Méthode de prévision

Etude de la série : Nombre de créations d'entreprises - Micro-entrepreneurs - Activités spécialisées, scientifiques et techniques et activités de services administratifs et de soutien - France

Keroudine BELLADJO

Université de Lille

Plan

I - Introduction : présentation de la série

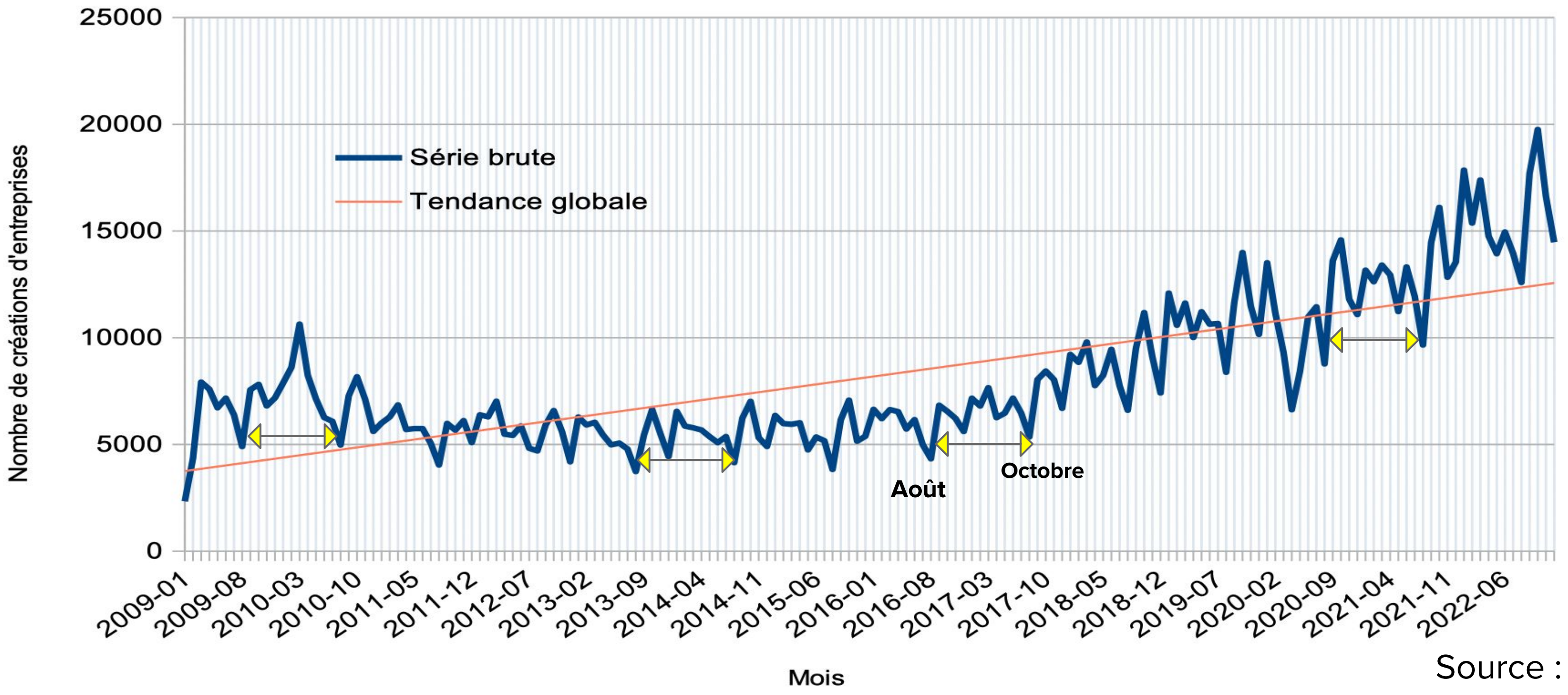
II - Lissage

III - Box-Jenkins

IV- Conclusion

I - INTRODUCTION : Présentation de la série

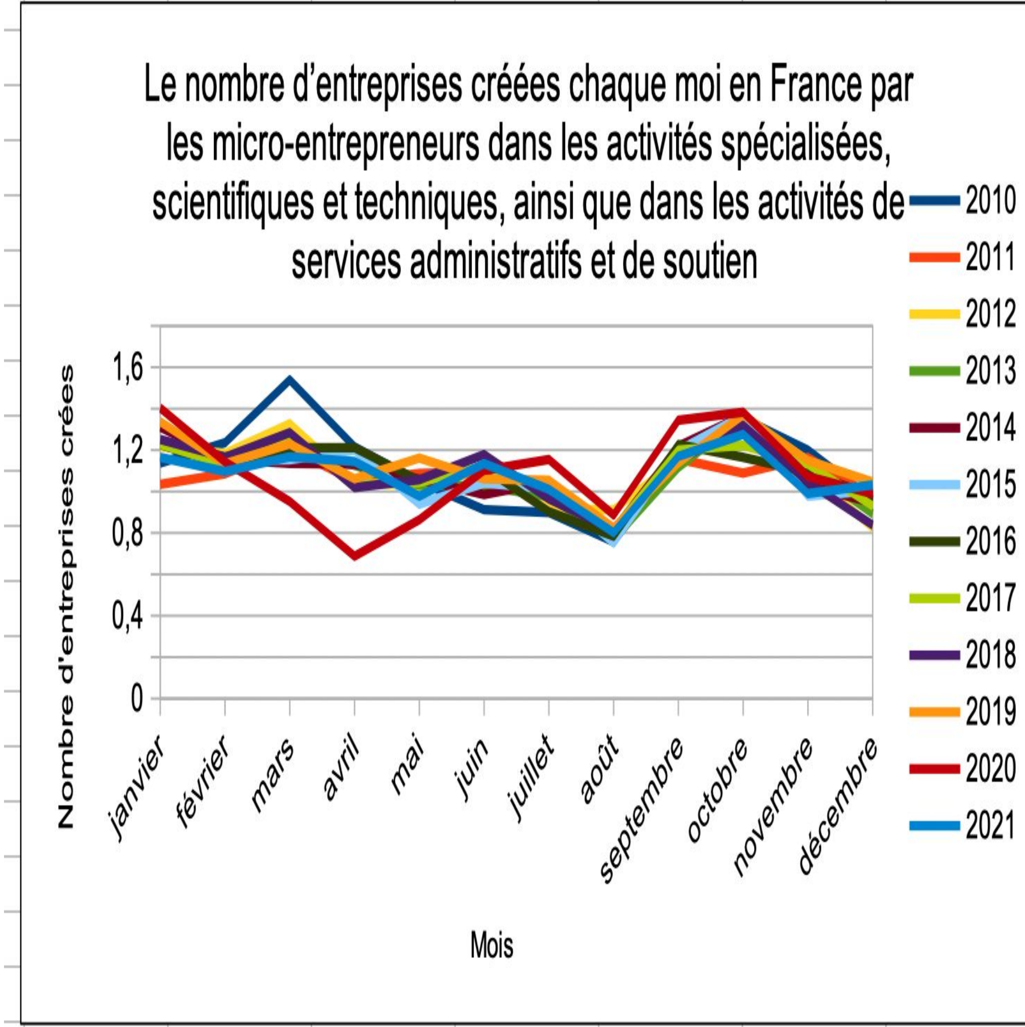
Le nombre d'entreprises créées chaque moi en France par les micro-entrepreneurs dans les activités spécialisées, scientifiques et techniques, ainsi que dans les activités de services administratifs et de soutien sur la période 2009-2022



Source : INSEE

1 - Graphique du TABLEAU DE BUYS-BALLOT

- Des creux tous les mois d'Août ,
- Des pics en octobre
- Des variations inattendues en mars 2010 et en avril 2020.

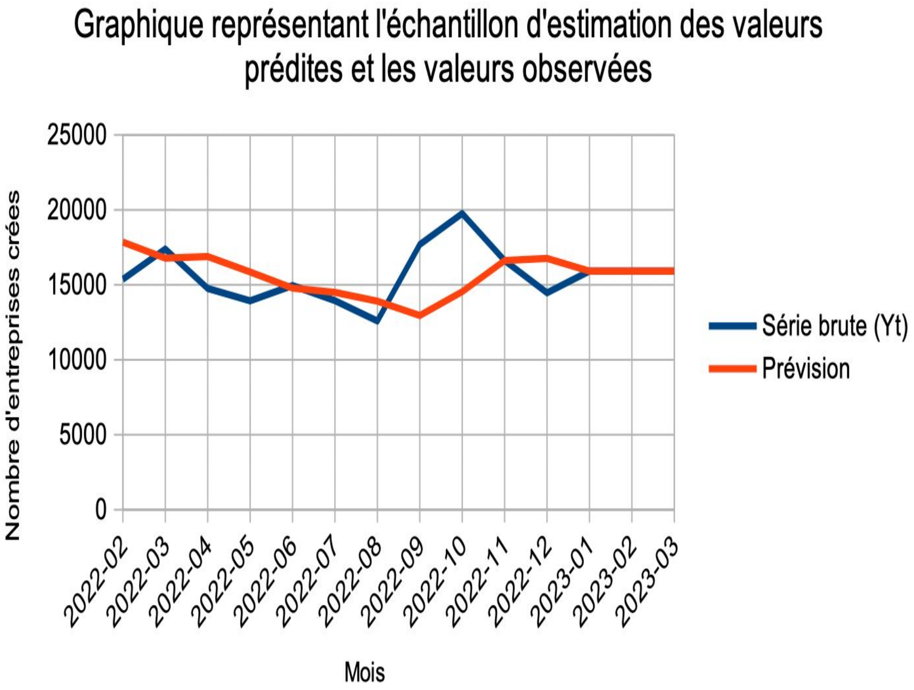


II - Lissage

Dans cette partie nous appliquerons différentes méthode de lissage : la méthode de Holt, la méthode de Holt-Winters sur notre série temporelle

Objectif : Déterminer la meilleure méthode de lissage pour cette série

1 - La méthode de Holt

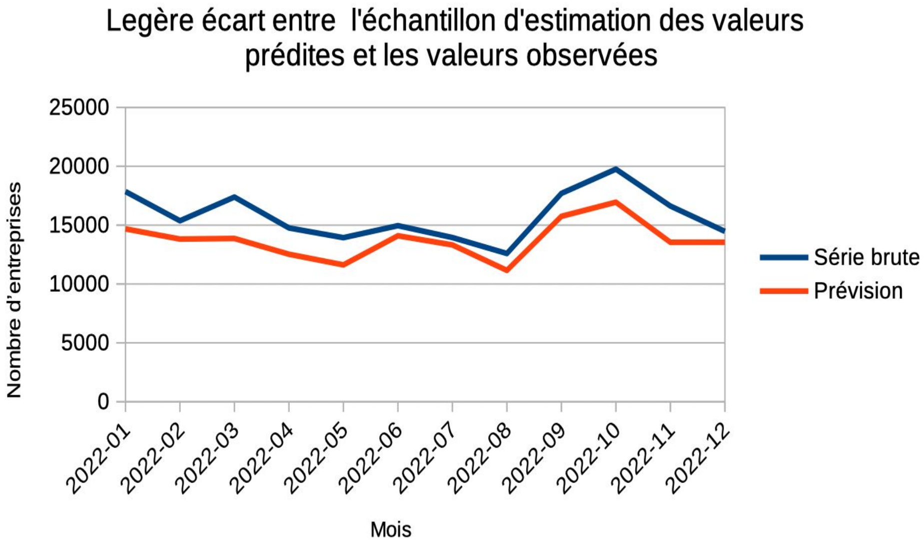


Le MAPE obtenu avec cette méthode :

MAPE	11,04 %

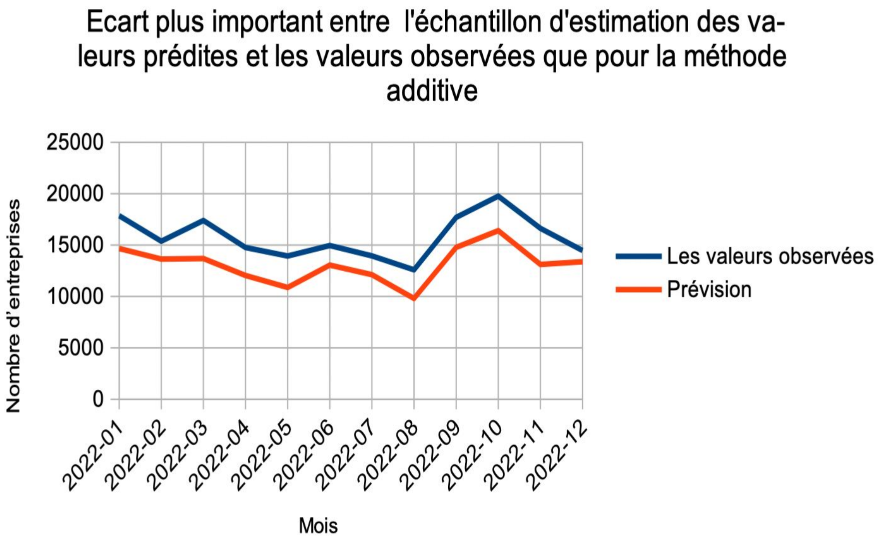
2 - La méthode de Holt-Winters

La méthode additive:



MAPE	12,63718509

La méthode multiplicative:



MAPE	16,74479238

La meilleure méthode entre les deux est donc la méthode additive

Comparaison entre La méthode de Holt-Winters additive et la La méthode de Holt

La méthode de Holt :

MAPE	11,04 %

La méthode de Holt-Winters additive :

MAPE	12,63718509

- → **La méthode de Holt a la plus petite MAPE.**
- → **C'est donc la meilleure méthode pour cette série parmi les méthode de lissage**

III-Méthode de Box-Jenkins

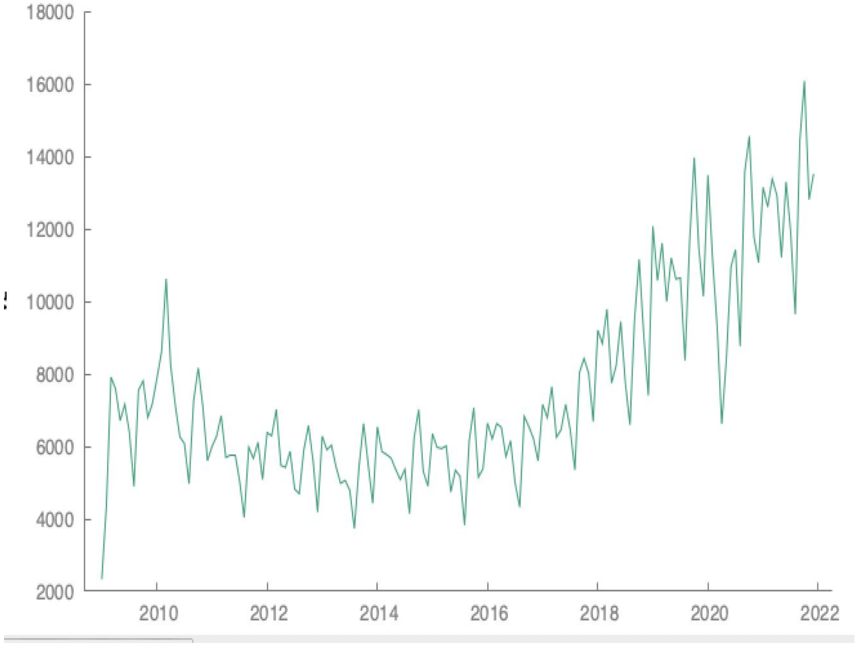
La méthode de Box-Jenkins est une approche en plusieurs étapes pour identifier et modéliser une série chronologique stationnaire avec un processus auto-régressif et une moyenne mobile (ARMA).

Première étape :Analyse de données

Étape ayant pour objectif la stationnarisation de la série

1 - La série Brute (v2)

La série est saisonnière



Test de Dicker-Fuller

H0 = La série admet de racines unitaires

H1 = La série n'admet pas de racines unitaires

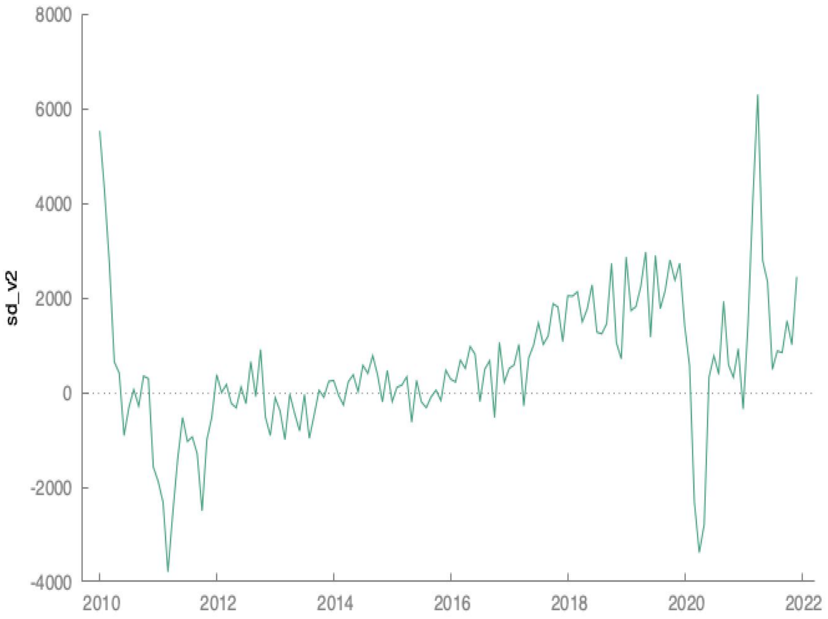
```
Test de Dickey-Fuller augmenté pour v2
test à reculons à partir de 13 retards, suivant le critère AIC
taille de l'échantillon 143
hypothèse nulle de racine unitaire : a = 1

  test avec constante
  avec 12 retards de (1-L)v2
  modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  valeur estimée de (a - 1): 0.0379423
  statistique de test: tau_c(1) = 1.09701
  p. critique asymptotique 0.9975
  Coeff. d'autocorrélation du 1er ordre pour e: 0.021
  différences retardées: F(12, 129) = 28.081 [0.0000]
```

La p-critique ou p-value 0.9975 > α = 0.05 , on ne rejette donc pas H0.

La série n'est donc pas stationnaire

2 - Série en différence saisonnière (Sd_v2)

La série n'est pas saisonnière



Test de Dicker-Fuller

H0 = La série admet de racines unitaires

H1 = La série n'admet pas de racines unitaires

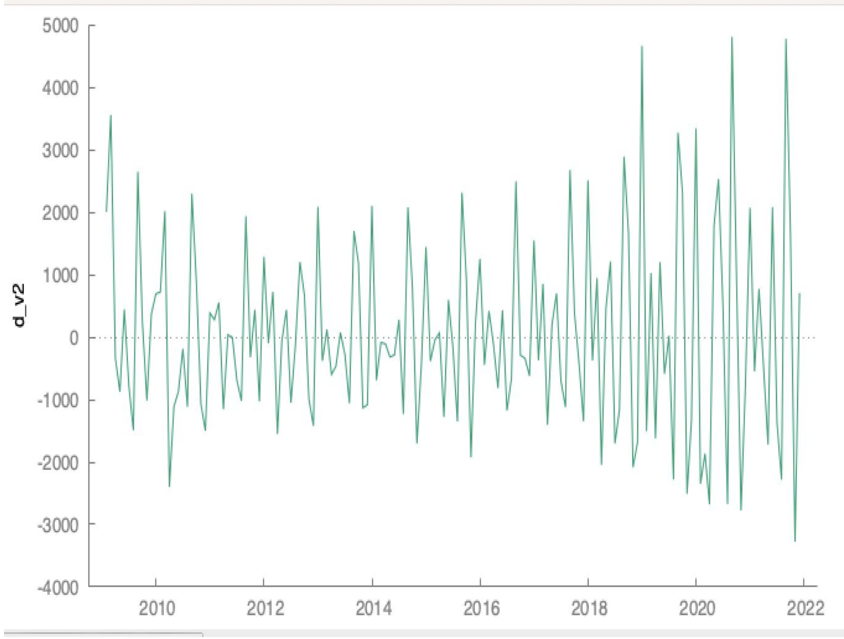
```
Test de Dickey-Fuller augmenté pour sd_v2
test à reculons à partir de 13 retards, suivant le critère AIC
taille de l'échantillon 130
hypothèse nulle de racine unitaire : a = 1

  test avec constante
  avec 13 retards de (1-L)sd_v2
  modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  valeur estimée de (a - 1): -0.148233
  statistique de test: tau_c(1) = -1.74347
  p. critique asymptotique 0.4093
  Coeff. d'autocorrélation du 1er ordre pour e: -0.009
  différences retardées: F(13, 115) = 2.852 [0.0014]
```

La p-critique 0.4093 > α = 0.05 , on ne rejette donc pas H0.

La série n'est donc pas stationnaire

3 - Série en différence première (d_v2)

La série est saisonnière



Test de Dicker-Fuller

H0 = La série admet de racines unitaires

H1 = La série n’admet pas de racines unitaires

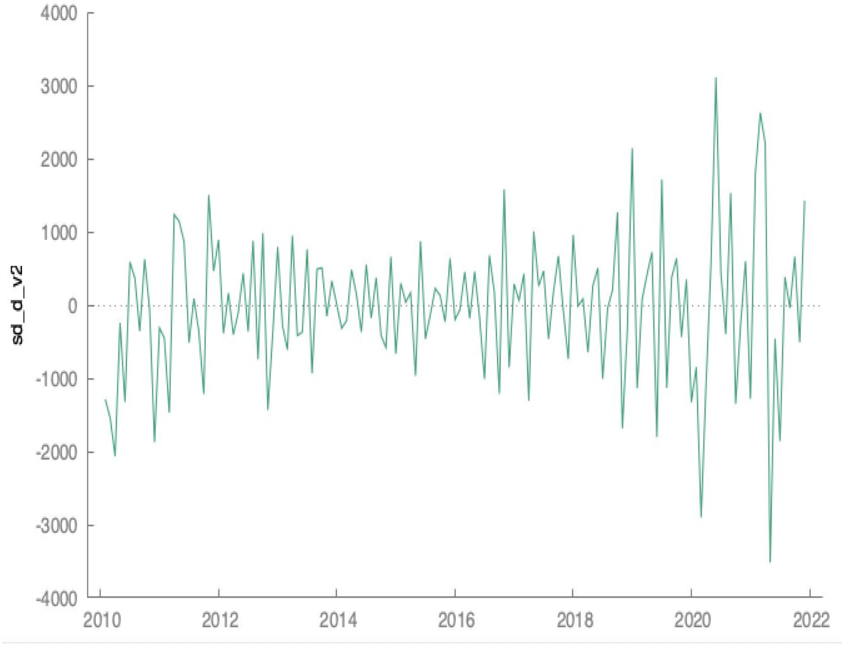
```
Test de Dickey-Fuller augmenté pour d_v2
test à reculons à partir de 13 retards, suivant le critère AIC
taille de l'échantillon 143
hypothèse nulle de racine unitaire : a = 1

  test avec constante
  avec 11 retards de (1-L)d_v2
  modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  valeur estimée de (a - 1): -2.29016
  statistique de test: tau_c(1) = -3.76651
  p. critique asymptotique 0.003285
  Coeff. d'autocorrélation du 1er ordre pour e: 0.015
  différences retardées: F(11, 130) = 30.770 [0.0000]
```

La p-critique 0.003282 < α = 0.05 , faible, on rejette donc H0.

La série est donc stationnaire

4 - Série en différence première et saisonnière (sd_d_v2)

La série n'est pas saisonnière



Test de Dicker-Fuller

H0 = La série admet de racines unitaires

H1 = La série n'admet pas de racines unitaires

```
Test de Dickey-Fuller augmenté pour sd_d_v2
test à reculons à partir de 13 retards, suivant le critère AIC
taille de l'échantillon 130
hypothèse nulle de racine unitaire : a = 1

   test avec constante
   avec 12 retards de (1-L)sd_d_v2
   modèle: (1-L)y = b0 + (a-1)*y(-1) + ... + e
   valeur estimée de (a - 1): -2.72201
   statistique de test: tau_c(1) = -5.13918
   p. critique asymptotique 1.051e-05
   Coeff. d'autocorrélation du 1er ordre pour e: -0.020
   différences retardées: F(12, 116) = 4.249 [0.0000]
```

La p-critique est très petite , on rejette donc H0.

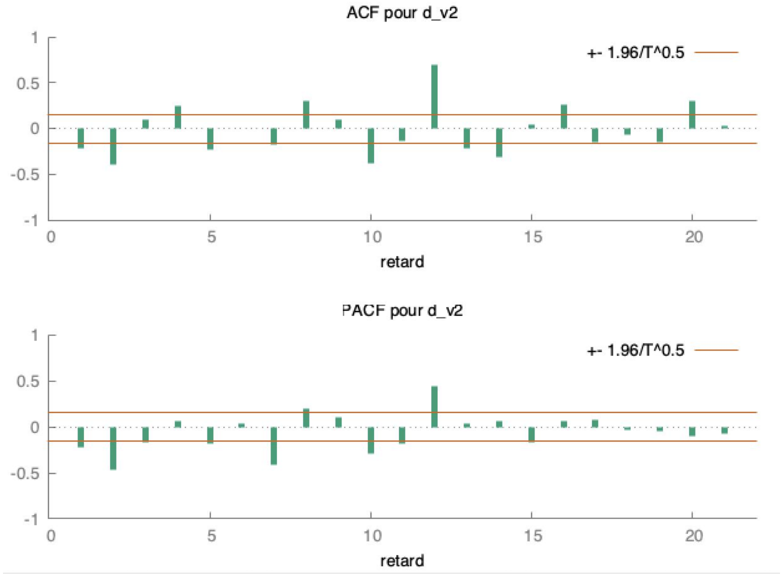
La série est donc stationnaire .

Deuxième étape

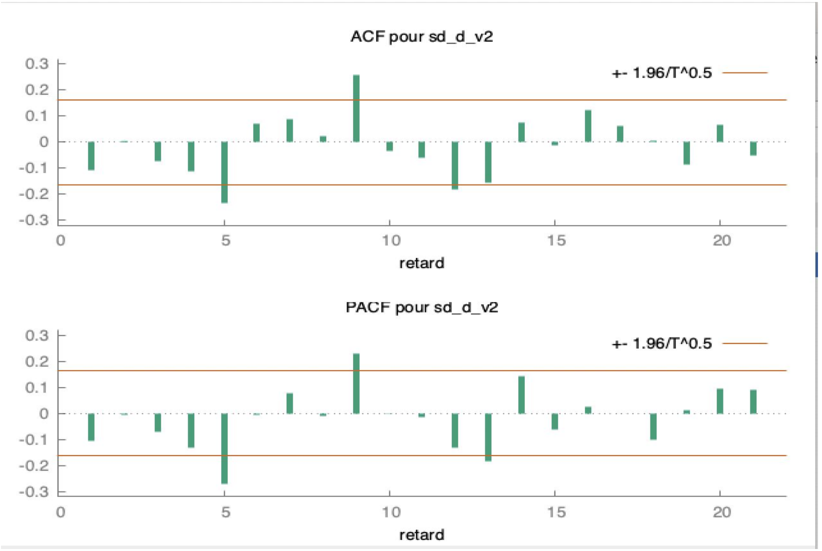
Identification, estimation, validation et prévision du modèle.

Corrélogramme des deux séries stationnaires

Série en différence premières (d_v2):



Série en différence première et saisonnière (sd_d_v2):



❖ La série sd_d_v2 a le corrélogramme le plus significatif, on choisit donc cette série.

Estimation du modèle et son écriture

Test individuel : Test de significativité du modèle

```
Évaluations de la fonction : 88
Évaluations du gradient : 31

Modèle 7: ARIMA, utilisant les observations 2010:02-2021:12 (T = 143)
Estimation par AS 197 (MV exact)
Variable dépendante: (1-L)(1-Ls) v2
Écarts-types basés sur la matrice hessienne

             coefficient    éc. type       z       p. critique
  ------------------------------------------------------------
  const        3.12569     20.0433       0.1559    0.8761
  phi_9        0.236105     0.0937563    2.518     0.0118     **
  phi_22      -0.316383     0.114289    -2.768     0.0056     ***
  theta_1     -0.186766     0.0919645   -2.031     0.0423     **
  theta_5     -0.277744     0.0909111   -3.055     0.0022     ***
  Theta_1     -0.513717     0.109569    -4.689     2.75e-06   ***

Moyenne var. dép.    -21.58741   Éc. type var. dép.    981.3808
Moyenne innovations  -55.03916   Éc. type innovations  814.8222
R2                    0.916353   R2 ajusté             0.913928
Log de vraisemblance -1165.059   Critère d'Akaike      2344.117
Critère de Schwarz    2364.857   Hannan-Quinn          2352.545

                        Réel  Imaginaire    Modulo   Fréquence
  ------------------------------------------------------------
  AR
   Ordre   1           0.7843     0.6760     1.0354     0.1132
   Ordre   2           0.7843    -0.6760     1.0354    -0.1132
   Ordre   3          -1.0460    -0.1677     1.0594    -0.4747
   Ordre   4          -1.0460     0.1677     1.0594     0.4747
   Ordre   5           1.0408    -0.1318     1.0491    -0.0200
```

- Tous les coefficients sont significatifs.
- Le modèle est donc significatif

Ecriture du modèle estimé:

Le modèle est significatif, on peut donc l'écrire:

$$(1-\varphi_9 B^9 - \varphi_{22} B^{22})(1-B)(1-B^{12})v2 = (1-\theta_5 B^5)(1-\Theta B^{12})\epsilon_t$$

Validation du modèle

```
Fonction d'autocorrélation résiduelle
***, **, * indiquent une significativité au seuil de 1%, 5% et 10%
sur la base d'écart-type 1/T^0.5
```

RETARD	ACF		PACF		Q	[p. critique]
1	-0.0075		-0.0075			
2	-0.0536		-0.0536			
3	-0.0439		-0.0448			
4	-0.0077		-0.0114			
5	0.0223		0.0174			
6	0.0844		0.0823		1.8780	[0.171]
7	0.0039		0.0071		1.8803	[0.391]
8	0.0493		0.0607		2.2543	[0.521]
9	0.0232		0.0331		2.3378	[0.674]
10	-0.0251		-0.0173		2.4363	[0.786]
11	-0.0403		-0.0372		2.6915	[0.846]
12	0.0678		0.0610		3.4186	[0.844]
13	-0.1388	*	-0.1497	*	6.4921	[0.592]
14	0.0531		0.0457		6.9457	[0.643]
15	0.0900		0.0780		8.2589	[0.604]
16	0.0534		0.0523		8.7248	[0.647]
17	0.0008		0.0152		8.7249	[0.726]
18	-0.0445		-0.0329		9.0536	[0.769]
19	-0.0357		-0.0057		9.2670	[0.814]
20	0.1039		0.0893		11.0868	[0.746]
21	-0.0630		-0.0809		11.7608	[0.760]

Test PORTMANTEAU

H_0 = Toutes les autocorrélations des résidus du modèle valent zéro

H_1 = Au moins une est différente de zéro

La p-critique la plus petite des autocorrélations de notre modèle est 17,1% > 5% , on ne rejette donc pas H0, Les résidus sont des bruits blancs. Notre modèle est donc valide

Prévision

Calcul des prévisions

	v2	prédiction	éc. type	intervalle de 95%
2021:01	13167.00	14045.70		
2021:02	12625.00	11860.34		
2021:03	13403.00	12406.41		
2021:04	12944.00	11618.32		
2021:05	11226.00	12933.87		
2021:06	13318.00	13929.85		
2021:07	11944.00	12812.68		
2021:08	9663.00	9061.33		
2021:09	14458.00	13562.02		
2021:10	16107.00	15813.53		
2021:11	12831.00	14128.20		
2021:12	13546.00	13043.02		
2022:01	17853.00	17264.25	814.822	15667.22 – 18861.27
2022:02	15372.00	15612.77	1050.252	13554.31 – 17671.22
2022:03	17388.00	15497.56	1241.822	13063.63 – 17931.48
2022:04	14762.00	13514.07	1407.556	10755.31 – 16272.83
2022:05	13936.00	12791.31	1555.734	9742.13 – 15840.49
2022:06	14955.00	15028.15	1615.763	11861.31 – 18194.99
2022:07	13948.00	13824.83	1673.641	10544.55 – 17105.10
2022:08	12590.00	11603.66	1729.583	8213.74 – 14993.58
2022:09	17693.00	16611.76	1783.771	13115.63 – 20107.88
2022:10	19757.00	18151.04	1891.328	14444.11 – 21857.97
2022:11	16624.00	15627.10	1982.047	11742.36 – 19511.84
2022:12	14470.00	15056.72	2068.792	11001.96 – 19111.48

Statistiques sur la qualité de la prévision utilisant 12 observations

Moyenne Erreur	730.4
Moyenne Erreur Carrée (racine)	1037.9
Moyenne Erreur Absolue	880.51
Moyenne Pourcentage Erreur	4.4734
Moyenne Pourcentage Erreur Absolue	5.4918
U2 de Theil	0.41107
Proportion de biais, UM	0.49528
Proportion des régressions, UR	0.0029592
Proportion des perturbations, UD	0.50176

Critères de qualité prédictive :

MAPE (Moyenne Pourcentage Erreur Absolue) = 5,4918%

Les modèles identifiés et leur comparaison

Xt = v2

Modèles	Coefficients significatifs	Test Portmanteau	Le critère d'Akaike (AIC)	MAPE
$(1-\varphi_9 B^9 - \varphi_{22} B^{22})(1-B)(1-B^{12})X_t = (1-\theta_5 B^5)(1-\Theta B^{12})\epsilon_t$	OUI	17% - 84%	2344,177	5,49
$(1-\varphi_1 B)(1-\Phi B^{12})(1-B)(1-B^{12})X_t = (1-\theta_5 B^5-\theta_9 B^9-\theta_{13} B^{13})\epsilon_t$	OUI	33,2%- 99,7%	2347,947	11,409
$(1-\varphi_1 B - \varphi_{13} B^{13})(1-\Phi B^{12})(1-B)(1-B^{12})X_t = (1-\theta_5 B^5-\theta_9 B^9)\epsilon_t$	OUI	32,5%-99,6%	2345,956	11,035
$(1-\varphi_1 B - \varphi_9 B^9)(1-\Phi B^{12})(1-B)(1-B^{12})X_t = (1-\theta_5 B^5)\epsilon_t$	OUI	43,5%-97,1%	2351.378	10,476
$(1-\varphi_1 B - \varphi_5 B^5- \varphi_{12} B^{12})(1-B)(1-B^{12})X_t = (1-\theta_9 B^9-\theta_{13} B^{13})\epsilon_t$	OUI	44,7%-97,2%	2348,249	11,435

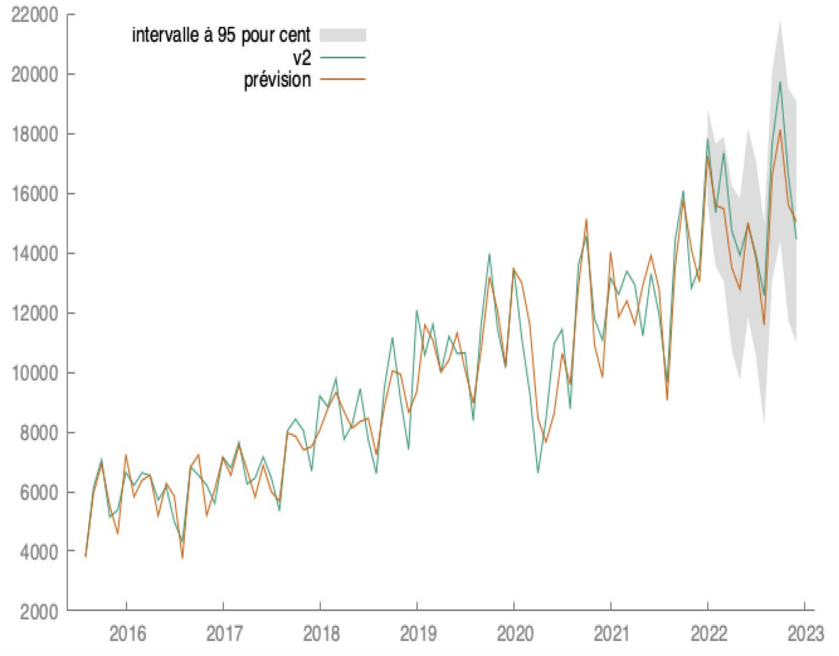
IV - Conclusion

Le MAPE du meilleurs modèle de la **méthode de lissage (méthode de Holt) : 11,04%**

Le MAPE du meilleurs modèle de la **méthode de Box-Jenkins : 5,49%**

Pour cette série la meilleure méthode est donc la méthode de Box-Jenkins.

Graphique comportant les observations et les prévisions du meilleur modèle



Merci pour votre attention !