

MASTER 2 MODÉLISATION STATISTIQUE

PROJET DE SERIES TEMPORELLES

Mise en pratique des différentes méthodes de prévision de série temporelle

Réalisé par : Delwende Noaga Damien MASSIMBO

Supervisé par : **Mr. Yacouba BOUBACAR MAINASSARA**

Année : 2020 - 2021

Table des matières

Introduction	3
1 Lissage Exponentiel Simple	4
1.1 Mise en pratique de la methode LES	4
1.2 Fréquence Cardiaque du patient private764744641567	5
2 Lissage Exponentiel Simple sur les données brutes en série temporelle.	5
3 Modélisation d'une série : CVS vs Holt-Winters	8
3.1 Analyse préliminaire effectuée	8
3.2 Mise en oeuvre de la méthode CVS	10
3.2.1 Choix du modèle	10
3.2.2 Prévision	13
3.3 Mise en oeuvre de la méthode Holt-Winters	14
3.4 Petite conclusion sur les méthodes : CVS vs Holt-Winters	15
4 Modélisation d'une série par un ARMA	16
4.1 Analyse préliminaire	16
4.2 Mise en pratique de la méthode de Box et Jenkins	16
4.2.1 Différenciation de notre série	19
4.3 Identification du modèle	21
4.4 Estimation et la validation des modèles	22
4.5 Sélection du meilleur modèle	24
4.6 Prévision du nombre de touristes	24
5 Modélisation d'une série financière	26
5.1 Analyse préparatoire	26
5.2 Propriétés d'une série financière	27
5.2.1 Non stationnarité du cours	28
5.2.2 Autocorrélations des carrées des rendements du cours	29
5.2.3 Regroupements des extrêmes	30
5.2.4 Queues de distribution épaisses	30
5.2.5 Effet de levier	31
5.3 Modélisation de la série par : GARCH, TGARCH et EGARCH	31
5.4 Courbes d'impact des nouvelles	34
6 Conclusion	36
Références	37
Annexe	38

Introduction

Une série chronologique ou série temporelle est une suite d'observations numériques d'une variable X_t correspondant à des dates différentes pour $t = 1, 2, \dots, T$. Souvent, nos dates sont équidistantes (séries journalière, mensuelle, annuelle).

L'étude des séries chronologiques fait intervenir deux aspects, à savoir : la modélisation du phénomène dans un premier temps et dans un second temps la prévision à partir des informations du passé, une valeur pour une date à venir.

En effet, l'étude réalisée dans ce projet porte essentiellement sur quatre séries temporelles et à travers notre problématique, différentes méthodes vues en cours seront appliquées. Nous allons d'abord dans un premier lieu estimer nos modèles et ensuite, à partir de nos critères d'ajustement et de sélection, voir quel modèle pourrions nous retenir.

1 Lissage Exponentiel Simple

Les méthodes de lissage exponentiel, basées sur les travaux de Holt et Winters, demeure utile car simple à mettre en oeuvre ; En particulier le lissage exponentiel simple (LES). En effet, le lissage exponentiel simple est intuitif. C'est l'ancêtre des méthodes plus modernes de séries chronologiques. Les techniques de lissage sont utiles afin de motiver les nouveaux modèles, avec les outils vus jusqu'à maintenant.

Le lissage exponentiel simple (LES) permet d'effectuer des prévisions et s'applique à des séries chronologiques à tendance constante et sans saisonnalité. Soit \hat{Y}_T la prévision à la date T pour l'horizon h , c'est-à-dire pour la date T+h.

Dans le cadre de la methode du LES, la prévision est donnée par :

$$\hat{Y}_T(h) = \sum_{t=1}^{T-1} (1 - \alpha)^t Y_{T-t}$$

.

Où α est la constate de lissage compris entre 0 et 1.

On constate que :

- Plus la constate de lissage α est proche de 0, plus l'influence des observations passées remonte loin dans le temps et plus la prévision est rigide (peu sensible aux fluctuations conjoncturelles autrement dit aux fluctuations exceptionnelles).
- En revanche, plus α est proche de 1, plus la prévision est souple (ie influencée par les observations récentes).

Ecriture de la formule LES mise à jour

$$\hat{Y}_T(h) = \alpha Y_T + (1 - \alpha) \hat{Y}_{T-1}(h)$$

- $\alpha = 1$, la prévision est alors égale à la dernière valeur observée.
- $\alpha = 0$, la prévision est alors égale à la celle passée.

Choix du coefficient de lissage

La constate de lissage α peut être choisi de façon artisanale c'est à dire en prenant plusieurs valeurs de α et pour chaque valeur de α , on calcule les prévisions et à partir des observations et des prévisions, on calcule la Somme des Carrés des Erreurs en abrégé (SCE) en faisant la différence entre les valeurs observées et prédites de la série chronologique élevées au carré. Et, finalement on choisira le α qui donne la SCE la plus petite.

On utilise souvent un algorithme d'optimisation afin de minimiser la SCE pour obtenir une valeur de α bien plus précise.

1.1 Mise en pratique de la methode LES

Les données utilisées dans cette partie sont des données sur la fréquence cardiaque journalière du 05/07/2018 au 11/10/2018 d'un patient de PRADO INCADO (Patient suivi dans le cadre

du programme PRADO INCADO. Il bénéficie d'une télésurveillance sur une période de 2 à 6 mois en fonction du niveau de leur IC). Nous ferons une modélisation en série temporelle dans la suite de notre étude sur les données de la fréquence cardiaque de ce patient ayant comme id : **private764744641567** parmi les patients de PRADO INCADO.

1.2 Fréquence Cardiaque du patient **private764744641567**

Avant d'effectuer nos modélisations, nous ferons une statistique descriptive de la variable d'intérêt "Frequence_cardiaque" du patient **private764744641567**. Dans R, la fonction `summary` nous donne les résultats suivants :

```
Frequence_cardiaque
Min.   : 52.00
1st Qu.: 71.75
Median : 77.00
Mean   : 78.77
3rd Qu.: 81.00
Max.   :125.00
```

TABLE 1 – Statistiques sur la fréquence cardiaque du patient **private764744641567**

la fréquence cardiaque moyenne du patient **private764744641567** est de 78.77 pour un minimum de 52 et un maximum de 125.

2 Lissage Exponentiel Simple sur les données brutes en série temporelle.

Nous présentons le graphique de la série.

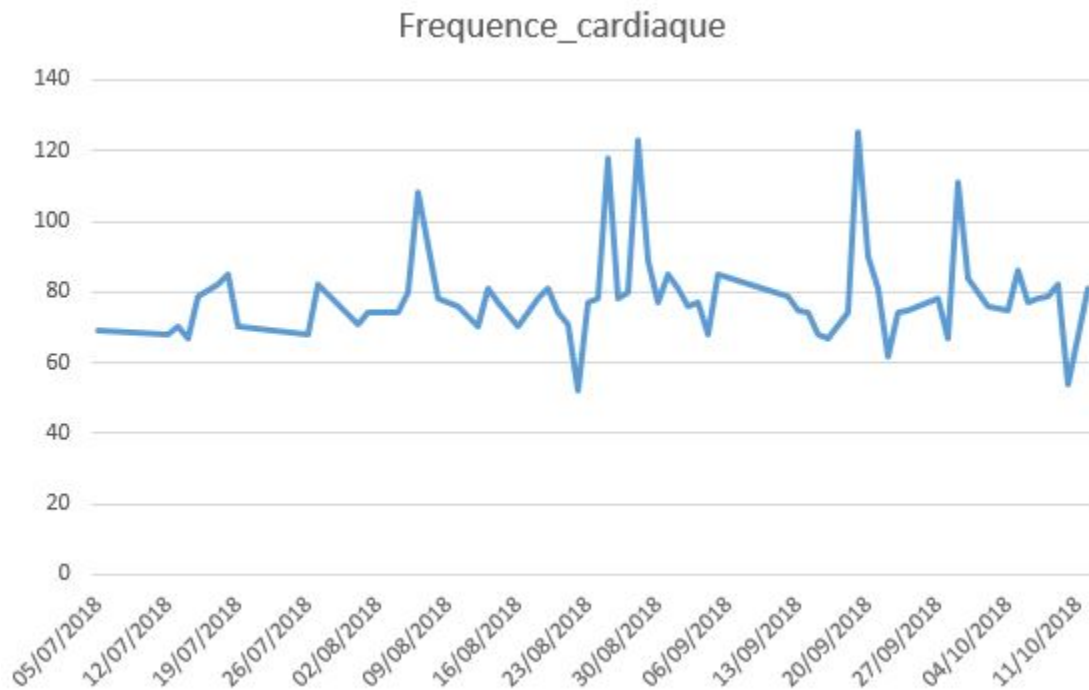


FIGURE 1 – Fréquence cardiaque du patient **private764744641567** du 05/07/2018 au 11/10/2018.

A travers le graphique de la série (figure 1), on constate qu'on a ni tendance, ni un mouvement saisonnier. On peut donc appliquer le lissage exponentiel simple. La méthode du lissage exponentiel simple a été introduit par Brown en 1992. Cette méthode s'applique à des séries sans variations saisonnières ni tendance avec ou sans changement de niveau. La série est caractérisée par des variations irrégulières autour d'un niveau localement constant.

Nous calculons donc la prévision pour $\alpha = 0.4$, ainsi que les différents erreurs de prévision et aussi l'erreur absolue moyenne en pourcentage en abrégé (MAPE) on obtient les resultats suivante :

Nous donnons les résultats des 7 derniers résultats de nos calculs vu que le tableau global est immense.

Date	Frequence_cardiaque	Previsions_FC	et	et/yt	MAPE	RMSE	alpha
05/10/2018	86	78,42762993	7,57237007	0,088050815	0,08805081	57,3407885	0,4
06/10/2018	77	81,45657796	-4,45657796	-0,05787764	0,05787764	19,8610871	
07/10/2018	78	79,67394677	-1,67394677	-0,02146086	0,02146086	2,8020978	
08/10/2018	79	79,00436806	-0,00436806	-5,5292E-05	5,5292E-05	1,908E-05	
09/10/2018	82	79,00262084	2,99737916	0,036553404	0,0365534	8,98428184	
10/10/2018	54	80,2015725	-26,2015725	-0,48521431	0,48521431	686,522402	
11/10/2018	81	69,7209435	11,2790565	0,139247611	0,13924761	127,217115	
		74,2325661					
					0,11624577	13707,5531	

TABLE 2 – Prévisions et les erreurs pour $\alpha = 0.4$

La prévision de la mesure de la fréquence cardiaque du patient private764744641567 pour la date correspondant au 12/10/2018 est de 74.2325.

Pour choisir la meilleure valeur de α , nous allons choisir la valeur de α qui minimise le MAPE en utilisant le solveur de l'Excel. Avec le solveur d'Excel, on obtient une constante de lissage égale à 0.0379 pour un MAPE correspondant à 9.66% . Finalement, on obtient la prévision suivante :

Date	Frequence_cardiaque	Previsions_FC	et	et/yt	MAPE	RMSE	alpha
04/10/2018	75	79,00030492	-4,00030492	-0,0533374	0,0533374	16,0024394	0,0379
05/10/2018	86	78,8483194	7,1516806	0,083159077	0,08315908	51,1465354	
06/10/2018	77	79,12003666	-2,12003666	-0,02753294	0,02753294	4,49455542	
07/10/2018	78	79,03948908	-1,03948908	-0,01332678	0,01332678	1,08053755	
08/10/2018	79	78,99999527	4,7301E-06	5,98753E-08	5,9875E-08	2,2374E-11	
09/10/2018	82	78,99999545	3,00000455	0,036585421	0,03658542	9,0000273	
10/10/2018	54	79,11397607	-25,1139761	-0,46507363	0,46507363	630,711794	
11/10/2018	81	78,15980865	2,84019135	0,035064091	0,03506409	8,06668688	
		78,26771742			0,09660557	12544,779	

TABLE 3 – Prévisions et les erreurs pour $\alpha = 0.0379$

La prévision de la mesure de la fréquence cardiaque du patient private764744641567 pour la date correspondant au 12/10/2018 est de 78.2677 pour la constante de lissage égale à 0.0379 obtenue avec le solveur Excel.

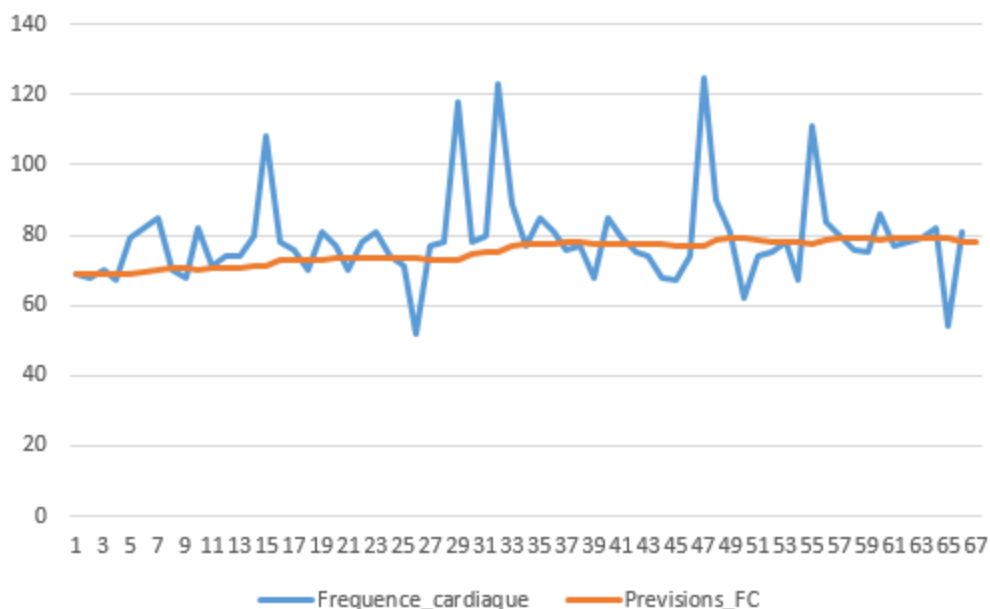


FIGURE 2

3 Modélisation d'une série : CVS vs Holt-Winters

3.1 Analyse préliminaire effectuée

Les données utilisées dans le cadre de ce projet sont des données sur les ventes en millions de dollars de bière, de vin et de boissons alcoolisées distillées (Tout ce qui n'est pas vin, bière ou cidre et qui titre plus de 15 % Vol d'alcool se nomme boisson alcoolisée distillée ou encore spiritueux. Dans cette catégorie, on rassemble tous les apéritifs, les alcools et les liqueurs) aux Etats-unis par les marchands grossistes. Nous disposons donc de données mensuelles sur les ventes des boissons alcoolisées en millions de dollars sur plusieurs années débutant à partir de Janvier 1992 à Decembre 2003. Avant d'effectuer nos modélisations, nous ferons une statistique descriptive de la variable d'intérêt "Alcohol_sales" désignant la vente d'alcool. Dans R, la fonction summary nous donne les résultats suivants :

	Alcohol_sales
Min.	:3031
1st Qu.	:4376
Median	:5060
Mean	:5259
3rd Qu.	:6155
Max.	:8540

TABLE 4 – Statistiques sur la vente des boissons alcoolisées.

La vente moyenne des boissons alcoolisées pendant ces 12 années s'élèvent autour de 5259 millions de dollars pour un minimum de 3031 millions dollars et un maximum de 8540 millions de dollars.

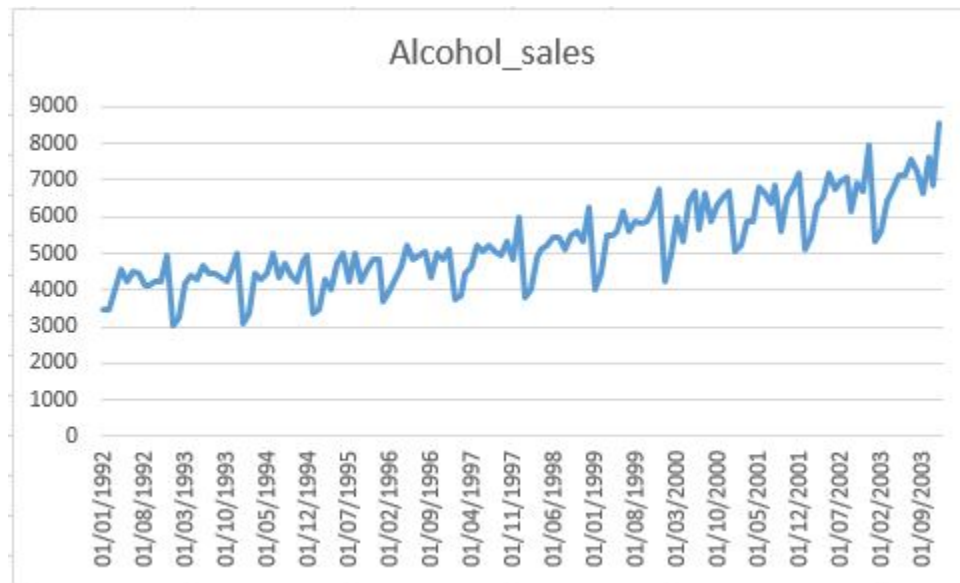


FIGURE 3 – Vente des boissons alcoolisées en millions de dollars en fonction des années.

A travers le graphique de la série (figure 1), on constate l'existence d'une tendance croissante et un mouvement saisonnier de période 12. Cette croissance peut se justifier par une vente accrue des boissons alcoolisées. Des pics sont observés tous les mois de : Décembre, Juin, Aout avec des creux observés notamment durant les mois de : Janvier, Février. Ces pics peuvent s'expliquer par les fêtes de fin d'année en décembre (la fête de Noël et du nouvel an) et également et également les vacances qui touchent le mois d'Aout. Nous émettons donc l'hypothèse suivante sur la saisonnalité : afin de mieux rentabiliser, on note une vente assez élevée des boissons alcoolisées par les marchands grossistes durant ces périodes. En revanche, on constate une baisse de la vente des boissons alcoolisées et ceci stabilise peu les ventes à cause des mois hors saison favorable.

Dans la suite de notre étude, dû à l'examen visuel effectué précédemment, nous appliquerons sur notre série les deux méthodes suivantes : la Correction des Variations Saisonnières (CVS) et le lissage de Holt-Winters. Et ceux, à partir des critères de mesure de la qualité de prédiction qui seront définis nous allons pouvoir dire, parmi ces deux méthodes, celle qui conduira à l'obtention des meilleures prévisions.

3.2 Mise en oeuvre de la méthode CVS

Présentation de nos données sous la forme d'un tableau de Buys-Ballot :

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	Moy
Janv	3459	3031	3075	3370	3699	3758	3796	4032	4243	5082	5117	5326	3999
Févr	3458	3261	3377	3484	3983	3825	4019	4435	4952	5216	5442	5609	4255,08
Mars	4002	4160	4443	4269	4262	4454	4898	5479	6008	5893	6337	6414	5051,58
Avril	4564	4377	4261	3994	4619	4635	5090	5483	5353	5894	6525	6741	5128
Mai	4221	4307	4460	4715	5219	5210	5237	5587	6435	6799	7216	7144	5545,83
Juin	4529	4696	4985	4974	4836	5057	5447	6176	6673	6667	6761	7133	5661,17
Juillet	4466	4458	4324	4223	4941	5231	5435	5621	5636	6374	6958	7568	5436,25
Aout	4137	4457	4719	5000	5062	5034	5107	5889	6630	6840	7070	7266	5600,92
Sep	4126	4364	4374	4235	4365	4970	5515	5828	5887	5575	6148	6634	5168,42
Oct	4259	4236	4248	4554	5012	5342	5583	5849	6322	6545	6924	7626	5541,67
Nov	4240	4500	4784	4851	4850	4831	5346	6180	6520	6789	6716	6843	5537,5
Dec	4936	4974	4971	4826	5097	5965	6286	6771	6678	7180	7975	8540	6183,25
Moy	4199,75	4235,08	4335,08	4374,58	4662,08	4859,33	5146,58	5610,83	5944,75	6237,83	6599,08	6903,67	

TABLE 5 – Tableau de Buys-Ballot.

On constate que les pics sont en moyenne en Juin, Août, et Decembre (en bleue) et les creux sont en rouge.

3.2.1 Choix du modèle

Nous traçons les deux courbes joignant les maxima (minima) distants de période 12 et voyons quel est le modèle le mieux adapté ?



FIGURE 4 – Vente des boissons alcoolisées en millions de dollars en fonction des années.

A travers le graphique ci-dessus, on constate qu'un modèle multiplicatif semble mieux adapté.

Nous superposons en pointillés l'équation de la tendance dans le graphe ci-dessus, après estimation des coefficients dans celle-ci.

Nous calculons les valeurs des tendances, à partir de l'équation $T_t = 22.74t + 3610.4$.

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Janv	3633,14	3906,02	4178,9	4451,78	4724,66	4997,54	5270,42	5543,3	5816,18	6089,06	6361,94	6634,82
Févr	3655,88	3928,76	4201,64	4474,52	4747,4	5020,28	5293,16	5566,04	5838,92	6111,8	6384,68	6657,56
Mars	3678,62	3951,5	4224,38	4497,26	4770,14	5043,02	5315,9	5588,78	5861,66	6134,54	6407,42	6680,3
Avril	3701,36	3974,24	4247,12	4520	4792,88	5065,76	5338,64	5611,52	5884,4	6157,28	6430,16	6703,04
Mai	3724,1	3996,98	4269,86	4542,74	4815,62	5088,5	5361,38	5634,26	5907,14	6180,02	6452,9	6725,78
Juin	3746,84	4019,72	4292,6	4565,48	4838,36	5111,24	5384,12	5657	5929,88	6202,76	6475,64	6748,52
Juillet	3769,58	4042,46	4315,34	4588,22	4861,1	5133,98	5406,86	5679,74	5952,62	6225,5	6498,38	6771,26
Aout	3792,32	4065,2	4338,08	4610,96	4883,84	5156,72	5429,6	5702,48	5975,36	6248,24	6521,12	6794
Sep	3815,06	4087,94	4360,82	4633,7	4906,58	5179,46	5452,34	5725,22	5998,1	6270,98	6543,86	6816,74
Oct	3837,8	4110,68	4383,56	4656,44	4929,32	5202,2	5475,08	5747,96	6020,84	6293,72	6566,6	6839,48
Nov	3860,54	4133,42	4406,3	4679,18	4952,06	5224,94	5497,82	5770,7	6043,58	6316,46	6589,34	6862,22
Dec	3883,28	4156,16	4429,04	4701,92	4974,8	5247,68	5520,56	5793,44	6066,32	6339,2	6612,08	6884,96

TABLE 6 – Valeur des tendances, $T_t = 22.74t + 3610.4$, mises sous la forme du tableau de Buys-Ballot.

Comme nous sommes dans le cas multiplicatif, nous faisons Y_t/T_t Pour mettre nos données sous la forme d'un tableau de Buys-Ballot sans la tendance.

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	moyennes	Calcul It
Janv	0,95207	0,77598	0,73584	0,757	0,78291	0,75197	0,72025	0,72736	0,72952	0,83461	0,80431	0,80273	0,7812135	0,78029
Févr	0,94587	0,83003	0,80373	0,77863	0,83899	0,76191	0,75928	0,7968	0,8481	0,85343	0,85235	0,8425	0,8259693	0,825
Mars	1,08791	1,05276	1,05175	0,94924	0,89347	0,8832	0,92139	0,98036	1,02497	0,96063	0,98901	0,96014	0,9795689	0,97842
Avril	1,23306	1,10134	1,00327	0,88363	0,96372	0,91497	0,95343	0,9771	0,90969	0,95724	1,01475	1,00566	0,9931547	0,99199
Mai	1,13343	1,07756	1,04453	1,03792	1,08376	1,02388	0,9768	0,99161	1,08936	1,10016	1,11826	1,06218	1,0616212	1,06037
Juin	1,20875	1,16824	1,1613	1,08948	0,99951	0,98939	1,01168	1,09174	1,12532	1,07484	1,04407	1,05697	1,0851082	1,08383
Juillet	1,18475	1,10279	1,00201	0,9204	1,01644	1,0189	1,0052	0,98966	0,94681	1,02385	1,07073	1,11766	1,0332668	1,03205
Aout	1,09089	1,09638	1,08781	1,08437	1,03648	0,9762	0,94058	1,03271	1,10956	1,09471	1,08417	1,06947	1,058611	1,05737
Sep	1,0815	1,06753	1,00302	0,91396	0,88962	0,95956	1,01149	1,01795	0,98148	0,88902	0,93951	0,97319	0,9773192	0,97617
Oct	1,10975	1,03049	0,96908	0,978	1,01677	1,02687	1,01971	1,01758	1,05002	1,03993	1,05443	1,115	1,0356348	1,03442
Nov	1,09829	1,08869	1,08572	1,03672	0,97939	0,9246	0,97239	1,07093	1,07883	1,07481	1,01922	0,9972	1,0355656	1,03435
Dec	1,27109	1,19678	1,12237	1,02639	1,02456	1,13669	1,13865	1,16874	1,10083	1,13264	1,20613	1,24038	1,1471038	1,14575
												moy2	1,0011781	1

TABLE 7 – Les données sans tendance : Y_t/T_t , mises sous la forme du tableau de Buys-Ballot

Pour obtenir les coefficients saisonniers I_t , on calcule la moyenne des données sans tendance pour chaque mois et on corrige ensuite les coefficients obtenus pour qu'ils aient une somme égale à $(\sum_{i=1}^{12} I_t = 12)$. Les coefficients saisonniers I_t sont obtenus donc, en divisant par m les moyennes mensuelles. Les coefficients I_t vont s'appliquer de manière égale tous les ans.

Nous obtenons donc le tableau de la série corrigée des variations saisonnières CVS ci-dessous en faisant Y_t/I_t .

	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
Janv	4432,94	3884,43	3940,82	4318,88	4740,52	4816,13	4864,83	5167,28	5437,69	6512,93	6557,78	6825,63
Févr	4191,53	3952,74	4093,35	4223,04	4827,89	4636,38	4871,53	5375,77	6002,44	6322,44	6596,38	6798,81
Mars	4090,28	4251,77	4541,01	4363,17	4356,02	4552,25	5006,05	5599,87	6140,54	6023	6476,79	6555,49
Avril	4600,87	4412,36	4295,42	4026,27	4656,32	4672,44	5131,12	5527,3	5396,24	5941,62	6577,71	6795,46
Mai	3980,68	4061,78	4206,07	4446,55	4921,86	4913,37	4938,83	5268,91	6068,63	6411,9	6805,16	6737,26
Juin	4178,69	4332,78	4599,42	4589,27	4461,95	4665,85	5025,69	5698,3	6156,86	6151,33	6238,06	6581,28
Juillet	4327,31	4319,55	4189,72	4091,85	4787,55	5068,55	5266,21	5446,44	5460,97	6176,05	6741,91	7332,97
Aout	3912,56	4215,19	4462,98	4728,73	4787,37	4760,89	4829,93	5569,5	6270,3	6468,91	6686,43	6871,8
Sep	4226,73	4470,54	4480,78	4338,39	4471,56	5091,33	5649,64	5970,28	6030,72	5711,1	6298,09	6795,95
Oct	4117,3	4095,06	4106,66	4402,48	4845,25	5164,27	5397,25	5654,4	6111,66	6327,24	6693,63	7372,27
Nov	4099,2	4350,57	4625,14	4689,92	4688,95	4670,58	5168,48	5974,78	6303,49	6563,56	6492,99	6615,77
Dec	4308,08	4341,25	4338,63	4212,07	4448,6	5206,18	5486,34	5909,65	5828,48	6266,62	6960,48	7453,61

TABLE 8 – La série corrigée des variations saisonnières CVS

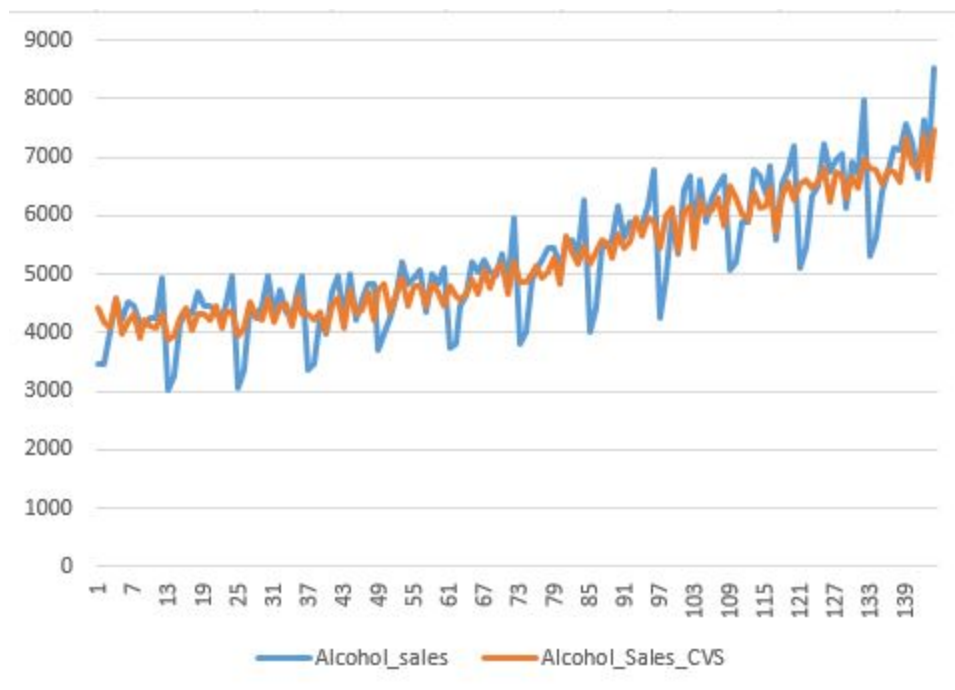


FIGURE 5 – Vente des boissons alcoolisées en millions de dollars en fonction des années.

On constate une disparition des pics et des creux et en ré-évaluant la tendance de la série corrigée des variations saisonnières, on obtient une nouvelle équation de la tendance :

$$T_t = 21.936t + 3662.6.$$

Nous calculons à nouveau les coefficients saisonniers et on constate qu'un nouveau calcul de ceux-ci n'apporte pas de modification sensible à deux décimales. Les coefficients saisonniers obtenus sont :

$$I_{Janvier} = 0.7795, I_{Février} = 0.8243, I_{Mars} = 0.9777, I_{Avril} = 0.9913,$$

$$I_{Mai} = 1.0601, I_{Juin} = 1.0835, I_{Juillet} = 1.0321, I_{Août} = 1.0576,$$

$$I_{Septembre} = 0.9764, I_{Octobre} = 1.0351, I_{Novembre} = 1.0351, I_{Decembre} = 1.1468.$$

3.2.2 Prédiction

Pour l'obtention d'une prédiction pour l'année suivante, nous calculons la tendance pour $T+1, T+2, \dots, T+12$, puis, on multiplie la valeur obtenue par le coefficient saisonnier correspondant : $\hat{Y} = T_t * I_t$. Nous avons à notre disposition des observations de l'année 2004.

Nous proposons un recapitulatif des prévisions dans le tableau suivant, ainsi que le graphe.

	Alcohol_sales	Prédiction
01/01/2004	5629	5334,37
01/02/2004	5898	5659,04
01/03/2004	7045	6733,61
01/04/2004	7094	6849,02
01/05/2004	7333	7347,63
01/06/2004	7918	7533,58
01/07/2004	7289	7198,84
01/08/2004	7396	7399,9
01/09/2004	7259	6853,17
01/10/2004	7268	7287,88
01/11/2004	7731	7310,59
01/12/2004	9058	8124,64

TABLE 9 – Prédiction 2004 de la vente des boissons alcoolisées en millions de dollars en fonction des années. Cas du modèle multiplicatif.

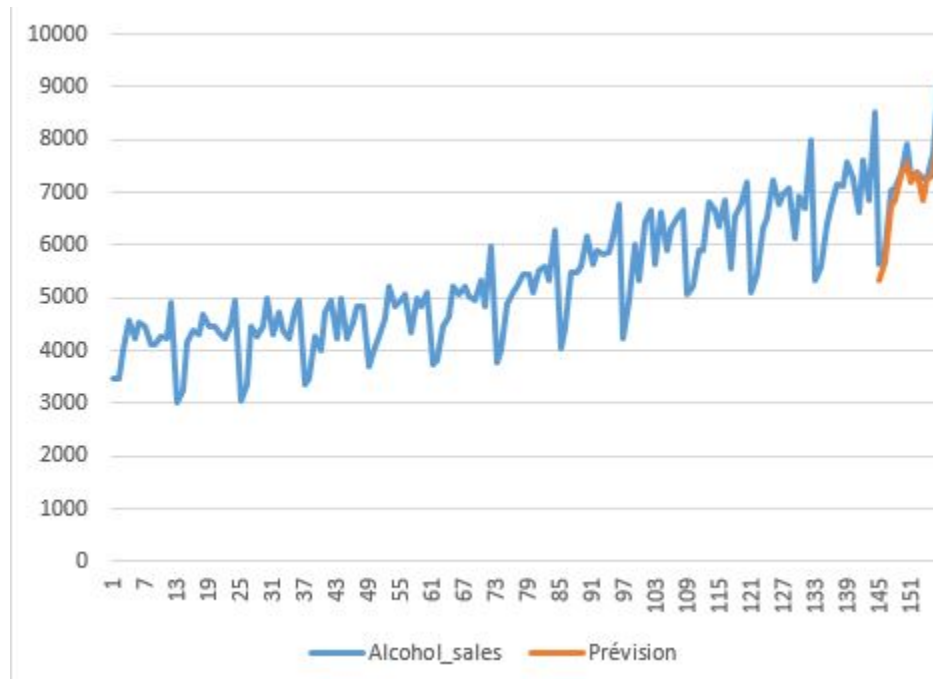


FIGURE 6 – Prédiction 2004 de la vente des boissons alcoolisées en millions de dollars en fonction des années. Cas du modèle multiplicatif.

Nous donnons dans le tableau ci-dessous l'erreur moyenne absolue en pourcentage (MAPE) et l'erreur quadratique moyenne (Root Mean Square Error en abrégé RMSE) pour la méthode CVS sous un modèle multiplicatif.

MAPE	4,70%
RMSE	300,816

TABLE 10 – Valeurs RMSE et MAPE pour la méthode CVS sous un modèle multiplicatif

3.3 Mise en oeuvre de la méthode Holt-Winters

Comme le modèle multiplicatif était mieux adapté à nos données, nous allons présenter la méthode de Holt-Winters seulement dans le cas multiplicatif dans cette partie. Nous présentons le graphe avec les prévisions réalisés pour l'année 2004.

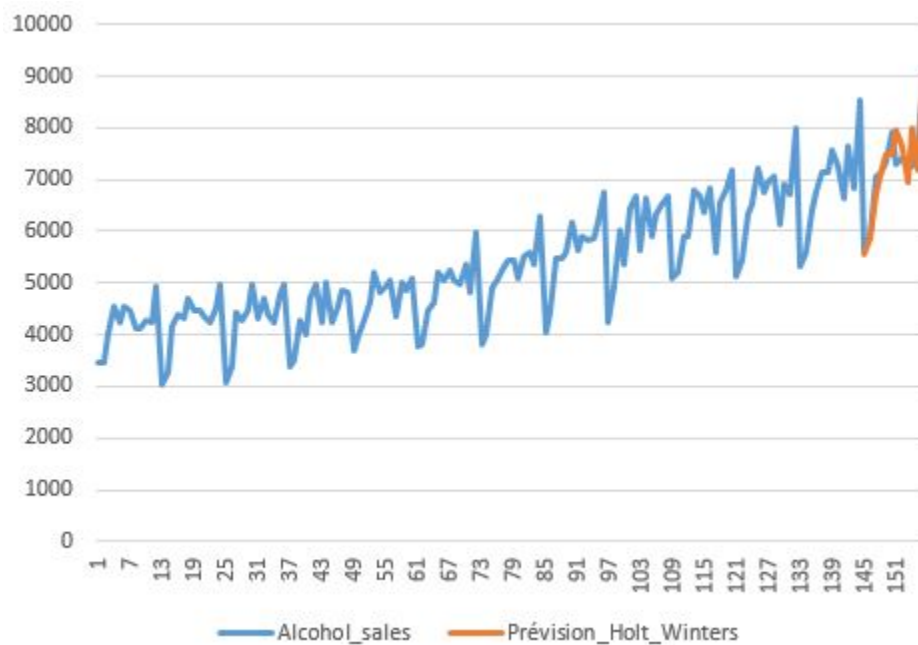


FIGURE 7 – Prévision 2004 de la vente des boissons alcoolisées en millions de dollars en fonction des années. Cas du modèle multiplicatif.

Avec le solveur du logiciel Excel nous avons déterminé d'une part les valeurs optimales pour les constantes de lissage selon le MAPE et de d'autre part selon le RMSE.

Ci-dessous le tableau :

		alpha	gamma	delta
MAPE	3,77%	0,01135784	0,39587552	0,99999777
RMSE	265,612949	0,01841035	0,30379976	0,99546422

TABLE 11 – Valeurs optimales pour les constantes de lissage avec le solveur du logiciel Excel.

3.4 Petite conclusion sur les méthodes : CVS vs Holt-Winters

En considérant le MAPE et le RMSE comme critère, on constate que, parmi les deux modélisations : modélisations CVS et Holt-Winters, la modélisation Holt-Winters montre une meilleure performance sur notre jeu de données.

4 Modélisation d'une série par un ARMA

4.1 Analyse préliminaire

Les touristes se demandent souvent pourquoi ne pas vivre à Portland (plus grande ville de l'Oregon, État du nord-ouest des États-Unis) après avoir visité cette ville tranquille et conviviale, réputée pour ses parcs soignés, sa vie nocturne trépidante, ses petites brasseries et distilleries. Cette fois ci, notre étude portera sur le nombre de touristes de passage mensuel pour le système de transport public de Portland. Dans la suite de notre étude, nous ferons une statistique descriptive de la variable "Passagers" à partir des observations de 1960 : 01 - 1965 : 08 soit (56 observations valides). Dans Eviews, la fonction summary nous donne les résultats suivants :

	PASSAGERS
Mean	893.6250
Median	876.5000
Maximum	1209.000
Minimum	613.0000
Std. Dev.	185.5714

TABLE 12 – Statistiques sur le nombre de touristes

Le nombre maximale de touristes dans la plus grande ville de l'Oregon pendant ces 5 années est de 1209 pour un minimum de 613. Dans la suite de notre étude, nous nous servirons de la démarche de la méthode de Box-Jenkins pour mieux capter le meilleur modèle qui sera utilisé pour effectuer nos prévisions. Avant toute tentative de modélisation, nous trouvons qu'il serait bien de présenter cette méthodologie de Box-Jenkins.

La méthodologie de Box et Jenkins comporte essentiellement les étapes suivantes :

- Transformation des données afin de stabiliser la variance (log, etc) et différenciation des données pour les stationariser.
- Visualisation des fonction d'autocorrélation et des fonction d'autocorrélation partiel empiriques pour identifier les paramètres p et q appropriés.
- Estimation des paramètres du modèle sélectionné.
- Diagnostique et tests d'adéquation du modèle.
- Prévision : La dernière étape consiste à la prévision des valeurs futures à travers le modèle retenu.

4.2 Mise en pratique de la méthode de Box et Jenkins

Représentation graphique de notre série brute

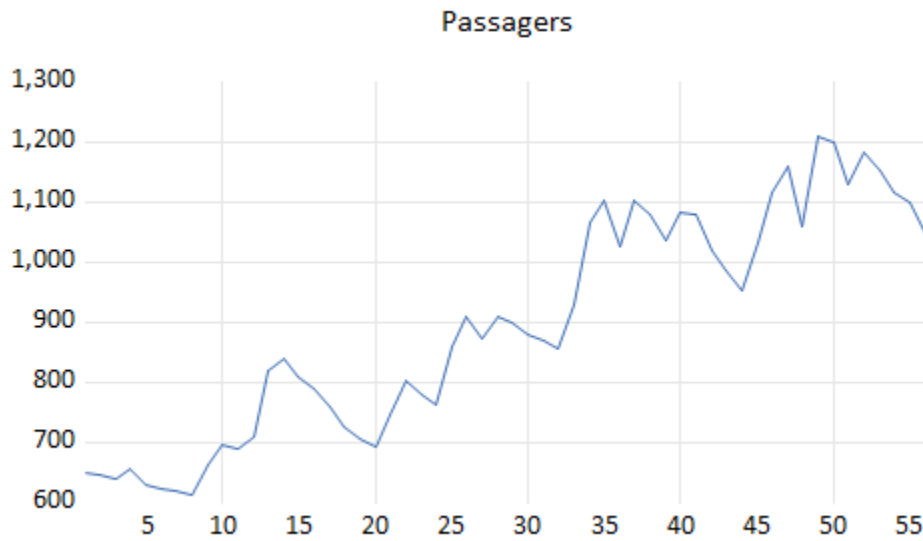


FIGURE 8 – Série brute

Notre série présente une tendance globalement croissante et il se peut qu'on se dise dans ce cas que notre série n'est pas stationnaire. Pour ne pas se laisser tromper par une analyse graphique, nous faisons appel au test de Dickey-Fuller. Nous nous intéressons également à la tendance dans ce test.

Stationnarisation

En ce qui concerne l'exploration et la préparation des données, on commencera par aborder la notion de stationnarité, concept essentiel en analyse temporelle. Au même titre que la régression linéaire, où nous regardons la corrélation des variables avant d'effectuer la régression pour éviter des estimations trompeuses, ici avant de modéliser notre série, nous effectuons au préalable la stationnarisation de notre série.

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-2.937746	0.1592
Test critical values:		
1% level	-4.133838	
5% level	-3.493692	
10% level	-3.175693	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(PASSAGERS)
Method: Least Squares
Date: 12/25/20 Time: 10:46
Sample (adjusted): 2 56
Included observations: 55 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
PASSAGERS(-1)	-0.327791	0.111579	-2.937746	0.0049
C	203.7849	66.48529	3.065113	0.0034
@TREND("1")	3.408621	1.296398	2.629302	0.0112

R-squared	0.145275	Mean dependent var	7.200000
Adjusted R-squared	0.112401	S.D. dependent var	52.73533
S.E. of regression	49.68327	Akaike info criterion	10.70221
Sum squared resid	128358.2	Schwarz criterion	10.81171
Log likelihood	-291.3109	Hannan-Quinn criter.	10.74456
F-statistic	4.419127	Durbin-Watson stat	1.720569
Prob(F-statistic)	0.016885		

FIGURE 9 – Test de Dickey Fuller

Le test augmenté de Dickey-Fuller ou test ADF est un test statistique qui vise à savoir si une série temporelle est stationnaire. La série a une racine unitaire puisque la p-value associée au test est supérieure à 5 % , cela implique que la série n'est pas stationnaire. De plus, on voit aussi que la tendance est significative ($p\text{-valeur}=0.0112 < 0.05$). Nous voyons bien que la série n'est pas stationnaire et pour la rendre stationnaire, on va d'abord estimer les coefficients de la tendance de la série **Passagers** puis on la retire, pour cela on utilise une nouvelle série nommée **TPassagers**.

On utilise encore le test Dickey-Fuller pour faire le constat de la disparition de la tendance.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
TPASSAGERS(-1)	-0.327791	0.111579	-2.937746	0.0049
C	0.083870	13.59734	0.006168	0.9951
@TREND("1")	-0.097234	0.423258	-0.229727	0.8192
R-squared	0.145275	Mean dependent var	-3.495386	
Adjusted R-squared	0.112401	S.D. dependent var	52.73533	
S.E. of regression	49.68327	Akaike info criterion	10.70221	
Sum squared resid	128358.2	Schwarz criterion	10.81171	
Log likelihood	-291.3109	Hannan-Quinn criter.	10.74456	
F-statistic	4.419127	Durbin-Watson stat	1.720569	
Prob(F-statistic)	0.016885			

FIGURE 10 – Test de Dickey Fuller de la série tpassagers

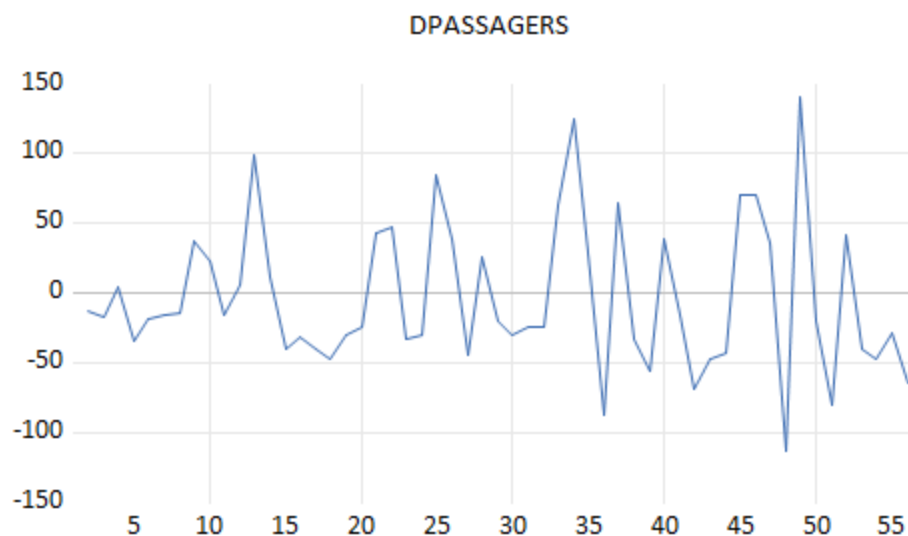
Les deux p-valeurs pour la tendance et la constante sont largement supérieur a 5% . D'où la tendance et la constante sont non significatives. La série n'est toujours pas stationnaire, nous allons résoudre ce problème dans la suite.

Dans le but de rendre notre série stationnaire, nous allons procéder à une différenciation et on nommera la série **dpassagers**.

$$dpassagers = tpassagers - tpassagers(-1)$$

4.2.1 Différenciation de notre série

Nous donnons une représentation graphique de notre série **dpassagers** différencié.

**FIGURE 11** – Série brute

On se base à nouveau sur le test de le test Dickey-Fuller pour vérifier la stationnarité de

notre série nommée **dpassagers**. Nous obtenons les résultats suivant pour ce test :

	t-Statistic	Prob.*		
Augmented Dickey-Fuller test statistic	-7.239341	0.0000		
Test critical values:				
1% level	-4.137279			
5% level	-3.495295			
10% level	-3.176618			
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DPASSAGERS)				
Method: Least Squares				
Date: 12/25/20 Time: 15:59				
Sample (adjusted): 3 56				
Included observations: 54 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DPASSAGERS(-1)	-1.024279	0.141488	-7.239341	0.0000
C	3.013739	15.34818	0.196358	0.8451
@TREND("1")	-0.224425	0.472612	-0.474860	0.6369
R-squared	0.507273	Mean dependent var	-0.962963	
Adjusted R-squared	0.487951	S.D. dependent var	75.62493	
S.E. of regression	54.11541	Akaike info criterion	10.87407	
Sum squared resid	149352.4	Schwarz criterion	10.98457	
Log likelihood	-290.5998	Hannan-Quinn criter.	10.91668	
F-statistic	26.25282	Durbin-Watson stat	1.991742	
Prob(F-statistic)	0.000000			

FIGURE 12 – Test de Dickey Fuller de la série **dpassagers**

En se basant sur le test de Dickey-Fuller, on rejette l'hypothèse nulle (H_0 : série non stationnaire), cela confirme bel et bien que la série obtenue est une série stationnaire puisque ($p\text{-value} < 0.05$). Donc la série **dpassagers** est stationnaire à notre seuil pris.

4.3 Identification du modèle

Comme notre série est stationnaire, nous procédons à la recherche d'un modèle ARMA (p,q). Nous nous servons du corrélogramme de la série stationnaire **dpassagers** Pour connaître les ordres du modèle ARMA (p,q). Ainsi, pour identifier le processus qui représente au mieux notre série, on examine les autocorrélations simples et l'autocorrélation partiel.

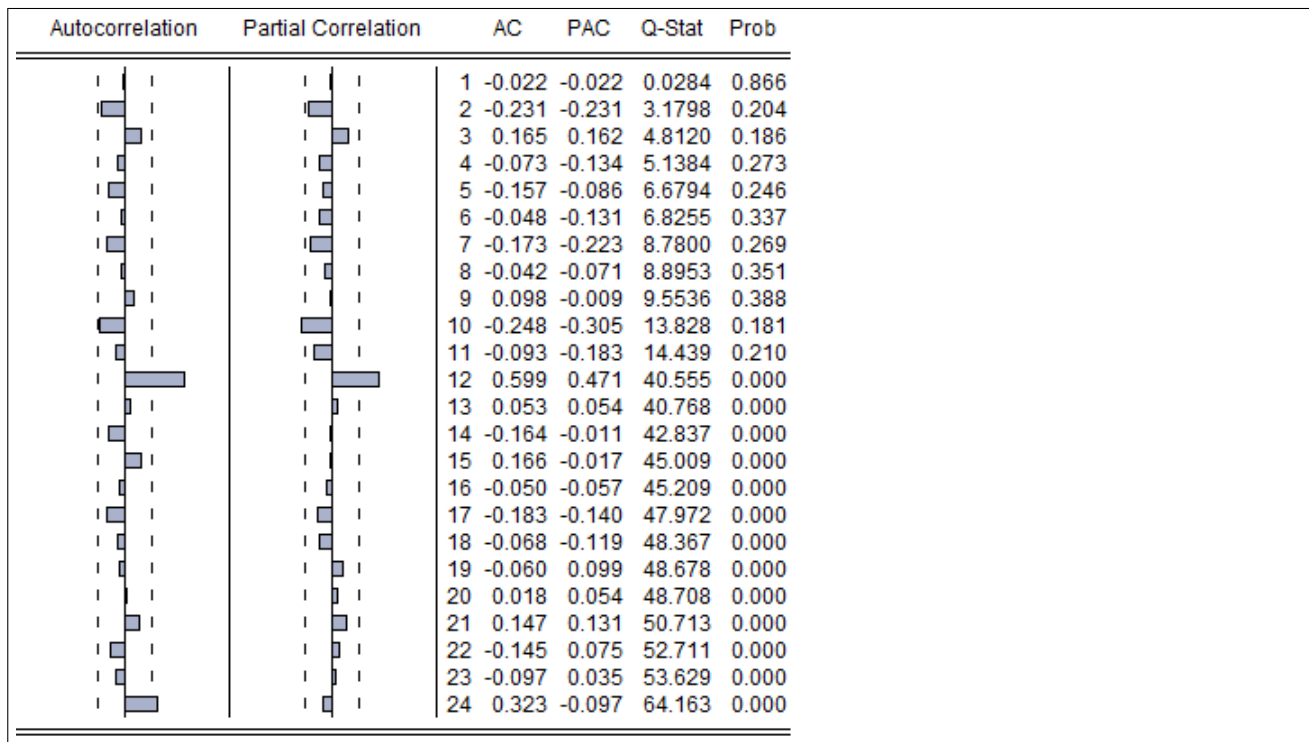


FIGURE 13 – Corrélogramme de la série dpassagers

Le corrélogramme de la fonction d'autocorrélation partielle montre qu'on a une perte d'information aux niveaux 10, 12.

Le corrélogramme de la fonction d'autocorrélation montre qu'on a une perte d'information au niveau 12.

Nous pouvons donc proposer trois modèles dans le but d'effectuer une prévision après :

- AR(12)
- ARMA(10,12)
- ARMA(12,12)

4.4 Estimation et la validation des modèles

Nous allons donc maintenant procéder à l'estimation des paramètres pour les différents modèles choisis. Pour cela, nous commençons d'abord avec AR(12)

Modèle AR(12)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	0.820892	0.081720	10.04512	0.0000
SIGMASQ	957.2084	180.8189	5.293742	0.0000
R-squared	0.649432	Mean dependent var		-3.495386
Adjusted R-squared	0.642818	S.D. dependent var		52.73533
S.E. of regression	31.51713	Akaike info criterion		10.01909
Sum squared resid	52646.46	Schwarz criterion		10.09208
Log likelihood	-273.5248	Hannan-Quinn criter.		10.04731
Durbin-Watson stat	1.812354			
Inverted AR Roots	.98	.85+.49i	.85-.49i	.49-.85i
	.49+.85i	.00+.98i	-.00-.98i	-.49-.85i
	-.49+.85i	-.85+.49i	-.85-.49i	-.98

FIGURE 14 – Estimation du modèle du modèle AR(12)

On constate que les coefficients sont significativement différents de 0, car on a t-statistic largement supérieur à 1.96 et aussi on voit que la probabilité liée est inférieure à 5%. On garde alors ce modèle pour le test sur les résidus.

Pour le modèle ARMA(10,12) :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(10)	-0.266747	0.145114	-1.838189	0.0717
MA(12)	1.000000	5786.328	0.000173	0.9999
SIGMASQ	1034.126	2991833.	0.000346	0.9997
R-squared	0.621262	Mean dependent var		-3.495386
Adjusted R-squared	0.606695	S.D. dependent var		52.73533
S.E. of regression	33.07246	Akaike info criterion		10.27571
Sum squared resid	56876.94	Schwarz criterion		10.38520
Log likelihood	-279.5821	Hannan-Quinn criter.		10.31805
Durbin-Watson stat	2.025407			

FIGURE 15 – Estimation du modèle du modèle ARMA(10,12)

Nous constatons que les probabilités liées à AR(10) et MA(12) sont supérieures à 5% d'où les coefficients ne sont pas significatives. Nous n'allons pas dans la suite de notre étude garder ce modèle pour les tests sur les résidus.

Pour le modèle ARMA(12,12) :

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(12)	0.856570	0.115178	7.436948	0.0000
MA(12)	-0.106193	0.216326	-0.490893	0.6256
SIGMASQ	948.1689	193.4866	4.900437	0.0000
R-squared	0.652743	Mean dependent var	-3.495386	
Adjusted R-squared	0.639387	S.D. dependent var	52.73533	
S.E. of regression	31.66814	Akaike info criterion	10.05105	
Sum squared resid	52149.29	Schwarz criterion	10.16054	
Log likelihood	-273.4040	Hannan-Quinn criter.	10.09339	
Durbin-Watson stat	1.846008			

FIGURE 16 – Estimation du modèle du modèle ARMA(12,12)

Nous constatons que la probabilité liée à MA(12) est supérieur à 5% d'où le coefficient n'est pas significative. Nous n'allons pas dans la suite de notre étude garder ce modèle pour les tests sur les résidus également.

4.5 Sélection du meilleur modèle

On a un seul modèle valide AR(12), donc on retient bel et bien ce modèle.

Nous allons maintenant vérifier l'hypothèse du bruit blanc des résidus. Pour cela, on utilisera le test de Ljung-Box.

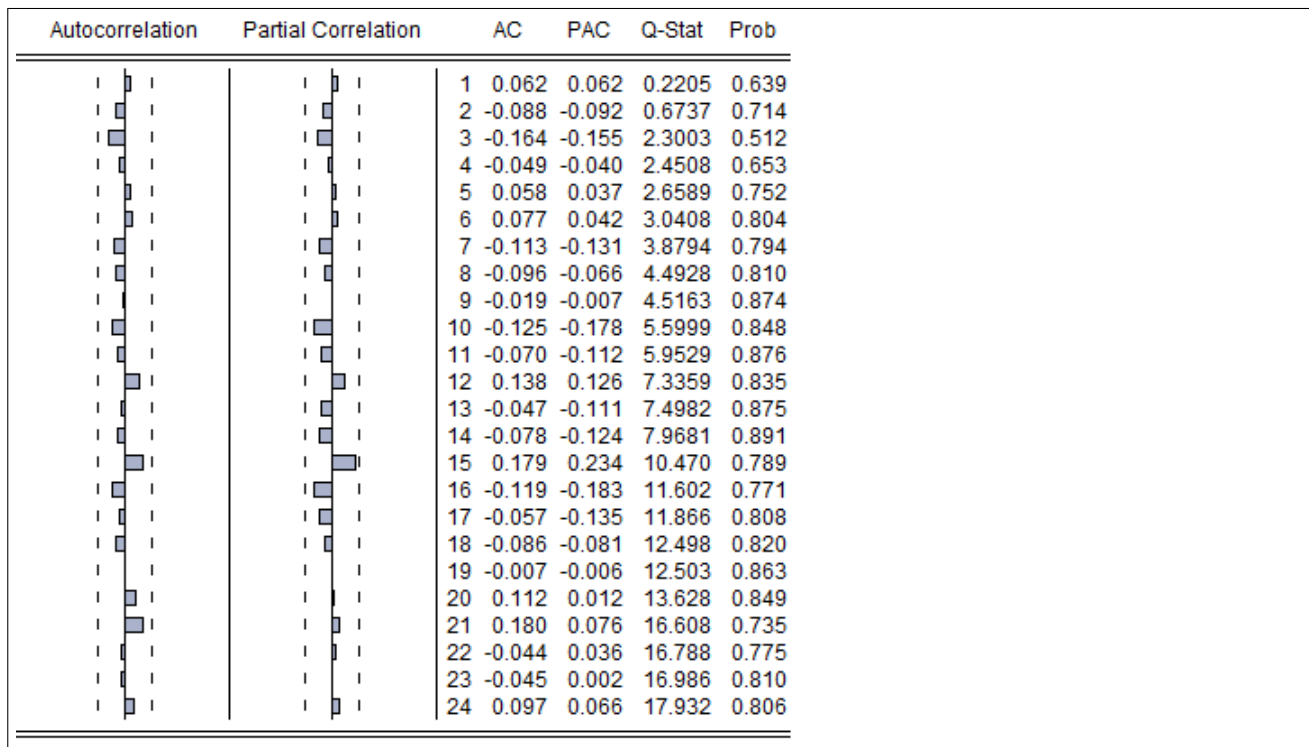


FIGURE 17 – Autocorrélogramme des résidus du modèle AR(12)

On constate qu'il y a pas des pics dans l'autocorrélogrammes i.e : il y a absence d'autocorrélation des erreurs ce qui montre que ce modèle est significatif. Donc les résidus forment un bruit blanc.

4.6 Prévision du nombre de touristes

En appliquant la méthode de Box-Jenkins aux données observées sur le nombre total mensuel de touristes pour le système de transport public de Portland de 1960 : 01 - 1965 :08, nous sommes bel et bien en mesure maintenant d'obtenir les prévisions du nombre touristes.

Ainsi, nous avons la prévision des 4 mois de : 1965 :09 - 1965 :12 sur le tableau ci-dessous :

Année	Mois	Prévision
1965	Septembre	1057
	Octobre	1073
	Novembre	1067
	décembre	1091

5 Modélisation d'une série financière

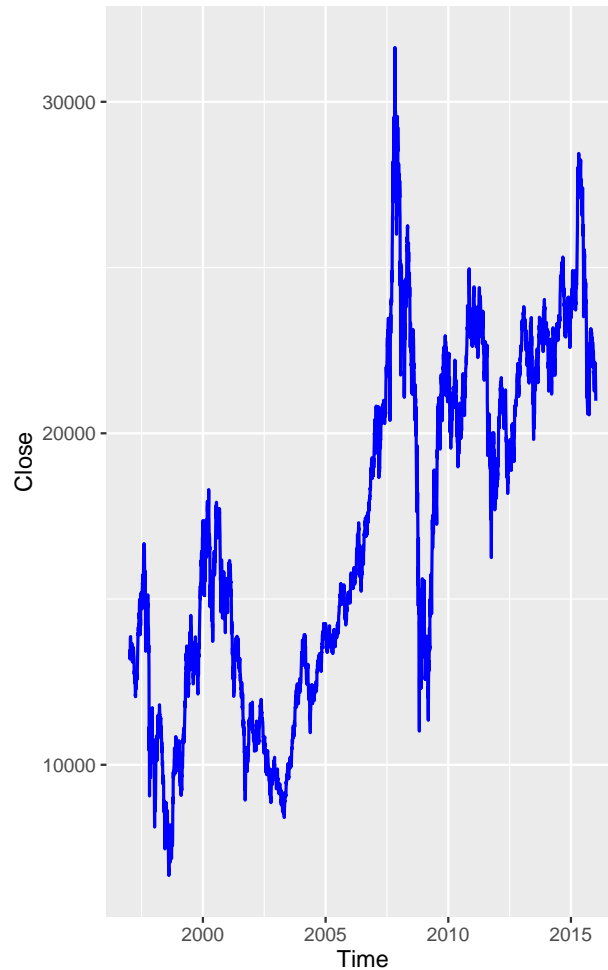
Modèle GARCH

Les modèles ARCH et GARCH (Autoregressive Conditional Heteroscedasticity and Generalized Autoregressive Conditional Heteroscedasticity) sont des modèles qui jouent un rôle primordial au niveau de la description des séries financières vu le comportement hétérosédastique de leur variance, chose qui a toujours été mal considérée par les modèles ARIMA dans lesquels on suppose que la variance est inconditionnelle par rapport au temps. En effet, face aux anomalies des représentations ARMA pour les problèmes monétaires et financiers, Engle (1982) a mis, à la disposition de l'ensemble des acteurs du marché financier, une nouvelle catégorie de modèles autorégressifs conditionnellement hétéroscédastiques adaptés à capter le comportement de la volatilité dans le temps. Cette dernière est un paramètre de mesure du risque du rendement et du prix. La volatilité sert également aux calculs pour optimiser la diversification des portefeuilles d'actifs financiers (MEDAF). Les séries monétaires et financières sont caractérisées par le regroupement (clustering) de la volatilité, à savoir les périodes de forte volatilité alternent avec les périodes de faible volatilité. Ce phénomène, que nous appelons aussi l'hétéroscédasticité conditionnelle.

5.1 Analyse préparatoire

Créé en novembre 1969, le Hang Seng (ou HSI) est l'indice phare des sociétés cotées à la Bourse de Hong-Kong. Il est détenu par une filiale de Hang Seng Bank et se compose de 50 sociétés.

Nous disposons du jeu de données journaliers de cet indice du 07/01/1997 jusqu'au 06/01/2016 et pour notre analyse nous allons nous intéresser au cours fixé à la fermeture (clôture du cours).

**FIGURE 18 – HSI cours**

L'indice représente environ 60% de la capitalisation des entreprises de Hong-Kong. Il s'agit du principal outil de mesure de performance du marché pour l'ancienne colonie britannique. En raison des liens entre la Chine et Hong Kong, l'indice est également un bon baromètre des fluctuations du marché chinois dont les entreprises sont nombreuses sur l'indice (Bank of China, China Mobile, China Tower). On constate rapidement que l'émergence de ces entreprises a attiré beaucoup d'investisseurs permettant d'atteindre un grand pic durant le début de l'année 2008. Le Hang Seng est aussi connu sous le nom de Hang Seng 50, HK50 ou HSI50. Les performances de l'indice sont de plus en plus corrélées du fait de la globalisation financière, avec celles des autres grands indices mondiaux : le CAC 40, le Footsie, le Dax, le Nasdaq, etc.

5.2 Propriétés d'une série financière

Les séries (financières) décrivant les marchés financiers possèdent des caractéristiques qui leur sont propres, ce qui les rend particulièrement complexes à modéliser. Notre but dans cette étude est d'en exhiber les propriétés dénommées "faits stylisés" des séries financières qui sont repérables sur le cours du Hang Seng. Nous abordons par la suite la non stationnarité du cours.

5.2.1 Non stationnarité du cours

Nous appliquons le test de Dickey Fuller sur notre série pour justifier l'absence de la stationnaire, on obtient ce résultat :

```
Augmented Dickey-Fuller Test
data: dhsi$close
Dickey-Fuller = -3.1501, Lag order = 16, p-value = 0.09646
alternative hypothesis: stationary
```

FIGURE 19 – Test de Dickey Fuller

On voit clairement que $p\text{-value} = 0.096 > 0,05$, nous rejettons donc que la série est stationnaire. Soit p_t le cours d'un actif Hang Seng à la date t . On constate que la série du rendement $\epsilon_t = \log(p_t/p_{t-1})$ du cours p_t semble compatible avec la stationnarité au second ordre :

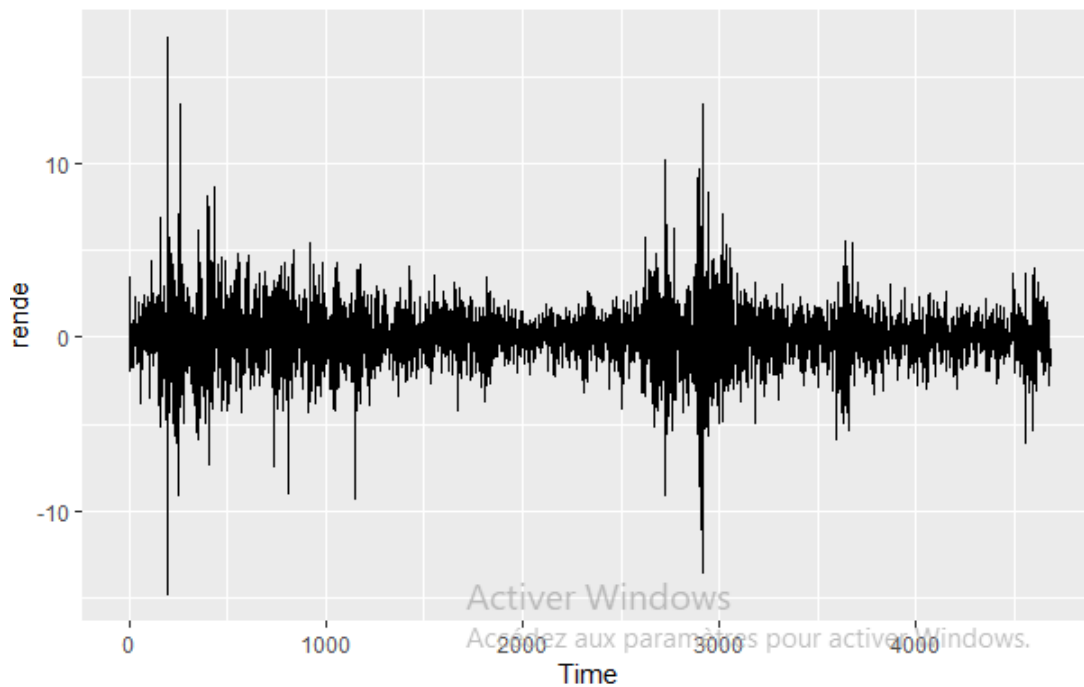


FIGURE 20 – Rendement du Hang Seng

On observe qu'il y a une fluctuation autour de 0, autrement dit cette série oscille autour de zéro, de variance très différente d'une date à l'autre des périodes agitées ainsi que des périodes calmes.

On effectue maintenant le test de Dicky-Fuller sur cette série pour bien confirmer la stationnarité.

Augmented Dickey-Fuller Test

```
data: rende
Dickey-Fuller = -15.853, Lag order = 16, p-value = 0.01
alternative hypothesis: stationary
```

FIGURE 21 – Test de Dickey Fuller sur le rendement

Le test de Dickey-Fuller sur le rendement nous confirme que cette série est stationnaire car on a une p-valeur inférieure au seuil de 5%.

5.2.2 Autocorrélations des carrées des rendements du cours

Nous présentons à travers notre Table 13, les autocorrélations de la série du rendement et la série des carrés.

m	1	2	3	4	5
$\rho(m)$	0.0038	-0.0251	0.0324	-0.0417	-0.029
$\rho_{\epsilon^2}(m)$	0.3878	0.2563	0.3024	0.2495	0.1735
	6	7	10	20	30
$\rho(m)$	0.0116	0.0057	-0.0194	0.0169	-0.0506
$\rho_{\epsilon^2}(m)$	0.1212	0.1342	0.1575	0.0616	0.0981

TABLE 13 – Autocorrélations du rendement ϵ_t et de son carré du Hang Seng

On constate que la série du rendement (ϵ_t) présente de très faibles autocorrélations, la rendant proche d'un bruit blanc. Cependant, la série des carrés (ϵ_t^2) est souvent autocorrélée. Cela permet d'en déduire que le bruit blanc n'est pas indépendant.

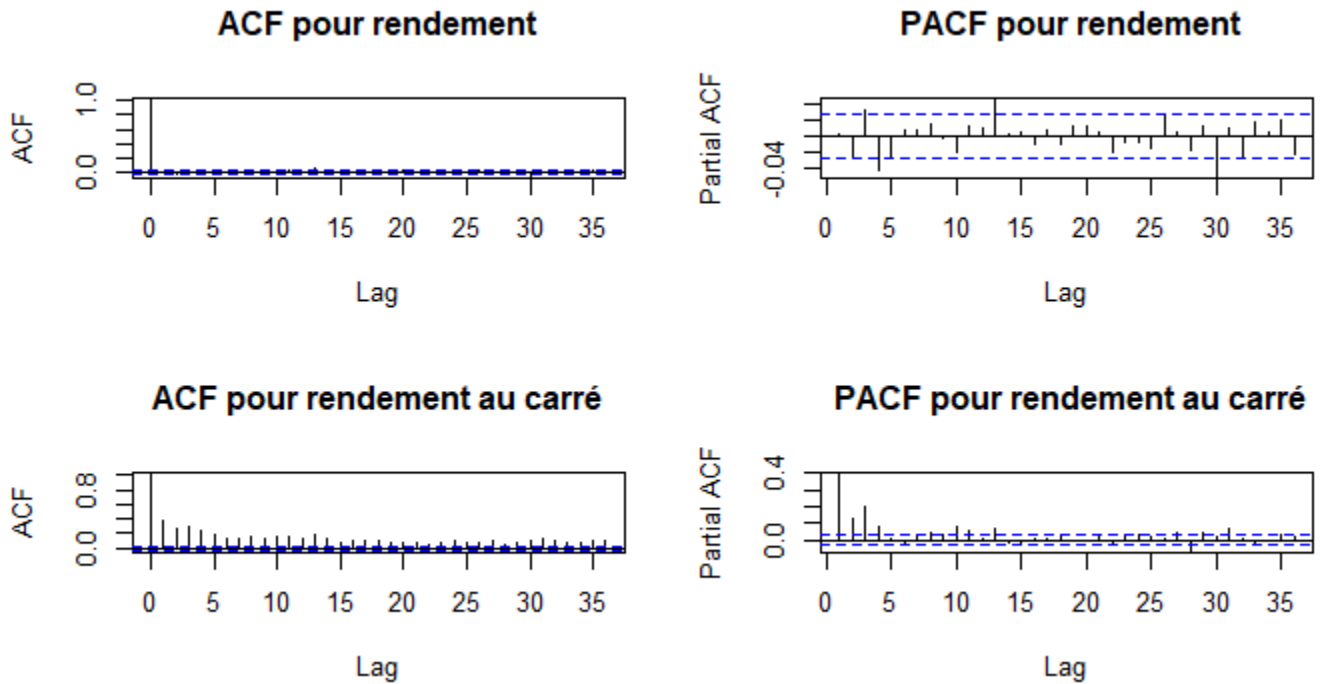


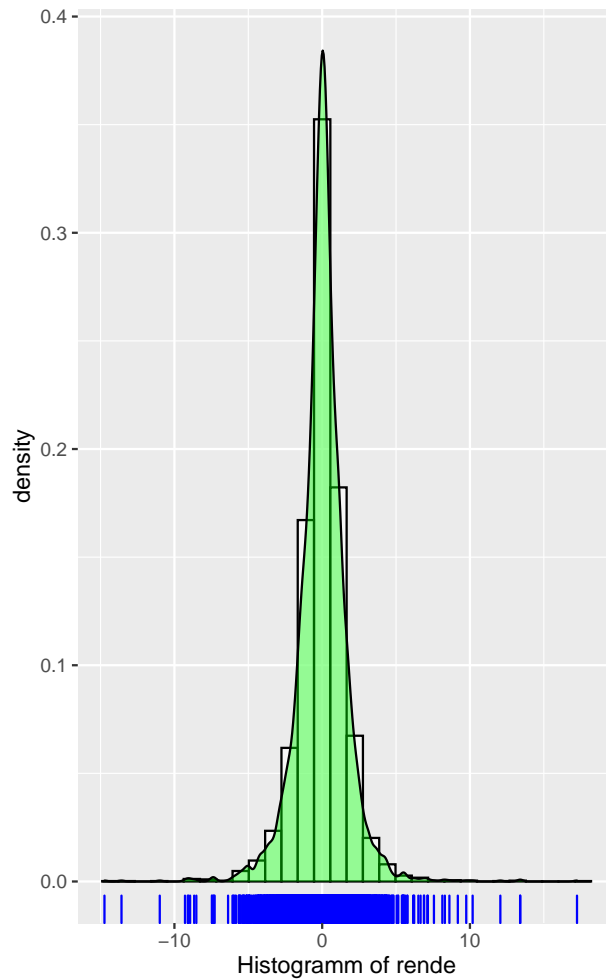
FIGURE 22 – Autocorrélogramme de ε_t et de son carré

5.2.3 Regroupements des extrêmes

On peut remarquer sur la figure 20 qu'il y a des paquets de volatilité. En fait, les grandes valeurs du rendement tendent à être suivies par des grandes valeurs et nous avons le même phénomène en ce qui concerne les petites valeurs. On remarque aussi des sous-période de forte volatilité, où il y a de forte agitation des prix (marché volatile) et également des sous périodes de faible volatilité, autrement dit des moments plus calmes.

5.2.4 Queues de distribution épaisses

Lorsqu'on considère la distribution de probabilité empirique de la série des rendements, on constate que celle-ci ne suit pas celle d'une gaussienne. A travers le test de Kolmogorov-Smirnov (test pour des échantillons de grande taille), on constate effectivement qu'on rejette l'hypothèse d'une distribution normale au seuil de 5% puisque $p\text{-value} < 2.2e - 16$. L'histogramme du rendement montre que la densité présente un grand pic en 0, cette densité est leptokurtique. Le kurtosis est égal à 9.721.

**FIGURE 23** – Histogramme du rendement

5.2.5 Effet de levier

C'est une propriété d'asymétrie des valeurs passées positives et négatives sur la volatilité à la date courante.

h	1	2	3	4	5	10	15	20
$\hat{\rho}_\epsilon(h)$	0.0038	-0.0251	0.0324	-0.0417	-0.029	-0.0194	0.0040	0.0169
$\hat{\rho}_{\epsilon^2}(h)$	0.3878	0.2563	0.3024	0.2495	0.1735	0.1575	0.1305	0.0616
$\hat{\rho}(\epsilon_{t-h}^+, \epsilon_t)$	0.0513	0.1103	0.1024	0.0771	0.0911	0.1322	0.0791	0.0992
$\hat{\rho}(-\epsilon_{t-h}^-, \epsilon_t)$	0.2766	0.2195	0.2781	0.2579	0.2227	0.1486	0.1554	0.1161

TABLE 14 – Autocorrélations du rendement du Hang Seng

On observe une corrélation qui est plus forte pour $\epsilon^- = \min(\epsilon_t, 0)$ et $|\epsilon_{t+h}|$ que pour $\epsilon^+ = \max(\epsilon_t, 0)$ et $|\epsilon_{t+h}|$ à travers notre tableau ci-dessus.

5.3 Modélisation de la série par : GARCH, TGARCH et EGARCH

Modélisation GARCH(1,1)

Par la modélisation GARCH(1,1) nous faisons des prévisions sur la volatilité du cours du Hang Seng. Les paramètres estimés valent dans les conditions optimales sont dans notre table ci-dessous :

Coefficient	W	α	β
estimation	0.018609	0.074669	0.919106

On voit bien que à travers les coefficients de α et β estimés que α est inférieur à β .



FIGURE 24 – Prédiction de la volatilité du rendement

Modélisation de TGARCH(1,1)

Avant de faire la prédiction, rappelons l'écriture du modèle TGARCH(1,1) :

$$\sigma_t = w + \alpha \epsilon_{t-1}^+ + \gamma (-\epsilon_{t-1}^-) + \beta \sigma_{t-1}$$

On a : α , γ et β sont des constantes positives et $w > 0$ est une constante strictement positive.

Nous résumons dans le tableau suivant les estimations de ces constantes dans les conditions optimales :

TABLE 15 – Estimation des coefficients

Coefficient	W	α	γ	β
estimation	0.022067	0.075844	0.497152	0.926549

**FIGURE 25** – Prédiction de la volatilité du rendement

Modélisation de EGARCH(1,1)

Avant de faire la prédiction, rappelons l'écriture du modèle EGARCH(1,1) :

$$\log(\sigma_t^2) = \omega + \alpha\eta_{t-1} + \gamma(|\eta_{t-1}| - E(|\eta_{t-1}|)) + \beta\log(\sigma_{t-1}^2)$$

Dans les conditions optimales les paramètres estimés dans le tableau ci-dessous valent :

TABLE 16 – Estimation des coefficients

Coefficient	W	α	γ	β
estimation	0.014351	-0.065591	0.144702	0.984783

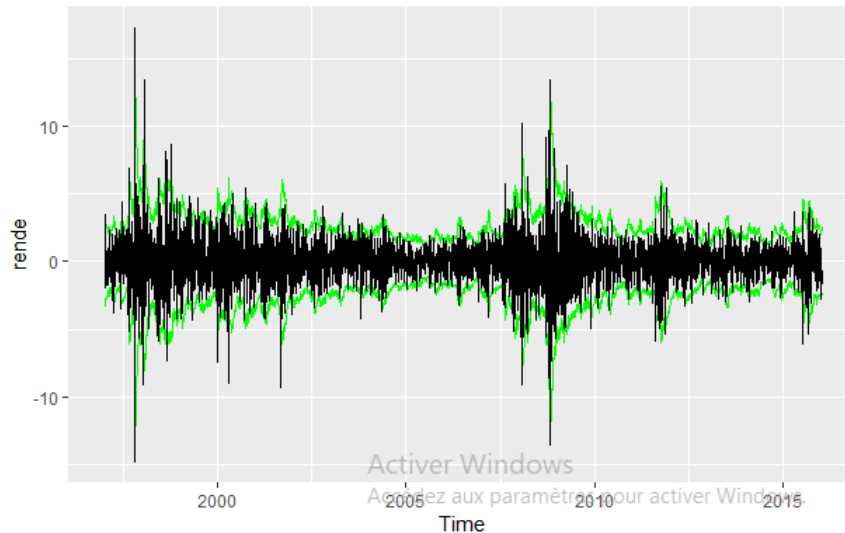


FIGURE 26 – Prédiction de la volatilité du rendement

5.4 Courbes d'impact des nouvelles

Les courbes d'impact des nouvelles du Hang Seng illustrent comment la volatilité agit sur les rendements positifs (partie droite du graphique) et les rendements négatifs (partie gauche du graphique).

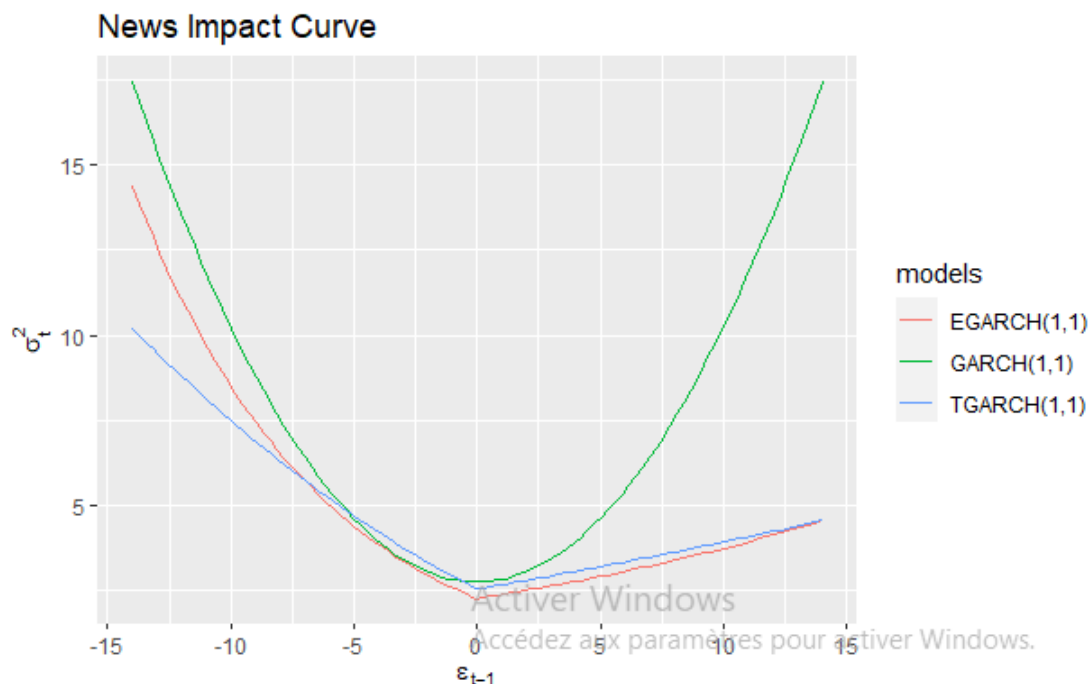


FIGURE 27 – Courbes d'impact des nouvelles

Le constat que l'on peut faire à travers ces courbes est que le modèle GARCH ne prend pas en compte l'asymétrie qui se présente entre les rendements positifs et négatifs. Contrairement aux deux autres modèles TGARCH et EGARCH. Ces deux modèles traitent les rendements positifs différemment des rendements négatifs. Ils permettent aux grandes valeurs négatives du rendement d'avoir plus d'impact sur la volatilité que les rendements positifs. Nous procédons

dans la suite à la sélection du meilleur modèle.

Sélection du meilleur modèle

Nous donnons les AIC des trois modèles obtenus et validés dans le tableau ci-dessous et nous baserons sur le critère AIC pour la sélection du meilleur modèle. Nos AIC étant normalisés. On revient à une forme connue du AIC en multipliant par le nombre des observations qui dans notre cas est de 4682 observations :

Modèle	AIC/obs
GARCH(1, 1)	3.4830
TGARCH(1, 1)	3.4606
EGARCH(1, 1)	3.4623

On conclut que le modèle TGARCH(1, 1) est le meilleur puisqu'il a un AIC plus faible.

6 Conclusion

Dans ce projet, nous avons utilisé différentes séries afin d'appliquer une méthode de prévision à chaque série. Cependant le projet en soi, nous a permis d'avoir une meilleure compréhension de l'analyse temporelle des séries. Nous nous sommes servi des trois logiciels pour la conception de ce projet que sont : Excel, Eviews, R. Nous avons effectué cinq types de modélisation et à l'issue de ses modélisations, il en découlait que d'abord pour le **lissage exponentiel simple**, nous avons une constate de lissage égale à 3.79 % pour un MAPE correspondant à 9.66 % . Ensuite, pour les analyses suivantes effectuées, la modélisation par **Holt-Winters** est meilleure par rapport à celle de **CVS** et enfin, les modèles **AR(12)** (le seul modèle validé) et **TGARCH(1,1)** ont été retenus en raison de leur performance assez satisfaisante dans la qualité d'ajustement des données.

Références

- [1] Kibala Kuma *Prévision par l'approche méthodologique de Box et Jenkins : Cas d'une Série Non Saisonnière et Non Stationnaire du type TS (Pratique sur EViews et Stata)*
<https://hal.archives-ouvertes.fr/cel-01771600/document>.
- [2] V. Monbet (2011), *Cours Modélisation de séries temporelles*.
- [3] zakariaelfaiz.https://zakariaelfaiz.blogspot.com/2016/02/analyse-de-la-stationnarite-dune-serie_24.html.
- [4] H. S. Mahamat, « *Estimation de la volatilité des données financières à haute fréquence : une approche par le Modèle Score-GARCH* »

Annexe : Code GARCH

```

library(data.table)
library(ggplot2)
library(dplyr)
library(tseries)
library(e1071)
library(lubridate)
library(rugarch)

hsi=fread("C:/Users/USER/Desktop/Master_2/Serie temporelle/ARMA/~HSI.csv")
attach(hsi)
library(ggplot2)

hsi[,5]<- as.numeric(hsi$Close)
#hsi[,1]<- as.numeric(hsi$Date)

#install.packages("tseries")
#library(tseries)

#Close<-na.omit(Close)
#Close<-as.numeric(Close)
#adf.test(Close)

dhsi <- hsi %>% select("Date","Close")
summary(dhsi)
dhsi$Close[dhsi$Close== "0"]<-NA
dhsi<-dhsi[!is.na(dhsi$Close),]

summary(dhsi)
dim(dhsi)
####bien

#Close <- as.numeric(dhsi$Close)
adf.test(dhsi$Close) #Test de Dickey-Fuller
## Effet levier, autocorrélation sur série négative puis en valeur absolue ----
## Graphique temporel ----
Base <- as.data.frame(dhsi)
gg <- ggplot(Base, aes(x=ymd(dhsi$Date), y=dhsi$Close)) +

```

```

    xlab("Time") + ylab("Close") + geom_line(size = 0.7, col="blue")
gg

Close<-(dhsi$Close)

adf.test(dhsi$Close)

n <- length(Close)
rende <-rep(0,n); rende[2:n] <- log(Close[2:n]/Close[1:(n-1)])*100

## Graphique temporel ----
v <- rende ; ind <- 0:(length(v)-1)
Base <- data.frame( vals=v, index=ind)
colnames(Base) <- c("vals","ind")
gg1 <- ggplot(Base, aes(x=Base$ind, y=Base$vals)) +
  geom_line(size = 0.7) + xlab("Time") + ylab("rende")
gg1

adf.test(rende)

###Fonction qui retourne les valeurs positives et negatives

PN <- function(rende) {
  res=matrix(0,length(rende),2)
  for (i in 1:length(rende)){
    if (rende[i] >= 0) {
      res[i,1]<-rende[i]
    }
    else {
      res[i,2]<-rende[i]
    }
  }
  return(res) }

## Les rendements positifs, négatifs et en valeurs absolues

RP <- PN(rende)[,1]

```

```

RN <- PN(rende)[,2]
Rend_abs <- abs(rende)

## Autocovariance et Autocorrélation d'une série univariée ----

acf.univ <- function (ser, h){
  n <- length(ser)
  autocov <- c(0)
  autocor <- c(0)
  h <- abs(h)
  centre <- ser-mean(ser)
  autocov <- sum(centre[1:(n-h)]*centre[(h+1):n])/n
  autocor <- autocov/(sum(centre[1:n]*centre[1:n])/n)
  list(autocov=autocov, autocor=autocor) }

## Fonction qui calcule les corrélations, centrées ----

Cor <- function(h,rend1,rend2){
  tmp1=0
  tmp2=0
  rend1 <- rend1-mean(rend1)
  rend2 <- rend2-mean(rend2)
  for(i in 1:(length(rend1)-h)){
    tmp1 <- tmp1+(rend1[i])*(rend2[i+h]) }
  gammah <-tmp1/length(rend1)
  gamma0rend1 <- acf.univ(rend1, 0)$autocov
  gamma0rend2 <- acf.univ(rend2, 0)$autocov
  gamma0 <- sqrt(gamma0rend1*gamma0rend2)
  return(gammah/gamma0) }

## Calcul des autocorrélations ----
h=30
Autocorr_rend <- rep(0,h) ; Autocorr_rend_carre <- rep(0,h)
for(i in 1:h){
  Autocorr_rend[i]=Cor(i,rende,rende)
  Autocorr_rend_carre[i]=Cor(i,rende^2,rende^2) }
Autocorr_rend; Autocorr_rend_carre

par(mfrow=c(2,2))

```

```

acf(rende,main="ACF pour rendement",rendlab="retard")
pacf(rende,main=" PACF pour rendement",rendlab="retard")
acf(rende^2,main="ACF pour rendement au carré",rendlab="retard")
pacf(rende^2,main=" PACF pour rendement au carré",rendlab="retard")

#plot.ts(rende^2,main="Carrée du rendement du cours")
#plot.ts(Rend_abs,main="Valeur absolu du rendement du cours ")

## Queues épaisses ----

kurtosis(rende)
gg <- ggplot(Base, aes(x=Base$vals)) +
  geom_histogram(aes(y=..density..), col="black", fill="white",alpha=.6) +
  geom_density(alpha=.4, fill="green") +
  geom_rug(col="blue") + xlab("Histogramm of rende")
gg

## Test de normalité : Kolmogorov-Smirnov(rend) ----
ks.test(rende,"pnorm",mean(rende),sd(rende))

## Effet de levier ----
positivecor<-rep(0,8)
negativecor<-rep(0,8)
for (h in c(1,2,3,4,5,10,15,20)) {
  positivecor[h]=Cor(h,RP,Rend_abs)
  negativecor[h]=Cor(h,-RN,Rend_abs)
}
positivecor[c(1,2,3,4,5,10,15,20)]
negativecor[c(1,2,3,4,5,10,15,20)]

## Estimation des coefficients du modèle GARCH ----

base <- cbind.data.frame(Date=dhsi$Date,rende)
GARCH <- ugarchspec(variance.model =list(model="fGARCH",submodel ="GARCH",
                                           garchOrder = c(1, 1)))

modele1 <- ugarchfit(rende,spec=GARCH)
show(modele1)

gg <- ggplot(base,aes(x=ymd(base$Date), y=base$rende)) +

```

```

    geom_line(aes(y=2*modele3@fit[["sigma"]]), col="green")+
    geom_line(aes(y=-2*modele3@fit[["sigma"]]), col="green")+
    xlab("Time") + ylab("rende") + geom_line(size = 0.7)
gg

TGARCH <- ugarchspec(variance.model = list(model="fGARCH",submodel ="TGARCH",
                                           garchOrder = c(1, 1)))

modele2 <- ugarchfit(rende,spec=TGARCH)
show(modele2)
EGARCH <- ugarchspec(variance.model = list(model="eGARCH", garchOrder = c(1, 1)))
modele3 <- ugarchfit(rende, spec=EGARCH)
show(modele3)

## News impacts curves----
ni=newsimpact(z=NULL, modele1); ni2=newsimpact(z=NULL, modele2)
ni3=newsimpact(z=NULL, modele3)
n1 <- cbind.data.frame(zx=ni$zx, zy=ni$zy, models=rep("GARCH(1,1)",length(ni$zy)))
n2 <- cbind.data.frame(zx=ni2$zx, zy=ni2$zy, models=rep("TGARCH(1,1)",length(ni$zy)))
n3 <- cbind.data.frame(zx=ni3$zx, zy=ni3$zy, models=rep("EGARCH(1,1)",length(ni$zy)))
NIC <- rbind(n1,n2,n3)
gg <- ggplot(NIC) + aes(x=zx, y=zy, col=models) + geom_line(size = 0.7)+
  ggtitle("News Impact Curve") + xlab(ni$xexpr) + ylab(ni$yexpr)
gg

```