. Test de cointégration selon Engle-Granger	32
7. Conclusion	41
8. Discussion et limites	44
9. Bibliographie	46
10. Annexes	49

1. Introduction

Aujourd'hui, l'ensemble des applications de Google sont devenus des outils indispensables et omniprésents dans le quotidien de chaque internaute. Par la création massive de traces digitales des utilisateurs dans l'historique de recherche, Google récolte ces données par le biais de Google Trends afin de les exploiter. L'application permet d'analyser les tendances de requête faite sur le moteur de recherche qui pourraient être pertinentes pour prévoir le comportement des consommateurs avec un accès libre pour le publique.

Nombreuses études ont été réaliser en s'appropriant ces données de Google Trends, par exemple celui de l'INSEE ¹ C. Bortoli et S. Combes qui ont réaliser un article sur l'analyse de prédiction des consommations des ménages français. Ils concluent que pour certaines catégories de biens Google Trends améliore la prévision d'achat. L'étude de H.Choi et H.Varian² montre l'intérêt des données provenant de google trend sur la prévision de bien (voitures et pièces détachées).

Cette étude est une occasion d'évaluer le lien entre les données provenant de google Trends et le nombre de Tourisme en Polynésie française au cours de la période de janvier 2007 et août 2018. Google Trend pourrait devenir un facteur important sur les décisions politiques ou sur le développement du tourisme en Polynésie Française. Il faudrait comparer avec d'autres variables pour avoir une relation entre les variables explicatives (nombre de touriste) et celle à expliquer dans le cas où les données de google trends ne pourraient pas définir le nombre de touriste. Nous utiliserons des séries comme le nombre de chambre vendus dans le luxe ou hors luxe.

 $^{^1}$: Bortoli C. et Combes S., « Apports de Google Trends pour prévoir la conjoncture française : des pistes limitées », INSEE, mars 2015. Téléchargeable : \underline{link} . (Consulté le 15 janvier 2020)

²: Choi.H et Varian.H « Predicting the Present with Google Trends », ©Google Inc., avril 2009, téléchargeable: https://static.googleusercontent.com/media/www.google.com/fr//googleblogs/pdfs/google_predicting_the_present.pdf (consulté le 15 janvier 2020)

2. Analyse économique du sujet

La prévision touristique était un sujet d'étude pour de nombreux chercheur. Coulomb F.³, réalise des prévision sur les recettes et dépenses touristiques française permettant d'améliorer et expliciter l'ensemble des prévisions sur le commerce extérieur en utilisant les dépenses du pays, recettes en dollars et les recettes en autres devises dans les payse de l'Ocdé (1973-1985). Une étude a été réaliser par Claveria et Torra (2014)⁴ pour fournir les déterminants les plus clair pour fournir l'information la plus précise sur le nombre de touriste par des modèle simple de ARIMA.

La Polynésie française est un territoire d'outre-mer se trouvant dans l'océan Pacifique avec un statut d'autonomie, dont la population élit son propre gouvernement et son président. L'économie du territoire est basé essentiellement sur le secteur du tourisme⁵. En 2013, les touristes ont dépensé 41 milliards de Franc Pacifique (environ 341,6 million d'euros ⁶). Les américains du Nord et les européens sont les touristes les plus répandu en Polynésie française avec une partition de 76% en 2013 contre 23% pour les touriste provenant d'Asie-Pacifique. Depuis l'apparition de vol low-cost sur le territoire avec la compagnie « French Bee » en 2017 ⁷, cela a permis de doper la fréquentation touristique et contribuer au PIB avec une hausse de 29.7% par rapport à l'année précédente.

³François C... "L'économétrie des échanges touristiques de la France," Économie et Prévision, 1988, Programme National Persée, vol. 84(3), pages 41-59.

⁴Claveria F., et Torra S., "Forecasting tourism demand to Catalonia: Neural networks vs. time series models", 2014, *Economic Modelling*, 36, issue C, p. 220-228.

⁵Rapport, « Le tourisme en Polynésie française. En quête d'un second souffle », 2015, Institut d'émission d'outre-mer (IEOM), téléchargeable: https://www.ieom.fr/IMG/pdf/ ne147 eclairage tourisme pf.pdf (consulté le 15 janvier 2020).

⁶ taux de change janvier 2020 : 1€ = 119,52 franc pacifique.

⁷ Rapport, « Polynésie Française », 2018, IEOM, https://www.ieom.fr/IMG/pdf/ra2018 polyne sie f.pdf (consulté le 15 janvier 2020)

Avec l'utilisation de données de google Trends, nous essayerons de trouver une relation avec la variable du nombre de touriste. Cette variable permet d'étudier le comportement des internautes sur le site de recherche Google.

Dans un premier temps dans notre étude, nous ferons une analyse économique sur la description de nos choix des séries. Puis une description graphique de chaque série. Ensuite, nous réaliserons une prévision par la méthode Box-Jenkins avec des modèles ARIMA sur la série du nombre de touriste (touriste). Pour terminer, des test de cointégration selon Engle-Granger pour trouver des relations entre la variable à expliquer Y et les variables explicatives X.

2.1. Définition de la variable à expliquer

La variable à expliquer Y est le nombre de tourisme arrivé durant le mois entre 2007 et 2018 que nous renommons « touriste ». Ce sont des données quantitative provenant des enquêtes de fréquentation touristique sur le territoire mené par l'institut de statistique de la Polynésie française (ISPF)⁸. C'est une variable pertinente pour expliquer le phénomène du tourisme et procéder à une prévision sur la fréquentation. Le nombre de tourisme est un bon indicateur sur la performance du secteur touristique, ainsi plus le nombre de tourisme est élevé et plus l'économie de ce secteur sera bonne.

La prévision de cette variable est un bon moyen aussi d'anticiper les comportement sur la fréquentation touristique par rapport au décision des entreprises locales, aériennes et gouvernementales. Cette méthode pourrait être utiliser aussi sur le vote des budgets et sur les dépenses du territoire dans le secteur du tourisme.

2.2. Définition des variables indépendantes

⁸Données https://www.data.gouv.fr/fr/datasets/tourisme-en-polynesie-francaise/

Google Trends

L'intérêt d'utiliser Google Trends est de suivre les tendances et les comportement des consommateurs (ou touristes) par rapport à leur recherche de requête sur le site. Les données se révèlent être très pertinentes pour prévoir l'arrivée des touristes sur le territoire polynésien et être un indicateur d'intention de s'y rendre aussi. L'exploitation de ces nouvelles données permettent d'élargir les possibilités sur des modèles de prévisions conjoncturelles performantes.

La variable à expliquer représente la tendance de recherche mensuel dans le moteur de recherche Google pour mots clés : « Hotel »+ « tahiti » entre la période du janvier 2007 jusqu'au septembre 2018. Pour faciliter l'interprétation des résultats, l'application réalise une mise en échelle avec des valeurs qui varient entre 0 et 100, dont la valeur 100 représente le record de trafic mensuel dans la période pour une requête de la recherche des mots clés.

Tout d'abord, nous sélectionnons une recherche des tendances en ciblant géographiquement dans le monde ⁹, puis uniquement en France ¹⁰ et en Etats-Unis ¹¹. Ainsi, les choix des pays ont été choisi en se référant aux enquêtes de fréquentation touristique de ISPF ¹² lequel les français et les américains sont les plus nombreux. Nous avons alors trois séries temporelles que nous allons comparer dans la suite de l'étude afin de choisir la série la plus pertinente.

<u>date=2006-01-01%202018-09-01&q=hotel%20Tahiti</u> (consulté le 20 Janvier 2019)

<u>date=2006-01-01%202018-09-01&geo=FR&q=hotel%20Tahiti</u> (consulté le 20 Janvier 2019).

date=2006-01-01%202018-09-01&geo=US&q=hotel%20Tahiti (consulté le 20 Janvier 2019).

⁹ Google Trends: <u>https://trends.google.fr/trends/explore?</u>

¹⁰Google Trends: <u>https://trends.google.fr/trends/explore?</u>

¹¹ Google Trends: https://trends.google.fr/trends/explore?

¹² ISPF: http://www.ispf.pf/docs/default-source/publi-pc-tourisme/eft-2018-05.pdf?sfvrsn=8 (consulté le 20 Janvier 2019)

Nombre de chambre vendu Luxe et hors luxe

Nous choisissons comme variable explicative le nombre de chambre vendu ¹³ dans les hôtellerie de luxe (catégoriser à quatre et cinq étoiles) et puis des chambre vendu hors luxe. C'est une variable explicative approprié pour comprendre la relation entre le nombre de touriste et les chambre vendu dans les deux catégories.

Ainsi, nous ajoutons aussi le ratio entre le nombre de chambre vendu sur le nombre de offert durant le mois, ainsi nous avons un ratio de l'offre et la demande de chambre mensuel afin de comprendre comment elle varie selon la fréquentation touristique. Cela permet aussi de comparer comment les hôtels arrivent à satisfaire les besoins des touristes et comme arrivent-t-ils à vendre leurs chambres disponibles.

D'après le rapport de l'ISPF, le nombre de remplissage des hôtels de luxe est plus importantes que des hôtels intermédiaires. Ainsi, les touristes venant en Polynésie française (PF) ont plutôt une préférence sur le marché de haute gamme ¹⁴. En effet, la plus part des touristes n'ont pas de plafond sur le budget de dépenses, ISPF estime les dépenses d'un touriste de 331 000 F CFP (soit 2750 euros) pour les croisiériste et 235 000 F CFP (soit 1958 euros) pour touriste simple en 2013 ¹⁵.

2.3. Présentation de la base de données

La base de données se présente sous forme d'observation mensuelle entre janvier 2007 et septembre 2018, lors de la prévision du nombre de touriste nous isolerons une observation (septembre 2018) afin de comparer la marge d'erreur entre la valeur prédite et la valeur actuelle.

¹³ Données provenant de l'ISPF : http://www.ispf.pf/Publications.aspx?Collection=Tourisme. (consulté le 20 Janvier 2019).

¹⁴ Rapport ISPF, 2017 : http://www.ispf.pf/docs/default-source/tb-tourisme/tb-2017-t1.pdf? sfvrsn=10 (consulté le 20 Janvier 2019).

¹⁵ Rapport IOEM, 2015 : https://www.ieom.fr/IMG/pdf/ne147_eclairage_tourisme_pf.pdf (consulté le 20 Janvier 2019).

Le nombre de touriste et le nombre de chambre de vendu dans la Polynésie française proviennent des « open data » mise à disposition du publique que nous avons extraite du site de l'ISPF et <u>www.data.gouv.fr</u>.

La variable google trends vient directement du site de google qui aussi libre d'accès.

3. Analyse des variables

Dans cette partir, nous allons analyser les variables les tendances par la description des graphiques de chacune des variables.

3.1. Description de l'évolution de chaque variable et graphique

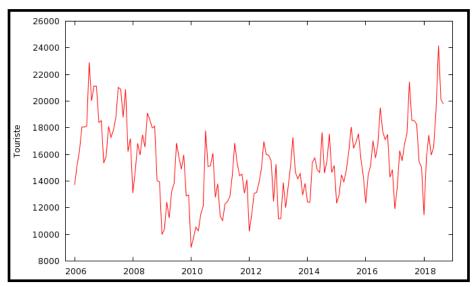


Figure 1 : graphique série temporelle Touriste

Source : données provenant de l'enquête touristique de l'ISPF (2006-2018) https://www.data.gouv.fr/fr/datasets/tourisme-en-polynesie-francaise/

Nombre de touriste : Touriste

Dans la figure 1, le graphique nous montre les variations du nombre de tourisme en Polynésie française entre 2007-2018. Nous constatons des pics et des creux saisonniers pour chaque année. Les pics apparaissent durant les périodes estivales (saison chaude), ce qui explique une fréquentation plus élevées durant cette période. Puis nous avons

une creux durant les saisons froides (automne-hiver), ce qui explique une plus forte fréquentation des touristes provenant de l'hémisphère nord comme les américains du nord ou les européens. En effet, la fréquentation touristique suit les mêmes saisonnalités de cette même hémisphère malgré la stagnation de la température chaude durant toute l'année en Polynésie. Nous constatons aussi une tendance, ainsi dans la période de 2007 à 2010, nous voyons une décroissance sur le nombre de tourisme puis croissance faible entre 2010 et 2018. Dans la même période, l'écart entre le pic du bas et le haut est de plus en plus importante lorsque nous nous rapprochons de la fin de la série. Cette hausse soudaine durant l'année 2017 peut s'expliquer par l'apparition de la nouvelle compagnie aérienne « French Bee » qui encourager la hausse du nombre de touriste lequel la compagnie proposer des prix très compétitif sur le marché.

Google Trend: google, googleFR, googleUSA

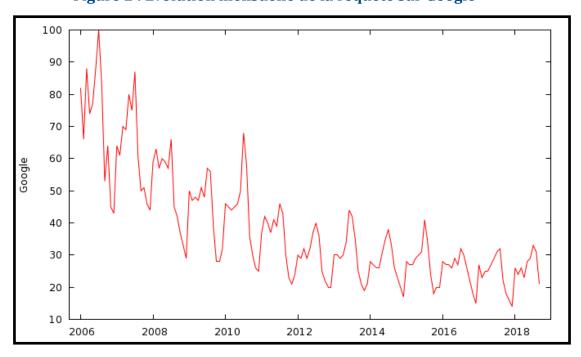


Figure 2 : Evolution mensuelle de la requête sur Google

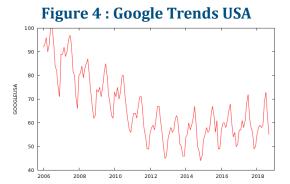
Source : Google Trends : https://trends.google.fr/trends/explore?
date=2006-01-01%202018-09-01&q=hotel%20Tahiti (consulté le 20 Janvier 2019)

- Google :

Dans la figure 1, le graphique montre d'important pic pour le mois de juillet 2007 de valeur 100, qui représente le mois où le pic de recherche de la requête est le plus élevé entre 2007 et 2018. Nous constatons une tendance de décroissance lente durant toute la période et l'écart entre les pics et les creux dans chaque période reste pratiquement la même en se déplaçant au fil du temps.

Nous voulons comparer les deux autre série de google trend sur des requêtes étudier en France et aux Etats-Unis.





Source : Google Trends - https://trends.google.fr/ trends/explore?

date=2006-01-01%202018-09-01&geo=FR&q=hotel%2

Source: Google Trends - https://trends.google.fr/
trends/explore?
date=2006-01-01%202018-09-01&geo=US&a=hotel%2

Nous remarquons une grande différence de variation entre les deux séries. Pour la variable GoogleFR, nous avons une décroissance entre 2007 et 2011et puis une légère croissance sur la recherche de requête dans le reste de la période. Tandis que le googleUSA, nous avons une décroissance assez forte de 2007 à 2013 puis une faible augmentation jusqu'au 2018. Nous trouvons que les trois séries google sont saisonnière à cause de la répétition de pic et de creux dans la même période de l'année.

Chambre d'hôtel vendu : chambre vendu de luxe (VL), chambre vendu hors luxe (VT), ratio de chambre vendu sur chambre offerte de luxe (RL) et ration de chambre vendu sur chambre offerte hors luxe (RT).

Tout d'abord, nous nous intéressons à la série VL, la courbe montre une tendance vers une décroissance dans toute la période. Puis une baisse soudaine durant le premier mois de l'année 2010. Alors que pour la série VT, nous avons tendance qui stagne entre 2007 et 2017. Puis une brusque augmentation durant l'été 2017, qui s'expliquent toujours par l'apparition de la nouvelle compagnie aérienne qui a impact important sur le nombre de touriste en PF sur le variable VT qui représente les chambres vendus hors luxe placer dans des catégories intermédiaires. Par conséquent, la PF par le biais de prix de billet plus attractif attire de nouvelle gamme de clientèle ayant des budget modéré et accède à des chambre d'hôtel moins cher.

En regardant l'échelle du nombre de chambre vendu, pour VL elle varie entre 14000 et 28000 alors que VT, elle ne varie que entre 1000 et 6000 durant la période 2007 et 2018. En effet, la PF reste une destination de haute gamme à cause du nombre élevé de vente chambre de luxe.

Figure 5: graphique temporelle VL

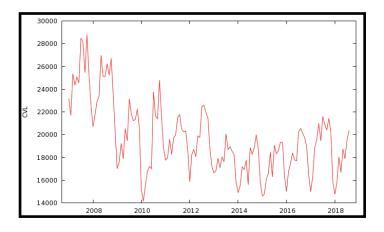
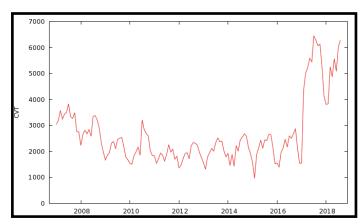


Figure 6: graphique temporelle VT



Source : Gretl, dossier Série temporelle

Source : Gretl, dossier Série temporelle

Par rapport aux séries de RL (en rouge) et RT (en bleu), entre 2007 et 2011, RT domine avec des pics plus élevés que ceux de RL puis il y a des variations inverse entre 2011 et 2018. Ainsi l'écart entre les pics et les creux sont plus importantes pour la variable RT dans toute la période alors que pour la variable RL, l'écart est beaucoup plus réduit. Nous pouvons dire que les hôtels de luxe ont la capacité de vendre la majorité des offres de chambre par rapport aux hôtels intermédiaires.

0.85
0.8 CMRL
CMRT
0.75
0.7 0.65 0.6 0.55 0.5 0.45 0.45 -

0.4

0.35

0.3

2008

2010

Figure 7: série RL et RT entre 2007 et 2018

Source : Gretl, dossier Série temporelle

2014

2016

2018

2012

4. Présentation de la méthodologie ARIMA

La théorie ARIMA est le modèle le plus général pour prévoir une série temporelle qui doit être rendu stationnaire par un degré de différenciation si nécessaire. L'acronyme anglais ARIMA signifie « Auto-Regressive Integrated Moving Average » dont nous pouvons traduire AR par modèle « auto-régressif » , MA pour « moyenne mobile » et I pour le nombre d'intégration.

Pour pouvoir réaliser les prévisions sur notre série Touriste, il est important de trouver un modèle SARIMA selon les paramètre p, d et q dans la partie classique mais P, D et Q dans ce cas d'une série saisonnière. Il est très important de réaliser des vérifications sur les modèles d'ARIMA pour éviter d'avoir des résultats biaisé selon le choix des paramètres et d'avoir une importante marge d'erreur sur le prévision de notre série. La méthode de Box-Jenkins enseigne plusieurs critères de sélection pour avoir un modèle performant sur la prévision basé sur plusieurs hypothèse.

Box, Hilimer et Tiao en 1978¹⁶ démontrent l'importance de cette méthodologie d'ARIMA par Box lui même sur l'analyse des série saisonnières macroéconomique. Les paramètres choisis devaient satisfaire certaines critères pour avoir des prévisions fiables.

Dans la suite de l'étude, nous allons présenter comme rendre une série stationnaire, puis identifier un modèle ARIMA selon la méthode de BOX-Jenkins. Ensuite, nous terminons la présentation de la méthodologie ARIMA avec un modèle multiplicatif, c'est-à-dire que ARIMA présente une parti classique et une autre parti saisonnière.

¹⁶Box P., Hillmer S. Et Tiao G. "Analysis and Modeling of Seasonal Time Series,". Seasonal Analysis of Economic Time Series, édition de Zellner, Arnold. 309U.S. Department of Commerce, pp 301-339, 1978. Téléchargeable: https://www.nber.org/chapters/c4329.pdf (consulté le 09 Fev 2020)

4.1. Stationnarité

Dans cette partie, nous parlerons de la stationnarisation d'une série. Une série stationnaire signifie que la moyenne et la variance d'une série sont constante dans le temps. Dans le cas où une tendance ou une saisonnalité sont présentes dans la série, il faudrait l'éliminer par une différenciation selon un certain nombre de degrés, nous parlons d'intégration pour le paramètre d . Ainsi, nous remplaçons la série brute par une série des différences adjacentes (z_t) de degré 1^{17} :

$$z_t = y_t - y_{t-1}$$
 (1)

Généralement, nous pouvons réaliser une deuxième différenciation (d=2) mais rarement le degrés est supérieur à 2:

$$z_t = y_t - 2y_{t-1} + y_{t+2} (2)$$

En effet, il faudrait vérifier si la série différencié est stationnaire selon les conditions suivantes :

Condition (3)
$$E[y_y] = m$$

$$V[y_t] = \sigma^2$$

$$cov[y_t, y_{t+j}] = \gamma_Y(j)$$

Nous avions dit que la moyenne et la variance devaient être constant au cours du temps et la covariance ne dépend que l'intervalle de temps. Nous refusons la stationnarité de la série si la moyenne de la série varie à court terme ou que la série possède plusieurs variations en cours du temps. Ainsi, il est possible aussi de proposer des changements de série par une transformation logarithmique ou bien la racine carré si la série présente une hétéroscédasticité. Il en est de même pour la variance dont nous refusons la stationnarité si la variance n'est pas constante, si il existe une forte volatilité (variance élevé) pour des observations élevées et faible volatilité dans le cas contraire.

¹⁷Hatanaka M. « Unit Roots and Co-Integrations », Time-Series-Based Econometrics, edition Oxford University press, Chapter 1 : « Difference Equations » pp. 2-7, (consulté le 10 Fev 2020)

Une fois avoir stationnarisé la série, l'étape suivante consiste à analyse la fonction d'autocorrélation et la fonction d'autocorrélation partiel pour confirmé la stationnarisation de la nouvelle série.

4.2. Corrélogrammes

La fonction d'autocorrélation que nous nommons FAC et FACP pour la fonction d'autocorrélation partiel. Les corrélogrammes sont des graphes de la FAC et FACP qui permettent d'identifier visuellement quel paramètre choisir par rapport à p et q qui ne dépassent pas généralement l'ordre de 2 et de voir si les séries sont stationnaires.

La FAC est composé de l'ensemble de l'autocorrélation tel que :

$$\rho_k = corr(z_t, z_{t+k})$$
, avec k qui représente le nombre de décalage. (4)

$$\rho_k = \frac{cov(z_t, z_{t+k})}{\sqrt{V(z_t) \cdot V(z_{t+k})}} \, ^{18}$$

Nous avons l'hypothèse nulle si la corrélation des erreurs est nulle et :

$$t_c = \frac{|r_k|}{\sqrt{1 - r_k^2}} \sqrt{n - 2}$$

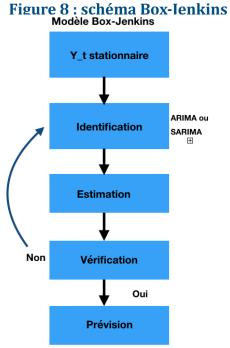
 t_c qui obéit à une loi de Student de degré (n-2), H0 est rejeté si $t_c > t_{n-2}^{1-\alpha/2}$ En effet, la FAC représenta la corrélation d'une observations avec ses valeurs passés tandis que la FACP, elle mesure la corrélation partielle entre z_t et z_{t-h} influençant les autres variables décalées de h périodes $(z_{t-1}, ..., z_{t+h-1})$ ayant été retirée. Nous avons deux matrices importante dont on fait le rapport deux déterminants.

Ainsi, la FAC permettra d'identifier le paramètre q de MA et FACP pour le paramètre p de AR selon la significativité des coefficients d'autocorrélation (FAC et FACP).

¹⁸ Bourbonnais R., Terraza M., « Analyse des séries temporelles, applications à l'économie et à la gestion », Dunod, 2ème édition, 2008, pp. 15-19

4.3. Construction du modèle de Box-Jenkins

Pour réaliser une prévision sur notre série, Box et Jenkins en 1970¹⁹ propose une démarche générale qui est fondé sur le processus ARIMA. il nous faut que la série soit évidemment stationnaire. Tout d'abord, l'approche de Box-Jenkins se fait par plusieurs étapes important avant de réaliser la prévision comme indiquer dans la **figure 8**. Après la stationnarisation de la série Y, nous nous intéressons à l'identification des paramètres ARIMA(p,d,q) ou SARIMA(p,d,q)x(P,D,Q) si nous avons une série saisonnière.



Source : dossier Série Temporelle

Pour comprendre l'identification des paramètres, nous devons décomposer l'ARIMA. Yule (1927) avait supposé que l'observation actuel dépend de l'action du passé (le processus AR) par une processus aléatoire.

¹⁹:Box, George., Box and Jenkins: « Time Series Analysis, Forecasting and Control ». 10.1057/9781137291264_6. Consulté le 10 Fev 2020

Par contre, la partie MA est constitué de combinaison linéaire des valeurs passés des variables aléatoires qui sont les bruits blancs (ϵ_t). Ces bruits blancs doivent répondre aux conditions suivantes :

$$E(\epsilon_t) = 0$$

$$\nu(\epsilon_t) = \sigma^2$$

$$cov(\epsilon_t, \epsilon_t') = 0$$
(5)

D'où:

- AR(p) représente la partie auto-régressif qui construite à partie de l'idée que l'observation au temps t s'expliquer linéairement par les observations précédentes. On a la formule suivante :

$$z_t = \epsilon_t + \sum_{j=1}^p \phi_j z_{t-j}$$
 (6)²⁰, où ϵ_t est le bruit blanc t centré

de variance σ^2 et ϕ_i est le coefficient d'ordre p de AR.

- MA(q) est la moyenne mobile d'ordre q qui est une fonction linéaire où les valeurs de la série stationnaire s'exprime par la combinaison linéaire de l'erreur aléatoire par rapport à q observations précédentes, nous avons le modèle MA(q):

$$z_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \ldots + \theta_a \epsilon_{t-a} (7)^{20}$$
, où

 $\epsilon_j, \ \, \forall \ \, t-q \leq j \leq t \, \text{et} \, \, \theta_q \, \text{est le coefficient d'ordre q}.$

 - I(d) est le nombre de différenciation qui ne dépasse pas généralement l'ordre de 2 pour des séries économiques, autrement cela peut conduire à une non-inversibilité du modèle ARIMA (voir partie estimation).

Pour trouver les paramètres ARIMA(p,d,q), nous devons analyser les corrélogrammes (étape : identification) après avoir stationnarisé la série, nous utilisons la FAC et la FACP qui donnent l'ordre de p et de d. Ainsi, par exemple si nous identifions les deux premiers coefficients de corrélation sont significatives dans la FAC et forment des pics sur le graphique alors nous pouvons proposer comme modèle MA(2). (MA(q))

²⁰Gourieroux C., Monfort A. « Séries temporelles et modèles dynamiques » 2eme edition Economica, chapitre 6 : « Prévision par la méthode Box-Jenkins » pp. 230-238

Et dans le cas où dans la FACP, les premiers coefficients de corrélation présentent des valeurs significatives avec une rupture alors, nous pouvons proposer le même ordre pour le paramètre p de AR(p).

Ensuite, nous nous intéressons à la partie estimation après avoir identifier les paramètres de ARIMA(p,d,q). L'estimation des paramètres est réalisé par la méthode non linéaire d'estimation des maximums de vraisemblance (terme anglais : maximum likehood estimator) d'où nous avons cité les formules (6) et (7). Nous devrons garder les coefficients θ et ϕ qui sont significatives et vérifier si ils respectent les deux conditions de stationnarité et d'invisibilité (**figure 9**) 21 , sinon il faudra changer les paramètres ARIMA.

Figure 9: deux conditions:

Condition de d'invisibilité	Condition de stationnarité
$MA(2) \Rightarrow AR(\infty)$	$ \phi_1 < 1$
$MA(1)$ $ \phi_1 < 1$	AR(2)
$MA(2)$ $\phi_1 + \phi_2 < 1$	$\phi_1 + \phi_2 < 1$ $\phi_2 - \phi_1 < 1$
$\phi_2 - \phi_1 < 1$ $ \phi_2 < 1$	$ \phi_2 < 1$

Source : Dossier Série Temporelle

Puis, nous arrivons à l'étape importante avant la prévision selon la méthode de Box-Jenkins : la vérification du modèle ARIMA consistant à vérifier la pertinence des choix des paramètres. Nous devons se centrer sur l'étude des résidus de la série ARIMA(p,d,q). Toujours à l'aide de la FAC et FACP des résidus, nous observons si le modèle est bien ajusté et que les résidus se comporte comme un processus de bruit blanc. Il faudrait que les auto-corrélations et les auto-corrélations partielles des résidus deviennent non significatives.

²¹ Gourieroux C., Monfort A. « Séries temporelles et modèles dynamiques » 2eme edition Economica, chapitre 6 : « Prévision par la méthode Box-Jenkins » pp. 230-238

Pour vérifier la significativité des coefficients de la FAC et FACP, nous utiliserons un test nommé « test Portemanteau » par le biais de la statistique Q lequel il doit vérifier certaine condition (voir partie : 5.2 Identification, estimation et vérification).

Si nous trouvons plusieurs modèles qui vérifient le test « test Portemanteau » présentant le même niveau d'ajustement à la série mais aussi par rapport au test de Portemanteau. Nous devons appliquer le principe de parcimonie lequel il faut choisir un modèle avec le moins de paramètres possibles. Cependant, si nous fessons face un dilemme par rapport au même nombre de paramètres, il faut faire appelle à d'autres critères comme choisir la plus petite variance parmi les modèles proposés ou selon le critère d'Akaike (AIC) préconisant de sélectionner la valeur la plus faible.

Pour terminer, lorsque toutes les conditions sont réunis pour pouvoir valider le choix du modèle, nous pouvons procéder sur la prévision du modèle en appliquant la combinaison du modèle AR et MA sur la série différencié si nécessaire. En réalisant la prévision, il est possible de comparer les résultats de plusieurs modèles qui ont été vérifier en amont avec le taux d'erreur afin de choisir le modèle le plus pertinent par rapport aux critères de la variance ou celui d'AIC.

4.4. Modèles multiplicatifs

Dans cette partie du modèle multiplicatif, nous nous focalisons sur le mixage d'une partie classique et d'une partie saisonnière, ainsi nous ajoutons de nouveau paramètre (P, D, Q) et nous avons un modèle SARIMA(p, d, q)x(P, D, Q). Le paramètre D représente la différenciation saisonnière : $w_t = y_t - y_{t-12}$, (w_t la nouvelle série)

Pour commencer, il est important d'identifier si une série brute est saisonnière ou pas, elle se présente par l'apparition de pics ou de creux régulier durant chaque période. Par exemple dans notre série y_p , nous remarquons des pics réguliers durant chaque saison estival (printemps - été) et des creux pendant les saisons hivernales. De plus, il est important de stationnarisé avant de commencer à traiter la série en utilisant les méthodes vu précédemment (**4.1 Stationnarité**)

Nous pouvons l'identifier plus facilement à l'aide des corrélogrammes en adaptant le retard selon la saisonnalité si la série est trimestrielle, mensuelle ou autre. La saisonnalité est détectable lors de l'apparition sur les graphiques de coefficient de corrélation qui sont fortement significatives et régulier, par exemple dans pour une série saisonnière annuelle, nous trouverons des pics important au 12ème et 24 ème coefficients dans le graphique de le FAC.

Ensuite, il est important de présenter le choix des paramètres dans la parti saisonnière 22 . Nous sélectionnons les paramètres de la même manière dans la partie classique avec les corrélogrammes de la FAC et FACP dont nous devons prendre une grande attention sur les périodes récurrentes comme dans le cas annuel. Par exemple, si dans la FAC nous trouvons des coefficients de corrélations significatives au 12ème et 13ème, alors nous choisissons $SMA(2)_{s=12}$ avec s=12 si nous avons une série mensuelle.

Après l'identification des paramètres $ARIMA(p,d,q) \times (P,D,Q)_{s=k}$, nous réalisons l'estimation du modèle multiplicatifs, c'est-à-dire la combinaison de la partie classique et saisonnière. Par exemple $SAR(1)_{s=12}$ et $SMA(1)_{s=12}$ pour SARIMA(0,1,0)(1,0,1):

$$\begin{split} a_t &= z_t (1 - \Phi B^{12}) \,, \quad z_t = a_t (1 - \Theta_1 A^{12}) \\ \Rightarrow z_t &= a_t + \Phi_1 z_{t-12} - \Theta_1 a_{t-12} \end{split}$$

Les nouveaux coefficients doivent aussi vérifier les mêmes conditions de stationnarité et d'inversibilités. Puis il faudra procéder les mêmes principes de la méthode de Box-Jenkins pour la vérification du modèle afin de réaliser les prévisions.

 $^{^{22}}$ Bourbonnais R., Terraza M., « Analyse des séries temporelles, applications à l'économie et à la gestion », Dunod, 2ème édition, 2008, pp. 186-192

5. Modèle ARIMA pour Y_t

Dans cette partie, nous réaliserons la méthodologie ARIMA afin de faire une prévision dans la période de janvier 2007 à aout 2018 dont nous gardons une observation pour pouvoir comparer le résultat trouvé avec la prévision.

5.1 Stationnarité de la série Y

Premièrement, nous devons déterminer si notre série est stationnaire ou pas en observant l'évolution dans le temps de la variable Y. Nous nous intéressons à l'espérance mathématique constant de Y qui ne dépend pas du temps, Y est stationnaire si la variable progresse autour de la valeur moyenne avec une fluctuation constante et la variance qui sert d'indicateur de dispersion.

Dans la figure 1, la représentation graphique correspond à une série qui n'est pas stationnaire, l'allure de la série est décroissante entre 2007 et 2010 puis croissante entre 2010 et 2018. De plus nous remarquons aussi que la série est saisonnière. La saisonnalité se montre par des pics et des creux pour la même saison d'année en année avec une concentration plus forte durant les périodes estivales et plus faible en automne-hiver, ce qui montre une attirance touristique plus grande durant les saisons chaudes.

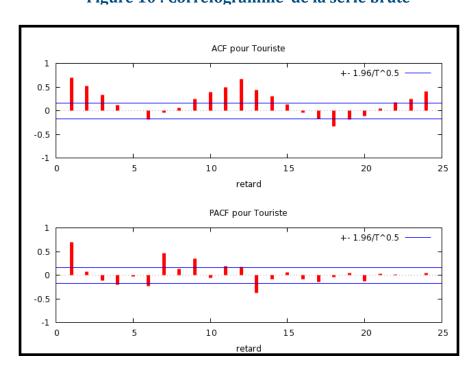
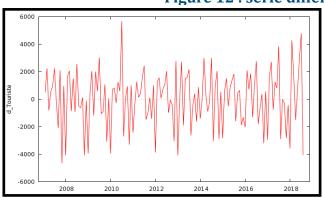


Figure 10 : Corrélogramme de la série brute

Page 22 sur 57

Pour confirmer que la série n'est pas stationnaire, il est nécessaire de réaliser des corrélogrammes de la série Y par FAC et FACP. Dans **la figure 10** pour la FAC, nous distinguons une variation périodique des retards, c'est-à-dire une diminution au cinquième retard et une augmentation jusqu'au douzième. Le mouvement se répète afin d'atteindre la 24ème retard, ce qui confirme que la série est saisonnière à l'aide du corrélogramme avec l'apparition de pics important de retard à 12 et 24 ème retard. En effet, les deux premiers retards et les pics cités précédemment sont fortement significatives et dans la parti FACP, nous avons le premier et le 12 ème retard qui sont aussi fortement significatifs. A partir des résultat, nous pouvons conclure que la série n'est pas stationnaire et il est primordial de procéder à une transformation de la série Y pour la rendre stationnaire.

Nous réalisons une première différenciation sur Y. Dans la figure 12, nous remarquons que la nouvelle série évolue autour d'une moyenne mobile, ce qui nous mène à penser que la série est stationnaire. Nous nous devons aussi de réaliser un corrélogramme pour confirmer la stationnarité. Le graphique de la FAC montre que le premier retard est significatif négativement, puis le 12 ème et 24 ème retards sont aussi significatives qui montrent bien que la série possède une partie saisonnière. En outre, nous ne retrouvons plus les mêmes problèmes de variation que dans la série brute. D'après le FACP, nous avons le premier retard qui est faiblement significative par rapport à la série brute. Ces résultats nous indiquent que la série est bien stationnaire avec une différenciation. (d=1)



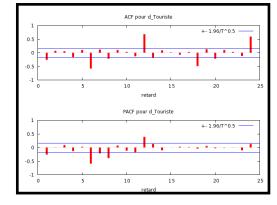


Figure 12 : série différencié et corrélogramme

5.2 Identification, estimation et vérification

A. Identification:

Tout d'abord, nous devons déterminer les paramètres de notre modèle ARIMA pour poursuivre l'étude à l'aide du corrélogramme de la première différenciation de Y. Nous avions constaté avec FACP (figure 12) un décalage du premier retard qui était significative et négative, ce qui nous pousse à choisir un modèle AR avec p max =1 à cause de la premier rupture dans FACP. Ayant une partie saisonnière aussi, FACP montre des pics périodique au 12 ème et 24 ème retard qui sont fortement significatives donc nous choisissons une partie saisonnière SAR avec P max =1. De même dans le graphique FAC indiquant le même comportement des retards que la FACP, ainsi nous avons alors une partie MA avec q max = 1 et une partie saisonnière Q max = 1 avec un retard.

Nous proposons comme modèle après identification : $ARIMA(1,1,1) \times (1,0,1)_{s=12}$ avec des valeurs maximum vu précédemment. Ainsi, nous pourrons proposer plusieurs modèles afin de les comparer et de choisir le meilleur modèle selon plusieurs critères, mais les paramètres ne doivent pas être supérieurs aux valeurs choisies après l'identification (valeur max). De plus, nous utiliserons la règles de Parcimonie pour réduire le nombre de paramètre pour estimer un meilleur modèle.

B. Estimations:

Dans cette partie consiste à réaliser l'estimation du modèle après identification des paramètres de ARIMA, nous utiliserons GRETL sur la série Y après une différenciation de premier degrés. Dans la figure 13 (dans l'annexe p.49) montre la première estimation du modèle $ARIMA(1,1,1)\times (1,0,1)_{s=12}$ et nous constatons que l'estimation à un coefficient constant non significative donc nous décidons de le ré-estimer en retirant celui-ci et nous obtenons la nouvelle estimation dans la figure 14. En regardant les coefficients AR de la partie classique et saisonnière, les deux coefficient respectent la condition de stationnarité car ils sont inférieurs à 1, mais Φ_1 est proche de 1 (Φ_1 =0.97). De même pour la condition d'invisibilité pour les coefficients de MA qui sont respectés car θ_1 et Θ_1 sont inférieurs à 1.

Figure 14: Estimation modèle

```
Evaluations de la fonction :
Évaluations du gradient : 20
Modèle 5: ARMA, utilisant les observations 2007:02-2018:08 (T = 139)
Estimated using AS 197 (MV exacte)
Variable dépendante: d Touriste
Écarts type basés sur la matrice hessienne
                                                 p. critique
            coefficient erreur std.
  phi 1
              0.410746 0.150196
                                        2.735
                                                  0.0062
                                                             ***
                                                             ***
  Phi 1
              0.976388
                           0.0149793
                                        65.18
                                                  0.0000
             -0.782844
                           0.105171
                                                  9.80e-14
                                        -7.444
  theta_1
                                                             ***
  Theta_1
             -0.568178
                           0.118923
                                        -4.778
                                                  1.77e-06
  Moy. var. dép.
                            34.21583
  Éc. type var. dép.
                            2033.741
  Moyenne des innovations
                            127.1472
  Ec. type des innovations
                            979.7617
                            0.772336
  R2
  R2 ajusté
                            0.767277
  Log de vraisemblance
                           -1165.742
  Critère d'Akaike
                            2341.485
  Critère de Schwarz
                            2356.157
  Hannan-Ouinn
                            2347.447
```

Source : Gretl, dossier Série

Ensuite, nous nous intéressons au corrélogramme des résidus pour confirmer l'indépendance entre les résidus par rapport à la covariance et de vérifier l'hypothèse du bruit blanc si elle est respectée. La **figure 15** indique le corrélogramme des résidus de ARIMA, nous remarquons que le 20 ème retard est significative à 5% car sa valeur dépasse le seuil de la borne pour FAC et FACP. Autrement, il faudrait être attentif par rapport aux deux premiers retard (k=1 et k=2) qui sont très importantes pour affirmer la validation du modèle. Ainsi, les deux premiers retards sont éloignés du seuil de significativité, nous pouvons accepter l'hypothèse de covariance nulle des résidus.

Puis, grâce au corrélogramme qui indique le Q_stat des résidus permettant de voir si les résidus de la série sont indépendant et identiquement distribués (iid). Nous pouvons utiliser le test Portemanteau qui évalue la corrélation existant entre les résidus avec pour hypothèse nulle (H0) si les résidus sont idd et il faut que la p-value soit supérieur à 0.05 et que la valeur de Q soit croissante quand le retard augmente pour accepter H0. La **figure 15** montre bien que pour tout les retards les p-values sont supérieur à 0.05 et que la valeur de Q est croissante donc on accepte H0 d'où il n'y a aucune corrélation entre les erreurs. Nous devrons insister sur le dernier retard (k=24) pour la

confirmation des hypothèses. Ainsi, nous pouvons confirmer que les résidus vérifient bien l'hypothèse de non-corrélation associée au bruit blanc.

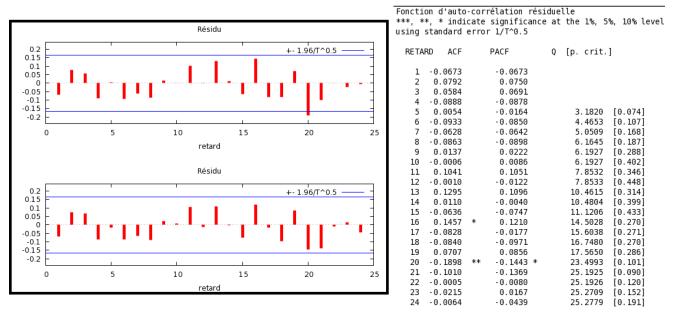


Figure 15 : Corrélogramme des résidus $ARIMA(1,1,1) \times (1,0,1)_{s=12}$

Source: Gretl, dossier Série

Pour la suite de la vérification du modèle, nous nous intéressons à la normalité des résidus pour valider l'hypothèse du bruit blanc avec le test de Chi deux. Le test indique si p-value est supérieur à 5% alors les résidus suivent une loi normale. Dans la figure 16, l'histogramme nous indique la distribution des résidus. Visuellement, nous pouvons dire que les barres forment une quasi-symétrique entre elles donc que les résidus pouvaient potentiellement suivent une loi normale, nous devons justifier avec le test de Chi deux. En effet, nous devons justifier avec la p-value valant 0.23 qui est supérieur à 5%, donc nous validons la normalité des résidus.

Dans **la figure 16**, le graphique Q-Q plot permet de vérifier si les résidus suivent bien une loi normale. Nous concluons que la majorité des points des résidus se placent sur la droite de Henry et qu'il y a un seul point en extrémité qui très peu éloigné de la droite, nous pouvons affirmer alors que les résidus suivent bien une loi normale par ce graphique.

0.0005 Graphique Q-Q pour uhat6 Test de normalité: Chi-deux(2) = 2.902 [0.2343] relative frequency N(127.15,985.76) 4000 0.00045 0.0004 3000 0.00035 2000 0.0003 1000 0.00025 0.0002 0 0.00015 -1000 0.0001 -2000 5e-05 -3000 -2000 -1000 1000 -3000 2000 3000 -3000 -2000 -1000 1000 2000 3000 uhat!

Figure 16 : Normalité des résidus $ARIMA(1,1,1) \times (1,0,1)_{s=12}$

Source: Gretl, dossier Série Temporelle

C. Prévision

Après vérification du modèle, nous réalisons la prévision de ARIMA $(1,1,1) \times (1,0,1)_{s=12}$. La base de données était des observations mensuelles de janvier 2007 à août 2018, nous avons isoler une observation de septembre 2018 afin de comparer la prévision du modèle estimé avec cette valeur et calculer la marge d'erreur.

Tout d'abord, ayant un modèle ARIMA avec une partie classique et saisonnière, nous devons utiliser un ARIMA multiplicatif. Ainsi, pour chaque paramètre p, q, P et Q avec s=12 car nous avons des saisonnalité annuel, on a les formules suivantes :

avec z_t qui est la différenciation de la série brute tel que : $z_t = y_t - y_{t-1}$

AR(1):
$$a_t = z_t (1 - \phi B), \quad z_{t-1} = B z_t$$

 $SAR(1)_{s=12}$: $a_t = z_t (1 - \Phi B^{12})$

MA(1):
$$z_t = a_t(1 - \theta A), \quad a_{t-1} = A a_t$$

SMA(1):
$$z_t = a_t(1 - \Theta_1 A^{12})$$

Dans notre modèle estimé, nous avons :

$$z_{T+1} = \phi_1 z_T + \Phi_1 z_{T-1} - \phi_1 \Phi_1 z_{T-12} + a_{T+1} - \Theta a_{T-11} - \theta_1 a_T + \theta_1 \Theta_1 a_{T-12}$$

Puis Y_{T+1} représente la valeur estimé pour le mois de septembre 2018 selon le modèle ARIMA(1,1,1)(1,0,1) avec T qui est le nombre de total d'observation dont 139

$$Z_{T+1} = Y_{T+1} - Y_T \Rightarrow Y_{T+1} = z_{T+1} + Y_T$$

Application numérique ²³: $Y_{09/18} = 19809$

$$Z_{T+1} = -1248.99$$

$$Y_{T+1} = 18861.00$$

Nous voulons calculer le taux d'erreur de la prévision ::

$$\frac{Y_{T+1} - Y_{09/18}}{Y_{09/18}} = -0.047,$$

Avec notre modèle estimé, nous avons un taux d'erreur 4.7% qui est très faible. Nous pouvons dire que le modèle ARIMA(1,1,1)(1,0,1) est de bonne qualité. Nous constatons que ce modèle prédit une baisse pour le mois de septembre 2018.

Cependant, nous avons dit précédemment que nous allons réaliser plusieurs modèle pour les comparer afin de choisir un modèle avec une meilleur prévision selon différentes caractéristiques.

$$\begin{split} \phi_1 &= -0.41, & \theta_1 &= -0.78, & a_{139} &= -645.66, & a_{128} &= 288.99, & a_{127} &= -280.32 \,, \\ \Phi_1 &= 0.97, & \Theta_1 &= 0.568, & Z_{139} &= -4058, & Z_{128} &= -24, & z_{127} &= -2885 \end{split}$$

²³ Valeur numérique:

5.3 Vérification et choix du modèle

Dans l'annexe, nous avons réaliser sur d'autres modèles de méthodologie de Box-Jenkins des 3 étapes : identification, estimation et vérification de ARIMA et nous résumons dans le tableau 1 (annexe p.49). Nous sélectionnions nos modèles selon la significativité de tous les coefficients dans une estimation. Par conséquent, nous retenons les modèles 1, 2 et 3 où nous lançons les prévisions sur ces trois modèles (figure 17-18 annexe p.50-51).

Le tableau 2 indique les modèles retenus (1,2 et 3), pour choisir le meilleur modèle il faudrait procéder la règle de Parcimonie qui dit de garder un modèle avec le moins de paramètre possible et un écart-type le plus faible. Ces critères nous guident à prendre le modèle 2 car l'écart-type vaut 991.913 et possède le moins de paramètre par rapport au modèle 1. Nous pouvons ignorer le critère du R2 qui n'a pas une grande différence entre les trois modèles.

Tableau 2 de modèle :

	Forme complete	Coefficient sans constante	Ecart- type	P_value (Q)	Test de normalité	P_value normalité	Critères AKAIKE	R2
Modele 1	ARIMA (1,1,1) (1,0,1)	4 coef significatives	985.764	25.2779 [0.191]	Oui	0.23	2341.485	0.77
Modele 2	ARIMA (0,1,1) (1,0,1)	3 coef significatives	991.913	24.5899 [0.265]	Oui	0.13	2341.434	0.76
Modele 3	ARIMA (1,1,0) (1,0,1)	3 coef significatives	1011.16	25.0853 [0.243]	Oui	0.10	2344.519	0.76

En outre, il est possible d'utiliser le critère AIC (Akaike Information Criterion) qui mesure la qualité du modèle statistique et permet de pénaliser les modèles en fonction du nombre de paramètres pour satisfaire la règle de Parcimonie en choisissant la valeur la plus faible. Ainsi, le modèle a le critère AIC le plus faible parmi les trois modèles ce qui nous pousse à choisir le modèle 2.

5.4 Prévision à une période

Pour confirmer notre choix de modèle, nous réalisons un tableau de prédiction du mois de septembre 2018 afin de les comparer et valider notre choix de modèle. D'après le **tableau 3**, le modèle initial 1 a la marge d'erreur la plus élevé, on ne retient pas ce modèle. Nous pouvons conclure que la règle de parcimonie, l'écart-type et le critère de Akaike confirme le choix du meilleur modèle avec une marge d'erreur la plus faible (0.8%) alors que le modèle 3 a une marge d'erreur de 3.5%.

Le modèle $ARIMA(0,1,1)(1,0,1)_{s=12}$ est le modèle le plus performant pour faire une prévision dans le futur pour le mois de septembre 2018, le modèle estime un écart avec la vrai valeur ($Y_{09/18}=19809$) de 174 touriste.

Tableau 3: Prévision et taux d'erreur 24

	Forme complete	Ecart-type	Critères AKAIKE	R2	Prévision	Taux d'erreur
Modele 1	ARIMA(1,1,1) (1,0,1)	985.764	2341.485	0.77	18861.00	4.7%
Modele 2	ARIMA(0,1,1) (1,0,1)	991.913	2341.434	0.76	19635.62	0.8%
Modele 3	ARIMA(1,1,0) (1,0,1)	1011.16	2344.519	0.76	20517,89	3.5%

De plus, dans la figure 19, nous pouvons vérifier la qualité des trois modèles avec les valeurs prédites durant la période entre 2007 et 2018. Nous constatons que le modèle 3 est très différent par rapport en trois autres courbes (la série brute et la prédiction du modèle 1 et 2) alors que le modèle 1 et 2 ont tendance à se confondre avec la série brute. Le graphique nous permet de confirmer le choix du modèle 2 pour la prévision du modèle.

²⁴ Voir calcul des prévision et application numérique pour le modèle 2 et 3 dans l'annexe

Touriste (gauche) preditemodele2 (gauche) preditemodele1 (gauche) preditemodele3 (drpite) -1000 -2000 -3000

Figure 19 : graphique de prédiction des modèles

Source : Gretl, dossier Série Temporelle

6. Test de cointégration selon Engle-Granger

Après avoir réaliser une prévision selon méthode de Box-Jenkins, nous voulons dans la suite de l'étude voir si il existe une relation à long terme entre notre série Touriste avec d'autres série comme le nombre de chambre vendu (Luxe et touriste), la variable de Google Trend ou le ratio de l'offre et la demande de chambre (Luxe et touriste) pouvant expliquer les variations de la série Y. Nous utiliserons le test de cointégration selon Engle-Granger qui est une méthode statistique pouvant détecter cette relation à long terme. L'intérêt de la cointégration est de voir si la régression entre les série Y et X ne sont pas fallacieuse.

6.1 Relation de corrélation de long terme (Gretl)

Dans cette partie, nous nous intéressons à la relation de corrélation de long terme (LT) entre nos série Y_t et X_t . Des variables cointégrées sont le résultat d'une jonction entre une approche économétrique classique et les séries temporelles de Box et Jenkins. Engle et Granger (1987) proposaient la cointégration sur des propriétés stochastiques (aléatoire) pour des variables à proposer, c'est-à-dire d'étudier le comportement de variables économiques tendancielles et voir si ils existeraient une relation à long terme entre les séries.

Dans un premier temps, pour pouvoir étudier la relation de corrélation de LT entre les séries, les séries X doivent être intégrées dans le même ordre que la série Y. Ainsi, toutes les série sont stationnaires avec le même de degré de différenciation :

$$y_t \sim I(d)$$
, $X_t \sim I(d)$, d: degré de différenciation

Nous verrons dans la partie suivante (6.2 Test de Dickey-Fuller augmenté ADF) comment vérifier la stationnarité des série avec le test ADF.

Après avoir vérifier que les séries sont intégrés du même ordre, il est nécessaire de réaliser une régression linéaire simple par la méthode de MCO entre les séries Y et X afin de montrer que la régression n'est pas **fallacieuse**, un problème associé aux variables non stationnaires. Nous avons le modèle linéaire suivant :

 $y_t = cnst + \beta_1 X_t + \epsilon_t$, 25 cnst la constante , β_1 le coefficient de X_t et ϵ_t bruit blanc.

Nous devons vérifier si le terme d'erreur e_t est stationnaire avec le test ADF pour confirmer la relation de corrélation à LT entre les série Y et X.

6.2 Test de Dickey-Fuller augmenté ADF (test de stationnarité)

Le test AFD 26 est un bon indicateur pour montrer la stationnarité des séries. Nous avons l'hypothèse nulle que la racine unitaire est présente dans la série, c'est-à-dire que la série n'est pas stationnaire. Ce test consiste à comparer la valeur estimée du t de student associé au coefficient de la série par rapport aux valeurs tabulées du test Student (dans la table statistique de Student). Nous rejetons H0 au seuil de 5% lorsque la valeur observée du t de Student est inférieur à la valeur critique tabulée par MacKinnon (1996).

En effet, le test réalise sous trois options : sans constance (A), avec constante (B), avec constante et tendance temporelle (C) dans le processus AR(1).

$$y_{t} = \phi_{1}y_{y-1} + \epsilon_{t}$$

$$y_{t} = \phi_{1}y_{t-1} + \beta_{1} + \epsilon_{t}$$

$$y_{t} = \phi_{1}y_{t-1} + \beta_{1} + \mu \cdot t + \epsilon_{T}$$
(A)
$$y_{t} = \phi_{1}y_{t-1} + \beta_{1} + \mu \cdot t + \epsilon_{T}$$
(B)

Avec β_1 qui est la constante, ϕ_1 coefficient de la valeur passé, ϵ_t le terme d'erreur et μ . t la tendance temporelle.

Le test ADF utilise AR(1) afin de détecter la racine unitaire pour les processus de type AR(p).

²⁵ Hamilton D., « Time series Analysis », Princeton, 1994, « 19. Cointegration », pp. 544-548

²⁶ Bourbonnais R., Terraza M., « Analyse des séries temporelles, applications à l'économie et à la gestion », Dunod, 2ème édition, 2008, pp. 157-166

Etude du sujet :

Tout d'abord, nous réaliserons le test ADF sur l'ensemble des séries explicative X (google googleFR, googleUSA, VL, VT, ratio offre/demande : RL, RT) avec un degré 1 pour la différenciation. Ainsi, nous garderons les séries qui sont stationnaire de degrés 1 dans la suite de l'étude.

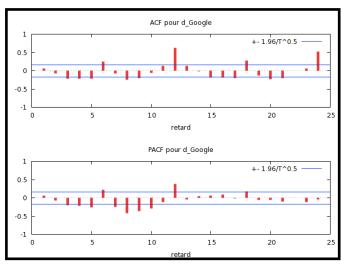
Nous commençons avec la première série google pour la réalisation du test ADF avec un ordre de retard à 12 car nous avons une série mensuelle. Dans la figure 20, nous constatons que la p-value critique asymptotique est inférieur à 5 % dans les trois cas : sans constante, avec constante et avec constante et tendance temporelle. Nous avons la série Google qui est stationnaire. Nous voulons vérifier la stationnarité de la série à l'aide de corrélogramme (figure 21), nous remarquons que la série est saisonnière par rapport à la forte significativité des coefficients de corrélation au 12ème et 24ème retard. Cependant, nous avons aussi d'autres coefficients qui sont significatives en seuil ce qui fait douter sur la stationnarité de la série Google.

Figure 20: test ADF série Google

```
Test de Dickey-Fuller augmenté pour d Google
testing down from 12 lags, criterion AIC
taille de l'échantillon 127
hypothèse nulle de racine unitaire : a = 1
 test sans constante
 avec 11 retards de (1-L)d Google
 modèle: (1-L)y = (a-1)*y(-1) + ...
 valeur estimée de (a - 1): -2.34431
 statistique de test: tau_nc(1) = -3.78518
 p. critique asymptotique 0.0001536
 Coeff. d'autocorrélation du 1er ordre pour e: -0.056
 différences retardées: F(11, 115) = 20.746 [0.0000]
 test avec constante
 avec 11 retards de (1-L)d Google
 modèle: (1-L)y = b0 + (a-1)*y(-1) +
 valeur estimée de (a - 1): -3.26255
 statistique de test: tau c(1) = -4.48428
 p. critique asymptotique 0.000205
 Coeff. d'autocorrélation du 1er ordre pour e: -0.029
 différences retardées: F(11, 114) = 21.965 [0.0000]
 avec constante et tendance temporelle
 avec 11 retards de (1-L)d_Google
 modèle: (1-L)y = b0 + b1*\bar{t} + (a-1)*y(-1) + ... + e
 valeur estimée de (a - 1): -4.4724
 statistique de test: tau ct(1) = -5.4068
 p. critique asymptotique 2.661e-05
 Coeff. d'autocorrélation du ler ordre pour e: -0.009
 différences retardées: F(11, 113) = 23.936 [0.0000]
```

Source : Gretl, dossier Série Temporelle

Figure 21 :corrélogramme google diff 1



Source : Gretl, dossier Série Temporelle

Dans l'annexe, figure 22 (p. 54), nous avons réalisé le test ADF avec les autres séries et que nous allons résumé dans le tableau suivant. De plus, nous avons vérifier à l'aide des graphiques des séries X si elles n'étaient pas déjà stationnaire pour éviter d'avoir une série à I(0) car nous avions dit que la cointégration s'effectuer avec dans le même ordre de différenciation.

Tableau 4 : Test ADF sur l'ensemble des séries

	Test sans constante	Test avec constante	Test avec constante et tendance temporelle	
Google	OUI	OUI	OUI	
GoogleFR	OUI	NON	OUI	
GoogleUSA	OUI	NON	NON	
VL	OUI	OUI	OUI	
VT	OUI	OUI	OUI	
RL	OUI	NON	NON	
RT	OUI	OUI	OUI	

D'après **le tableau 4,** nous remarquons que toutes les séries refusent H0 si nous prenons pas en compte le test sans constante (séries sont stationnaires). Nous gardons la constante que dans le cas si elle est significative (**annexe p. 55 figure 23**). Nous résumons dans l**e tableau 4** et nous avons la constante qui est significative pour la variable google lorsque nous effectuons le test avec constante. Puis, nous avons les séries : google, googleFR, VL et RT qui ont la constante et la tendance temporelle étant significatives alors que pour VT et RL, aucune des deux n'est significatives.

Ensuite la prochaine étape est de réaliser les test ADF sur les résidus de régression linéaire simple entre la série brute Y et les séries brutes X par la méthode de MCO afin de montrer si il existe une relation à LT entre les séries. D'où la premier estimation de MCO entre la série Y et la série Google est montré dans la figure 24

Figure 24 : MCO entre série Y et google

Modàla 1, MCO	utilicant '	lac obcar	vations 2	007.01.2019.	00	T - 140)	ŀ
Modèle 1: MCO, utilisant les observations 2007:01-2018:08 (T = 140)							
Variable dépendante: Touriste							
CO	efficient	erreur	std t	de Student	n.	critique	
	CITICICITE	CIICUI	Jea. c	ac stauciit	ρ.	CITCIQUE	
	936.0	594.76	8	23.43	6.	03e-50	***
Google	34.0120	15.36	45	2.214	0.	0285	**
, and the second							
Moy. var. dép.	1	5156.84	Éc. type	var. dép.		2672.503	
Somme carrés résidus			,,9		on		
R2		. 034292	R2 ajusté			0.027294	
F(1, 138)		. 900367	p. critique (F)			0.028493	
		300.415	Critère d'Akaike			2604.830	
Critère de Schwarz		610.713	Hannan-Quinn			2607.221	
rho	0	.706671	Durbin-W	atson (0.597357	

Source: Gretl, dossier Série Temporelle

D'après les résultats de la MCO, nous constatons que la constante et le coefficient sont statistiquement significatifs. Mais, le R2 est très faible avec 3% donc le modèle a un très faible pouvoir explicatif pour décrire le nombre de Touriste. Ensuite, nous devons vérifier la stationnarité des résidus avec le test ADF (**figure 24**) avec un retard de 12. Nous trouvons que les résidus ne sont pas stationnaires alors il n'y a pas de relation de

Figure 24 : Test ADF sur les résidus 1

```
est de Dickey-Fuller augmente pour uhatl
esting down from 12 lags, criterion AIC
aille de l'échantillon 127
ypothèse nulle de racine unitaire : a = 1
 test sans constante
 avec 12 retards de (1-L)uhat1
 modèle: (1-L)y = (a-1)*y(-1) + ... + e valeur estimée de (a-1): -0.0308241 statistique de test: tau_nc(1) = -0.461667
 p. critique asymptotique 0.5156
 Coeff. d'autocorrélation du 1er ordre pour e: -0.106
 différences retardées: F(12, 114) = 21.785 [0.0000]
 test avec constante
 avec 12 retards de (1-L)uhat1
 modèle: (1-L)y = b0 + (a-1)*y(-1) + \dots + e valeur estimée de (a-1): -0.0257812 statistique de test: tau_{-}(1) = -0.37861
 p. critique asymptotique 0.9105
 Coeff. d'autocorrélation du ler ordre pour e: -0.106
différences retardées: F(12, 113) = 21.641 [0.0000]
 avec constante et tendance temporelle
 avec 12 retards de (1-L)uhat1
 avec 12 rectars de (1-1)diac1
modèle: (1-1)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -0.177555
statistique de test: tau_ct(1) = -2.27157
p. critique asymptotique 0.449
 Coeff. d'autocorrélation du 1er ordre pour e: -0.057
 différences retardées: F(12, 112) = 22.921 [0.0000]
```

Source : Gretl, dossier Série Temporelle

corrélation à LT entre la série Y et la série Google. Donc nous avons une régression fallacieuse.

Dans **l'annexe**, **figure 25** (**p.55-56**, **figure 25**), nous avons réaliser la MCO et le test ADF sur tout les résidus des séries. Toutes les séries de google et VL ne sont pas stationnaire.

Tableau 5 de stationnarité des résidus

Série	Stationnarité des résidus
Google	NON
Google FR	NON
Google USA	NON
CVL	NON
CVT	OUI
RL	OUI
RT	OUI

Ainsi, les séries VT, RL et RT ont une relation de corrélation à LT(**figure 26**). Donc il est possible d'interpréter les coefficient de la MCO sur les trois séries. Tout d'abord, pour la série VT, nous avons un coefficient équivalent à 1.4, nous pouvons dire lorsque une chambre (hors luxe) est vendu, il y a alors une augmentation de 1 ou 2 touriste.

Pour la série RL, lorsque le ratio d'offre et demande de chambre de luxe augmente de 1% alors le nombre de touriste augmente de 18546. Avec la série RT, pour un 1% le nombre de touriste augmente 17753. Nous constatons une différence de 793 touriste entre les deux coefficients des deux séries.

Figure 26: MCO VT, RL et RT

	coefficie	ent erreur	std. t de	Student	p.	critique	
const	11412.7	420.543	3 2	7.14	3	. 46e - 57	**
CVT	1.409	958 0.144	103	9.782	1	. 73e - 17	**
Moy. var.	dép.	15156.84	Éc. type v	ar. dép.		2672.503	:
Somme carr	és résidus	5.86e+08	Éc. type d	e régress	ion	2061.161	
R2		0.409457	R2 ajusté	-		0.405177	,
(1, 138)		95.68313	p. critique	e (F)		1.73e-17	,
og de vra	isemblance	-1265.988	Critère d'	Akaike		2535.975	i
ritère de	Schwarz	2541.859	Hannan-Oui	nn		2538.366	i
rho		0.619204	Durbin-Wat			0.761394	ļ

Modèle 17: MCO, utilisant les observations 2007:01-2018:08 (T = 140) Variable dépendante: Touriste							
	coefficient	erreur	std. t	de Student	р.	critique	
const	3463.37	1083.	81	3.196	0	.0017	***
CMRL	18546.7	1698.	72	10.92	2	. 17e - 20	***
Moy. var. dé	p.	15156.84	Éc. typ	e var. dép.		2672.503	
Somme carrés	résidus	5.33e+08	Éc. typ	e de régress	sion	1964.654	
R2		0.463463	R2 ajus	té		0.459575	
F(1, 138)		119.2048	p. crit	ique (F)		2.17e-20	
Log de vrais	emblance -	1259.274	Critère	d'Akaike		2522.548	
Critère de S	chwarz	2528.432	Hannan-	Quinn		2524.939	
rho		0.590697	Durbin-	Watson		0.817361	

	MCO, utilisa pendante: To		ervation	s 2007:01-201	L8:08	(T = 140)
	coefficier	nt erreur	std.	t de Student	р.	critique	
const	5385.26	797.0	976	6.756	3	.64e-10	***
CMRT	17753.2	1420.4	41	12.50	1	. 90e - 24	***
Moy. var. d	lép.	15156.84	Éc. ty	oe var. dép.		2672.503	
Somme carré	s résidus	4.66e+08	Éc. ty	oe de régress	sion	1836.930	
R2		0.530957	R2 aju	sté		0.527558	
F(1, 138)		156.2159	p. cri	tique (F)		1.90e-24	
Log de vrai	semblance	-1249.863	Critèr	e d'Akaike		2503.727	
Critère de	Schwarz	2509.610	Hannan	-Quinn		2506.117	
rho		0.517219	Durbin	-Watson		0.957335	

Source : Gretl, Dossier Série Temporelle

Nous trouvons des R2 de bonne qualité pour la série RT (53%) et de moyenne qualité pour les séries RL (46%) et VT (41%). Par exemple pour la série RT, la variable explicative explique à 53% les variations du nombre de touristes en Polynésie Française.

Après cette étape, nous sommes capable de vérifier si ils ont une relation à court terme avec le modèle à correction d'erreurs.

6.3 Modèles à Correction d'Erreur MCE

En effectuant une corrélation à long terme entre deux séries, nous pouvant réaliser le modèle à correction d'erreur MCE pour vérifier une relation à court terme. Tout d'abord avant de commencer les estimations MCE sur notre étude, nous expliquez son fonctionnement. Les MCE ont étés introduits par Hendry (dans les années 90). L'idée des MCE est de considérer des relations entre les séries avec le même ordre de différenciation.

Par exemple, nous prenons la première différenciation sur deux série, soient Y et X puis nous réalisons une deuxième MCO tel que :

$$\Delta Y_t = \lambda \Delta X_t + \mu [Y_{t-1} - \beta X_{t-1}] + a_t \quad ,$$

Avec $Y_{t-1} - \beta X_{t-1}$ est le résidu de la RLS classique entre la série Y et X qui est stationnaire.

 ΔY et ΔX sont les séries différenciés d'ordre 1

 a_t est le résidu de la régression puis λ et μ qui sont les coefficients respectives des variables ΔX et de $Y_{t-1} - \beta X_{t-1}$.

Le coefficient μ doit être négative et significative pour confirmer la relation à court terme entre les deux séries Y et X. Dans le cas contraire, on rejette l'hypothèse de la MCE alors nous nous retrouvons avec une régression fallacieuse si l'hypothèse n'est pas validé. En outre, le coefficient μ indique le pourcentage de vitesse d'ajustement de la relation à court terme vers l'équilibre à long terme.

6.4 Coefficients de cointégration : Interprétation

Nous allons tester la relation à CT avec les séries qui validaient la relation à long terme. (VT, RL, RT). Nous réaliserons de nouveaux une MCO avec les séries de première différenciation et les résidus de la première MCO. D'après **le tableau de stationnarité des résidus (p.37)**, nous devrons estimer avec les séries VT, RL et RT.

Figure 27: MCE entre Touriste et VT

Modele 25: MCO, utilisant les observations 2007:02-2018:08 (T = 139) Variable dépendante: d_Touriste écarts type HAC, largeur de bande 3 (Kernel Bartlett)						
coeffici	ient erreur	std.	t de Student	р. с	ritique	
const -11.7783	3 131.564		-0.08953	0.9	288	
d CVT 1.9927	0.488	895	4.076	7.7	5e-05	***
residCVT_1 -0.3658	325 0.063	3454	-5.775	4.9	9e-08	***
Moy. var. dép.	34.21583	Éc.	type var. dép.		2033.7	41
Somme carrés résidus			type de régres			
R2	0.384506	R2 8	ajusté		0.3754	155
F(2, 136)	25.89578	р. (ritique (F)		2.96e-	10
Log de vraisemblance	-1221.851	Cri	tère d'Akaike		2449.7	702
Critère de Schwarz	2458.506	Hanı	nan-Quinn		2453.2	280
rho	-0.121888	Du ri	oin-Watson		2.2120	800

Source : Gretl, dossier Série Temporelle

La figure 27 indique la MCO entre les série différencié touriste et VT, le coefficient μ est négative et fortement significative alors il existe une relation à court terme entre les deux séries donc nous devons accepter que la régression n'est pas fallacieuse. Cela veut dire la série va être corriger de l'ordre de 36%, la série pourrait retrouver lentement l'équilibre à long terme au bout de 7 années (voir calcul annexe). En regardant la figure 27, pour interpréter le coefficient de la série VT, nous pouvons dire que lorsqu'une chambre pour touriste (hors luxe) est vendu alors nous retrouvons deux touriste en plus sur le territoire polynésien. De plus, la qualité du modèle R2 représente 38.45%, ainsi le pouvoir explicatif du modèle est moyennement faible.

En outre, dans la relation à long terme, avec la première MCO **(figure 25 annexe p.56)**, nous trouvons un coefficient de VT qui égale à 1.40 et nous constatons une diminution par rapport au court terme.

Nous trouvons aussi des relations à court terme pour les séries RL et RT avec Y (**figure 28**). Ainsi, les régressions ne sont pas fallacieuse. Pour l'interprétation des résultats de la série RL, nous pouvons dire que la série va être corriger 26%, dont au bout de 8 ans et 6 mois, la série va retrouver l'équilibre du long terme. Pour rappel, RL est le rapport entre le nombre de chambre de luxe vendu sur le nombre de chambre offert. Lorsqu'on augmente le ratio de 1%, alors le nombre de touriste augmente de 17674. Pour la qualité du modèle, nous avons un R2 qui est égale 31% soit moyennement faible, l'augmentation du ratio de l'offre et la demande des chambres d'hôtel de luxe explique à 31% les variations de Y (nombre de touriste) en Polynésie française durant la période entre 2007 et 2018.

Pour la série RT, nous trouvons un coefficient μ qui est égale à 0.46, ainsi la vitesse de l'ajustement de la relation de CT est 46% pour qu'elle s'équilibre vers le LT au bout de 6 ans, la vitesse d'ajustement avec cette variable est plus rapide par rapport aux séries précédentes. RT est le ratio de l'offre et la demande de chambre d'hôtel hors luxe, en interprétant le coefficient RT, lorsqu'on augmente de 1% le ratio RT, nous avons le nombre de touriste qui augmente de 15284. Nous avons un R2 qui est bien meilleur par rapport aux deux modèles précédent (41%) comme dans la première MCO.

Figure 28 : MCE avec RL et RT

Modèle 23: MCO, util: Variable dépendante: écarts type HAC, larç	d_Touriste		18:08 (T = 13	39)
coeffici	lent erreur	std. t de Student	p. critique	e
const -3.70	154 129.15	52 –0.02866	0.9772	•
		9.566		**
		396123 -5.136		**
Moy. var. dép.	34.21583	Éc. type var. dép.	2033.74	41
Somme carrés résidus		Éc. type de régres		
R2	0.417167	R2 ajusté	0.40859	96
F(2, 136)	46.07993	p. critique (F)	5.25e-1	16
Log de vraisemblance	-1218.062	Critère d'Akaike	2442.12	23
Critère de Schwarz	2450.927	Hannan-Quinn	2445.70	91
rho	-0.097570	Durbin-Watson	2.18288	31

Modèle 24: MCO, utilis Variable dépendante: d écarts type HAC, large	_Touriste		. ,
coeffici	ent erreu	r std. t de Studen	t p. critique
const -0.34 d_CMRL 17674.0 residCVL_1 -0.26	2309.	7.653	3.25e-12 ***
Log de vraisemblance	3.92e+08 0.312918 45.28030	Éc. type var. dép. Éc. type de régres R2 ajusté p. critique (F) Critère d'Akaike Hannan-Quinn Durbin-Watson	sion 1698.126 0.302814 8.47e-16 2464.996

Source : Gretl, dossier Série temporelle

7. Conclusion

Le but de cette étude était de réaliser la méthode ARIMA sur une série temporelle en rapport avec le tourisme en Polynésie française et de tester la cointégration Engle - Granger entre des série temporelle. Nos séries comprenaient des observations mensuelle entre la période du janvier 2007 et septembre 2018 trouver sur le site de l'Institut Statistique de la Polynésie française (ISPF) et sur Google Trend.

Dans la premier partie, cette étude de cas nous a permis de pratiquer des techniques prévisionnelle avec le méthodologie de Box-Jenkins passant par plusieurs étapes : identification, estimation, vérification et prévision. En effet, il était important de stationnarisé la série avant de procéder à l'identification des paramètre ARIMA. Ainsi, la série présentait deux partie différentes, une partie classique et une partie saisonnière identifiable par le graphique temporel par la présence de pics et de creux périodique pour chaque année. Après, il fallait identifier les paramètres à l'aide de corrélogramme et vérifier les modèles ARIMA par plusieurs critères comme le test Portemanteau.

Nous avions proposé trois modèles d'ARIMA puis comparé par rapport à l'écart-type ou le critère AIC. En réalisant, la prévision sur les trois modèles sélectionnés, nous avions trouvé une meilleur prévision pour le modèle ARIMA(0,1,1)x(1,0,1) avec un taux d'erreur de 0.8% pour une valeur prédite de 19635.62 contre une valeur réel pour le mois de septembre 2018 : 19809. Nous remarquons que l'écart-type et l'indice AIC sont la plus faible parmi les trois modèles proposés, ce qui peut affirmer le choix du modèle.

Dans la deuxième partie, nous avions un test de cointégration selon Engle-Granger afin de trouver des relation à long terme et courte terme entre la variable à expliquer et les variables explicatives. Il est important de différencié les séries dans le même ordre dont d=1. Puis, il fallait réaliser des régression linéaire simple (MCO) entre les série Y et série X afin d'étudier les résidus si elles étaient stationnaires. Si elles étaient stationnaires, alors les séries avaient une relation à long terme. Ensuite, nous réalisons la relation à court terme dont les séries avaient une relation à LT. Nous avions utilisé les méthodes de correction d'erreur (MCE) en fessant une MCO avec les séries différenciés dans le même ordre et les résidus de la première MCO. L'attention se porte ici vers le coefficient du résidus de la première régression, ainsi il devait être négative et significative pour que les séries aient une relation à court terme mais aussi une cointégration des séries Y et X.

Nous retenons seulement trois cointégrations entre la série Y et les séries explicatives : RL, RT et VT. Tout d'abord, la cointégration avec la série VT montre des résultats assez cohérents par rapport aux coefficient de la variable dans la relation à court terme, nous trouvons lorsqu'**une chambre d'hôtel hors luxe est vendu** alors il y a deux touristes qui s'ajoutent. Cela peut s'expliquer que la Polynésie est devenu une destination pour les couples, par exemple pour des touristes américains, « touristes de noces » représentent à 40% de cette population qui vient sur le territoire. ²⁷

²⁷ Rapport ISPF 1999: http://www.ispf.pf/docs/default-source/publi-pr/Regards 09-Tourisme en P F.pdf?sfvrsn=0

Les « lunes de miels » sont le motif principal des voyages pour les couples. Dans une relation à long terme, avec la première MCO (figure 25, annexe p.54), nous trouvons un coefficient de VT qui est égale à 1.40 et nous constatons une valeur supérieur dans la relation à court terme (1.99). Ainsi, nous avions dit qu'au bout de 7 années, le modèle trouvait son équilibre à LT.

Ensuite, nous nous intéressons aux résultats des relation de cointégration de Y avec les séries RT et RL. Nous pouvons dire que les deux modèles ont le même niveau de représentation sur la qualité de R2, ainsi il est évident de faire une remarque sur la différence des coefficient des deux variables dans la relation à court terme et la relation à long terme mais aussi sur le coefficient de vitesse d'ajustement. La série RT a une vitesse d'ajustement (46%) qui est plus rapide que la série RL(26%). En se fiant à la vitesse d'ajustement, la préférence se fera plus vers la série RT pour étudier la relation de cointégration entre la série touriste et cette série.

Par exemple, pour la série RL le coefficient est plus élevé que la série RT dans les deux relations. En effet, l'activité de l'hôtellerie est plus importante dans le remplissage des hôtels catégoriser de luxe. Dans le première trimestre 2017 ²⁸, les hôtels de luxe avait gagner 1.1 point par rapport à l'année 2016 tandis que les hôtels hors luxe ont diminués de 3.2 point. Ce qui peut expliquer cette écart de coefficient dont les tourismes ont plus une préférence à des hôtels de luxe que les hôtels de gamme intermédiaires.

²⁸ Rapport ISPF 2017, « Point conjoncture Polynésie Française » : http://www.ispf.pf/docs/default-source/tb-tourisme/tb-2017-t1.pdf?sfvrsn=10 (Consulté le 10 fev 2020)

8. Discussion et limites

Par ailleurs, nous obtenons des résultats assez décevantes avec les séries de Google Trend (Google, GoogleFR, GoogleUSA) dont les internautes lancer des requêtes sur la recherche du terme : « Hotel Tahiti » où nous plaçons sur trois zones géographiques différentes, c'est-à-dire une dans le monde, en France puis seulement aux Etats-Unis. Aucune des séries ne pouvaient cointégrés avec la série Y (le nombre de touriste) car pendant la première MCO, aucun résidus n'étaient stationnaires.

Ces séries n'apparaissent pas comme des variables explicatives pour la série Y et en regardant la matrice de corrélation (**figure 29 annexe p.57**), nous remarquons les séries sont faiblement corrélé sauf pour la série Google FR avec plus de 50% ce qui peut expliquer la non cointégration des séries.

En outre, Tahiti étant une destination touristique l'une des plus cher au monde, très peu de personne peuvent se permettre de s'y rendre. Ces personne vont plutôt s'adresser directement à des agences de voyages à défaut de s'organiser par soi même son voyage. Le comportement des internautes est difficile à assurer par rapport aux requêtes faites sur le site de recherche Google mais il est plutôt préférable d'utiliser des variables comme le nombre de réservation réalisée en ligne sur des sites comme booking.com ou sur Airbnb (location entre particulier) afin d'étudier la relation avec le nombre de touriste.

Nous remarquons aussi que d'après le graphique (figure 30), il y a un écart sur les variations entre les série touriste et VL dans la période 2007-2014. Ainsi cette écart peut s'expliquer par la consommation de la population locale sur l'achat de chambre de d'hôtel de luxe. La série VL ne pourrait pas expliquer clairement sur le nombre de tourisme en Polynésie Française cause de l'achat de chambre faite par la population locale.

Il faudrait prendre avec précaution la variable Google Trend pour des modélisations économétriques pour expliquer des phénomènes tel que le tourisme. De plus, la série Google trend change en permanence au fil des jours d'où il fait réaliser une réestimation des modèles économétriques.

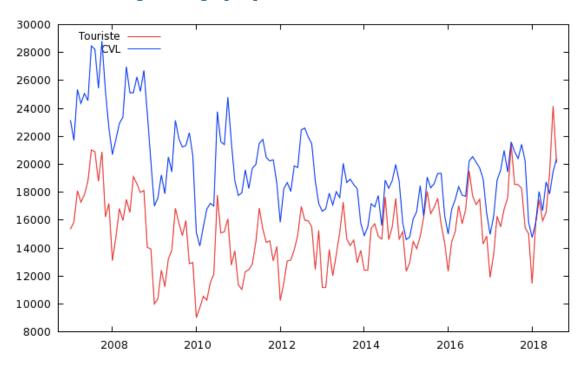


Figure 30: graphique série touriste et VL

Source : dossier Série Temporelle

9. Bibliographie

- ¹: Bortoli C. et Combes S., « Apports de Google Trends pour prévoir la conjoncture française : des pistes limitées », INSEE, mars 2015. Téléchargeable : link. (Consulté le 15 janvier 2020)
- ²: Choi.H et Varian.H « Predicting the Present with Google Trends », ©Google Inc., avril 2009, téléchargeable : https://static.googleusercontent.com/media/www.google.com/fr//googleblogs/pdfs/google_predicting_the_present.pdf (consulté le 15 janvier 2020)
- ³François C... "L'économétrie des échanges touristiques de la France," Économie et Prévision, 1988, Programme National Persée, vol. 84(3), pages 41-59.
- ⁴Claveria F., et Torra S., "Forecasting tourism demand to Catalonia: Neural networks vs. time series models", 2014, *Economic Modelling*, 36, issue C, p. 220-228.
- ⁵Rapport, « Le tourisme en Polynésie française. En quête d'un second souffle », 2015, Institut d'émission d'outre-mer (IEOM), téléchargeable: https://www.ieom.fr/IMG/pdf/ne147_eclairage_tourisme_pf.pdf (consulté le 15 janvier 2020).
- ⁷ Rapport, « Polynésie Française », 2018, IEOM, https://www.ieom.fr/IMG/pdf/ra2018_polyne_sie_f.pdf (consulté le 15 janvier 2020)
- ¹⁴ Rapport ISPF, 2017 : http://www.ispf.pf/docs/default-source/tb-tourisme/tb-2017-t1.pdf? sfvrsn=10 (consulté le 20 Janvier 2019).
- 15 Rapport IOEM, 2015 : https://www.ieom.fr/IMG/pdf/ne147_eclairage_tourisme_pf.pdf (consulté le 20 Janvier 2019).
- ¹6Box P., Hillmer S. Et Tiao G. "Analysis and Modeling of Seasonal Time Series,".
 Seasonal Analysis of Economic Time Series, édition de Zellner, Arnold. 309U.S.

Department of Commerce, pp 301-339, 1978. Téléchargeable : https://www.nber.org/chapters/c4329.pdf (consulté le 09 Fev 2020)

- ¹⁷Hatanaka M. « Unit Roots and Co-Integrations », Time-Series-Based Econometrics, edition Oxford University press, Chapter 1 : « Difference Equations » pp. 2-7, (consulté le 10 Fev 2020)
- ¹⁸ Bourbonnais R., Terraza M., « Analyse des séries temporelles, applications à l'économie et à la gestion », Dunod, 2ème édition, 2008, pp. 15-19
- ¹⁹:Box, George., Box and Jenkins: « Time Series Analysis, Forecasting and Control ».
 10.1057/9781137291264 6. Consulté le 10 Fev 2020
- ²⁰Gourieroux C., Monfort A. « Séries temporelles et modèles dynamiques » 2eme edition Economica, chapitre 6 : « Prévision par la méthode Box-Jenkins » pp. 230-238
- ²⁵ Hamilton D., « Time series Analysis », Princeton, 1994, « 19. Cointegration », pp.
 544-548
- ²⁶ Bourbonnais R., Terraza M., « Analyse des séries temporelles, applications à l'économie et à la gestion », Dunod, 2ème édition, 2008, pp. 157-166
- ²⁷ Rapport ISPF 1999: http://www.ispf.pf/docs/default-source/publi-pr/ Regards_09_- _Tourisme_en_P_F.pdf?sfvrsn=0
- ²⁸ Rapport ISPF 2017, « Point conjoncture Polynésie Française » : http://www.ispf.pf/docs/default- source/tb-tourisme/tb-2017-t1.pdf?sfvrsn=10 (Consulté le 10 fev 2020)

Table des matières

1. Introduction	4
2. Analyse économique du sujet	5
2.1. Définition de la variable à expliquer	6
2.2. Définition des variables indépendantes	6
2.3. Présentation de la base de données	8
3. Analyse des variables	9
3.1. Description de l'évolution de chaque variable et graphique	9
4. Présentation de la méthodologie ARIMA	-14
4.1. Stationnarité	15
4.2. Corrélogrammes	16
4.3. Construction du modèle de Box-Jenkins	17
4.4. Modèles multiplicatifs	
5. Modèle ARIMA pour	-22
5.1 Stationnarité de la série Y	22
5.2 Identification, estimation et vérification	24
5.3 Vérification et choix du modèle	29
5.4 Prévision à une période	30
6. Test de cointégration selon Engle-Granger	-32
6.1 Relation de corrélation de long terme (Gretl)	32
6.2 Test de Dickey-Fuller augmenté ADF (test de stationnarité)	33
6.3 Modèles à Correction d'Erreur MCE	38
6.4 Coefficients de cointégration : Interprétation	39
7. Conclusion	-41
8. Discussion et limites	-44
9. Bibliographie	-46
10. Annexes	

10. Annexes

Figure 13: Estimation modele ARIMA

Evaluations	de la fonction	: 1712			
Évaluations	du gradient :	716			
	RMA, utilisant		ons 2007:02	-2018:08 (T =	139)
	sing AS 197 (MV				
	oendante: d_Tou				
Écarts type	basés sur la m	atrice hessie	nne		
	coefficient	erreur std.	Z	p. critique	
const	-16.7430	199.439	-0.08395	0.9331	
phi_1	0.411230	0.149962	2.742	0.0061	***
Ph1_1	0.976340 -0.783367 -0.567852	0.0150122	65.04	0.0000	***
theta_1	-0.783367	0.104978	-7.462	8.51e-14	***
Theta_1	-0.567852	0.118962	-4.773	1.81e-06	***
Moy. var.	dép.	34.21583			
Ec. type	/ar. dép.	2033.741			
	es innovations				
	des innovations				
R2		0.772641			
R2 ajusté	aisemblance	0.767588			
	'Akaike				
	e Schwarz				
Hannan - Qu	inn	2350.630			

Tableau 1 de modèle :

	Forme complete	Coefficient sans constante	Ecart- type	P_value (Q)	Test de normalité	P_value normalité	Critères AKAIKE	R2
Modele 1	ARIMA (1,1,1) (1,0,1)	4 coef significatives	985.764	25.2779 [0.191]	Oui	0.23	2341.485	0.77
Modele 2	ARIMA (0,1,1) (1,0,1)	3 coef significatives	991.913	24.5899 [0.265]	Oui	0.13	2341.434	0.76
Modele 3	ARIMA (1,1,0) (1,0,1)	3 coef significatives	1011.16	25.0853 [0.243]	Oui	0.10	2344.519	0.76
Modele 4	ARIMA (0,1,0) (1,0,1)	2 coef significatives	1059.3	38.6298 [0.016]	Oui	0.49	2357.900	0.73
Modele 5	ARIMA (1,1,0) (1,0,0)	2 coef significatives	1111.62	49.2829 [0.001]	Oui	0.35	2364.364	0.70
Modele 6	ARIMA (1,1,1) (0,0,1)	1 coef significative	1554.01	144.1930 [0.000]	Oui	0.86	2447.036	0.46

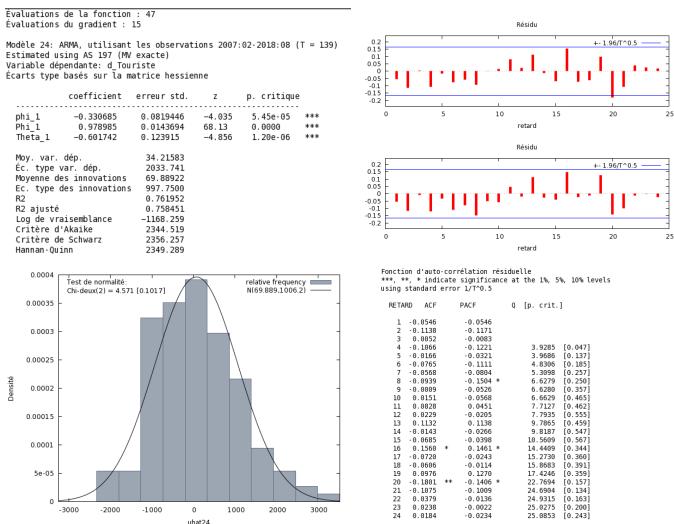
Évaluations de la fonction : Évaluations du gradient : 14 +- 1.96/T^0.5 0.2 0.15 0.1 0.05 0 -0.05 -0.1 -0.15 -0.2 Modèle 20: ARMA, utilisant les observations 2007:02-2018:08 (T = 139) Estimated using AS 197 (MV exacte) Variable dépendante: d Touriste Écarts type basés sur la matrice hessienne p. critique coefficient erreur std. 67.77 Phi 1 0.977978 0.0144300 0.0000 retard -0.403743 7.05e-06 theta 1 0.0898774 -4.492 Theta_1 0.126416 -4.526 6.01e-06 Résidu Moy. var. dép. 34.21583 0.2 0.15 0.1 0.05 0 +- 1.96/T^0.5 Éc. type var. dép. 2033.741 Moyenne des innovations 85.13127 984.8371 Ec. type des innovations 0.767821 -0.05 -0.1 R2 ajusté 0.764407 Log de vraisemblance -015 -1166.717 Critère d'Akaike 2341.434 Critère de Schwarz 2353.172 10 15 20 Hannan-Quinn 2346.204 Fonction d'auto-corrélation résiduelle 0.00045 * indicate significance at the 1%, 5%, 10% levels Test de normalité: relative frequency using standard error 1/T^0.5 Chi-deux(2) = 4.024 [0.1337] N(85.131.991.91) 0.0004 RETARD ACF PACE Q [p. crit.] 0.00035 0.0014 0.0014 -0.0105 0.0105 -0.0100 -0.0100 -0.1287 -0.1288 2.4335 [0.119] 0.0436 2.7018 [0.259] -0.1086 -0.1140 4.4405 [0.218] 0.00025 -0.0802 -0.0883 5.3946 [0.249] -0.0992 -0.1286 6.8664 [0.231] -0.0023 -0.0285 6.8672 [0.333] 0.0002 0.0122 -0.0333 6.8897 [0.440 11 0.1010 0.0642 8.4529 [0.391] 12 0.0300 8.5159 [0.483] 0.0202 0.00015 13 0.1231 0.1022 10.8738 [0.367] 14 0.0087 0.0190 [0.453] 10.8856 15 16 0.0525 0.0433 11.3205 [0.502] 0.0001 0.1489 0.1466 14.7427 [0.324] 15.4344 5e-05 18 19 -0.0659 0.0406 16.1382 [0.373] 0.0672 0.1105 16.8766 [0.394] 20 21 0.1855 0.1514 22.5424 [0.165] -0.1092 0.1061 24.5227 [0.139] 0 22 0.0141 0.0089 24.5558 [0.176] uhat20 23 0.0090 0.0061 24.5696 [0.218] 0.0109 0.0451 24.5899 [0.265]

Figure 17: estimation modèle 2 ARIMA(0,1,1)(1,0,1)

Source : Gretl, dossier Série

Dans le modèle 2, nous remarquons que tous les coefficients sont significatives et respectent les condition d'invisibilité et de stationnarité. D'après les corrélogramme, nous n'avons pas de problème de covariance et corrélation entre les termes d'erreurs car les deux premières retards sont éloignés des bornes. En regardant la dernière valeur de Q-stat, nous avons une p-value supérieur de 5% donc on peut confirmer l'hypothèse de non corrélation entre les erreurs associé aux bruits blanc. De plus, on trouve aussi que les résidus suivent une loi normale d'après le test de Chi deux. Nous pouvons procédé à la prévision du modèle.

Figure 18: estimation modèle 3 ARIMA (1,1,0)



Source : Gretl, dossier Série

Dans le modèle 3, on trouve aussi les mêmes résultats que le modèle 1 et 2, nous pouvons aussi procéder à la prévision et comparer les valeurs prédites.

Calcul des prévision :

 $Y_{09/18}=19809~,~Y_T=20110$, voir les extraits des résidus et la valeur différencié de la serie Y.

Modèle 2:

$$\begin{split} z_{T+1} &= \Phi_1 z_{T-11} + a_{T+1} - \Theta_1 a_{T-11} - \theta_1 a_T + \theta_1 \Theta_1 a_{T-12} \;, \quad T = 139 \\ z_{T+1} &= y_{T+1} - y_T \\ z_{T+1} &= -474.37 \Rightarrow \quad \widehat{Y}_{T+1} = 19635.62 \\ &\frac{\widehat{Y}_{T+1} - Y_{09/18}}{Y_{09/18}} = -0.008 \; \text{--> calcul marge d'erreur} \end{split}$$

$$\begin{split} &\Phi_1=0.97 \;,\;\; \theta_1=-\;0.40 \;,\;\; \Theta_1=-\;0.57 \\ &z_{128}=-\;24 \;\;,\;\;\; a_{128}=-\;84.79 \;\;,\;\; a_{127}=-\;396.96 \quad , \quad a_{139}=-\;780.65 \end{split}$$

Modèle 3:

$$z_{T+1} = \Phi_1 z_{T-11} + \phi_1 z_t - \phi_1 \Phi_1 z_{T-12} + a_{T+1} - \Theta_1 a_{T-11}$$

$$z_{T+1} = 407.8930 \Rightarrow \widehat{Y}_{T+1} = 20110$$

$$\frac{\widehat{Y}_{T+1} - Y_{09/18}}{Y_{09/18}} = 0.035 \longrightarrow \text{calcul marge d'erreur}$$

$$\begin{split} \phi &= -0.33 \quad \Phi_1 = 0.979 \,, \quad \Theta_1 = -0.60 \\ z_{127} &= -2285 \quad, \quad z_{128} = -24 \quad, \quad a_{128} = -282.56 \end{split}$$

Extrait des résultats des résidus du modèle 1, 2 et 3

Modèle 1 Modèle 2

Mois	d_touriste	ajusté	résidus	Mois	d_touriste	ajusté	résidus
2017:01	- 2948.00	-2050.11	-897.89	2017:01	- 2948.00	- 2343. 14	-604.86
2017:02	1654.00	1513.28	140.72	2017:02	1654.00	1309.82	344.18
2017:03	2717.00	1522.86	1194.14	2017:03	2717.00	1263.25	1453.75
2017:04	- 758.00	141.92	-899.92	2017:04	-758.00	-61.50	-696.50
2017:05	1259.00	-72.82	1331.82	2017:05	1259.00	41.12	1217.88
2017:06	817.00	502.92	314.08	2017:06	817.00	437.35	379.65
2017:07	3849.00	2092.22	1756.78	2017:07	3849.00	2238.67	1610.33
2017:08	- 2885.00	-2604.68	-280.32	2017:08	-2885.00	-2488.04	-396.96
2017:09	-24.00	-312.99	288.99	2017:09	-24.00	60.79	-84.79
2017:10	-268.00	423.00	-691.00	2017:10	-268.00	655.39	- 923.39
2017:11	-2822.00	-2346.20	-475.80	2017:11	-2822.00	-2157.36	-664.64
2017:12	-419.00	470.62	-889.62	2017:12	-419.00	493.20	-912.20
2018:01	-3573.00	-2087.90	- 1485. 10	2018:01	-3573.00	-2162.93	- 1410.07
2018:02	4290.00	2012.91	2277.09	2018:02	4290.00	1850.24	2439.76
2018:03	1705.00	1353.12	351.88	2018:03	1705.00	919.87	785.13
2018:04	- 1497.00	-362.44	- 1134.56	2018:04	- 1497.00	-323.98	- 1173.02
2018:05	604.00	649.58	- 45.58	2018:05	604.00	847.17	- 243. 17
2018:06	2813.00	990.49	1822.51	2018:06	2813.00	961.31	1851.69
2018:07	4796.00	2300.70	2495.30 *	2018:07	4796.00	2182.98	2613.02 *
2018:08	-4058.00	-3403.34	-654.66	2018:08	-4058.00	-3277.35	- 780 . 65

Modèle 3

d_touriste	ajusté	résidus
-2948.00	- 2535.18	-412.82
1654.00	1208.32	445.68
2717.00	1225.60	1491.40
-758.00	33.51	-791.51
1259.00	222.31	1036.69
817.00	444.35	372.65
3849.00	2428.54	1420.46
-2885.00	-2315.66	- 569.34
-24.00	249.80	- 273.80
- 268.00	602.18	-870.18
-2822.00	- 2220 . 49	-601.51
-419.00	351.81	-770.81
-3573.00	-2316.83	- 1256. 17
4290.00	1578.23	2711.77 *
1705.00	879.30	825.70
- 1497.00	49.97	- 1546.97
604.00	858.37	- 254.37
2813.00	783.45	2029.55
4796.00	2247.65	2548.35 *
-4058.00	-2821.69	-1236.31
	- 2948.00 1654.00 2717.00 - 758.00 1259.00 817.00 3849.00 - 2885.00 - 24.00 - 268.00 - 2822.00 - 419.00 - 3573.00 4290.00 1705.00 - 1497.00 604.00 2813.00 4796.00	- 2948.00

Figure 22 : Test ADF sur l'ensemble des séries

```
Test de Dickey-Fuller augmenté pour d_GoogleFR
testing down from 12 lags, criterion AIC
taille de l'échantillon 126
hypothèse nulle de racine unitaire : a = 1

test sans constante
avec 12 retards de (1-1)d_GoogleFR
modèle: (1-1)y = (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -1.62902
statistique de test: rad_nC(1) = 0.56714
p. critique asymptot que 0.00994
Coeff. d'autocorrélation du lor ordre pour e: -0.083
différences retardées: F(12, 113) = 78.894 [0.0000]

test avec constante
avec 12 retards de (1-1)d_GoogleFR
modèle: (1-1)y = b0 + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -1.68213
statistique de test: rad_c(1) = -2.5587
p. critique asymptot que 0.1017
Coeff. d'autocorrélation de la ordre pour e: -0.081
différences retardées: F(12, 112) = 78.268 [0.0000]

avec constante et tendance temporelle
avec 12 retards de (1-1)d_GoogleFR
modèle: (1-1)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -3.01061
statistique de test: rad_ct(1) = -3.48271
p. critique asymptot que 0.0419
Coeff. d'autocorrélation de la ordre pour e: -0.047
différences retardées: F(12, 111) = 81.457 [0.0000]
```

```
Test de Dickey-Futler augmente pour d_GOUGLEUSA testing down from 12 lags, criterion AIC taille de l'échantillon 126 hypothèse nulle de racine unitaire : a = 1

test sans constante avec 12 retards de (1-L)d_GOUGLEUSA modèle: (1-L)y = (a-1)*y(-1) + ... + e valeur estimée de (a - 1): -1.12374 statistique de test: tsancti) 2,31265 p. critique asymptotique 0,02004 Coeff. d'autocorrélative du ler estre pour e: -0.012 différences retardées: F(12, 113) = 37.978 [0.0000] test avec constante avec 12 retards de (1-L)d_GOUGLEUSA modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e valeur estimée de (a - 1): -1.36325 statistique de test: tau (1) = 0,404 p. critique asymptofique 0.1406 Coeff. d'autocorrélative du 12 ordre pour e: -0.009 différences retardées: F(12, 112) = 37.891 [0.0000] avec constante et tendance temporelle avec 12 retards de (1-L)d_GOUGLEUSA modèle: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e valeur estimée de (a - 1): -2.5542 statistique de test: tau_ct(1) = 0.24962 p. critique asymptotique 0.07489 Coeff. d'autocorrélation du les ordre pour e: 0.001 différences retardées: F(12, 111) = 39.432 [0.0000]
```

Google USA

```
Test de Dickey-Fuller augmente pour d_CVL testing down from 12 lags, criterion AIC taille de l'échantillon 126 hypothèse nulle de racine unitaire : a = 1

test sans constante avec 12 retards de (1-L)d_CVL modèle: (1-L)y = (a-1)*y(-1) + ... + e valeur estimée de (a - 1): -2.71837 statistique de test: tx mc(1) = -5,57779 p. critique asymptotique 0.0003423 Coeff. d'autocorrélation de las arde pour e: -0.037 différences retardées: F(12, 113) = 9.000 [0.0000]

test avec constante avec 12 retards de (1-L)d_CVL modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e valeur estimée de (a - 1): -3.0367 statistique de test: tr_c(1) = 7,741 p. critique asymptotique 0.003199 Coeff. d'autocorrélation du ler endre pour e: -0.031 différences retardées: F(12, 112) = 9.148 [0.0000]

avec constante et tendance temporelle avec 11 retards de (1-L)d_CVL modèle: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e valeur estimée de (a - 1): -4.07782 statistique de test: ta_c(t(1)) = -5,17502 p. critique asymptotique 8.116e-05 Coeff. d'autocorrélation de las ordre pour e: -0.028 différences retardées: F(11, 113) = 10.252 [0.0000]
```

```
Test de Dickey-Fuller augmente pour d_RL
testing down from 12 lags, criterion AIC
taille de l'échantillon 126
nypothèse nulle de racine unitaire : a = 1

test sans constante
avec 12 retards de (1-L)d_RL
modèle: (1-L)y = (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -2,72288
statistique de test: t.a.n(c1) = -3.9905
p. critique asymptotique 0.0004754
Coeff. d'autocorrélation de 1s. ordre pour e: -0.027
différences retardées: F(12, 113) = 8.845 [0.0000]

test avec constante
avec 12 retards de (1-L)d_RL
modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -2.78417
statistique de test: lac_c(1) = -3.52087
p. critique asymptotique 0.007494
Coeff. d'autocorrélation du les e dre pour e: -0.026
différences retardées: F(12, 112) = 8.798 [0.0000]

avec constante et tendance temporelle
avec 12 retards de (1-L)d_RL
modèle: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -2.88498
statistique de test: tac_c(1) = 3.52603
p. critique asymptotique 0.03659
Coeff. d'autocorrélation = 10 - .s. ne pour e: -0.024
différences retardées: F(12, 111) = 8.754 [0.0000]
```

```
VT
```

VL

Google FR

```
iest de Dickey-Fuller augmente pour d_CVI
testing down from 12 lags, criterion AIC
taille de l'échantillon 126
hypothèse nulle de racine unitaire : a = 1

test sans constante
avec 12 retards de (1-L)d_CVT
modèle: (1-L)y = (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): 1.00467
statistique de test: fau_nc(1) = 2.29885
p. critique asymptotique 0.02078
Coeff. d'autocorrélation du ler ordre pour e: 0.020
différences retardées: F(12, 113) = 4.754 [0.0000]

test avec constante
avec 12 retards de (1-L)d_CVT
modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -1.0555
statistique de test: tau_c(1) = 2.36092
p. critique asymptotique 0.1531
Coeff. d'autocorrélation du les ordre pour e: 0.021
différences retardées: F(12, 112) = 4.704 [0.0000]

avec constante et tendance temporelle
avec 12 retards de (1-L)d_CVT
modèle: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -1.52222
statistique de test: tau_t(1) = 2.71349
p. critique asymptotique 0.231
Coeff. d'autocorrélation du ler ordre pour e: 0.022
différences retardées: F(12, 111) = 4.732 [0.0000]
```

```
test de Dickey-Fuller augmente pour d_CMRI
testing down from 12 lags, criterion AIC
taille de l'échantillon 126
hypothèse nulle de racine unitaire : a = 1

test sans constante
avec 12 retards de (1-L)d_CMRT
modèle: (1-L)y = (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -4.05323
statistique de test: tau_nct(1) = **.29736
p. critique asymptotique 1.848e-05
Coeff. d'autocorrélation lu lor sare pour e: -0.028
différences retardées: F(12, 113) = 10.775 [0.0000]

test avec constante
avec 12 retards de (1-L)d_CMRT
modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -4.05894
statistique de test: tau_ct(1) = .27862
p. critique asymptotique 0.0004788
Coeff. d'autocorrélation du ler orare pour e: -0.027
différences retardées: F(12, 112) = 10.677 [0.0000]

avec constante et tendance temporelle
avec 12 retards de (1-L)d_CMRT
modèle: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -4.85547
statistique de test: tau_ct(1) = 4.77767
p. critique asymptotique 0.0004746
Coeff. d'autocorrélation du ler orare pour e: -0.029
différences retardées: F(12, 111) = 11.261 [0.0000]
```

RT

RL

Figure 23 : significativité de la

```
Test de Dickey-Fuller augmenté pour d_Google
testing down from 12 lags, criterion AIC
taille de l'échantillon 127
ypothèse nulle de racine unitaire : a = 1
  test avec constante
 test avec constante avec 11 retards de (1-L)d\_Google modèle: (1-L)y=b0+(a-1)*y(-1)+\dots+e valeur estimée de (a-1):-3.26255 statistique de test: tau_c(1)=-4.48428 p. critique asymptotique 0.000205 Coeff. d'autocorrélation du 1er ordre pour e: -0.029 différences retardées: F(11, 114)=21.965 [0.0000]
Régression du test de Dickey-Fuller augmenté
MCO, utilisant les observations 2008:02-2018:08 (T = 127)
Variable dépendante: d_d_Google
                           coefficient erreur std. t de Student p. critique
                                                  0.405318
                            -0.931813
                                                                          -2.299
                                                                                                0.0233
  const
  d_Google_1
d_d_Google_1
d_d_Google_2
                            -3.26255
                                                  0.727553
                                                                          -4.484
                                                                                                0.0002
                              1.96187
1.73746
                                                                           2.915
                                                   0.673092
                                                                                                0.0043
                                                                                                0.0051
                                                  0.608267
  d_d_Google_3
d_d_Google_4
d_d_Google_5
                              1.50188
                                                   0.544477
                                                                            2.758
                                                                                                0.0068
                                                                                                                   ***
                              1.23095
0.801244
                                                  0.484198
0.428212
                                                                            2.542
1.871
                                                                                                0.0639
  d_d_Google_6
d_d_Google_7
d_d_Google_8
                              0.667316
                                                  0.365046
                                                                            1.828
                                                                                                0.0702
                              0.533056
0.192675
                                                  0.297465
0.244482
                                                                            1.792
0.7881
                                                                                                0.0758
                                                                                                0.4323
  d_d_Google_9
d_d_Google_10
                             -0.119352
                                                  0.189695
                                                                           -0.6292
                                                                                                0.5305
                            -0.368688
-0.405602
                                                   0.134032
                                                                                                0.0069
                                                  0.0761460
                                                                                                5.09e-07
  d d Google 11
                                                                           -5.327
  AIC: 712.476 BIC: 749.451 HQC: 727.499
```

Source : Gretl, dossier Série

Figure 25 : Estimation MCO de l'ensemble des séries et test ADF sur les

```
odèle 14: MCO, utilisant les observations 2007:01-2018:08 (T = 140)
Variable dépendante: Touriste
            coefficient erreur std. t de Student
                                                        p. critique
              4997.51
                           1172.72
                                            4.261
                                                         3.74e-05
 GoogleFR
              149.575
                             17.0575
                                            8.769
                                                         5.99e-15
                                                                    ***
Moy. var. dép.
                       15156.84
                                  Éc. type var. dép.
                                                           2672.503
Somme carrés résidus
                        6.38e+08
                                  Éc. type de régression
                                                           2149.386
R2
                        0.357820
                                  R2 ajusté
                                                           0.353166
F(1, 138)
                        76.89298
                                   p. critique (F)
                                                           5.99e-15
Log de vraisemblance
                       -1271.855
                                   Critère d'Akaike
                                                           2547.711
Critère de Schwarz
                        2553.594
                                   Hannan-Ouinn
                                                           2550.102
rho
                        0.519693
                                  Durbin-Watson
                                                           0.960763
```

```
Modèle 15: MCO, utilisant les observations 2007:01-2018:08 (T = 140)
Variable dépendante: Touriste
             coefficient erreur std. t de Student p. critique
            10515.5
                          1184.02
                                         8.881
  const
                                                      3.15e-15
  G00GLEUSA
               72.2467
                            18.1249
                                          3.986
                                                      0.0001
Moy. var. dép.
                        15156.84
                                   Éc. type var. dép.
                                                           2672.503
Somme carrés résidus
                        8.90e+08
                                   Éc. type de régression
                                                           2539.935
R2
                        0.103247
                                                           0.096748
                                   R2 ajusté
F(1, 138)
                        15.88847
                                   p. critique (F)
                                                           0.000108
Log de vraisemblance
                       -1295.229
                                   Critère d'Akaike
                                                           2594.459
                        2600.342
                                   Hannan-Quinn
                                                           2596.849
Critère de Schwarz
                        0.704562
                                   Durbin-Watson
                                                           0.599719
rho
```

Google FR

```
Test de Dickey-Fuller augmente pour resGFR
testing down from 12 lags, criterion AIC
taille de l'échantillon 127
hypothèse nulle de racine unitaire : a = 1

test sans constante
avec 12 retards de (1-L)resGFR
modèle: (1-L)y = (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -0.0618151
statistique de test: tau_nc(1) = -0.581955
p. critique asymptotique 0.4655
Coeff. d'autocorrélation du ler ordre pour e: -0.087
différences retardées: F(12, 114) = 17.486 [0.0000]
```

Google USA

```
Test de Dickey-Fuller augmente pour residGUSA
testing down from 12 lags, criterion AIC
taille de l'échantillon 127
hypothèse nulle de racine unitaire : a = 1

test sans constante
avec 12 retards de (1-L)residGUSA
modèle: (1-L)y = (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -0.0278866
statistique de test: tau_nc(1) = -0.421616
p. critique asymptotique 0.5315
Coeff. d'autocorrélation du ler ordre pour e: -0.109
différences retardées: F(12, 114) = 20.166 [0.0000]
```

Modèle 16: MCO, utilisant les observations 2007:01-2018:08 (T = 140) Variable dépendante: Touriste coefficient erreur std. t de Student p. critique 4883.67 1126.39 4.336 2.78e-05 *** const 0.522160 0.0565307 4.09e-16 CVL 9.237 Moy. var. dép. 15156.84 Éc. type var. dép. 2672.503 Somme carrés résidus 6.13e+08 Éc. type de régression 2108.453 R2 ajusté 0.382047 0.377569 F(1, 138) 85.31778 p. critique (F) 4.09e-16 Log de vraisemblance Critère d'Akaike 2542.327 -1269.1642548.210 Hannan-Ouinn Critère de Schwarz 2544.718 0.715476 Durbin-Watson 0.579518

VL

Test de Dickey-Fuller augmenté pour residCVL
testing down from 12 lags, criterion AIC
taille de l'échantillon 127
hypothèse nulle de racine unitaire : a = 1

test sans constante
avec 12 retards de (1-L)residCVL
modèle: (1-L)y = (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): 0.0437327
statistique de test: tau_nc(1) = 0.716008
p. critique asymptotique 0.8697
Coeff. d'autocorrélation du ler ordre pour e: -0.071
différences retardées: F(12, 114) = 15.652 [0.0000]

Modèle 17: MCO, utilisant les observations 2007:01-2018:08 (T = 140) Variable dépendante: Touriste coefficient erreur std. t de Student p. critique 3463.37 1083.81 0.0017 const 3.196 18546.7 1698.72 10.92 2.17e-20 Moy. var. dép. 15156.84 Éc. type var. dép. 2672.503 Somme carrés résidus 5.33e+08 Éc. type de régression 1964.654 0.463463 R2 ajusté 0.459575 R2 F(1. 138) 119.2048 p. critique (F) 2.17e-20 Log de vraisemblance Critère d'Akaike -1259.2742522.548 Critère de Schwarz 2528.432 Hannan-Ouinn 2524.939 rho 0.590697 Durbin-Watson 0.817361

RL

Test de Dickey-Fuller augmenté pour residRL
testing down from 12 lags, criterion AIC
taille de l'échantillon 127
hypothèse nulle de racine unitaire : a = 1

test sans constante
avec 12 retards de (1-L)residRL
modèle: (1-L)y = (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -0.15532
statistique de test: tau_nc(1) = -2.00809
p. critique asymptotique 0.04278
Coeff. d'autocorrélation du ler ordre pour e: -0.093
différences retardées: F(12, 114) = 16.924 [0.0000]

```
dele 18: MCO, utilisant les observations 2007:01-2018:08 (T
Variable dépendante: Touriste
            coefficient erreur std. t de Student p. critique
                                                       3.46e-57 ***
            11412.7
                         420.543
                                           27.14
 const
               1.40958
                         0.144103
 CVT
                                           9.782
                                                       1.73e-17
Moy. var. dép.
                       15156.84
                                  Éc. type var. dép.
                                                         2672.503
Somme carrés résidus
                       5.86e+08
                                 Éc. type de régression
                                                         2061.161
                       0.409457
R2
                                  R2 ajusté
                                                         0.405177
F(1, 138)
                       95.68313
                                  p. critique (F)
                                                         1.73e-17
Log de vraisemblance
                                  Critère d'Akaike
                                                         2535.975
                      -1265.988
Critère de Schwarz
                       2541.859
                                  Hannan-Ouinn
                                                         2538.366
                                  Durbin-Watson
rho
                       0.619204
                                                         0.761394
```

VT

Test de Dickey-Fuller augmente pour residCVT
testing down from 12 lags, criterion AIC
taille de l'échantillon 127
hypothèse nulle de racine unitaire : a = 1

test sans constante
avec 12 retards de (1-L)residCVT
modèle: (1-L)y = (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -0.221019
statistique de test: tau_nc(1) = -2.35059
p. critique asymptotique 0.01812
Coeff. d'autocorrélation du ler ordre pour e: -0.042
différences retardées: F(12, 114) = 10.591 [0.0000]

```
Modele 19: MCO, utilisant les observations 2007:01-2018:08 (T = 140)
Variable dépendante: Touriste
             coefficient erreur std. t de Student p. critique
 const
              5385.26
                           797.076
                                            6.756
                                                        3.64e-10 ***
 CMRT
             17753.2
                                           12.50
                                                        1.90e-24
                                                                  ***
                           1420.41
Moy. var. dép.
                                  Éc. type var. dép.
                       15156.84
Somme carrés résidus
                       4.66e+08
                                  Éc. type de régression
                                                          1836.930
                                  R2 ajusté
                       0.530957
                                                          0.527558
F(1, 138)
                       156.2159
                                  p. critique (F)
                                                          1.90e-24
Log de vraisemblance
                                  Critère d'Akaike
                                                          2503.727
                      -1249.863
Critère de Schwarz
                                  Hannan-Quinn
                       2509.610
                                                          2506.117
                       0.517219
                                  Durbin-Watson
rho
                                                          0.957335
```

RT

Test de Dickey-Fuller augmenté pour residRT
testing down from 12 lags, criterion AIC
taille de l'échantillon 127
hypothèse nulle de racine unitaire : a = 1

test sans constante
avec 12 retards de (1-L)residRT
modèle: (1-L)y = (a-1)*y(-1) + ... + e
valeur estimée de (a - 1): -0.262249
statistique de test: tau_nc(1) = -2.16475
p. critique asymptotique 0.02927
Coeff. d'autocorrélation du ler ordre pour e: -0.009
différences retardées: F(12, 114) = 4.930 [0.0000]

Source : Gretl, dossier Série temporelle

Calcul du nombre d'années pour que la relation à court terme retrouve l'équilibre à la relation à long terme :

exemple pour la série RT :
$$139(1 - 0.36) = 88.96$$
 , $\frac{89}{12} = 7.41$

Figure 31 : matrice de corrélation

Coefficients de corrélation, utilisant les observations 2007:01 - 2018:08 5% valeur critique (bilatéral) = 0.1660 pour n = 140				
Touriste 1.0000	Google 0.1852 1.0000	GoogleFR 0.5982 0.6442 1.0000	G00GLEUSA 0.3213 0.9277 0.7121 1.0000	Touriste Google GoogleFR GOOGLEUSA