

2022-2023

Economie Appliquée

Parcours ingénierie des données et évaluations économétriques

Modélisation ARIMA pour la Prévision des Taux de Change Euro-Dollar

Box&Jenkins



GHEMMOUR Samy

Sous la direction de Jérôme DANO

02/10/2023

ENGAGEMENT DE NON-PLAGIAT

Je, soussigné Mr Samy GHEMMOUR

Déclare être pleinement conscient que le plagiat de documents ou d'une partie d'un document publiée sur toutes formes de support, numérique ou papier, constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée. En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour écrire ce document.



GHEMMOUR Samy,
Le 28/09/2023.

Table des matières

• INTRODUCTION :	5
• Description des données utilisés :	6
1. Analyse de la série des données brutes :	7
1.1. Graphique des données brutes :	7
1.2. Moyenne des données brutes :	8
1.3. Variance des données brutes :	9
1.4. ACF Des données brutes :	9
1.5. PACF :	10
1.6. Résumé analyse des données brutes :	10
2. Analyse de la série en DIF 1 :	11
2.1. Différenciation première :	11
2.2. Graphique de la série DIF 1 :	12
2.3. Graphique des moyennes DIF 1 :	12
2.4. Variance en DIF 1 :	13
2.5. ACF EN DIF 1 :	13
2.6. PACF en DIF 1 :	14
2.7. Résumé de la série en DIF1 :	15
3. Analyse de la série en DIF 2 :	16
3.1. Graphique des données brutes :	16
3.2. Moyenne en Dif 2 :	16
3.3. Variance en Dif 2 :	17
3.4. ACF en DIF 2 :	18
3.5. PACF en Dif 2 :	19
4. Choix de la différenciation :	20
4.1. Estimation des paramètres q et p :	22
4.2. Estimation des modèles :	23
4.2.1. En différence d'ordre 1 :	23
4.2.1.1. Estimation	23
4.2.1.2. Phase de tatonnement :	24
4.2.2. En différence d'ordre 2 :	26
4.2.2.1. Estimation	26
4.2.2.2. Phase de tâtonnement :	27
4.3. Sélection entre DIF (1) et DIF (2) :	29
4.3.1. AIC BIC RMSE :	29
4.3.2. Calcul du coefficient d'inégalité de Theil :	31

4.4. Prévvision en données brutes :	34
• Conclusion :	36

• INTRODUCTION :

Dans un monde financier en constante évolution, le taux de change entre les devises majeures, telles que l'euro et le dollar américain, est un indicateur clé des fluctuations du marché. La capacité à comprendre et à prédire ces mouvements est cruciale pour les acteurs économiques, car elle peut influencer de manière significative leurs décisions. Ce projet se propose d'explorer la modélisation et la prévision du taux de change quotidien entre l'euro et le dollar américain sur une période d'un an, du 8 septembre 2022 au 8 septembre 2023.

Nous adoptons la méthode Box-Jenkins, une approche éprouvée en analyse des séries temporelles. Plus précisément, nous utilisons des modèles ARIMA (AutoRegressive Integrated Moving Average) pour décomposer et analyser la série temporelle des taux de change, dans le but de faire des prévisions précises sur leur évolution future.

Le processus commence par une analyse approfondie des données brutes, où nous évaluons la stationnarité, la saisonnalité et d'autres caractéristiques de la série. Cette étape est cruciale pour déterminer le modèle ARIMA optimal. Ensuite, nous appliquons des différenciations successives pour rendre la série stationnaire et réduire les tendances et les structures non stationnaires.

Nous verrons ensuite comment choisir le meilleur modèle pour notre prévision en utilisant des critères tels que l'AIC, le BIC, le RMSE et le coefficient d'inégalité de Theil. Ces critères nous aident à sélectionner le modèle qui offre les meilleures prévisions, une fois le modèle ARIMA sélectionné, nous l'appliquons aux données pour générer des prévisions de taux de change pour le futur. Nous évaluons la précision de ces prévisions à l'aide de méthodes d'évaluation appropriées.

- Description des données utilisés :

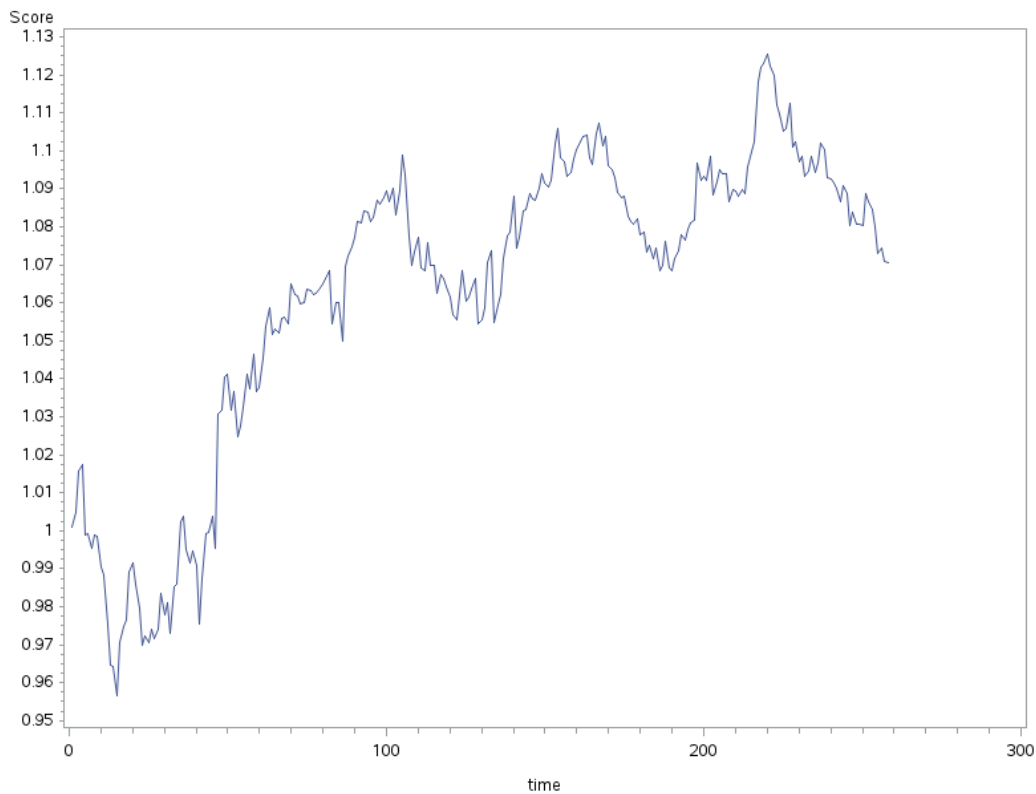
Les données utilisées dans cette étude sont une série de taux de change quotidiens entre l'euro et le dollar américain, s'étendant sur une période d'un an, du 8 septembre 2022 au 8 septembre 2023. Ces informations ont été obtenues à partir du site web de la Banque Centrale Européenne, une source reconnue pour sa fiabilité en matière d'information financière.

Cette base de données se compose de 258 observations, chacune correspondant au taux de change d'une journée spécifique. Pour faciliter l'analyse des tendances et de la saisonnalité dans les données, nous avons ajouté une variable 'Classe', qui regroupe les observations en catégories allant de 'A' à 'Z'. Chaque classe contient entre 11 et 12 observations.

En outre, nous avons inclus une variable 'Temps' pour indiquer la nature temporelle des données, ce qui est essentiel pour les analyses effectuées à l'aide de SAS. Cette variable permet à SAS de reconnaître que les observations sont ordonnées dans le temps, ce qui est crucial pour les méthodes d'analyse qui tiennent compte de la dépendance temporelle des observations.

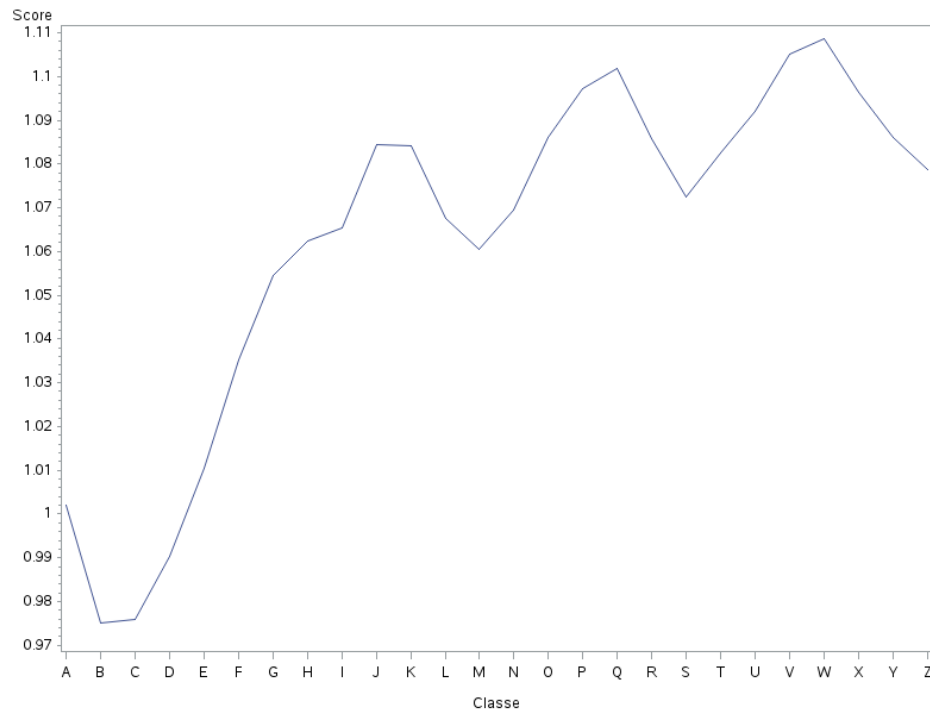
1. Analyse de la série des données brutes :

1.1. Graphique des données brutes :



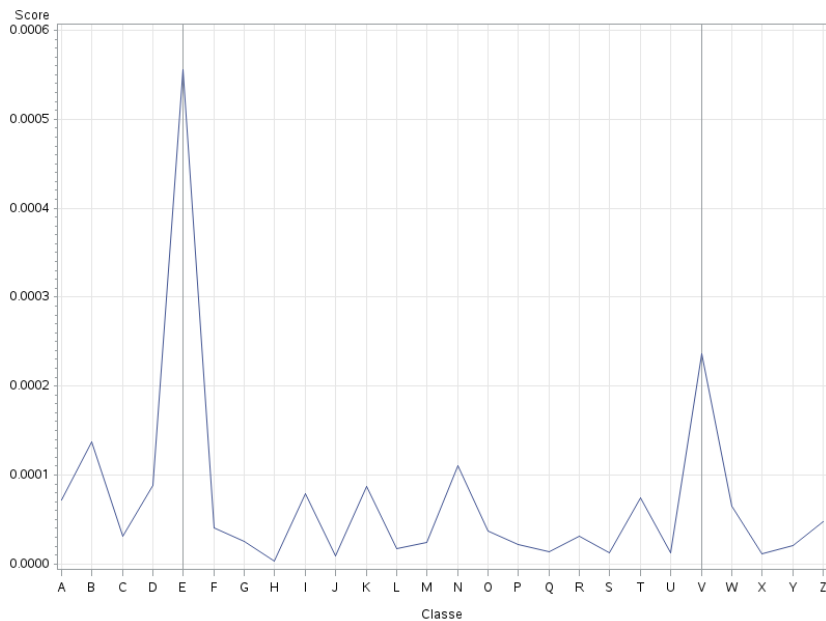
Au cours de cette période, le taux de change a connu des fluctuations, avec des périodes de hausse et de baisse. Cependant, il est important de noter que la série de données brutes n'est pas stationnaire, ce qui signifie que ses propriétés statistiques, comme la moyenne et la variance, changent avec le temps. De plus, il n'y a pas de saisonnalité apparente dans les données, c'est-à-dire qu'il n'y a pas de motifs ou de cycles qui se répètent à intervalles réguliers. Cela peut rendre l'analyse et la prévision du taux de change plus complexe, car les modèles statistiques classiques qui supposent la stationnarité et/ou la saisonnalité peuvent ne pas être appropriés. Par conséquent, il est possible que des transformations soient nécessaires pour rendre cette série stationnaire et faciliter sa prévision.

1.2. Moyenne des données brutes :



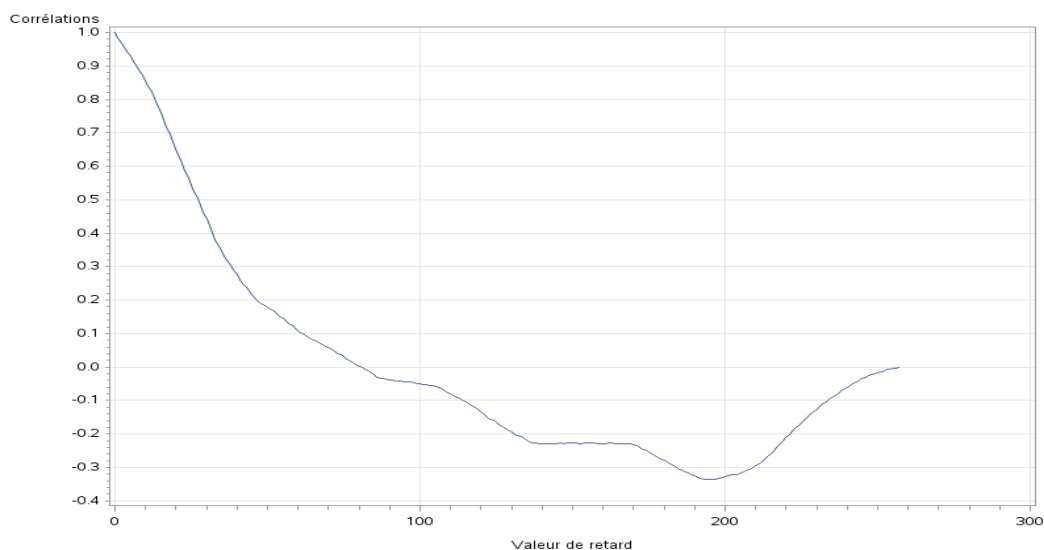
Le graphique des moyennes des classes mets en évidence que les moyennes des taux de changes n'est pas stationnaire en moyenne, des tendances à la hausse et à la baisse peuvent être identifiés, la série n'est donc pas stationnaire en moyenne.

1.3. Variance des données brutes :



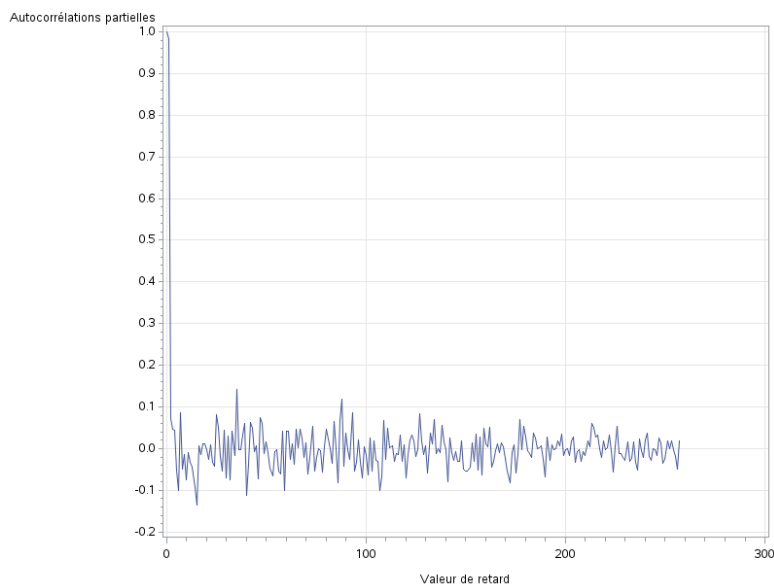
LA classes E présente une variance plus élevée que les autres classes, mais si l'on suppose que c'est une fluctuation conjoncturelle, la série semble stationnaire en variance, nous allons confirmer cela à travers les graphiques des variances de la série en DIF 1 et DIF 2, si nous observons à nouveau ce pique sur ces séries, nous allons alors le considérer comme une fluctuation conjoncturelle.

1.4. ACF Des données brutes :



La fonction d'autocorrélation (ACF) sur les données brutes ne montre pas de pics significatifs à des intervalles saisonniers réguliers, cela suggère que la série n'a pas de saisonnalité. En d'autres termes, il n'y a pas de motifs ou de cycles qui se répètent à intervalles réguliers dans les données, ce qui signifie que notre ARIMA ne sera pas saisonnier.

1.5. PACF :



La fonction d'autocorrélation partielle (PACF) décroît vers 0 et reste dans l'intervalle de confiance à 95%, cela indique que la série est stationnaire.

1.6. Résumé analyse des données brutes :

Stationnarité	DIF 0
Données	Non
Moyennes	Non
Variances	Oui/Non
ACF	Non
PACF	Oui

1. Nous avons constaté que la moyenne de la série varie avec le temps, ce qui signifie qu'elle n'est pas stationnaire
2. Nous avons envisagé que la variance de la série pourrait devenir stationnaire sous certaines conditions, en supposant que le pic observé dans la classe "E" soit le résultat d'une fluctuation conjoncturelle.
3. La fonction d'autocorrélation ACF est également non stationnaire.
4. Cependant, l'analyse du spectre des fluctuations de la fonction de corrélation partielle (PACF) a montré une tendance vers zéro, ce qui signifie que la PACF de la série est stationnaire.

À ce stade et compte tenu de ces constatations, il est devenu évident qu'une prévision fiable de l'avenir de la série n'est pas envisageable. Par conséquent, notre prochaine étape consistera à rendre la série stationnaire, ce qui nous permettra de mieux comprendre ses caractéristiques et de procéder à des prévisions plus précises.

2. Analyse de la série en DIF 1 :

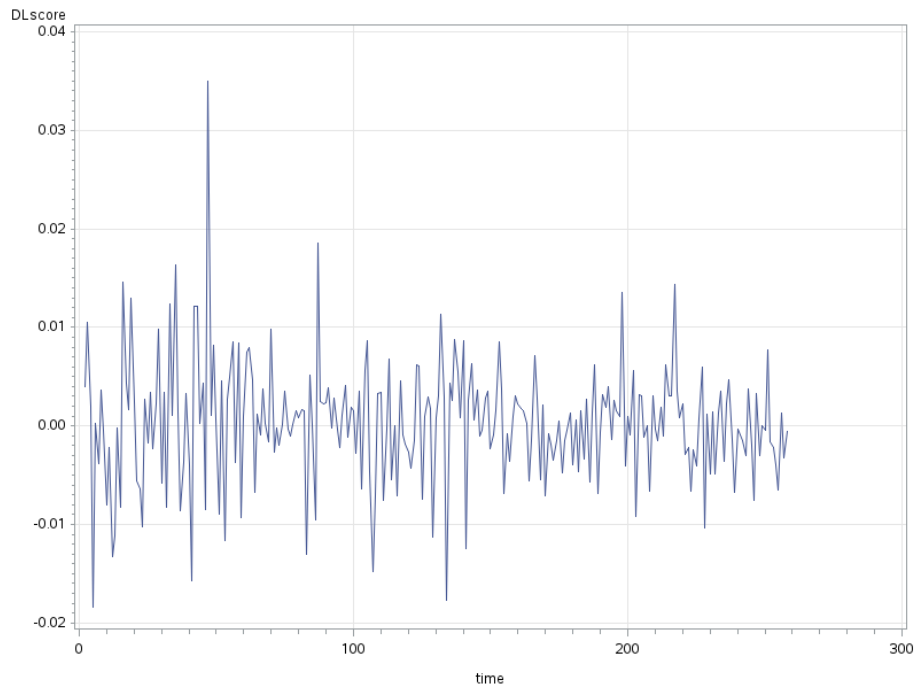
2.1. Différenciation première :

Pour corriger le problème de non-stationnarité des données, de la moyenne et de la variance, nous allons transformer la série en logarithme pour normaliser nos données et faire une différenciation première, ceci aidera à obtenir une moyenne constante.

Nous allons donc réaliser ces transformations et vérifier par la suite la stationnarité de nos données

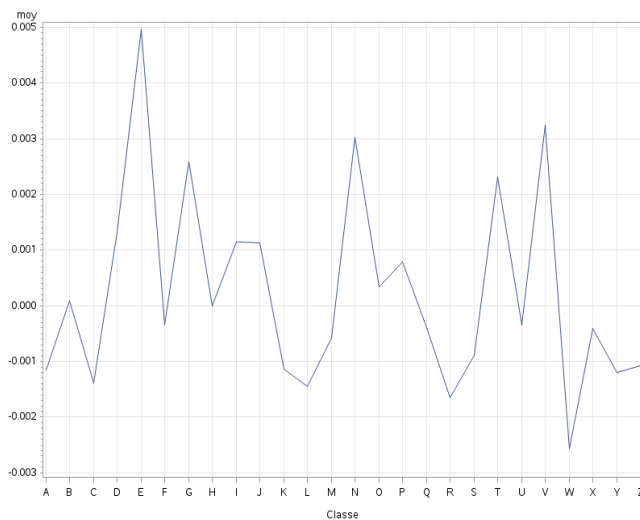
```
DATA base;  
SET ang;  
time=_n_;  
Lscore=log(score); /*passage des données en log*/  
DLscore=dif(Lscore); /*DIF1*/  
RUN;
```

2.2. Graphique de la série DIF 1 :



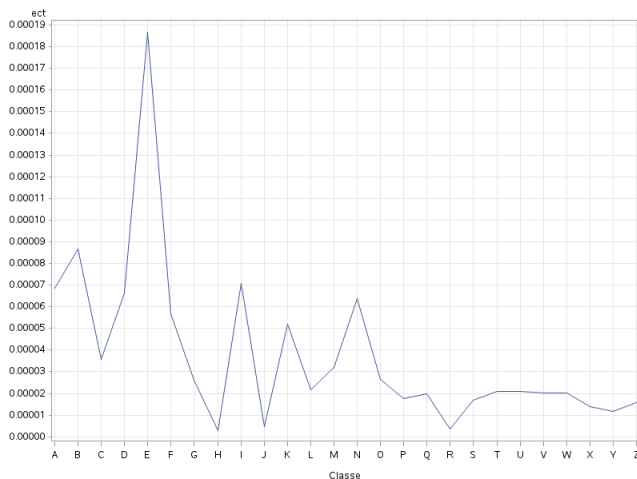
Le graphique de la série désaisonnalisés et différencié une fois, les différentes tendances que nous avons identifiés sur nos données brutes n'apparaissent plus et la série est stationnaire.

2.3. Graphique des moyennes DIF 1 :



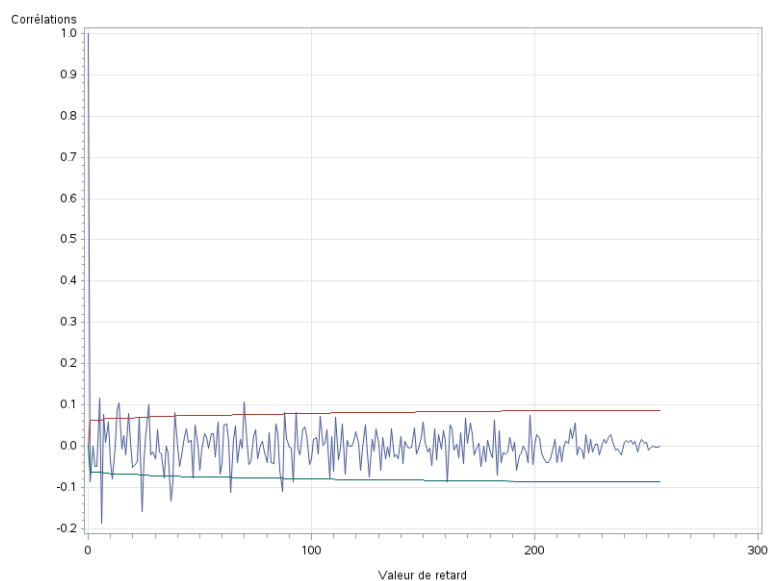
Les tendances précédemment observées ne sont plus visibles. De plus, les moyennes semblent être à peu près égales, ce qui est une indication que la série est maintenant stationnaire.

2.4. Variance en DIF 1 :



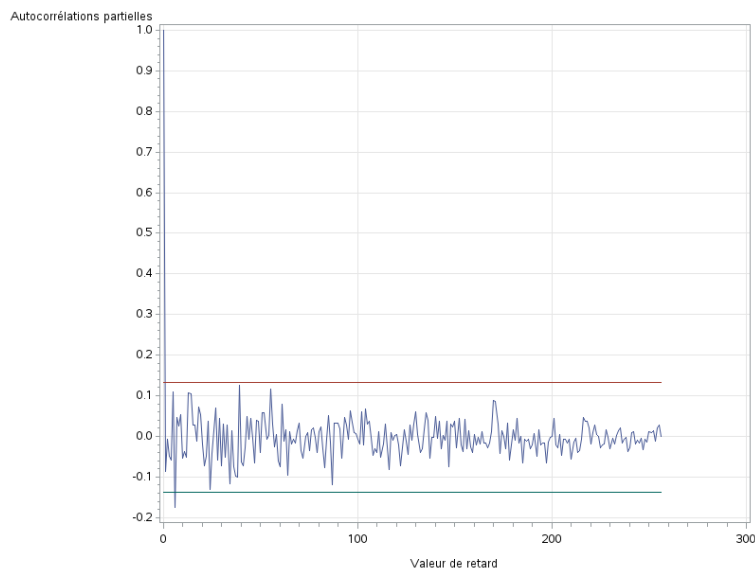
On peut voir que le pic E précédemment observé sur les variances des données brutes est toujours présent, en supposant que cette valeur est une fluctuation conjoncturelle, la série semble stationnaire, nous ne pouvons cependant toujours pas confirmer notre hypothèse sans avoir analysé le graphique des variances en différence seconde.

2.5. ACF EN DIF 1 :



La fonction d'autocorrélation tend vers 0 et ses valeurs sont bien comprise dans l'intervalle de confiance à 95%, nous considérons donc que L'ACF est stationnaire. Certaines valeur de la fonction sortent de l'intervalle de confiance, ces valeurs serviront à définir le paramètre q la moyenne mobile du modèle.

2.6. PACF en DIF 1 :



La fonction d'autocorrélation partielle tend vers 0 et ses valeurs sont bien comprise dans l'intervalle de confiance à 95%, nous considérons donc que L'PACF est stationnaire.

Certaines valeurs de la fonction sortent de l'intervalle de confiance, ces valeurs serviront à définir p le nombre de décalages à considérer pour le modèle autorégressif, le nombre d'ordres des termes AR.

2.7. Résumé de la série en DIF1 :

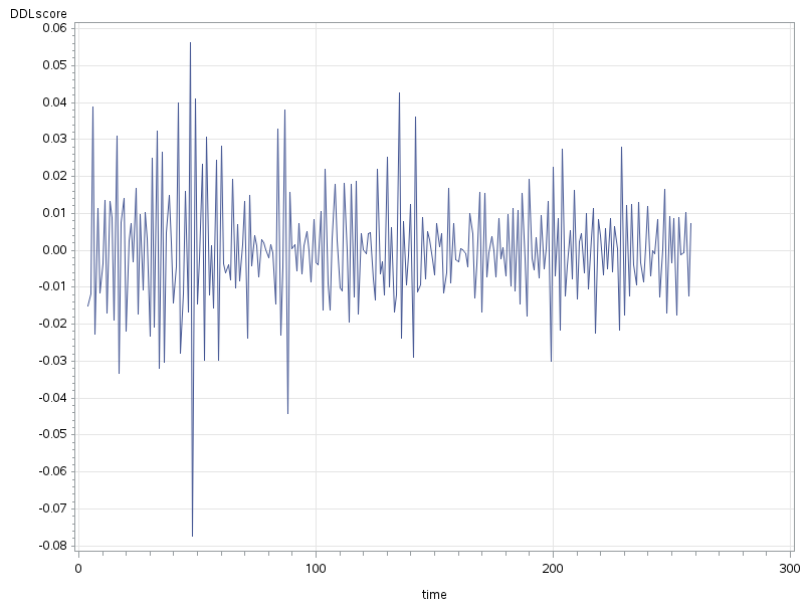
Stationnarité	DIF 1
Données	Oui
Moyennes	Oui
Variances	Oui/Non
ACF	Oui
PACF	Oui

Après avoir désaisonnalisé la série et appliqué une différenciation première tout en effectuant une transformation logarithmique des données brutes, nous avons observé que le graphique des données ne présente plus de tendances ni de saisonnalité. Les moyennes sont stationnaires, mais il semble que les variances ne le soient pas.

Nous avons émis l'hypothèse que le PIC présent pourrait être le résultat de phénomènes conjoncturels. Pour confirmer cela, nous devons appliquer une différenciation seconde à la série. Si les pics de variance persistent après la différenciation seconde, il serait alors acceptable de les ajuster artificiellement pour les ramener aux niveaux observés dans les autres périodes.

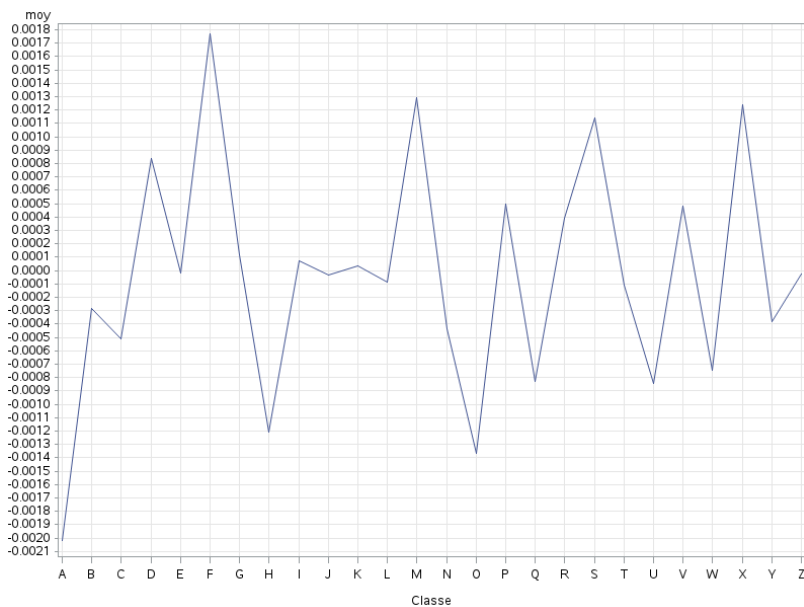
3. Analyse de la série en DIF 2 :

3.1. Graphique des données brutes :



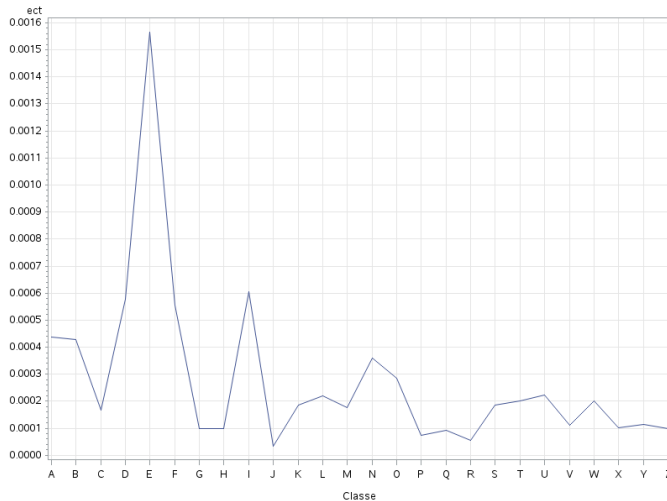
Le graphique de la série différenciés deux fois indique que la série est stationnaire, tout comme après une différenciation de premier ordre.

3.2. Moyenne en Dif 2 :



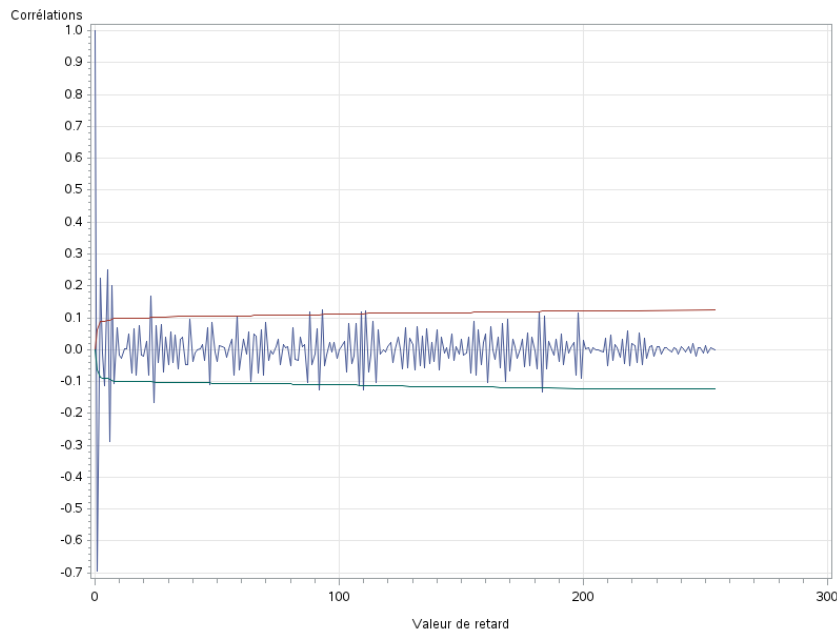
Les moyennes de la série semblent être constantes et encore plus stables que la différenciation première. En conséquence, nous pouvons conclure que la moyenne de la série est stationnaire.

3.3. Variance en Dlf 2 :



Le pic E de la variance que nous avons identifié précédemment dans les données brutes et leurs différences premières demeure présent. Notre hypothèse était que si ce pic se maintenait après une différenciation de premier et de second ordre, nous pourrions conclure qu'il s'agit d'une fluctuation conjoncturelle. En conséquence, nous sommes en mesure d'affirmer que les variances restent stationnaires pour les données brutes ainsi que pour leurs différences premières et secondes.

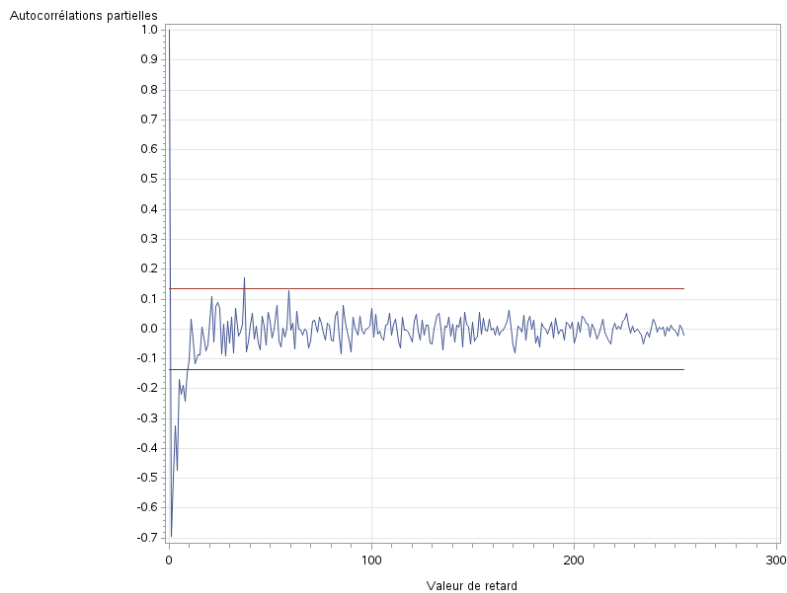
3.4. ACF en DIF 2 :



La fonction d'autocorrélation (ACF) montre une tendance à se rapprocher de zéro et ses valeurs se situent principalement à l'intérieur de l'intervalle de confiance à 95%. Par conséquent, nous concluons que l'ACF est stationnaire.

Cependant, certaines valeurs de la fonction d'autocorrélation dépassent l'intervalle de confiance. Ces valeurs atypiques seront utilisées pour déterminer le paramètre q , qui représente l'ordre de la moyenne mobile du modèle.

3.5. PACF en Dif 2 :



La fonction d'autocorrélation partielle (PACF) montre une tendance à se rapprocher de zéro, et la plupart de ses valeurs se situent à l'intérieur de l'intervalle de confiance à 95%. Par conséquent, nous concluons que la PACF est stationnaire.

Les valeurs de la PACF qui dépassent l'intervalle de confiance seront utilisées pour déterminer le paramètre p , qui représente le nombre de décalages à considérer pour le modèle autorégressif, c'est-à-dire l'ordre des termes AR dans le modèle.

4. Choix de la différenciation :

Stationnarité			
Série	DIF 0	DIF 1	DIF 2
Données	Non	Oui	Oui
Moyennes	Non	Oui	Oui
Variances	Oui	Oui	Oui
ACF	Non	Oui	Oui
PACF	Oui	Oui	Oui

- **Données brutes :**

- La série de données brutes n'est pas stationnaire en termes de moyennes, ce qui signifie que la moyenne de la série varie au fil du temps.
- Cependant, les variances des données brutes sont stationnaires, indiquant que la dispersion des données reste relativement constante.
- La fonction d'autocorrélation (ACF) montre des comportements non stationnaires, suggérant la présence de motifs temporels.
- La PACF est stationnaire en ce qui concerne les moyennes.

- **Différenciation de premier ordre (DIF 1) :**

- Après avoir appliqué une différenciation de premier ordre, la série devient stationnaire en termes de moyennes, ce qui signifie que la moyenne de la série reste relativement constante au fil du temps.
- Les variances restent stationnaires.
- À la fois l'ACF et la PACF sont stationnaires.

- **Différenciation de second ordre (DIF 2) :**

- Lorsque la série subit une différenciation de second ordre, elle maintient la stationnarité en termes de moyennes.
- Les variances demeurent stationnaires.
- Tant l'ACF que la PACF continuent d'afficher la stationnarité.

Le choix entre la série en DIF 1 ou DIF 2 n'est pas évident car les deux séries semblent stationnaires, cependant en analysant les graphiques plus en détail nous pouvons noter que :

Données brutes : DIF1 > DIF2, La DIF1 présente une dispersion moins disparate que la DIF2.

Moyenne : DIF2 > DIF 1, les moyennes de DIF 2 semblent plus concentrées autour de 0, de plus elles semblent plus constantes

Variances : DIF2 > DIF 1, La variance de la DIF2 est en moyenne plus faible que celle de la DIF1.

ACF : DIF2 > DIF1, La fin de la queue des graphiques en DIF1 présente une reprise plus importante que celle en DIF2.

PACF : DIF1 = DIF2, Les graphiques des ACF sont comparables, puisqu'ils tendent tous les deux vers 0 et présentent des reprises à la fin.

Le choix entre DIF 1 et DIF 2 n'est pas évident, les deux séries sont stationnaires, même si les données brutes de DIF1 semblent plus stationnaires que la DIF 2, les moyennes et variances de DIF 2 semblent aussi plus stationnaires que la DIF 1, il est difficile de choisir graphiquement quelle série est plus stationnaire que l'autre, c'est la raison pour laquelle nous allons estimer les modèles en DIF 1 et en DIF2, et utiliser les critères d'AIC BIC, et le coefficient d'inégalité de Theil, qui sont des critères plus précis pour choisir entre les deux séries.

4.1. Estimation des paramètres q et p :

Les retards du PACF qui se situent au-delà de notre intervalle de confiance seront utilisés pour définir les paramètres autorégressifs "p" (AR). Ces valeurs nous aideront à définir combien de décalages de temps passés sont nécessaires pour expliquer les motifs de la série temporelle.

De manière similaire, les paramètres de l'ACF qui se trouvent en dehors de l'intervalle de confiance seront utilisés pour définir les valeurs des moyennes mobiles, "q" (MA). Ces valeurs indiqueront combien de décalages de temps passés doivent être pris en compte pour tenir compte des variations aléatoires ou des erreurs dans la série temporelle.

Dif 1 :

Les processus autorégressifs retenus sortants de l'intervalle de confiance sont les suivants :

- $P = (6)$.

Les processus de moyennes mobiles qui sortent de l'intervalle de confiance sont relativement importants, nous avons donc choisi les 15 premiers paramètres sortant de l'intervalle de confiance, si cela ne suffit pas alors nous prendrons tous les retards :

- $q = (1\ 5\ 6\ 7\ 11\ 13\ 14\ 18\ 23\ 24\ 27\ 34\ 37\ 38\ 39)$.

Dif 2 :

Les processus autorégressifs retenus sortants de l'intervalle de confiance sont les suivants :

- $P = (1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 37)$.

Les processus de moyennes mobiles retenues sortants de l'intervalle de confiance sont les suivants :

- $q=(1\ 2\ 4\ 5\ 6\ 7\ 8\ 23\ 24\ 47\ 88\ 92\ 93\ 108\ 109\ 110\ 111\ 183)$

4.2. Estimation des modèles :

4.2.1. En différence d'ordre 1 :

4.2.1.1. Estimation

```
PROC ARIMA data=base;
IDENTIFY var=DLscore scan;
ESTIMATE q=(1 5 6 7 11 13 14 18 23 24 27 34 37 38 39) p=(6) method=ml;
FORECAST id=time printall out=b;
RUN;
```

Estimation du maximum de vraisemblance					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr > t	Retard
MU	0.0002311	0.0004108	0.56	0.5737	0
MA1,1	0.08776	0.06108	1.44	0.1508	1
MA1,2	-0.09080	0.06155	-1.48	0.1402	5
MA1,3	0.15613	0.14799	1.06	0.2914	6
MA1,4	-0.15151	0.06262	-2.42	0.0155	7
MA1,5	0.03703	0.06296	0.59	0.5564	11
MA1,6	-0.03205	0.06482	-0.49	0.6210	13
MA1,7	-0.18441	0.06303	-2.93	0.0034	14
MA1,8	-0.12251	0.06222	-1.97	0.0490	18
MA1,9	-0.12843	0.06346	-2.02	0.0430	23
MA1,10	0.12565	0.06423	1.96	0.0504	24
MA1,11	-0.16646	0.06421	-2.59	0.0095	27
MA1,12	0.07647	0.06310	1.21	0.2255	34
MA1,13	0.12197	0.06332	1.93	0.0541	37
MA1,14	0.12519	0.06490	1.93	0.0538	38
MA1,15	-0.08127	0.06299	-1.29	0.1970	39
AR1,1	-0.05186	0.15702	-0.33	0.7412	6

Estimation constante	0.000243
Estimation variance	0.000033
Erreur type Valeur estimée	0.005702
AIC	-1904.95
SBC	-1844.61
Nombre de résidus	257

ainsi nous avons estimé notre modèle ARIMA avec les différents retards que nous avons identifié, et notre exogène différencié une fois.

nous pouvons voir que la significativité de nos variables est relativement faible, pour améliorer notre modèle, les variables non significatives par tâtonnement, en

éliminant à chaque estimation la variable la moins significative, nous nous arrêteront quand toutes nos variables seront significative au seuil de 5%.

Cette approche nous permettra de simplifier notre modèle tout en conservant uniquement les variables qui ont une contribution statistiquement significative.

4.2.1.2. Phase de tâtonnement :

```

/***** PREVISION *****/
/***** PREVISION *****/
/* Dif 1 */
PROC ARIMA data=b ;
IDENTIFY var=Lscore(1) ;
ESTIMATE q=(5 6 7 14 23 27) method=ml;
FORECAST id=time printall out=b;
RUN;

/* GRAPHIQUE DE LA PREVISION */

PROC GPLOT data=b;
PLOT (Lscore forecast 195 u95)*time / overlay GRID;
SYMBOL interpol=join;
*WHERE time >300;
RUN;

/* Calcul de l'erreur */
DATA b;
SET b;
error = Lscore - forecast;
RUN;

/* Calcul du carré de l'erreur */
DATA b;
SET b;
sq_error = error*error;
RUN;

/* Calcul de la moyenne du carré de l'erreur */
PROC SQL;
CREATE TABLE rmse AS
SELECT AVG(sq_error) AS mean_sq_error
FROM b;
QUIT;

/* Calcul de la racine carrée de la moyenne du carré de l'erreur (RMSE) */
DATA rmse;
SET rmse;
rmse_value = SQRT(mean_sq_error);
RUN;

/* Affichage du RMSE */
PROC PRINT DATA=rmse;
RUN;

```

Estimation du maximum de vraisemblance					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr > t	Retard
MU	0.0001351	0.0005633	0.24	0.8105	0
MA1,1	-0.13576	0.05823	-2.33	0.0197	5
MA1,2	0.21515	0.05824	3.69	0.0002	6
MA1,3	-0.13843	0.05796	-2.39	0.0169	7
MA1,4	-0.19834	0.05915	-3.35	0.0008	14
MA1,5	-0.13002	0.05934	-2.19	0.0284	23
MA1,6	-0.23241	0.05980	-3.89	0.0001	27

Estimation constante	0.000135
Estimation variance	0.000033
Erreur type Valeur estimée	0.005776
AIC	-1909.5
SBC	-1884.66
Nombre de résidus	257

Obs.	mean_sq_error	rmse_value
1	.000033298	.005770463

En procédant par tâtonnement, nous avons obtenu un modèle dépourvu de processus autorégressifs (p) et avec les moyennes mobiles (q) suivantes :

- q = (5, 6, 7, 14, 23, 27).

Il est important de noter que les valeurs de l'AIC et du BIC du modèle ont considérablement diminué par rapport à notre modèle initial, de même que la racine de l'erreur quadratique moyenne (RMSE).

En utilisant les critères de sélection tels que l'AIC, le BIC et le RMSE, nous allons évaluer si la suppression du dernier retard que nous avons éliminé lors de l'estimation précédente, c'est-à-dire 'q=(38)' avec une significativité de 10%, est justifiée ou non. L'objectif est de minimiser ces trois critères, car cela nous permettra d'obtenir un modèle qui offre le meilleur ajustement aux données tout en maintenant la simplicité du modèle.

Estimation du maximum de vraisemblance					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr > t	Retard
MU	0.0001741	0.0005310	0.33	0.7430	0
MA1,1	-0.12015	0.05906	-2.03	0.0419	5
MA1,2	0.20830	0.05868	3.55	0.0004	6
MA1,3	-0.14293	0.05822	-2.45	0.0141	7
MA1,4	-0.20726	0.05957	-3.48	0.0005	14
MA1,5	-0.14026	0.05980	-2.35	0.0190	23
MA1,6	-0.22004	0.06051	-3.64	0.0003	27
MA1,7	0.09886	0.06028	1.64	0.1010	38

Estimation constante	0.000174
Estimation variance	0.000033
Erreur type Valeur estimée	0.005753
AIC	-1910.08
SBC	-1881.69
Nombre de résidus	257

En incluant q(38), l'AIC du modèle diminue ce qui signifie que le modèle est meilleur en termes de qualité d'ajustement, le BIC augmente légèrement qui veut dire que le modèle est moins bon en terme de simplicité et de parcimonie. pour départager les deux modèle nous allons également calculer le RMSE l'erreur quadratique moyenne pour voir qu'elle modèle minimise cette erreur :

RMSE modèle sans q(38) :

Obs.	mean_sq_error	rmse_value
1	.000033298	.005770463

RMSE du modèle incluant q (38) :

Obs.	mean_sq_error	rmse_value
1	.000033018	.005746162

L'ajout du retard q(38) au modèle a entraîné une réduction du RMSE, ce qui témoigne d'une amélioration de la précision du modèle, cela indique que le modèle est devenu plus efficace pour minimiser les écarts entre les valeurs prédites et les valeurs observées.

4.2.2. En différence d'ordre 2 :

4.2.2.1. Estimation

Estimation du maximum de vraisemblance					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr > t	Retard
MU	-9.3438E-6	0.00003438	-0.27	0.7858	0
MA1,1	-0.26241	1.25183	-0.21	0.8340	1
MA1,2	0.33056	0.80651	0.41	0.6819	2
MA1,3	0.18142	0.76787	0.24	0.8132	4
MA1,4	0.21555	0.77876	0.28	0.7819	5
MA1,5	0.37765	0.77940	0.48	0.6280	6
MA1,6	0.07200	0.88474	0.08	0.9351	7
MA1,7	-0.08105	0.49756	-0.16	0.8706	8
MA1,8	-0.04798	0.09662	-0.50	0.6195	23
MA1,9	-0.01675	0.11476	-0.15	0.8839	24
MA1,10	0.01270	0.06500	0.20	0.8451	47
MA1,11	-0.01526	0.06617	-0.23	0.8177	88
MA1,12	0.04756	0.10868	0.44	0.6617	92
MA1,13	0.03756	0.11348	0.33	0.7406	93
MA1,14	0.03266	0.11846	0.28	0.7828	108
MA1,15	0.02673	0.18826	0.14	0.8871	109
MA1,16	0.03920	0.18819	0.21	0.8350	110
MA1,17	0.01413	0.12632	0.11	0.9109	111
MA1,18	0.03568	0.09314	0.38	0.7016	183
AR1,1	-1.13576	1.21199	-0.94	0.3487	1
AR1,2	-0.86775	1.41153	-0.61	0.5387	2
AR1,3	-1.13426	1.41187	-0.80	0.4218	3
AR1,4	-0.96837	1.77495	-0.55	0.5854	4
AR1,5	-0.77371	1.54374	-0.50	0.6162	5

AR1,6	-0.42519	1.29418	-0.33	0.7425	6
AR1,7	-0.32715	0.63140	-0.52	0.6044	7
AR1,8	-0.25854	0.44260	-0.58	0.5591	8
AR1,9	-0.04341	0.24952	-0.17	0.8619	9
AR1,10	0.01065	0.03710	0.29	0.7740	37

Estimation constante	-0.00006
Estimation variance	0.000109
Erreur type Valeur estimée	0.010456
AIC	-1568.65
SBC	-1465.96
Nombre de résidus	255

Nous avons procédé à l'estimation de notre modèle ARIMA en incluant les différents retards que nous avons identifiés, ainsi que notre variable exogène différenciée deux

fois. Nous avons observé que la significativité de nos variables est relativement faible. Afin d'améliorer notre modèle, nous également faire une approche itérative pour éliminer les variables non significatives. À chaque itération, nous retirons la variable qui présente la moins bonne significativité et répéter cette approche jusqu'à ce que toutes nos variables atteignent une significativité inférieure à 5%.

4.2.2.2. Phase de tâtonnement :

Estimation du maximum de vraisemblance					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr > t	Retard
MU	9.62567E-8	2.51617E-6	0.04	0.9695	0
MA1,1	1.00000	0	Infini	<.0001	2
AR1,1	-1.64326	0.05875	-27.97	<.0001	1
AR1,2	-0.99851	0.11534	-8.66	<.0001	2
AR1,3	-0.54181	0.11819	-4.58	<.0001	3
AR1,4	-0.28726	0.07089	-4.05	<.0001	4
AR1,5	-0.08202	0.02312	-3.55	0.0004	7

Estimation constante	4.382E-7
Estimation variance	0.000046
Erreur type Valeur estimée	0.006813
AIC	-1804.35
SBC	-1779.56
Nombre de résidus	255

En procédant par tâtonnement, nous avons obtenu un modèle avec :
les processus autorégressifs (p) suivants :

- $p=(1\ 2\ 3\ 4\ 7)$

Les moyennes mobiles (q) suivantes :

- $q=(2)$

Comme nous l'avons fait pour le modèle différencié d'ordre 1, nous allons utiliser les mêmes critères pour voir si le dernier retard que nous avons supprimé dans l'itération précédente $q=(92)$ avec une significativité de 11%, est justifiée ou non. L'objectif est

de minimiser ces trois critères :

Estimation du maximum de vraisemblance					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr > t	Retard
MU	1.37285E-6	8.61373E-6	0.16	0.8734	0
MA1,1	0.91487	0.10921	8.38	<.0001	2
MA1,2	0.08513	0.05406	1.57	0.1153	92
AR1,1	-1.63076	0.06460	-25.24	<.0001	1
AR1,2	-1.03341	0.11595	-8.91	<.0001	2
AR1,3	-0.54838	0.12065	-4.55	<.0001	3
AR1,4	-0.26115	0.07204	-3.63	0.0003	4
AR1,5	-0.08183	0.02438	-3.36	0.0008	7

Estimation constante	6.254E-6
Estimation variance	0.000049
Erreur type Valeur estimée	0.006965
AIC	-1791.63
SBC	-1763.3
Nombre de résidus	255

En incluant le retard $q = (92)$, l'AIC et LE BIC augmentent, ce qui signifie que le modèle est moins bon en termes de qualité d'ajustement, le BIC augmente aussi qui veut dire que le modèle est moins bon en termes de simplicité et de parcimonie.

Calcul du RMSE modèle avec $q = (92)$:

Obs.	mean_sq_error	rmse_value
1	.000051116	.007149515

Calcul du RMSE modèle sans $q = (92)$:

Obs.	mean_sq_error	rmse_value
1	.000049194	.007013819

L'ajout du retard $q = (92)$ au modèle a entraîné une augmentation du RMSE, ce qui témoigne d'une baisse de la précision du modèle, cela indique que les écarts entre les valeurs prédites et les valeurs observées ont augmentés.

4.3. Sélection entre DIF (1) et DIF (2) :

4.3.1. AIC BIC RMSE :

DIF(1)

Estimation du maximum de vraisemblance					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr > t	Retard
MU	0.0001741	0.0005310	0.33	0.7430	0
MA1,1	-0.12015	0.05906	-2.03	0.0419	5
MA1,2	0.20830	0.05868	3.55	0.0004	6
MA1,3	-0.14293	0.05822	-2.45	0.0141	7
MA1,4	-0.20726	0.05957	-3.48	0.0005	14
MA1,5	-0.14026	0.05980	-2.35	0.0190	23
MA1,6	-0.22004	0.06051	-3.64	0.0003	27
MA1,7	0.09886	0.06028	1.64	0.1010	38

Estimation constante	0.000174
Estimation variance	0.000033
Erreur type Valeur estimée	0.005753
AIC	-1910.08
SBC	-1881.69
Nombre de résidus	257

Obs.	mean_sq_error	rmse_value
1	.000033018	.005746162

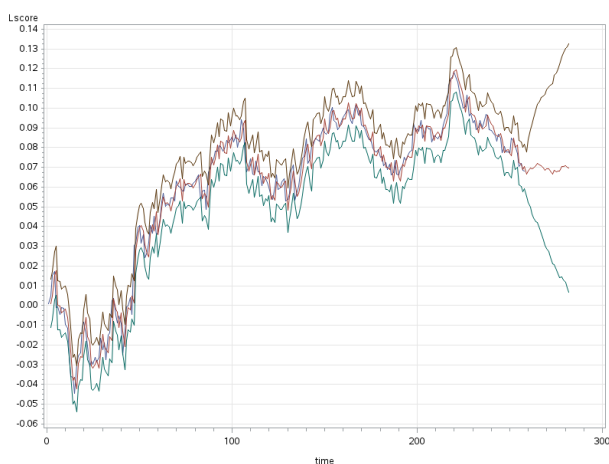
DIF(2)

Estimation du maximum de vraisemblance					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr > t	Retard
MU	9.62567E-8	2.51617E-6	0.04	0.9695	0
MA1,1	1.00000	0	Infini	<.0001	2
AR1,1	-1.64326	0.05875	-27.97	<.0001	1
AR1,2	-0.99851	0.11534	-8.66	<.0001	2
AR1,3	-0.54181	0.11819	-4.58	<.0001	3
AR1,4	-0.28726	0.07089	-4.05	<.0001	4
AR1,5	-0.08202	0.02312	-3.55	0.0004	7

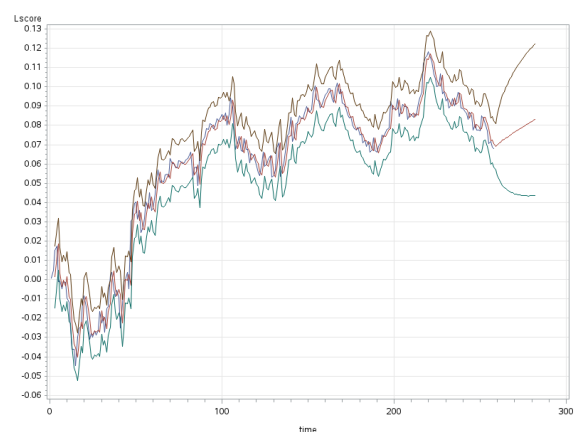
Estimation constante	4.382E-7
Estimation variance	0.000046
Erreur type Valeur estimée	0.006813
AIC	-1804.35
SBC	-1779.56
Nombre de résidus	255

Obs.	mean_sq_error	rmse_value
1	.000049194	.007013819

DIF (1)



DIF (2)



Le modèle en DIF1 présente des métriques d'évaluation de modèle meilleures que le modèle en DIF2. En particulier, le modèle en DIF1 affiche un AIC et un BIC plus faibles que le modèle en DIF2, ce qui suggère que le modèle en DIF1 est plus approprié en termes de parcimonie et de capacité à expliquer les données avec moins de complexité.

De plus, le modèle en DIF1 présente également un RMSE (Erreur Quadratique Moyenne) plus faible par rapport au modèle en DIF2. Cette réduction du RMSE signifie que le modèle en DIF1 est plus précis dans ses prédictions, avec des erreurs de prévision globalement plus petites.

Les mesures d'évaluation du modèle indiquent que le modèle en DIF1 est préférable au modèle en DIF2. Il est à la fois plus parcimonieux, mieux ajusté aux données et offre une meilleure précision dans ses prédictions.

Pour confirmer nos résultats et prendre une décision éclairée entre les deux modèles, nous allons également calculer le coefficient d'inégalité de Theil. Ce coefficient est un indicateur important pour évaluer la performance prédictive des modèles et leur capacité à minimiser les erreurs de prévision. Il nous permettra de quantifier la qualité relative des prévisions générées par chaque modèle.

La ligne rouge représente notre prévision, à la fois dans le passé et le futur. Cependant, il est essentiel de comprendre que cette prévision devient de moins en moins précise à mesure que nous avançons dans le temps. La méthode que nous utilisons est principalement destinée à des prévisions de court terme. Les valeurs prédites sur la courbe sont exprimées de manière logarithmique. Pour les interpréter correctement, nous devons les convertir en utilisant une fonction exponentielle, car le logarithme naturel a été appliqué précédemment.

Les courbes marrons au-dessus et vertes en dessous de notre prévision représentent l'intervalle de confiance à 95% de nos prévisions. Elles indiquent l'intervalle dans laquelle nous pouvons raisonnablement nous attendre à ce que les valeurs réelles se situent.

4.3.2. Calcul du coefficient d'inégalité de Theil :

DIF (1) :

```
DATA fiabilite;
SET new;
Um=((0.06012-0.06017)**2)/ 0.000033;
Us=((0.03871-0.03842)**2)/ 0.000033;
Uc=(2*(1-0.98889)*(0.0387146*0.0384178))/0.000033;
total=um+us+uc;
RUN;
```

Estimation constante	0.000174
Estimation variance	0.000033
Erreur type Valeur estimée	0.005753
AIC	-1910.08
SBC	-1881.69
Nombre de résidus	257

La procédure MEANS

Variable	Libellé	N	Moyenne	Ec-type	Minimum	Maximum
FORECAST	Prévision pour Lscore	257	0.0601229	0.0387146	-0.0422204	0.1194213
Lscore		257	0.0601733	0.0384178	-0.0444745	0.1182274

La procédure CORR

2 Variables : Lscore FORECAST

Statistiques simples

Variable	N	Moyenne	Ec-type	Somme	Minimum	Maximum	Libellé
Lscore	257	0.06017	0.03842	15.46454	-0.04447	0.11823	
FORECAST	257	0.06012	0.03871	15.45159	-0.04222	0.11942	Prévision pour Lscore

Coefficients de corrélation de Pearson, N = 257
Proba > |r| sous H0: Rho=0

	Lscore	FORECAST
Lscore	1.00000	0.98889 < .0001
FORECAST	0.98889 < .0001	1.00000
Prévision pour Lscore		

Um	Us	Uc
.000075758	.002548485	1.00147

DIF 2 :

```
DATA fiabilite;
SET new;
Um=((0.06002-0.06039)**2)/ 0.000038;
Us=((0.03818-0.03834)**2)/ 0.000038;
Uc=(2*(1-0.98686)*(0.0381763*0.0383364))/0.000038;
total=um+us+uc;
RUN;
```

Estimation constante	0.000015
Estimation variance	0.000038
Erreur type Valeur estimée	0.006154
AIC	-1866.63
SBC	-1841.82
Nombre de résidus	256

La procédure MEANS

Variable	Libellé	N	Moyenne	Ec-type	Minimum	Maximum
FORECAST	Prévision pour Lscore	256	0.0600222	0.0381763	-0.0399938	0.1169159
Lscore		256	0.0603893	0.0383364	-0.0444745	0.1182274

La procédure CORR

2 Variables : Lscore FORECAST

Statistiques simples

Variable	N	Moyenne	Ec-type	Somme	Minimum	Maximum	Libellé
Lscore	256	0.06039	0.03834	15.45965	-0.04447	0.11823	
FORECAST	256	0.06002	0.03818	15.36569	-0.03999	0.11692	Prévision pour Lscore

Coefficients de corrélation de Pearson, N = 256
Proba > |r| sous H0: Rho=0

	Lscore	FORECAST
Lscore	1.00000	0.98686 <.0001
FORECAST Prévision pour Lscore	0.98686 <.0001	1.00000

Um	Us	Uc
.003602632	.000673684	1.01215

Interprétation :

- Pour le modèle en DIF1 :
 - UM (Mauvaise prévision) : 0.000075758
 - US (Sous-estimation) : 0.002548485

- UC (Sous-estimation cumulative) : 1.00147
- Pour le modèle en DIF2 :
 - UM (Mauvaise prévision) : 0.003602632
 - US (Sous-estimation) : 0.000673684
 - UC (Sous-estimation cumulative) : 1.01215

Le modèle en DIF1 présente un UM beaucoup plus faible que le modèle en DIF2, ce qui signifie qu'il a de meilleures prévisions globales.

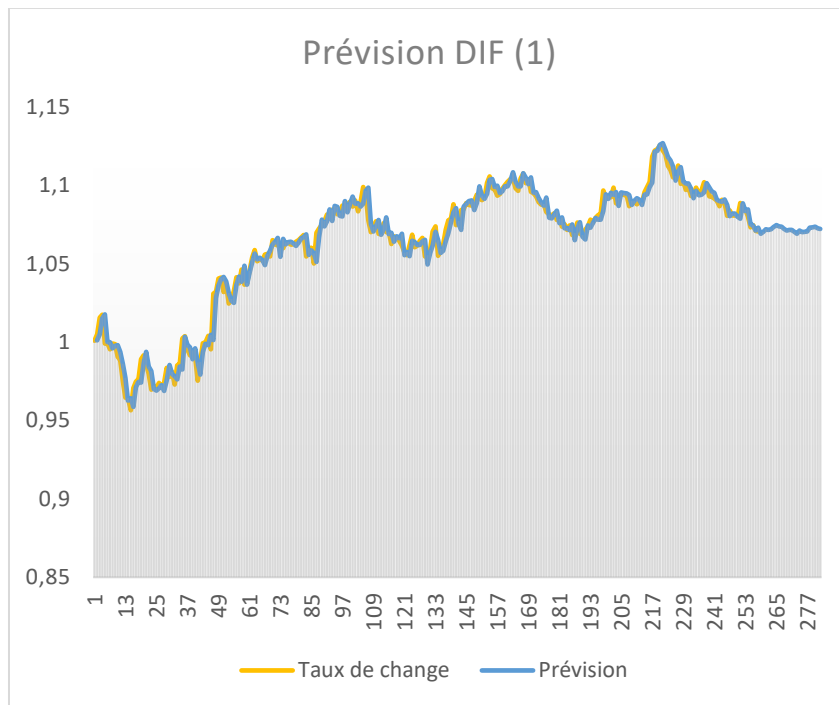
Le modèle en DIF1 a également un US plus faible, ce qui suggère qu'il sous-estime moins les valeurs réelles par rapport au modèle en DIF2.

En ce qui concerne l'UC, le modèle en DIF1 a une valeur légèrement inférieure à 1, ce qui peut indiquer une légère sous-estimation cumulative, tandis que le modèle en DIF2 a une valeur légèrement supérieure à 1, indiquant une légère sous-estimation cumulative.

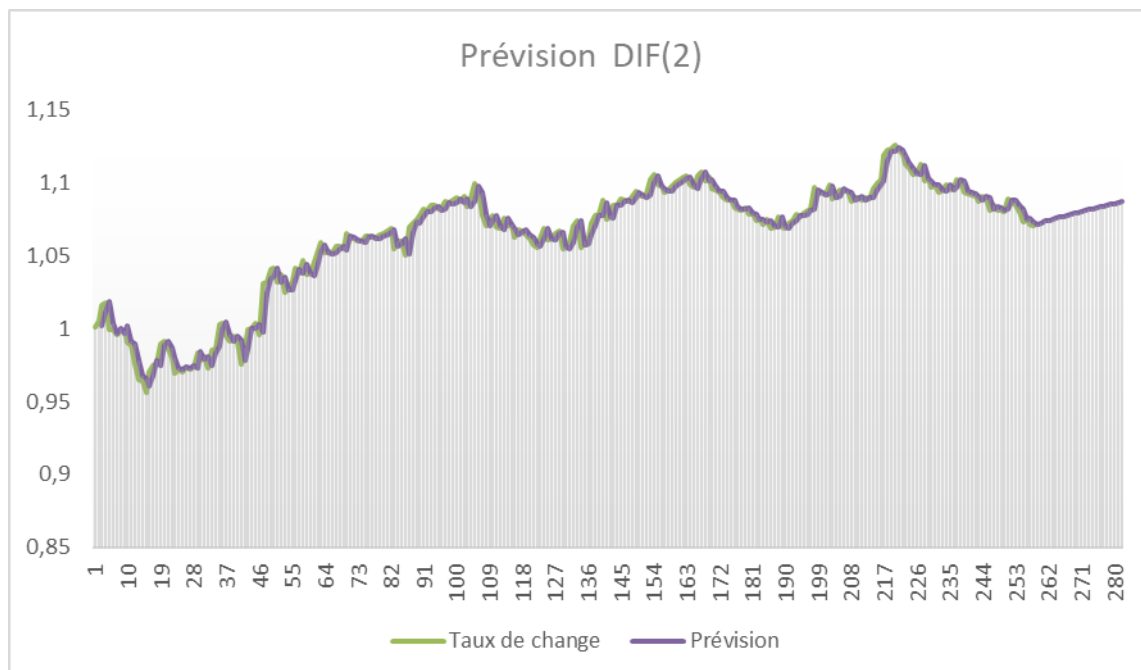
Nous pouvons donc conclure que le modèle en DIF (1) présente de meilleures prévisions que le modèle en DIF (2).

4.4. Prédiction en données brutes :

DIF (1)



DIF (2)



Nous avons utilisé une fonction inverse, la fonction exponentielle sur notre série logarithmique pour afficher nos prévisions en données brutes.

Il est remarquable que les deux modèles offrent de bonnes prévisions pour la période passée, reproduisant de manière cohérente les tendances et les variations observées dans nos données financières, mais les prévisions sur le futur sont relativement différentes.

Comme précédemment démontré, le modèle basé sur une différenciation d'ordre 1 (DIF 1) se révèle être le choix préférable en termes de qualité d'ajustement, de parcimonie et de précision. Ce modèle, qui a montré sa robustesse en étant capable de reproduire avec succès les données passées, propose également des prévisions plus crédibles pour l'avenir, nous accordons donc une plus grande crédibilité aux prévisions générées par le modèle en DIF 1.

• Conclusion :

Cette étude a été une occasion d'explorer l'analyse des séries temporelles et la prévision des taux de change entre l'euro et le dollar américain. En utilisant la modélisation ARIMA, il a été possible de démystifier la complexité de la série temporelle des taux de change et d'élaborer des prévisions fiables. Notre analyse approfondie des données brutes a révélé des tendances et des fluctuations dans les taux de change euro-dollar. Nous avons identifié des défis majeurs liés à la non-stationnarité des données, avec des moyennes et des variances changeantes au fil du temps. De plus, l'absence de saisonnalité apparente a nécessité des transformations pour rendre la série stationnaire.

Après avoir effectué des différenciations successives, nous avons choisi entre une différenciation d'ordre 1 (DIF1) et d'ordre 2 (DIF2) en utilisant des critères tels que l'AIC, le BIC, le RMSE et le coefficient d'inégalité de Theil. Cette étape cruciale nous a permis de sélectionner le modèle ARIMA optimal, en tenant compte à la fois de la qualité des ajustements et de la parcimonie du modèle.

Nos résultats ont montré que le modèle ARIMA en DIF1 était supérieur en termes de précision des prévisions et de simplicité par rapport au modèle en DIF2. Cela a été confirmé par le coefficient d'inégalité de Theil, qui a montré que le modèle en DIF1 avait de meilleures performances globales.

Bien que les prévisions futures de taux de change comportent toujours une part d'incertitude inhérente aux marchés financiers, le modèle a démontré sa capacité à fournir des informations précieuses.

Ce projet met en évidence l'importance de l'analyse des séries temporelles et des modèles ARIMA dans la prévision des taux de change. Il souligne également l'efficacité de la méthode Box-Jenkins. Il montre comment ces techniques peuvent être appliquées pour mieux comprendre et anticiper les tendances sur les marchés financiers internationaux. Cependant, il convient de noter que ce n'est qu'une approche parmi tant d'autres pour modéliser et prévoir les taux de change. D'autres méthodes pourraient également être explorées et comparées à l'approche adoptée dans ce projet pour une compréhension plus complète du sujet.