



# Méthode de Prévision sous SAS

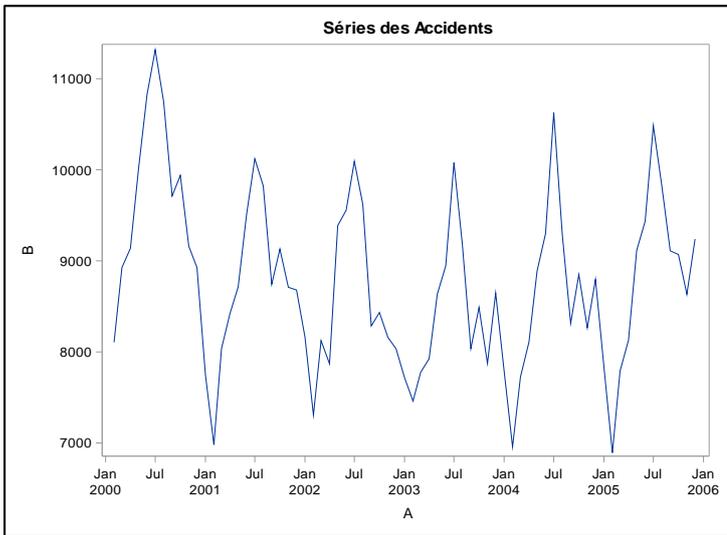
KAZEMI KOUROSH

04/05/2021

## Méthode de prévision : Document 1

### Q1/ Quel est le modèle le plus adapté pour générer des prévisions sur la série 'Accidents mortels' ?

On a récupéré des données sur le nombre d'accidents mortels entre le 1<sup>er</sup> janvier 2000 jusqu'au 1<sup>er</sup> décembre 2005 et on souhaite prédire nos données à l'aide de la méthode de lissage de Winter.

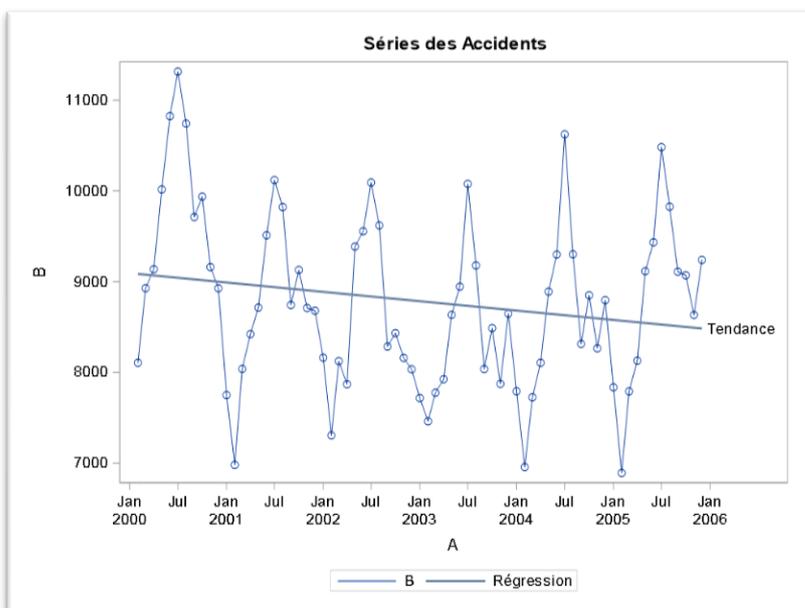


- On a une série chronologique qui s'étale entre le 1<sup>er</sup> janvier 2000 et le 1<sup>er</sup> décembre 2005 (71 observations) avec une fréquence mensuelle. On a un pic au mois de juillet et une baisse au mois de février. De plus on note que la

- On peut remarquer la présence d'une saisonnalité dans le modèle. La longueur de cette saisonnalité est égale à 12. Avec la présence de saisonnalité, seules les méthodes de Winters sont adéquates.

Afin de choisir entre la méthode de lissage de Winter multiplicatif et additif il existe trois méthodes : Méthode du profil, Méthode de la bande, Méthode analytique. On souhaite vérifier s'il existe analytiquement une tendance :

On calcule la moyenne et les écarts types pour chacune des périodes et on trace la droite des Moindres carrés.



- On remarque que la droite des Moindres Carrés n'est pas nul, or cela décrit une tendance dans notre série.  
- Le modèle le plus adapté serait celui du lissage de Winter Multiplicatif car ce dernier prend en compte la tendance décroissante du modèle.

## Q2/ Utiliser la méthode de décomposition multiplicative pour générer des prévisions de la variable 'Accidents mortels' pour les années 2004 et 2005.

- On divise nos données en deux échantillons, une partie apprentissage (jusqu'à 2004) ce qui permet d'estimer les paramètres de prévisions sur cet échantillon. Puis on a une partie test qui permet de vérifier la fiabilité de notre modèle de prévision en calculant différents critères de comparaison.

- PROC X11 permet d'outputer des sorties.

- On cite les tableaux qu'on veut sortir :

b1 = série observée

d10= composante saisonnière.

d11= série désaisonnalisée

d12= estimation de la tendance

d13= estimation de la composante aléatoire dans la série désaisonnalisée.

- A l'aide de la PROC X11 on a décomposé une série temporelle.

- La table 12 nous permet d'avoir une tendance qui n'est pas linéaire et qui est impactée par la composante cyclique.

- Pour chercher la composante 'tendance linéaire' je dois régresser la série sans les variations linéaire sur la variable temps.

### Procédure REG :

#### Prévisions avec les différentes méthodes de lissage

Nb d'observations lues	59
Nb d'obs. utilisées	59

Analyse de variance					
Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Modèle	1	6672349	6672349	53.37	<.0001
Erreur	57	7125952	125017		
Total sommes corrigées	58	13798301			

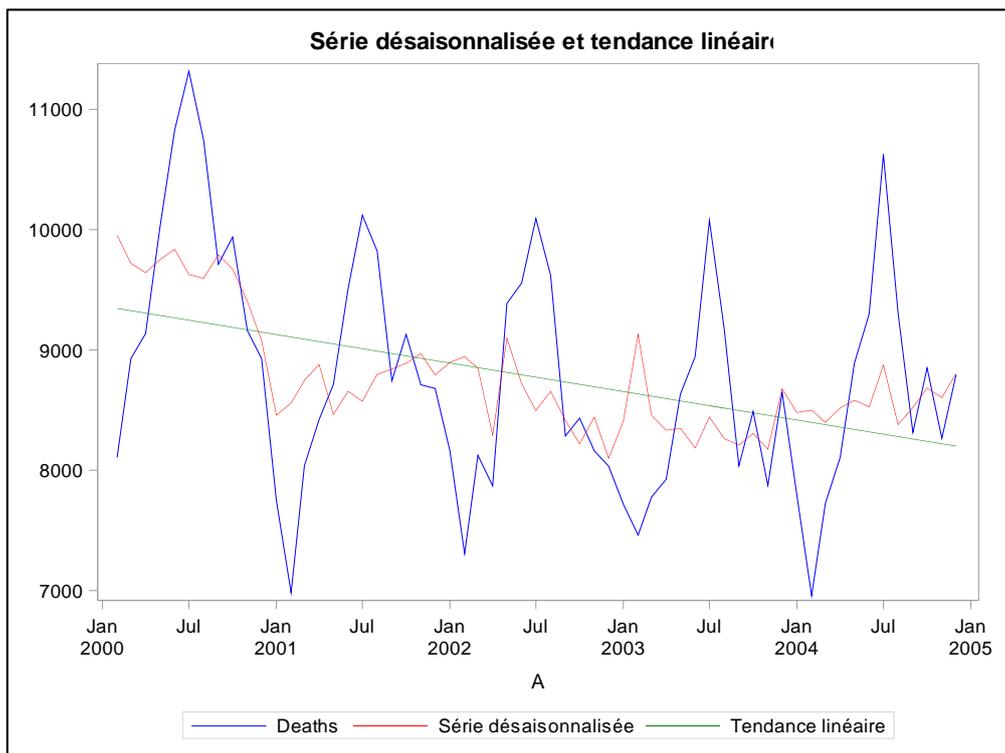
Root MSE	353.57701	R carré	0.4836
Moyenne dépendante	8773.15195	R car. ajust.	0.4745
Coeff Var	4.03022		

Résultats estimés des paramètres						
Variable	Libellé	DDL	Valeur estimée des paramètres	Erreur type	Valeur du test t	Pr >  t
Intercept	Intercept	1	18845	1379.46353	13.66	<.0001
DATE	A	1	-0.64886	0.08882	-7.31	<.0001

- Dans le tableau 'résultats des paramètres on a l'intercept(constante) et le coefficient de la pente qui sont fortement significatifs au seuil de 1%.
- On va utiliser ces paramètres pour estimer de nouveau la tendance linéaire.

**Etapes :**

- MACRO DATA.
- Calcul de la tendance linéaire.
- Représentation sur un même graphique.
- On supprime les options graphiques.
- On a la série des deaths en bleu, on a aussi la série corrigée des variations saisonnière en rouge et la tendance linéaire en vert.
- La procédure timeseries nous offre la possibilité de faire une décomposition multiplicative : produit des 3 composantes non aléatoires (T\*C\*S) ou additive (T+C+S).

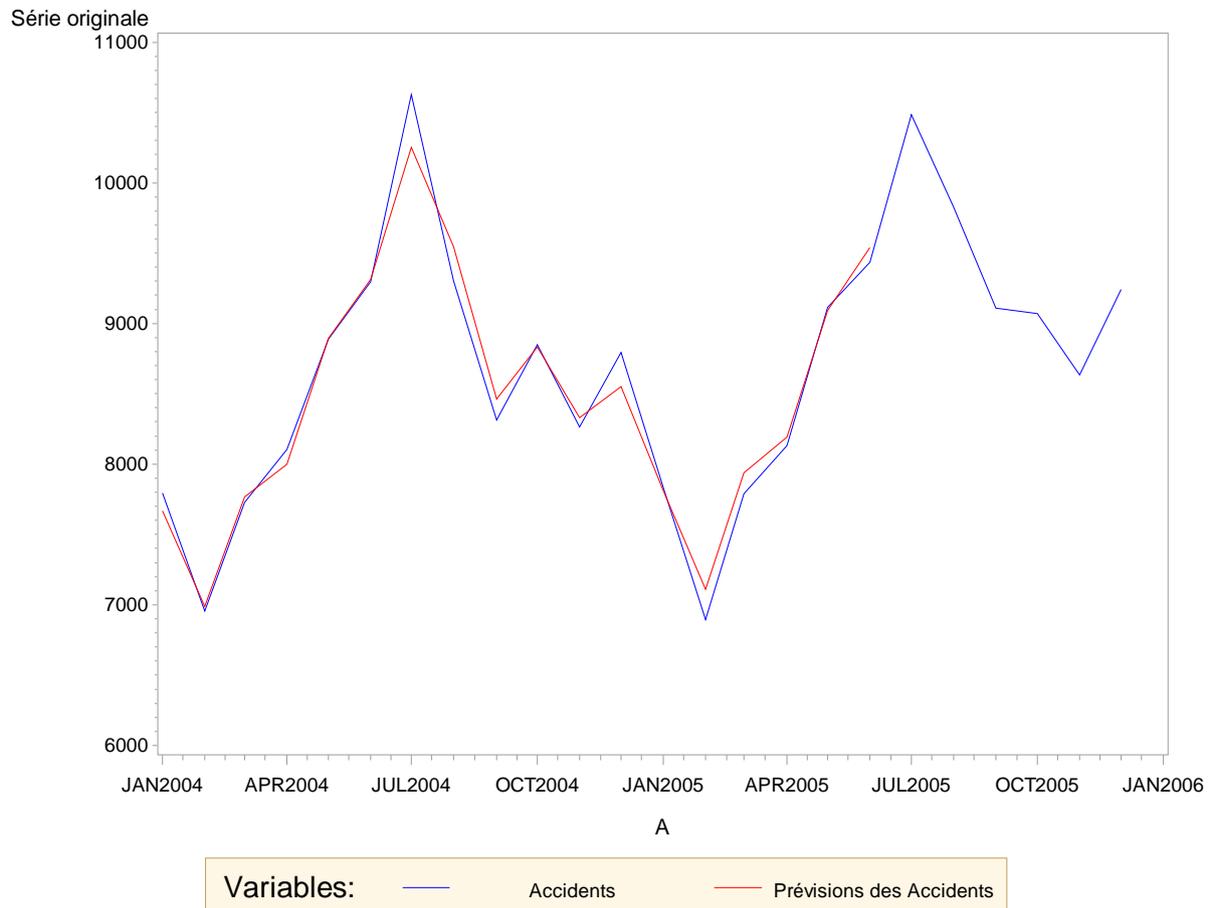


## Méthode de DECOMPOSITION MULTIPLICATIVE

La prédiction sera égale la composante saisonnière : la composante cyclique.

### Série désaisonnalisée et tendance linéaire

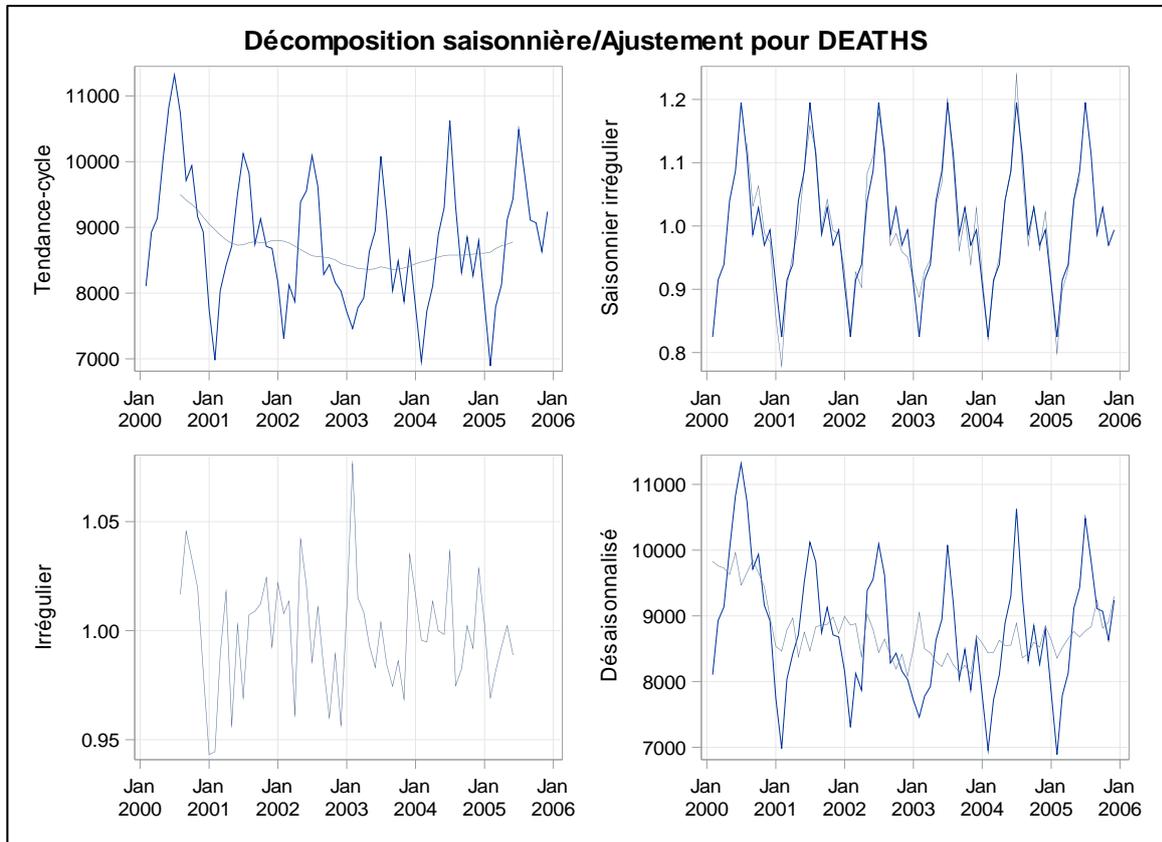
#### **Prévisions obtenues avec la méthode de décomposition (MULTIPLICATIVE)**



On remarque que la prévision est assez bien effectuée dans l'ensemble et que les points correspondent plus ou moins à notre modèle de départ.

### Q3/ Utiliser la méthode de décomposition additive pour générer des prévisions de la variable 'Accidents mortels' pour les années 2004 et 2005.

- La prévision sera calculée comme la somme de la tendance, de la composante saisonnière et de la composante cyclique.
- 'Mode' permet de préciser la méthode de décomposition qu'on souhaite générer.

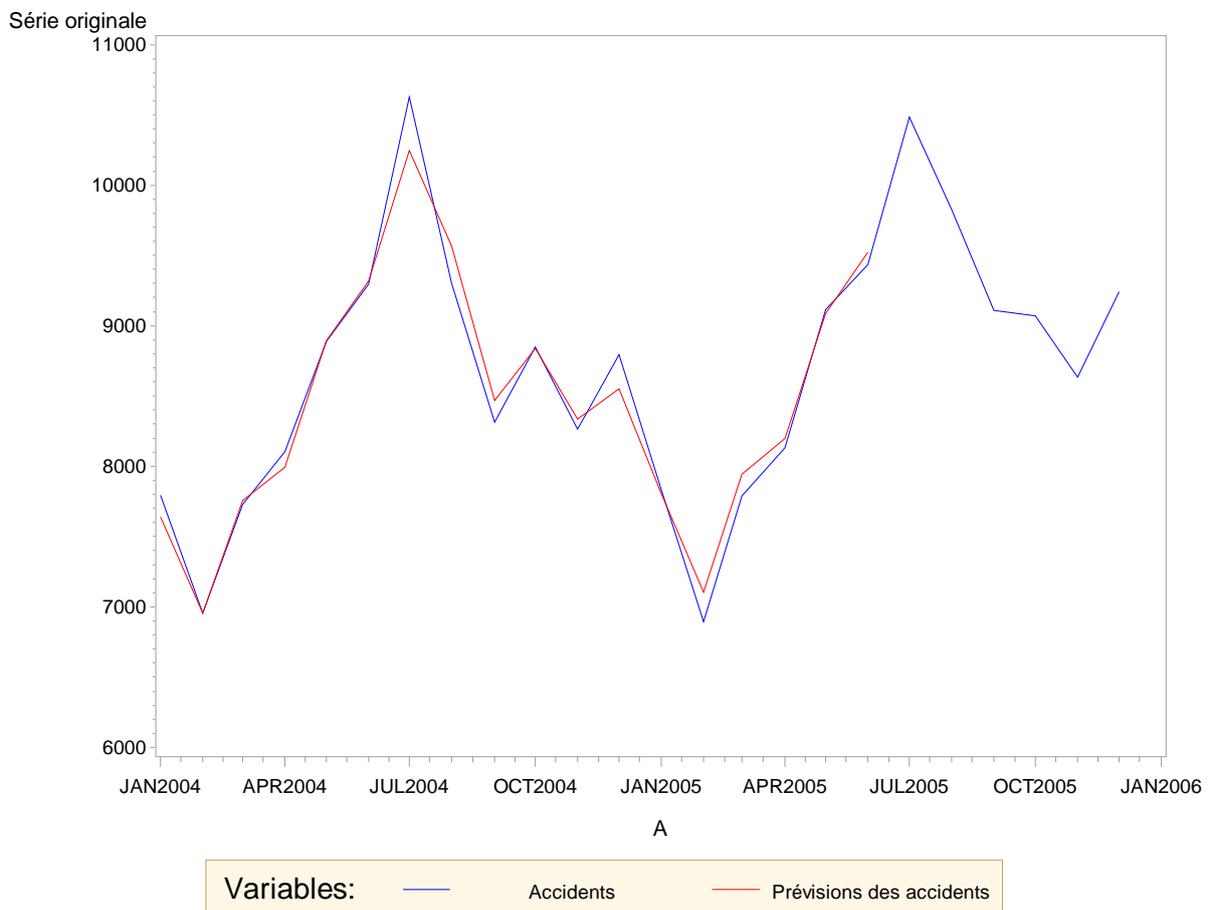


## Méthode de DECOMPOSITION ADDITIVE

- ON GENERE LES PREVISIONS POUR LES ANNEES 2004 & 2005 à l'aide de la proc x11

-Représentation graphique :

### Prévisions obtenues avec la méthode de décomposition (ADDITIVE)



Q4/ Comparer la qualité des prévisions générées par les deux méthodes précédentes en utilisant les critères MSE, RMSE, MAE et MAPE.

Prévisions avec les différentes méthodes de lissage exponentiel

	MSE	RMSE	MAE	MAPE
Décomposition Additive	418895.8	152.5516	111.8941	1.301584
Décomposition Multiplicative	397474.8	148.5999	110.3899	1.288556

Sachant que :

$$397\,474.8 < 418\,895.8$$

$$148.5999 < 152.5516$$

$$110.3899 < 111.8941$$

$$1.288556 < 1.301584$$

Nous savons que les critères RMSE, MAE, MSE et MAPE indiquent le degré d'erreur possible.

En effet, la RMSE représente l'erreur quadratique moyenne, la MSE représente le carré moyen des erreurs, la MAE représente la moyenne arithmétique des valeurs absolues des écarts, et, la MAPE représente l'erreur absolue moyenne en pourcentage.

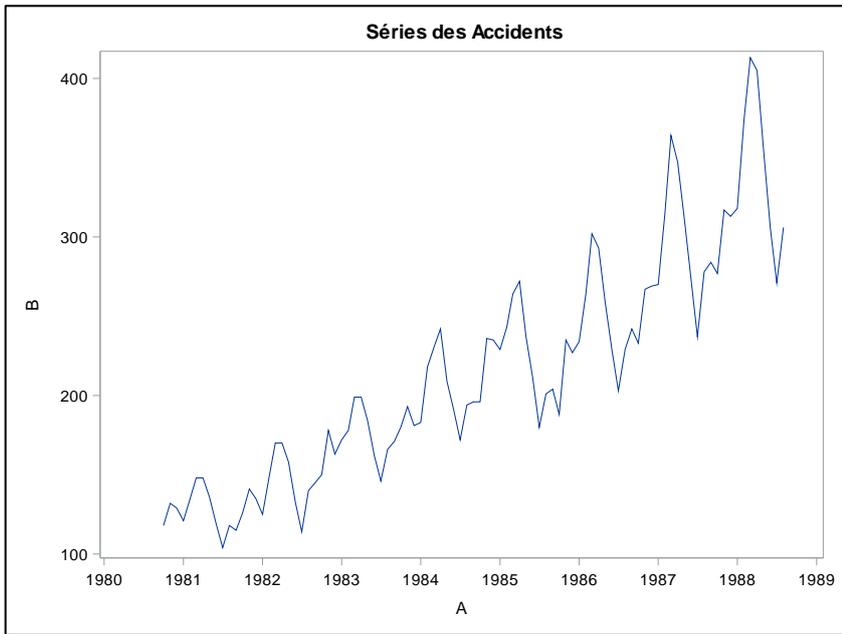
Ainsi, puisque la décomposition multiplicative a des critères RMSE, MAE, MSE, MAPE inférieure à celles de la décomposition additive.

On peut donc conclure que la décomposition multiplicative est plus précise et donc meilleure.

## Méthode de prévision : Document 2

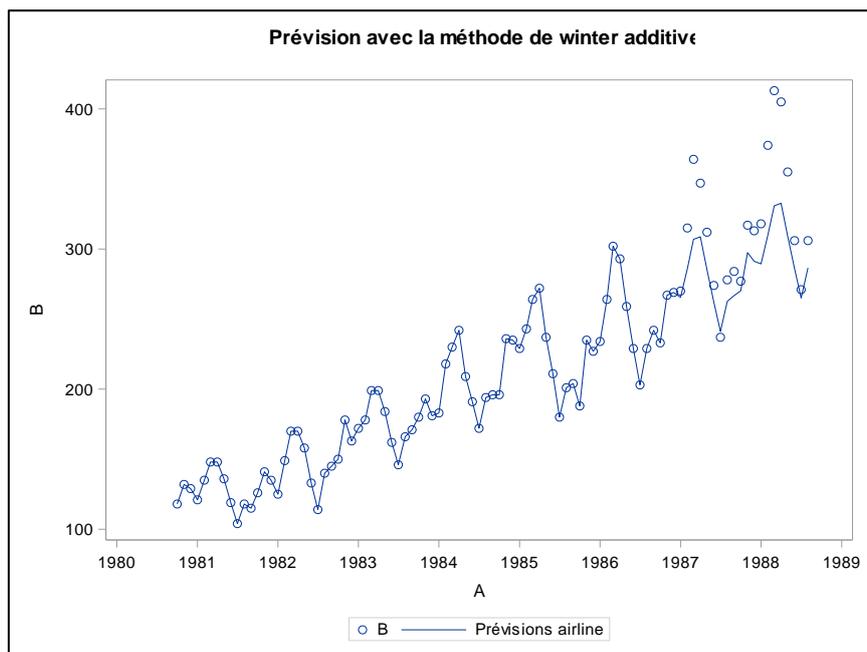
### Q1/ Générer des prévisions pour les années 1987 et 1988 à l'aide de la méthode de Winters additive

On a récupéré des données sur le Trafic aérien entre le 01/09/1980 et 01/08/1988. On souhaite prédire nos données à l'aide de la méthode de lissage de Winter.



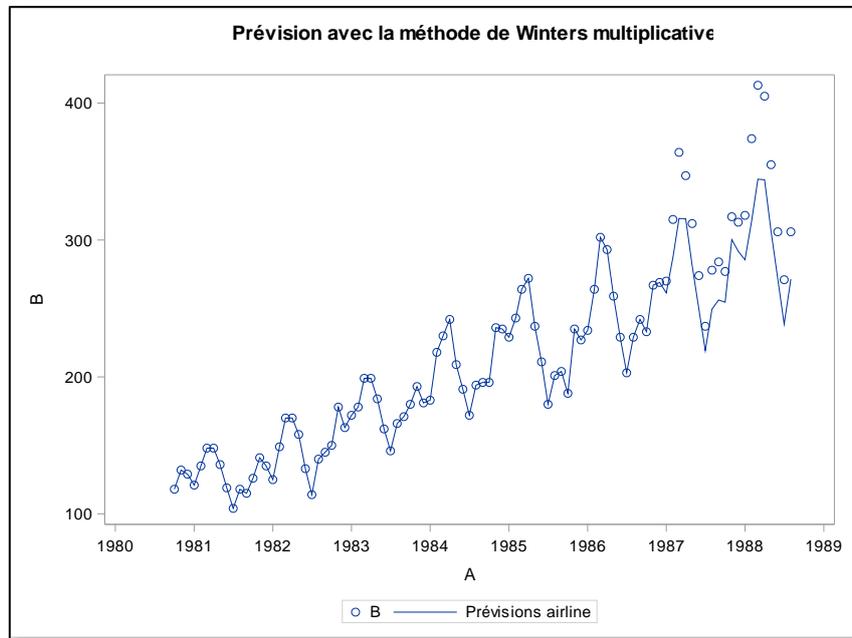
- La série possède une tendance haussière et il y a une saisonnalité dans cette série.
- On a une série chronologique qui s'étale entre le 1er septembre 80 et le 1er août 88 (95 observations)  
Avec une fréquence mensuelle la longueur de la saisonnalité est égale à 12 pour un an.
- Avec la présence de saisonnalité seules les méthodes de Winters sont adéquates.

- On peut effectuer les prévisions à l'aide de la proc forecast;
- Représentation graphique ;



- Graphiquement les prévisions ne sont pas très fiables.

### Q2/ Générer des prévisions pour les années 1987 et 1988 à l'aide de la méthode de Winters multiplicative

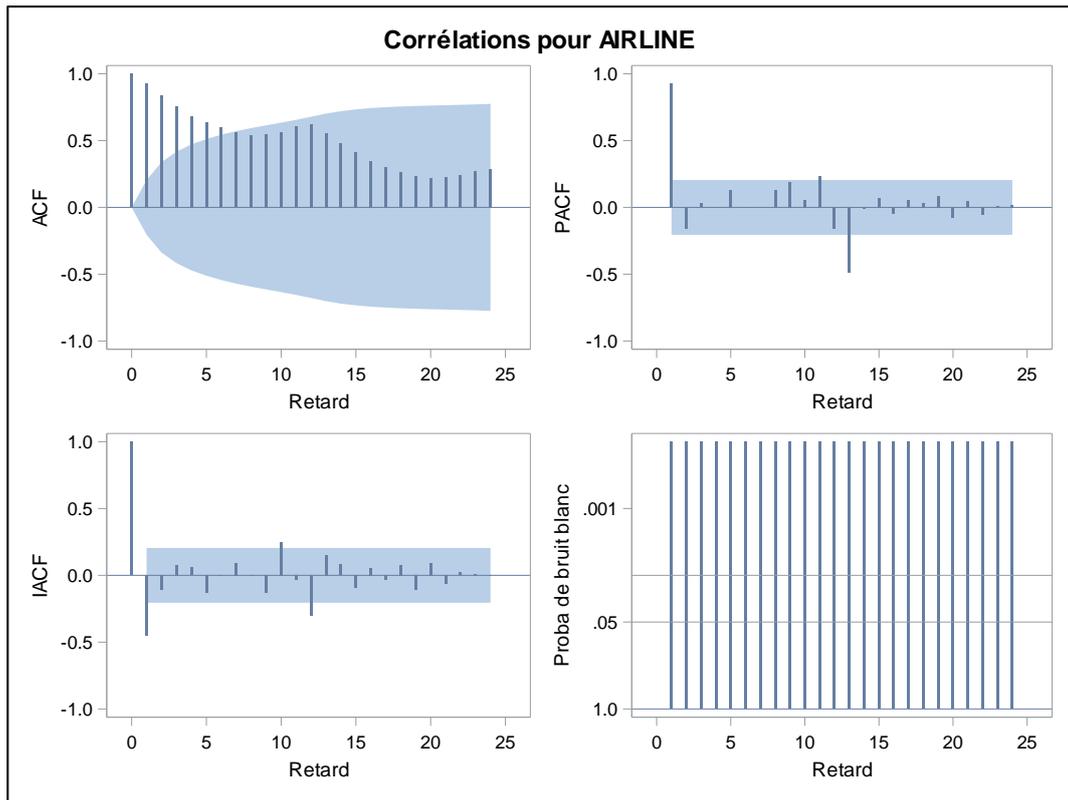


### Q3/ Générer des prévisions pour les années 1987 et 1988 à l'aide de la méthode de décomposition multiplicative

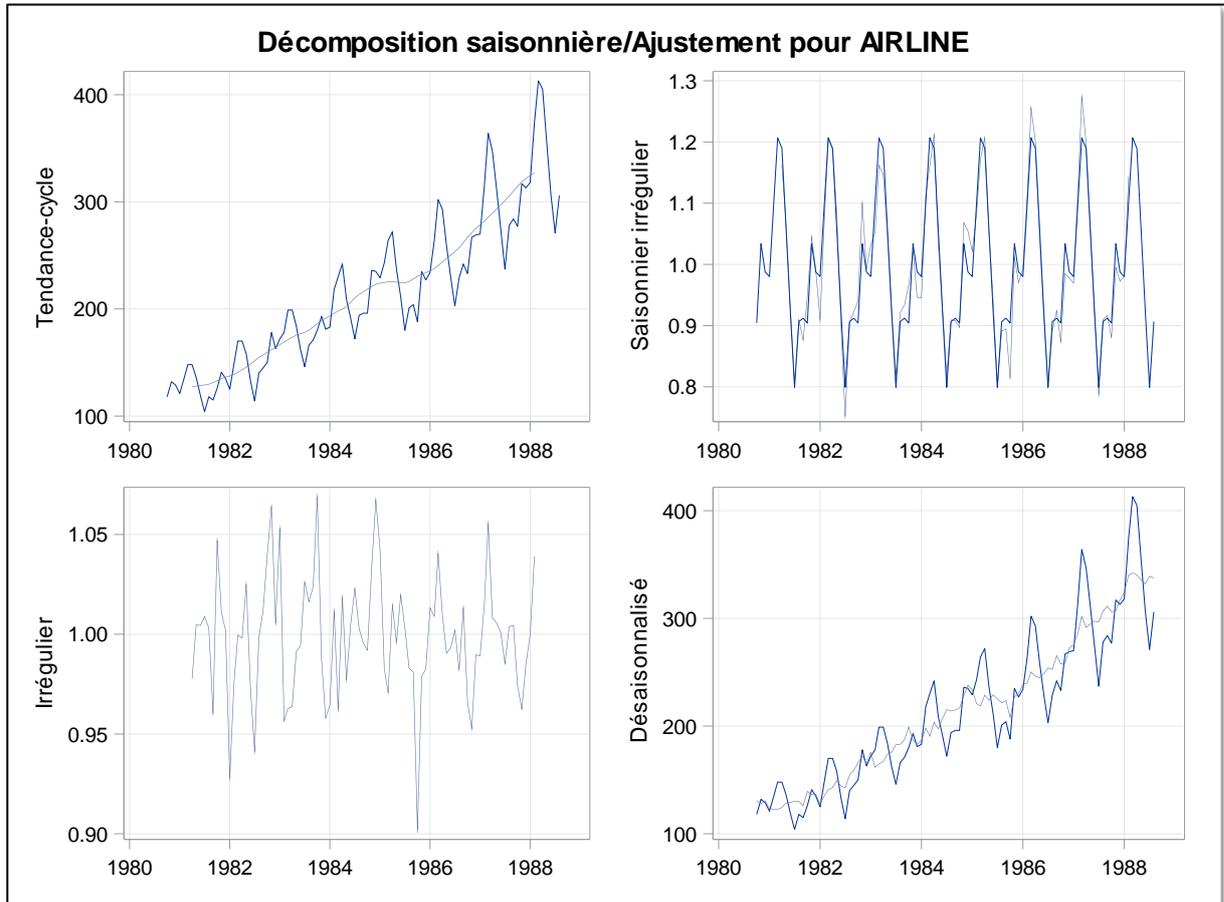
- La prédiction sera égale la composante saisonnière : la composante cyclique.

### Procédure TIMESERIES :

#### Prévisions avec la méthode de Winters multiplicative

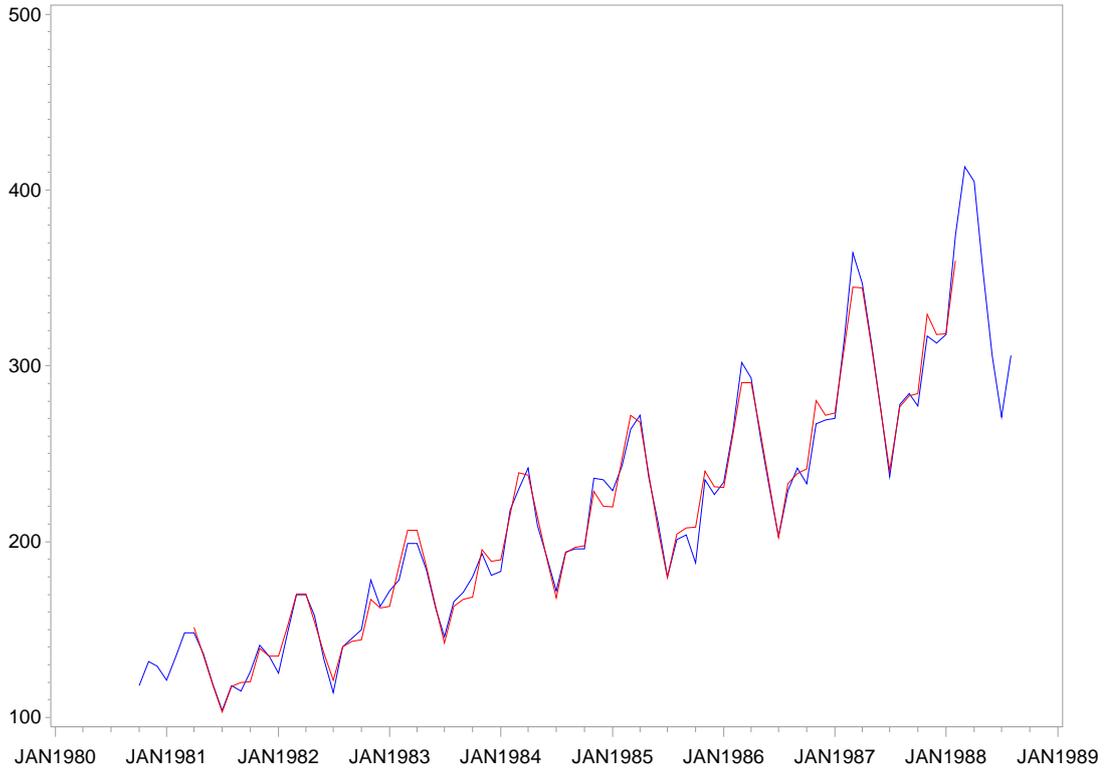


On remarque que nous sommes bien en présence d'une tendance haussière mais qu'il existe aussi une tendance qui se répète tous les 12 mois, c'est-à-dire tous les ans.



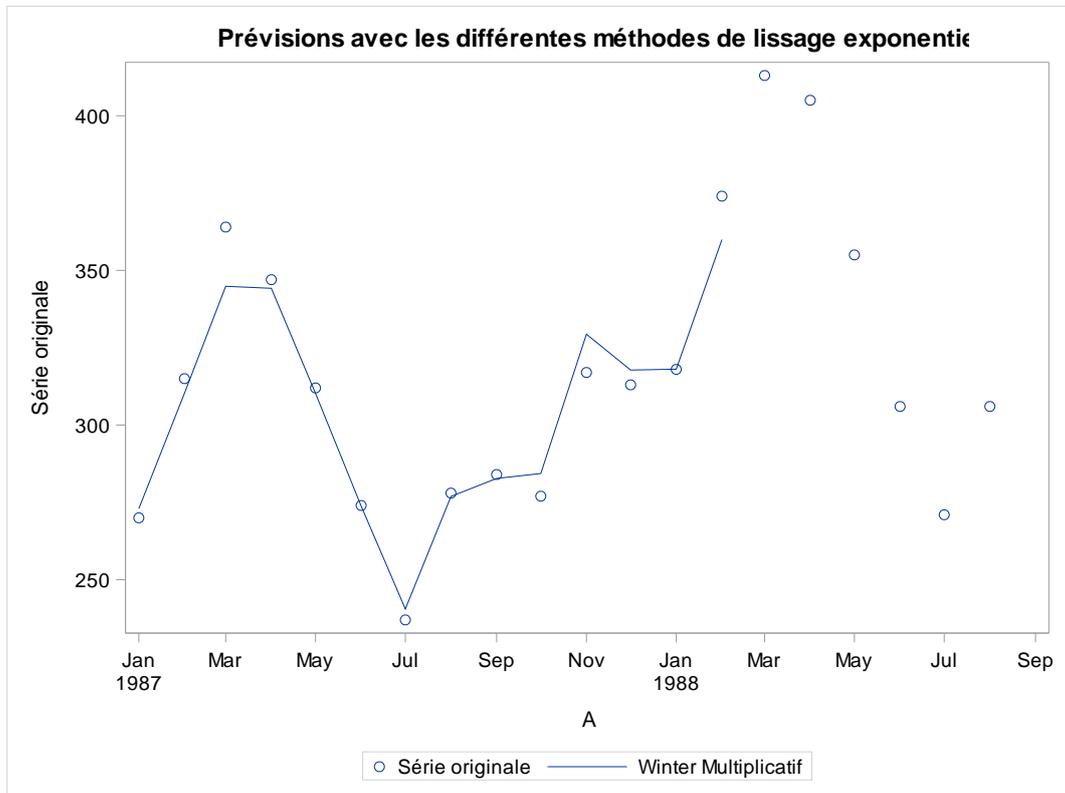
## Prévisions obtenues avec la méthode de décomposition (MULTIPLICATIVE)

Série originale



A

Variables: — Accidents — Prévisions du trafic aérien



A

○ Série originale — Winter Multiplicatif

## Q4/ Générer des prévisions pour les années 1987 et 1988 à l'aide de la méthode de décomposition additive

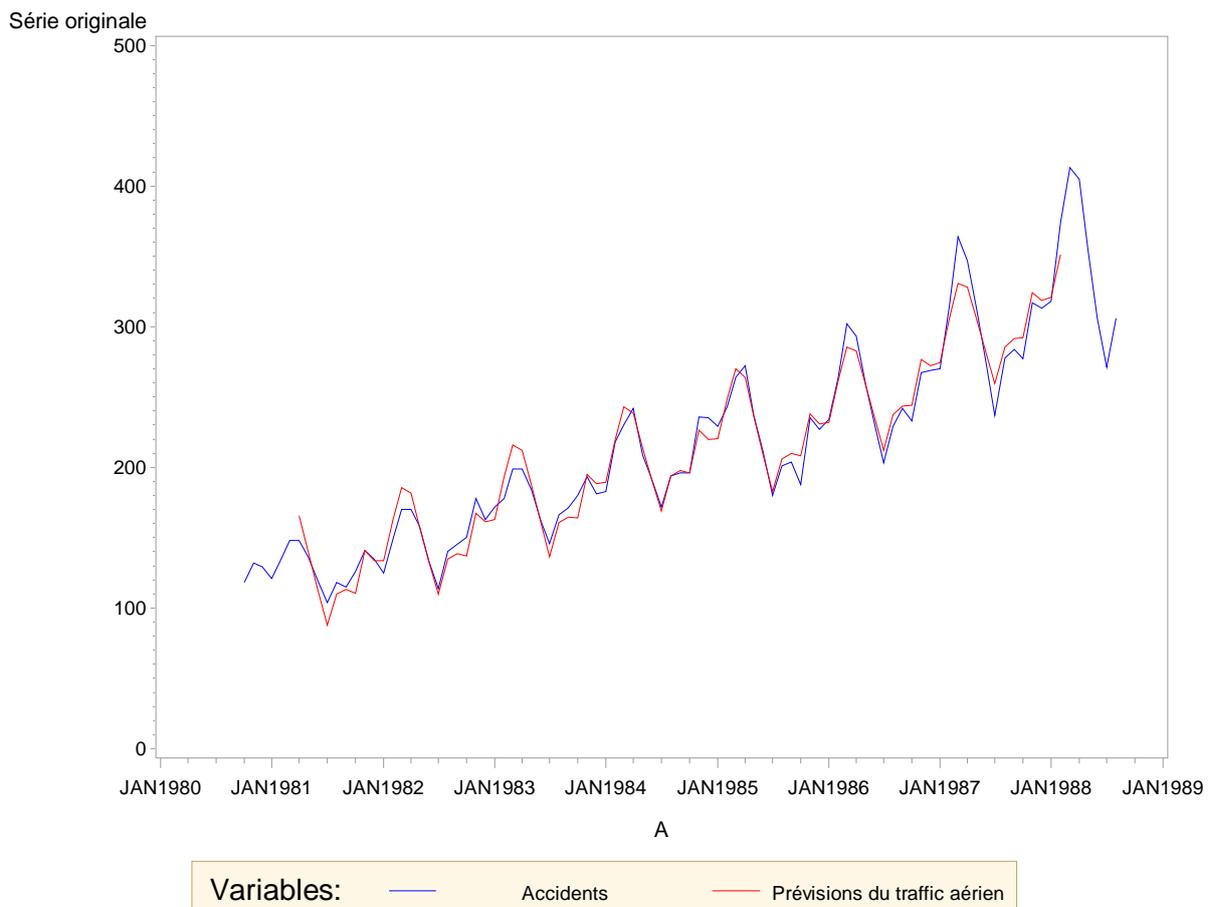
\*La prévision sera calculée comme la somme de la tendance et de la composante saisonnière et la composante cyclique.

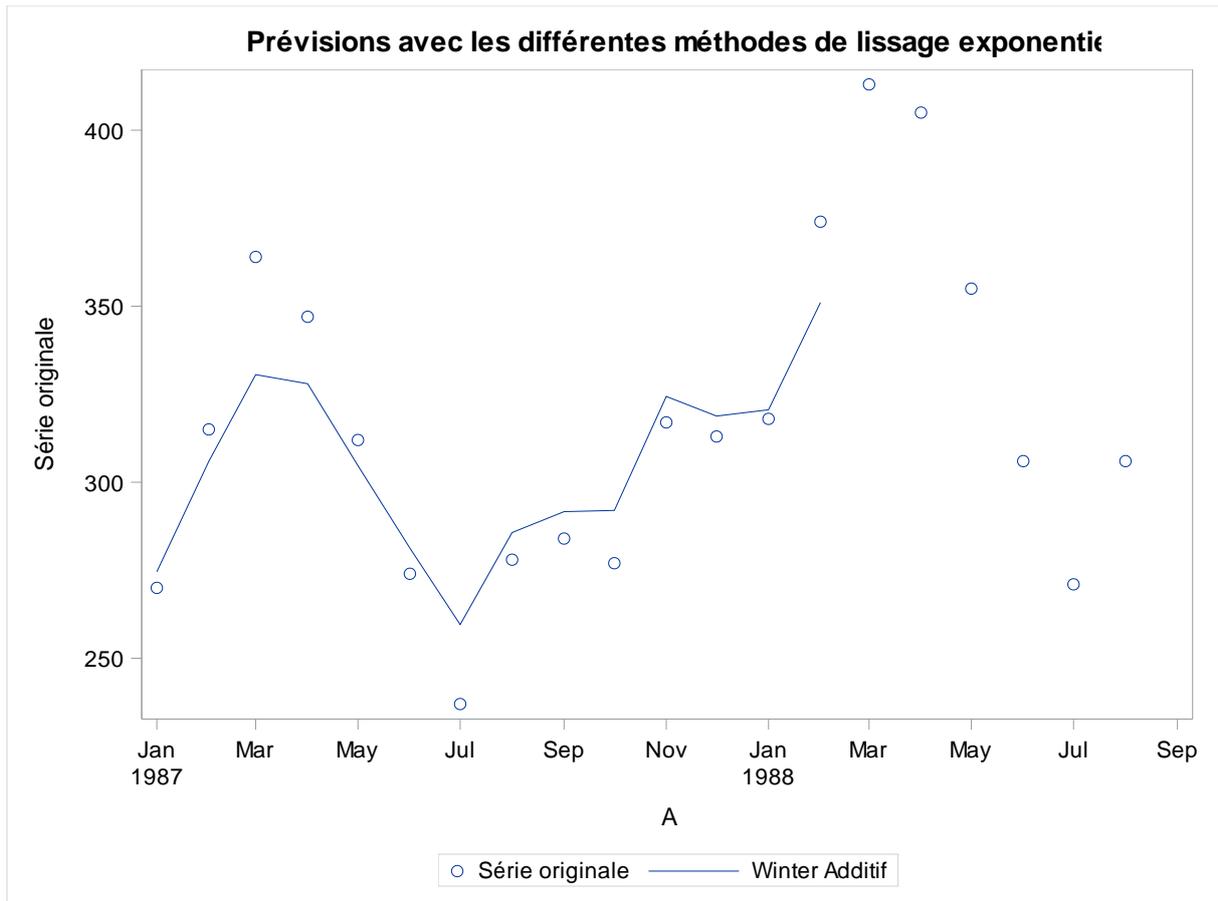
### Prévisions avec la méthode de Winters multiplicative

Table d'entrée	
Nom	WORK.TABLE
Libellé	
Variable ID temps	DATE
Intervalle de temps	MONTH
Longueur du cycle saisonnier	12

Les prévisions sont plus ou moins bonne, mais on peut remarquer que le modèle surestime la série au début pour ensuite la sous-estimer, donc les prévisions vont sans doute sous-estimer la valeur du Trafic aérien futur.

### **Prévisions obtenues avec la méthode de décomposition (ADDITIVE)**





Q5/Sélectionner la méthode qui donne les meilleures prévisions en utilisant les critères RMSE et MAPE.

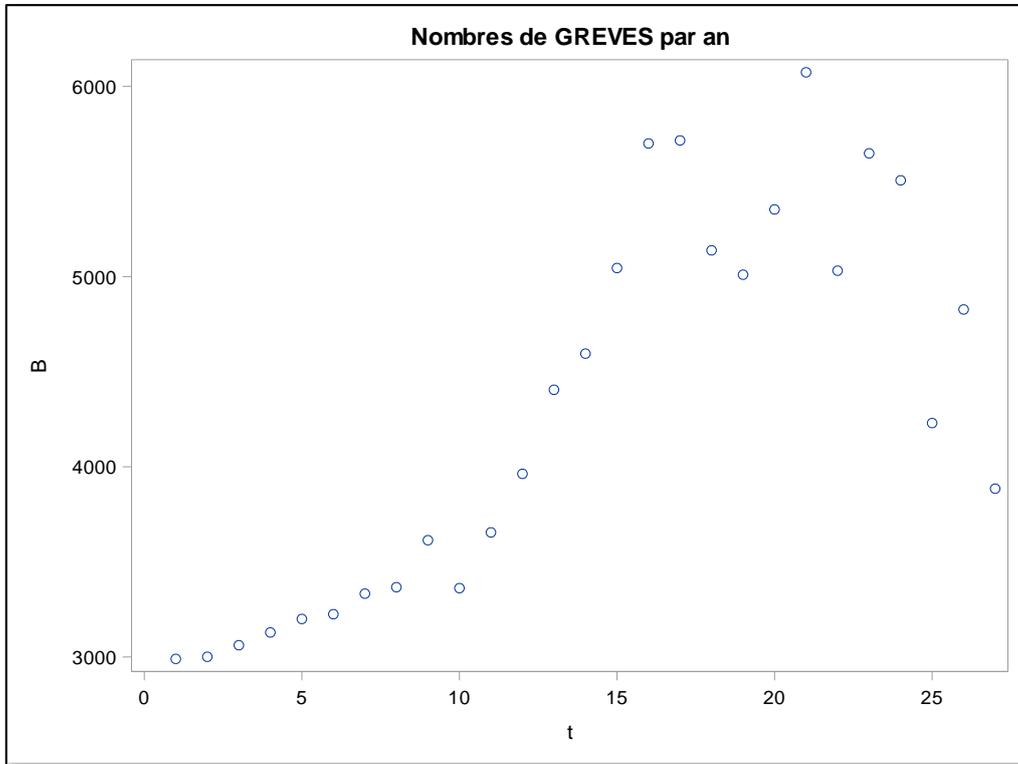
Prévisions avec les différentes méthodes de lissage exponentiel

	RMSE	MAPE
Winters Additive	17.08454	1.803475
Winters Multiplicative	17.06942	2.17577
Décomposition Additive	15.03034	3.983689
Décomposition Multiplicative	7.795538	1.677596

On remarque que la décomposition multiplicative est la plus adaptée car elle fournit la RMSE ET la MAPE la plus faible.

## Méthode de prévision : Document 3

Q1/ Estimer une tendance linéaire sur la série « grève » et l'utiliser pour générer 12 prévisions.



Nous avons récupéré des données concernant des grèves, puis nous les avons présentés sous forme de graphique. On remarque qu'il existe une tendance dans cette série et on souhaite l'estimer de la manière la plus exacte.

1) Prévision à l'aide de l'estimation linéaire du trend à l'aide de la proc REG

### Procédure REG :

#### Nombre de GREVES par an

Nb d'observations lues	14
Nb d'obs. utilisées	14

Analyse de variance					
Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F	Pr > F
Modèle	1	2759464	2759464	57.28	<.0001
Erreur	12	578109	48176		
Total sommes corrigées	13	3337573			

Résultats estimés des paramètres						
Variable	Libellé	DDL	Valeur estimée des paramètres	Erreur type	Valeur du test t	Pr >  t
Intercept	Intercept	1	-18.72177	3.49837	-5.35	0.0002
GREVE	B	1	0.00751	0.00099191	7.57	<.0001

La constante et la variable greve sont significatives à 1%.

## Q2/ Estimer une tendance quadratique sur la série « greve » et l'utiliser pour générer 12 prévisions.

### Procédure REG :

#### Nombre de GREVES par an

Résultats estimés des paramètres						
Variable	Libellé	DDL	Valeur estimée des paramètres	Erreur type	Valeur du test t	Pr >  t
Intercept	Intercept	1	-355.66822	35.01854	-10.16	<.0001
GREVE	B	1	0.12258	0.00993	12.35	<.0001

Les deux coefficients intercept et Grève sont tous les deux significatifs.

## Q3/ Estimer une tendance exponentielle sur la série « greve » et l'utiliser pour générer 12 prévisions.

Résultats estimés des paramètres						
Variable	DDL	Valeur estimée des paramètres	Erreur type	Valeur du test t	Pr >  t	
Intercept	1	-224.40696	26.55767	-8.45	<.0001	
greve_exp	1	28.45666	3.25839	8.73	<.0001	

Les deux coefficients sont significatifs, intercept et greve\_exp.

## Q4/ Quelle est la meilleure méthode pour ces données ?

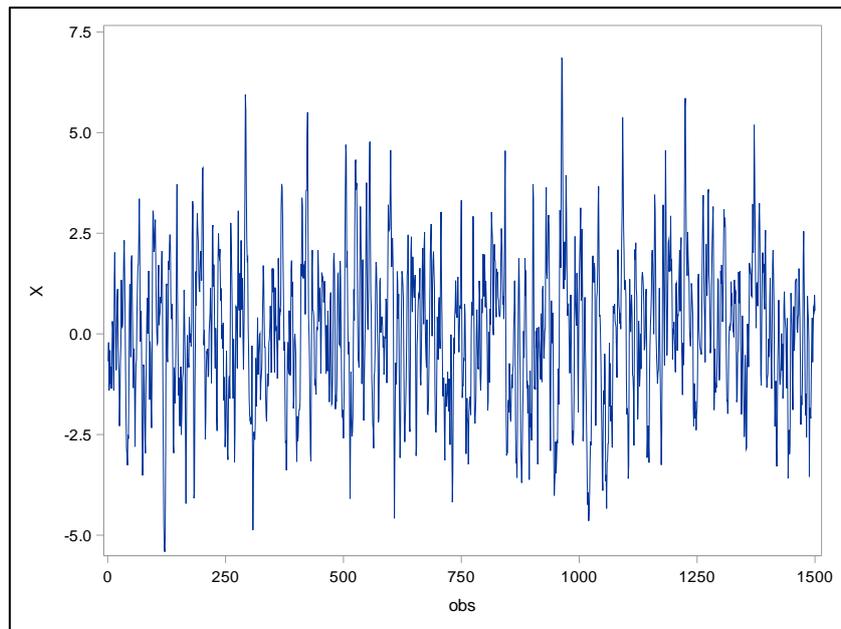
	RMSE	MAPE
Prévision Linéaire	5179.234	99.61691
Prévision Quadratique	4915.839	94.7258
Prévision exponentiel	5180.76	99.64242

En utilisant le critère de RMSE, qui permet de calculer l'erreur quadratique moyenne, on choisit de garder un trend quadratique, ceux qui pouvait être supposé dès le départ à cause de la forme de la série.

## Méthode de prévision : Document 4

### Q1/ Utiliser la méthode d'identification des processus ARIMA pour sélectionner le modèle adéquat pour modéliser la série X.

- On constitue notre échantillon TRAIN et notre échantillon TEST.
- Le graphique nous montre une stationnarité donc pas de trend et pas de saisonnalité à première vue.



### Procédure ARIMA :

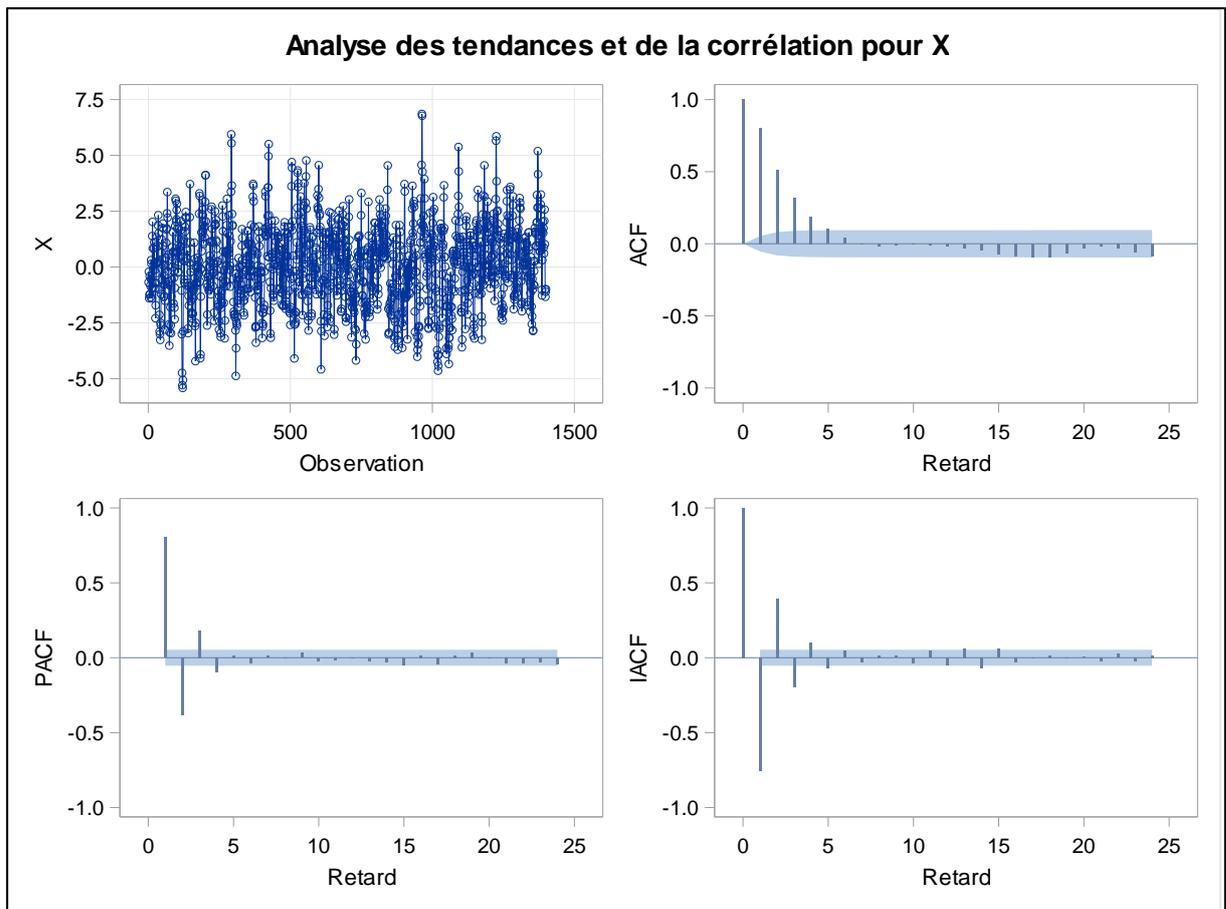
#### Systeme SAS

Nom de la variable = X	
Moyenne des séries de travail	0.063978
Ecart-type	1.825055
Nombre d'observations	1400

Vérification de l'autocorrélation pour le bruit blanc									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
6	1476.75	6	<.0001	0.803	0.510	0.314	0.189	0.103	0.039
12	1478.17	12	<.0001	-0.000	-0.017	-0.013	-0.006	-0.011	-0.019
18	1523.73	18	<.0001	-0.029	-0.045	-0.070	-0.086	-0.093	-0.091
24	1548.66	24	<.0001	-0.065	-0.033	-0.021	-0.034	-0.058	-0.084

Critère d'information minimum						
Lags	MA 0	MA 1	MA 2	MA 3	MA 4	MA 5
AR 0	1.201865	0.66418	0.298464	0.114497	0.032487	-0.00322
AR 1	0.168569	-0.02906	-0.02568	-0.02115	-0.01615	-0.01205
AR 2	0.013199	-0.02601	-0.02161	-0.01734	-0.01254	-0.00829
AR 3	-0.0166	-0.02149	-0.0174	-0.01237	-0.00865	-0.0036
AR 4	-0.02192	-0.01734	-0.0128	-0.00887	-0.00386	0.001051
AR 5	-0.01706	-0.01245	-0.00811	-0.00375	0.001008	0.00618

Modèle de série incorrect : AR(6)  
 Valeur de table minimale : BIC(1,1) = -0.02906



- Fonction d'autocorrélation décroît fortement de manière linéaire donc on peut supposer que pas de tendance dans la série.
- Pour vérifier on regarde dans trend, le Tau qui correspond à la statistique qui permet de savoir si on rejette ou non l'hypothèse d'une racine unitaire.
- $Pr(\tau) < 0.05$  On rejette  $H_0$  donc pas de tendance dans la série.
- D'après le test ADF on n'a pas de racine unitaire donc stationnarité du modèle.
- ADF  $H_0$  : Racine Unitaire,  $H_1$  : Pas de racine unitaire.
- On cherche à estimer plusieurs représentation ARIMA afin de trouver la plus adéquate en fixant un pmax et un qmax.

- On va estimer les modèles ARIMA compris entre ses valeurs et on va regarder les critères d'information qui sont le plus faible.

## Procédure ARIMA :

### Système SAS

Nom de la variable = X	
Moyenne des séries de travail	0.032427
Ecart-type	1.801088
Nombre d'observations	1500

Vérification de l'autocorrélation pour le bruit blanc									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
6	1550.32	6	<.0001	0.800	0.504	0.307	0.180	0.095	0.037
12	1551.21	12	<.0001	0.002	-0.012	-0.006	0.003	-0.006	-0.019
18	1593.19	18	<.0001	-0.031	-0.047	-0.070	-0.080	-0.083	-0.079
24	1616.79	24	<.0001	-0.060	-0.031	-0.020	-0.034	-0.056	-0.079

Critère d'information minimum				
Lags	MA 0	MA 1	MA 2	MA 3
AR 0	1.175554	0.635814	0.276056	0.093387
AR 1	0.157442	-0.03522	-0.03195	-0.02777
AR 2	0.005797	-0.03222	-0.02823	-0.02399
AR 3	-0.02129	-0.02798	-0.02402	-0.01958

**Modèle de série incorrect : AR(6)**  
**Valeur de table minimale : BIC(1,1) = -0.03522**

Estimation des moindres carrés conditionnels					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr >  t	Retard
MU	0.03243	0.04652	0.70	0.4859	0

Estimation constante	0.032427
Estimation variance	3.246081
Erreur type Valeur estimée	1.801688
AIC	6023.988
SBC	6029.301
Nombre de résidus	1500

Vérification de l'autocorrélation des résidus									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
6	1550.32	6	<.0001	0.800	0.504	0.307	0.180	0.095	0.037
12	1551.21	12	<.0001	0.002	-0.012	-0.006	0.003	-0.006	-0.019
18	1593.19	18	<.0001	-0.031	-0.047	-0.070	-0.080	-0.083	-0.079
24	1616.79	24	<.0001	-0.060	-0.031	-0.020	-0.034	-0.056	-0.079
30	1644.79	30	<.0001	-0.094	-0.081	-0.047	-0.014	0.007	0.022
36	1657.54	36	<.0001	0.035	0.046	0.052	0.042	0.022	0.006
42	1665.69	42	<.0001	0.009	0.020	0.024	0.030	0.037	0.045
48	1706.59	48	<.0001	0.053	0.061	0.066	0.071	0.078	0.066

Modèle pour la variable X	
Moyenne estimée	0.032427

- Le meilleur modèle est un ARMA(1,1) car il donne le critères d'information le plus faible.
- On a l'estimation de cet ARMA(1,1), on peut aussi regarder ou s'intéresser à l'autocorrélation des résidus.
- On remarque qu'on rejette  $H_0$  donc les résidus suivent bien un Bb.

## Q2 / Générer des prévisions avec le modèle sélectionné à la question précédente et les modèles ARIMA (1,0,0) et ARIMA (0,0,1)

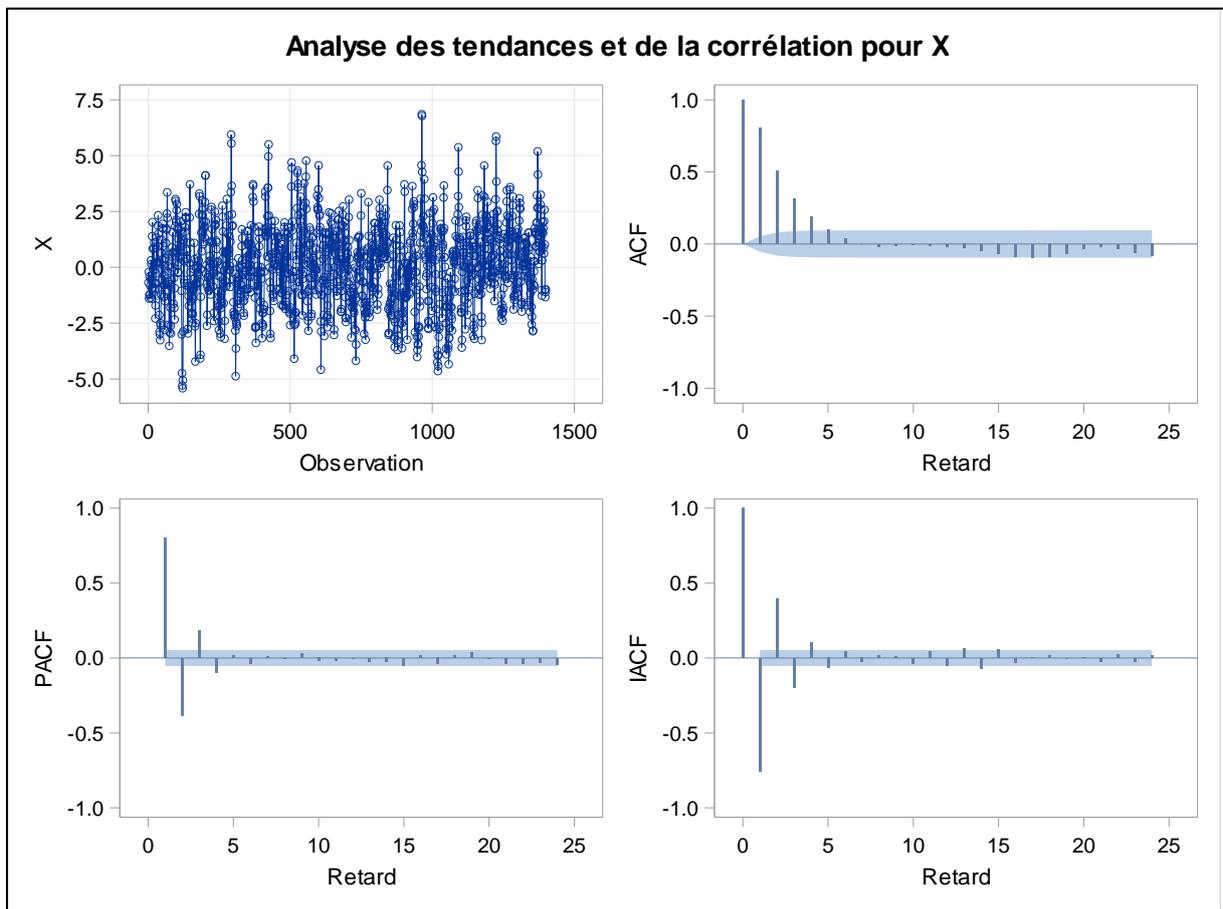
### Modèle de prévision pour ARIMA (1,0,0)

#### Procédure ARIMA :

## Systeme SAS

Nom de la variable = X	
Moyenne des séries de travail	0.063978
Ecart-type	1.825055
Nombre d'observations	1400

Vérification de l'autocorrélation pour le bruit blanc									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
6	1476.75	6	<.0001	0.803	0.510	0.314	0.189	0.103	0.039
12	1478.17	12	<.0001	-0.000	-0.017	-0.013	-0.006	-0.011	-0.019
18	1523.73	18	<.0001	-0.029	-0.045	-0.070	-0.086	-0.093	-0.091
24	1548.66	24	<.0001	-0.065	-0.033	-0.021	-0.034	-0.058	-0.084



Estimation des moindres carrés conditionnels					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr >  t	Retard
MU	0.05000	0.14669	0.34	0.7332	0
AR1,1	0.80370	0.01592	50.48	<.0001	1

Modèle pour la variable X	
Moyenne estimée	0.050004

Facteurs autorégressifs	
Facteur 1:	1 - 0.8037 B**(1)

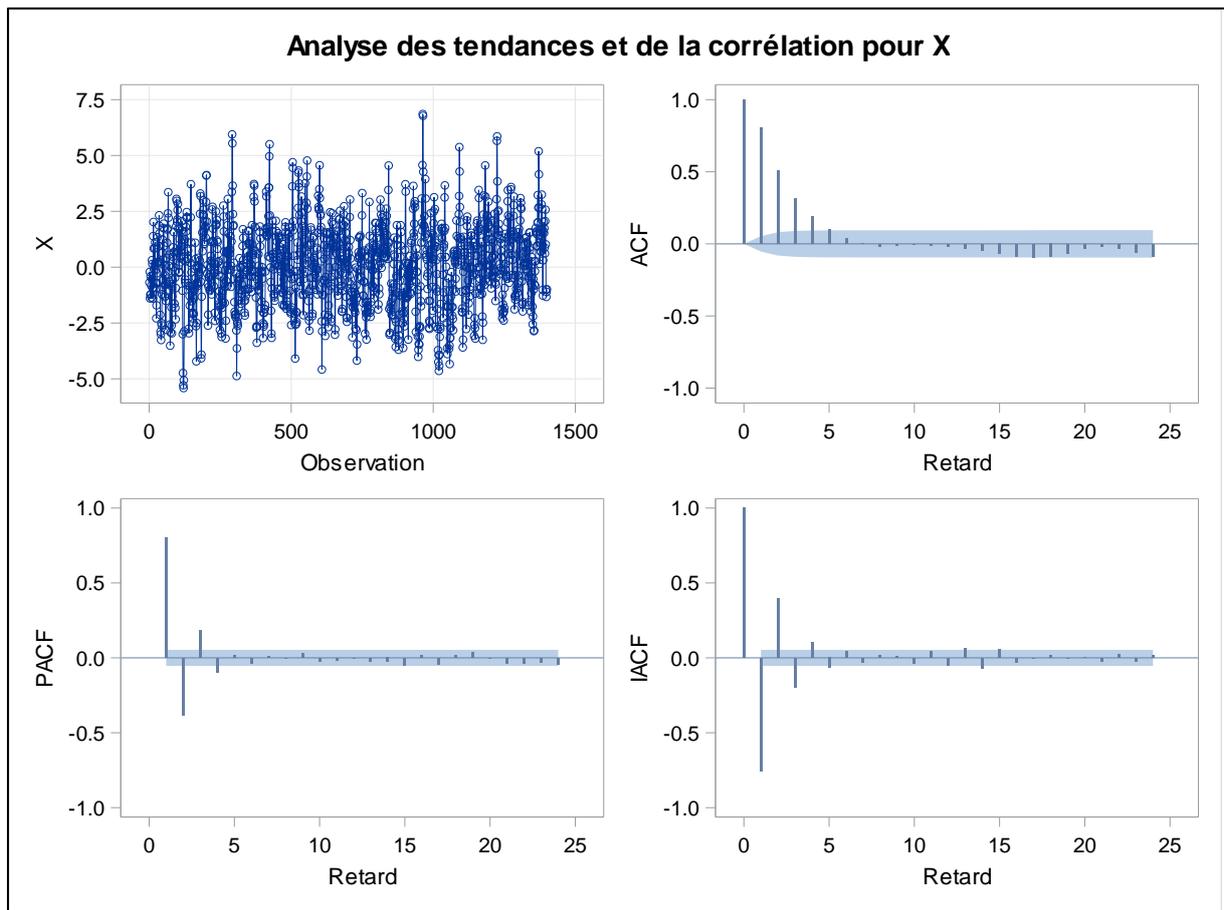
## Modèle de prévision pour ARIMA (0,0,1)

### Procédure ARIMA :

#### Système SAS

Nom de la variable = X	
Moyenne des séries de travail	0.063978
Ecart-type	1.825055
Nombre d'observations	1400

Vérification de l'autocorrélation pour le bruit blanc									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
6	1476.75	6	<.0001	0.803	0.510	0.314	0.189	0.103	0.039
12	1478.17	12	<.0001	-0.000	-0.017	-0.013	-0.006	-0.011	-0.019
18	1523.73	18	<.0001	-0.029	-0.045	-0.070	-0.086	-0.093	-0.091
24	1548.66	24	<.0001	-0.065	-0.033	-0.021	-0.034	-0.058	-0.084



Estimation des moindres carrés conditionnels					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr >  t	Retard
MU	0.06172	0.05734	1.08	0.2819	0
MA1,1	-0.81515	0.01554	-52.47	<.0001	1

Estimation constante	0.061724
Estimation variance	1.400178
Erreur type Valeur estimée	1.183291
AIC	4446.266
SBC	4456.754
Nombre de résidus	1400

AIC et SBC ne contiennent pas de Log déterminant

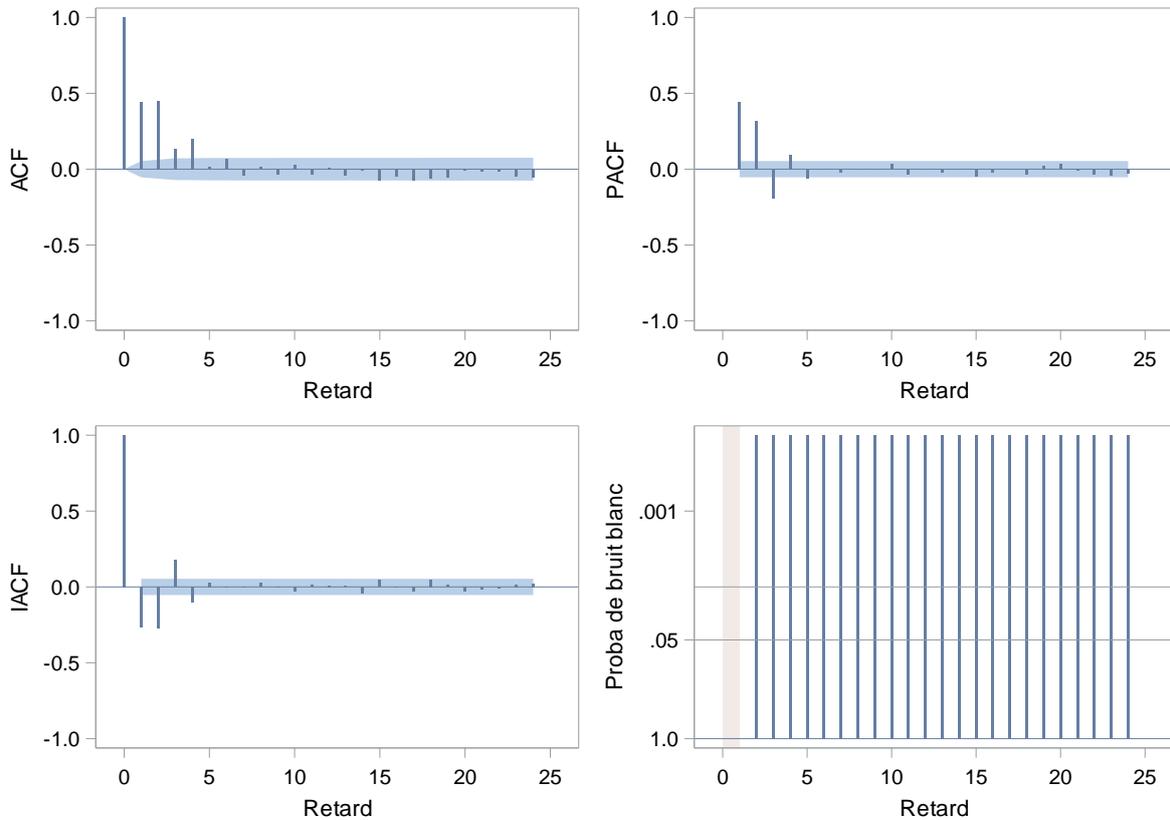
**Corrélations des résultats estimés du paramètre**

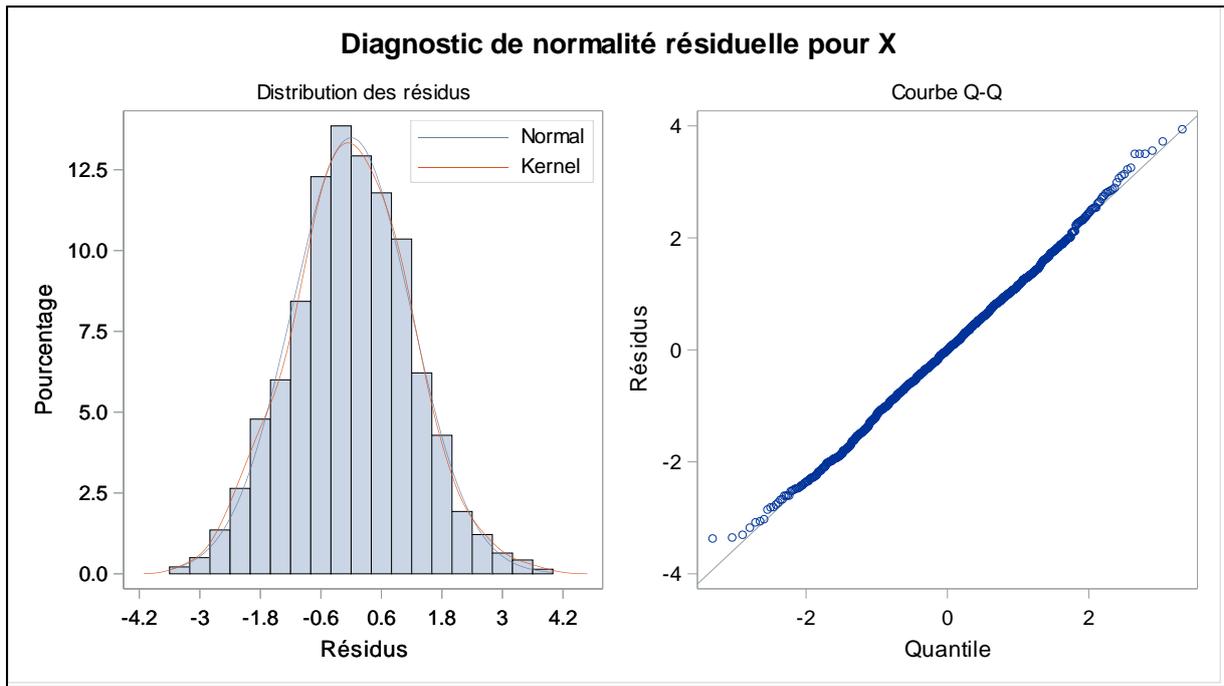
Paramètre	MU	MA1,1
<b>MU</b>	1.000	0.002
<b>MA1,1</b>	0.002	1.000

**Vérification de l'autocorrélation des résidus**

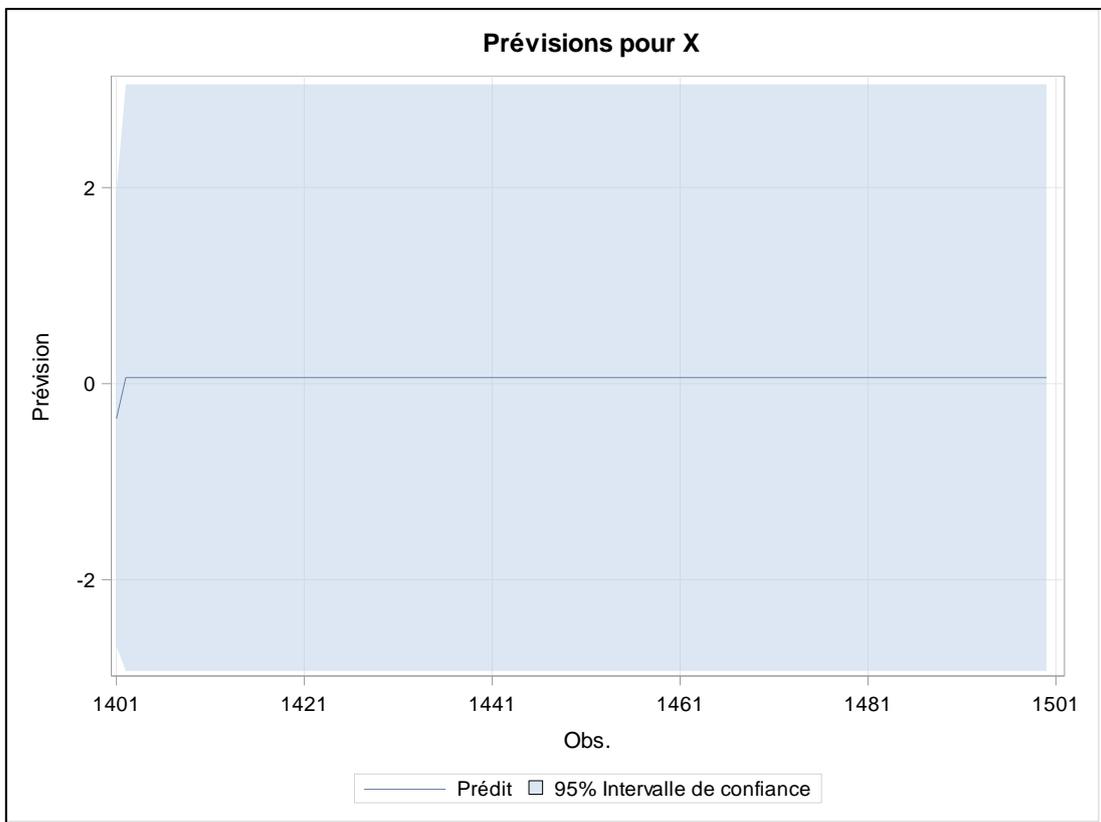
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
<b>6</b>	641.80	5	<.0001	0.440	0.449	0.133	0.198	0.017	0.066
<b>12</b>	649.03	11	<.0001	-0.039	0.013	-0.039	0.026	-0.033	0.010
<b>18</b>	677.24	17	<.0001	-0.043	-0.007	-0.073	-0.049	-0.078	-0.064
<b>24</b>	690.26	23	<.0001	-0.057	-0.012	-0.015	-0.018	-0.049	-0.053
<b>30</b>	711.16	29	<.0001	-0.090	-0.063	-0.047	-0.007	-0.000	0.020
<b>36</b>	720.55	35	<.0001	0.031	0.041	0.044	0.041	0.013	0.009
<b>42</b>	726.11	41	<.0001	-0.001	0.028	0.006	0.028	0.022	0.042
<b>48</b>	753.07	47	<.0001	0.034	0.054	0.046	0.058	0.063	0.071

**Diagnostic de corrélation résiduelle pour X**





<b>Modèle pour la variable X</b>	
<b>Moyenne estimée</b>	0.061724
<b>Facteurs de la moyenne mobile</b>	
<b>Facteur 1:</b>	$1 + 0.81515 B^{**}(1)$



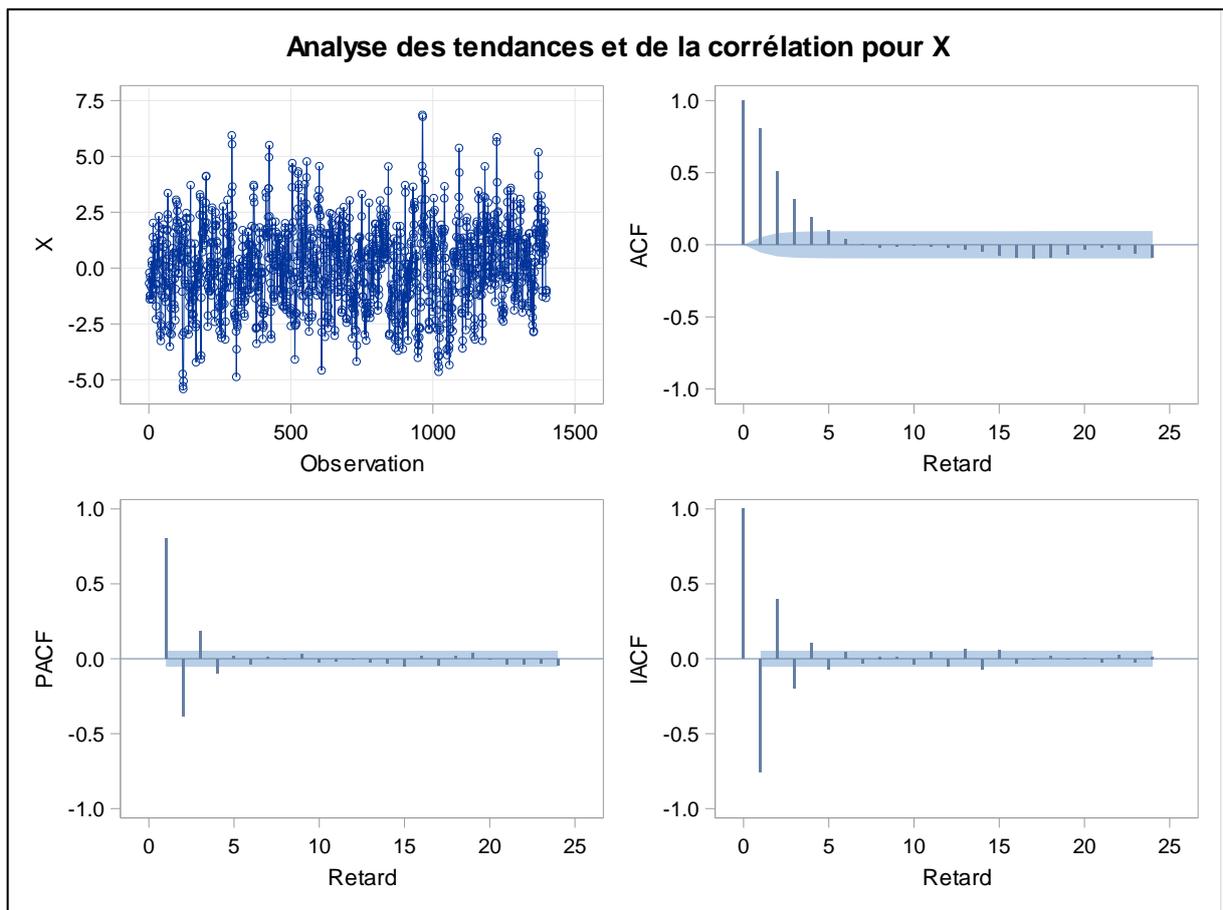
## Modèle de prévision pour ARIMA (1,0,1)

### Procédure ARIMA :

### Système SAS

Nom de la variable = X	
Moyenne des séries de travail	0.063978
Ecart-type	1.825055
Nombre d'observations	1400

Vérification de l'autocorrélation pour le bruit blanc									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
6	1476.75	6	<.0001	0.803	0.510	0.314	0.189	0.103	0.039
12	1478.17	12	<.0001	-0.000	-0.017	-0.013	-0.006	-0.011	-0.019
18	1523.73	18	<.0001	-0.029	-0.045	-0.070	-0.086	-0.093	-0.091
24	1548.66	24	<.0001	-0.065	-0.033	-0.021	-0.034	-0.058	-0.084



Estimation des moindres carrés conditionnels					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr >  t	Retard
<b>MU</b>	0.04603	0.11464	0.40	0.6881	0
<b>MA1,1</b>	-0.54387	0.02549	-21.33	<.0001	1
<b>AR1,1</b>	0.64889	0.02310	28.08	<.0001	1

<b>Estimation constante</b>	0.016162
<b>Estimation variance</b>	0.965769
<b>Erreur type Valeur estimée</b>	0.982735
<b>AIC</b>	3927.262
<b>SBC</b>	3942.995
<b>Nombre de résidus</b>	1400

AIC et SBC ne contiennent pas de Log déterminant.

Corrélations des résultats estimés du paramètre			
Paramètre	MU	MA1,1	AR1,1
<b>MU</b>	1.000	-0.000	-0.009
<b>MA1,1</b>	-0.000	1.000	0.473
<b>AR1,1</b>	-0.009	0.473	1.000

Modèle pour la variable X	
<b>Moyenne estimée</b>	0.046032

Facteurs autorégressifs	
<b>Facteur 1:</b>	1 - 0.64889 B**(1)

Facteurs de la moyenne mobile	
<b>Facteur 1:</b>	1 + 0.54387 B**(1)

## Q4/ Le modèle sélectionné à la première question donne-t-il les meilleures prévisions ?

### Statistique de Diebold et Mariano

#### Procédure ARIMA :

#### Système SAS

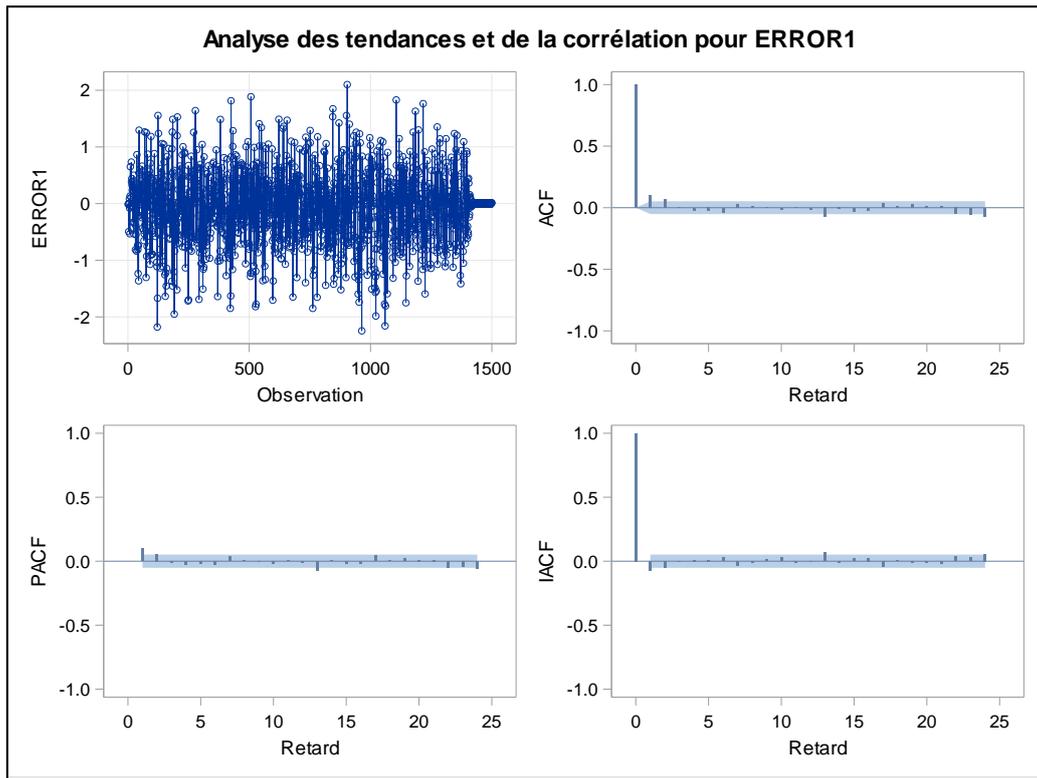
Nom de la variable = ERROR1	
Moyenne des séries de travail	-0.07405
Ecart-type	0.627354
Nombre d'observations	1500

Vérification de l'autocorrélation pour le bruit blanc									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
6	24.36	6	0.0004	0.098	0.065	0.003	-0.021	-0.025	-0.037
12	26.22	12	0.0100	0.028	0.010	0.001	-0.014	0.003	-0.013
18	40.05	18	0.0021	-0.076	-0.010	-0.031	-0.023	0.039	0.013
24	58.42	24	0.0001	0.032	0.013	0.013	-0.051	-0.053	-0.073

Critère d'information minimum				
Lags	MA 0	MA 1	MA 2	MA 3
AR 0	-0.9328	-0.93642	-0.93601	-0.93121
AR 1	-0.93753	-0.9348	-0.93115	-0.92701
AR 2	-0.93583	-0.93098	-0.92665	-0.92214
AR 3	-0.93106	-0.92681	-0.92235	-0.91838

Modèle de série incorrect : AR(3)

Valeur de table minimale : BIC(1,0) = -0.93753



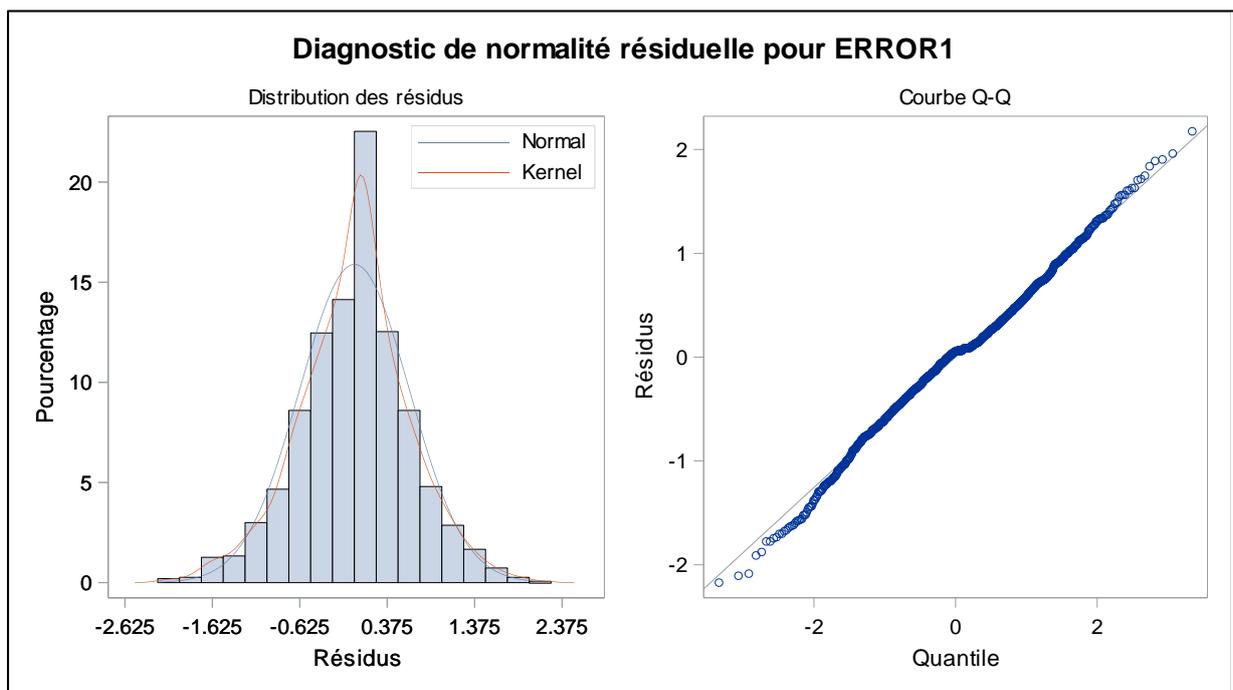
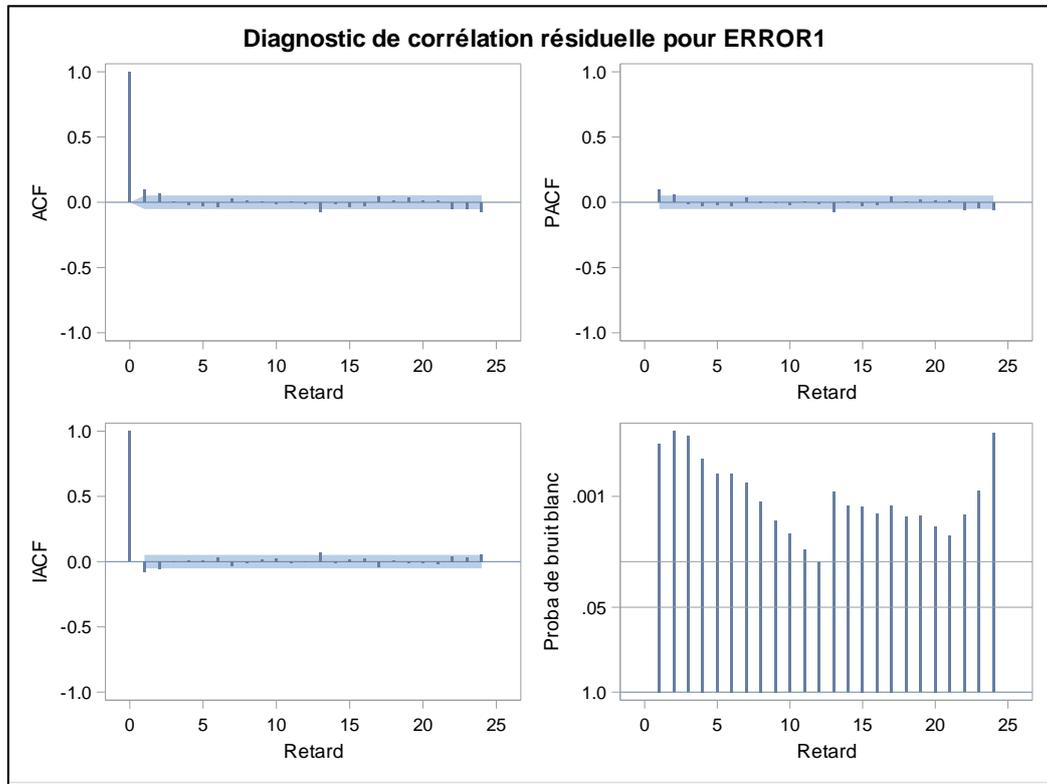
Estimation des moindres carrés conditionnels					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr >  t	Retard
MU	-0.07405	0.01620	-4.57	<.0001	0

Estimation constante	-0.07405
Estimation variance	0.393835
Erreur type Valeur estimée	0.627563
AIC	2860.081
SBC	2865.394
Nombre de résidus	1500

AIC et SBC ne contiennent pas de Log déterminant

Vérification de l'autocorrélation des résidus									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
				6	24.36	6	0.0004	0.098	0.065
12	26.22	12	0.0100	0.028	0.010	0.001	-0.014	0.003	-0.013
18	40.05	18	0.0021	-0.076	-0.010	-0.031	-0.023	0.039	0.013
24	58.42	24	0.0001	0.032	0.013	0.013	-0.051	-0.053	-0.073
30	66.26	30	0.0002	-0.029	-0.053	-0.036	-0.008	-0.007	-0.009
36	78.75	36	<.0001	-0.028	0.002	-0.018	0.030	0.059	0.051

Vérification de l'autocorrélation des résidus									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
42	82.83	42	0.0002	0.038	-0.001	0.026	-0.006	-0.002	0.022
48	92.49	48	0.0001	-0.026	-0.001	-0.002	-0.042	-0.036	-0.050



Modèle pour la variable ERROR1	
Moyenne estimée	-0.07405

- Moyenne estimée égale à -0.07405.
- Test de Diebold-Mariano pour une différence significative dans les MAEs.AR(3) est le meilleur modèle pour la variable error1.
- ➔ Puisque la moyenne du modèle AR(3) n'est pas significativement inférieure à zéro, nous concluons que la MAE de population de la première méthode de prévision est supérieure à la MAE de la population de la seconde méthode de prévision.

## Procédure ARIMA :

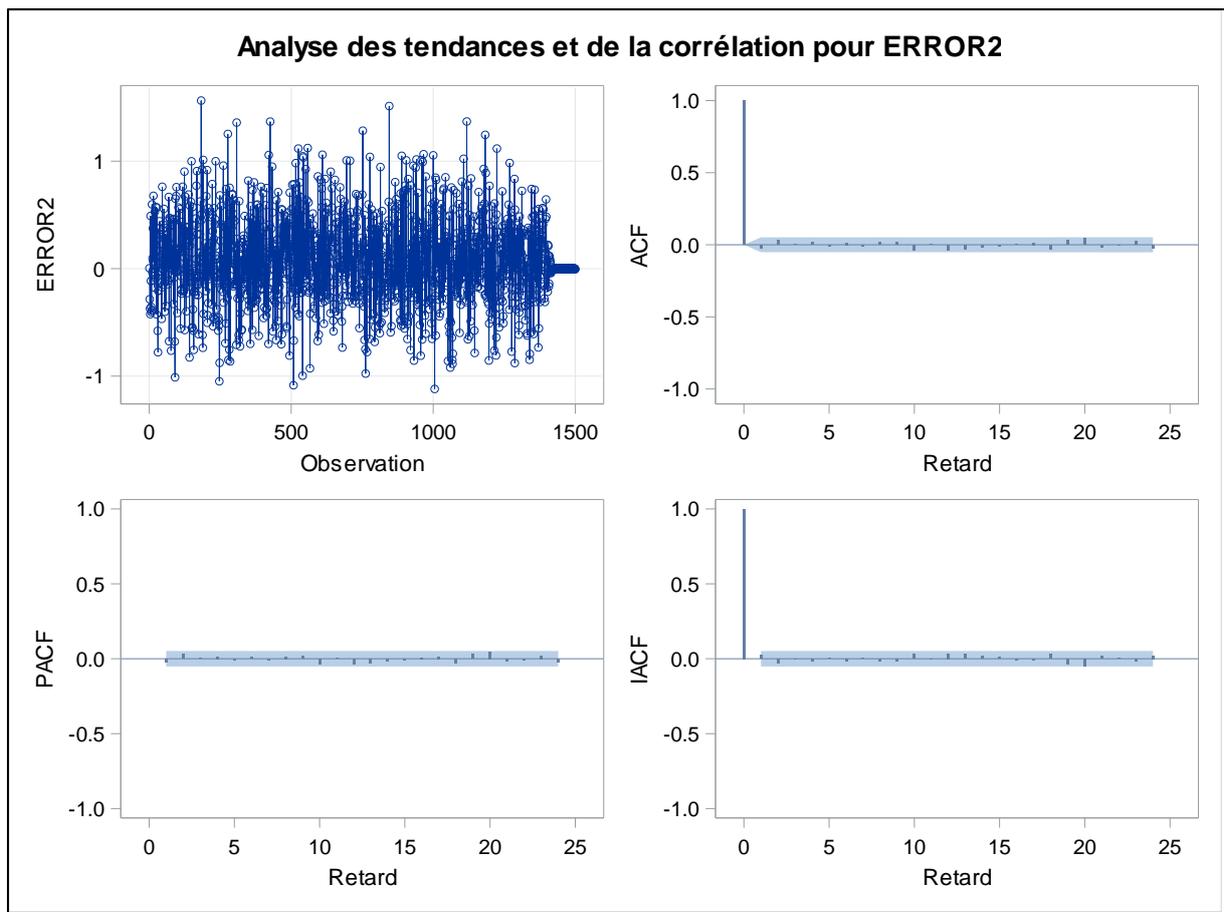
### Systeme SAS

Vérification de l'autocorrélation pour le bruit blanc									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
6	3.23	6	0.7792	-0.025	0.031	0.003	0.016	-0.010	0.014
12	8.19	12	0.7700	-0.011	0.015	0.016	-0.038	0.005	-0.035
18	12.47	18	0.8223	-0.032	-0.017	-0.011	0.006	0.014	-0.034
24	19.23	24	0.7398	0.033	0.045	-0.018	-0.005	0.023	-0.022

Nom de la variable = ERROR2	
Moyenne des séries de travail	0.072906
Ecart-type	0.394912
Nombre d'observations	1500

Critère d'information minimum				
Lags	MA 0	MA 1	MA 2	MA 3
AR 0	-1.8596	-1.85646	-1.85323	-1.84848
AR 1	-1.85649	-1.85181	-1.84838	-1.84376
AR 2	-1.85324	-1.84841	-1.84371	-1.839
AR 3	-1.8485	-1.84385	-1.83906	-1.83435

Modèle de série incorrect : AR(3)  
Valeur de table minimale : BIC(0,0) = -1.8596



Estimation des moindres carrés conditionnels					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr >  t	Retard
MU	0.07291	0.01020	7.15	<.0001	0

Estimation constante	0.072906
Estimation variance	0.156059
Erreur type Valeur estimée	0.395044
AIC	1471.537
SBC	1476.851
Nombre de résidus	1500

AIC et SBC ne contiennent pas de Log déterminant

Modèle pour la variable ERROR2	
Moyenne estimée	0.072906

- Moyenne estimée égale à 0.072906.

- Test de Diebold-Mariano pour une différence significative dans les MAEs.AR(3) est le meilleur modèle pour la variable error2.
- Puisque la moyenne du modèle AR(3) n'est pas significativement inférieure à zéro, nous concluons que la MAE de population de la première méthode de prévision est supérieure à la MAE de population de la seconde méthode de prévision(ARIMA101).

## Procédure ARIMA :

### Système SAS

Estimation des moindres carrés conditionnels					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr >  t	Retard
<b>MU</b>	-0.14695	0.01392	-10.56	<.0001	0

<b>Estimation constante</b>	-0.14695
<b>Estimation variance</b>	0.290473
<b>Erreur type Valeur estimée</b>	0.538955
<b>AIC</b>	2403.447
<b>SBC</b>	2408.76
<b>Nombre de résidus</b>	1500

AIC et SBC ne contiennent pas de Log déterminant

Modèle pour la variable ERROR3	
<b>Moyenne estimée</b>	-0.14695

- Moyenne estimé égale à -0.14695.
- Test de Diebold-Mariano pour une différence significative dans les MAEs.AR(3) est le meilleur modèle pour la variable error3.
- Puisque la moyenne du modèle AR(3) est significativement inférieure à zéro, nous concluons que la MAE de population de la première méthode de prévision ARIMA(101) est inférieure à la MAE de population de la seconde méthode de prévision.

## Statistique de Clark et McCracken

- Clark & McCracken : Une autre statistique t pour l'englobement des prévisions, appelée ENC-REG par Clark et McCracken (2001).
- Elle est calculée comme la statistique t associée à l'englobement des prévisions.
- L'hypothèse nulle est qu'il n'y a pas de différence entre la capacité de prévision du modèle restreint et du modèle non restreint.
- Si l'hypothèse nulle est rejetée, le modèle non restreint donne de meilleures prévisions.

## Procédure ARIMA :

### Système SAS

Nom de la variable = ERROR1	
Moyenne des séries de travail	-0.07405
Ecart-type	0.627354
Nombre d'observations	1500

Vérification de l'autocorrélation pour le bruit blanc									
Jusqu'au retard	Khi-2	DDL	Pr > khi-2	Autocorrélations					
6	24.36	6	0.0004	0.098	0.065	0.003	-0.021	-0.025	-0.037
12	26.22	12	0.0100	0.028	0.010	0.001	-0.014	0.003	-0.013
18	40.05	18	0.0021	-0.076	-0.010	-0.031	-0.023	0.039	0.013
24	58.42	24	0.0001	0.032	0.013	0.013	-0.051	-0.053	-0.073

Critère d'information minimum				
Lags	MA 0	MA 1	MA 2	MA 3
AR 0	-0.9328	-0.93642	-0.93601	-0.93121
AR 1	-0.93753	-0.9348	-0.93115	-0.92701
AR 2	-0.93583	-0.93098	-0.92665	-0.92214
AR 3	-0.93106	-0.92681	-0.92235	-0.91838

**Modèle de série incorrect : AR(3)**  
**Valeur de table minimale : BIC(1,0) = -0.93753**

Estimation des moindres carrés conditionnels					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr >  t	Retard
MU	-0.07405	0.01620	-4.57	<.0001	0

Modèle pour la variable ERROR1	
Moyenne estimée	-0.07405

→ Rejet Ho, le modèle non restreint donne de meilleures prévisions.

### Procédure ARIMA :

#### Systeme SAS

Estimation des moindres carrés conditionnels					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr >  t	Retard
MU	0.07291	0.01020	7.15	<.0001	0

Modèle pour la variable ERROR2	
Moyenne estimée	0.072906

→ Rejet Ho, le modèle non restreint donne de meilleures prévisions

### Procédure ARIMA :

#### Systeme SAS

Critère d'information minimum				
Lags	MA 0	MA 1	MA 2	MA 3
AR 0	-1.23737	-1.24814	-1.25409	-1.25074
AR 1	-1.2518	-1.2528	-1.25024	-1.24588
AR 2	-1.25487	-1.25005	-1.24567	-1.2412
AR 3	-1.25037	-1.24604	-1.24131	-1.23684

Modèle de série incorrect : AR(3)

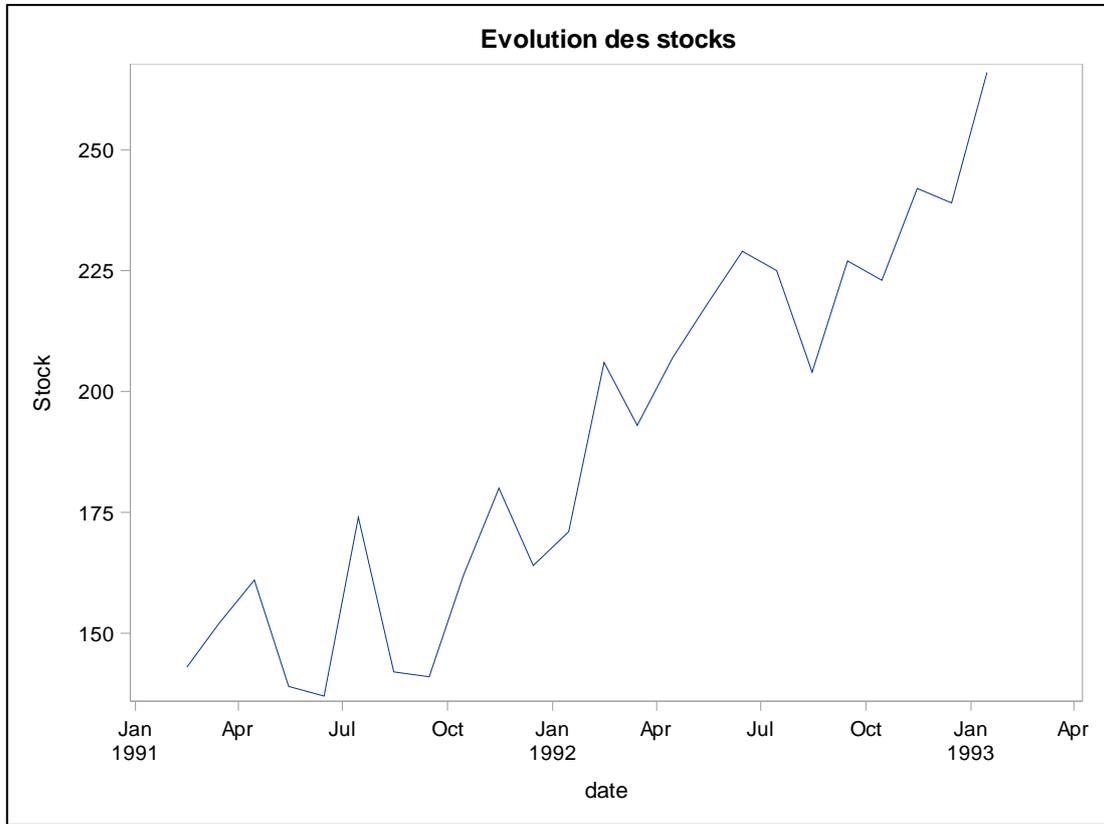
Valeur de table minimale : BIC(2,0) = -1.25487

Modèle pour la variable ERROR3	
Moyenne estimée	-0.14695

→ NON rejet Ho, le modèle restreint donne de meilleures prévisions.

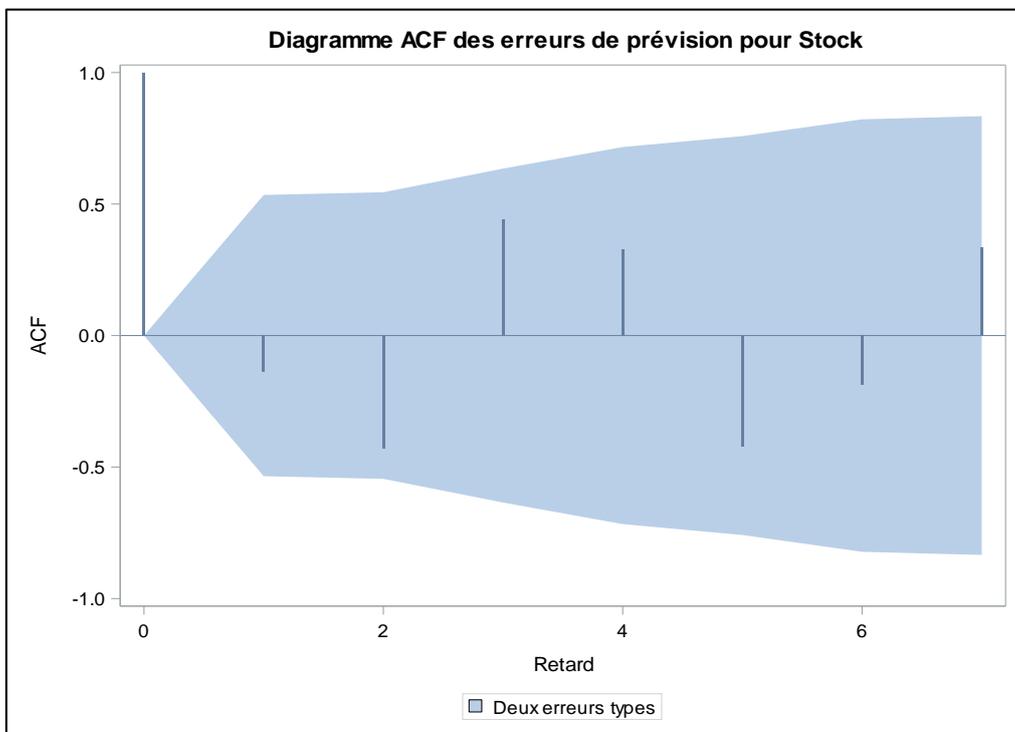
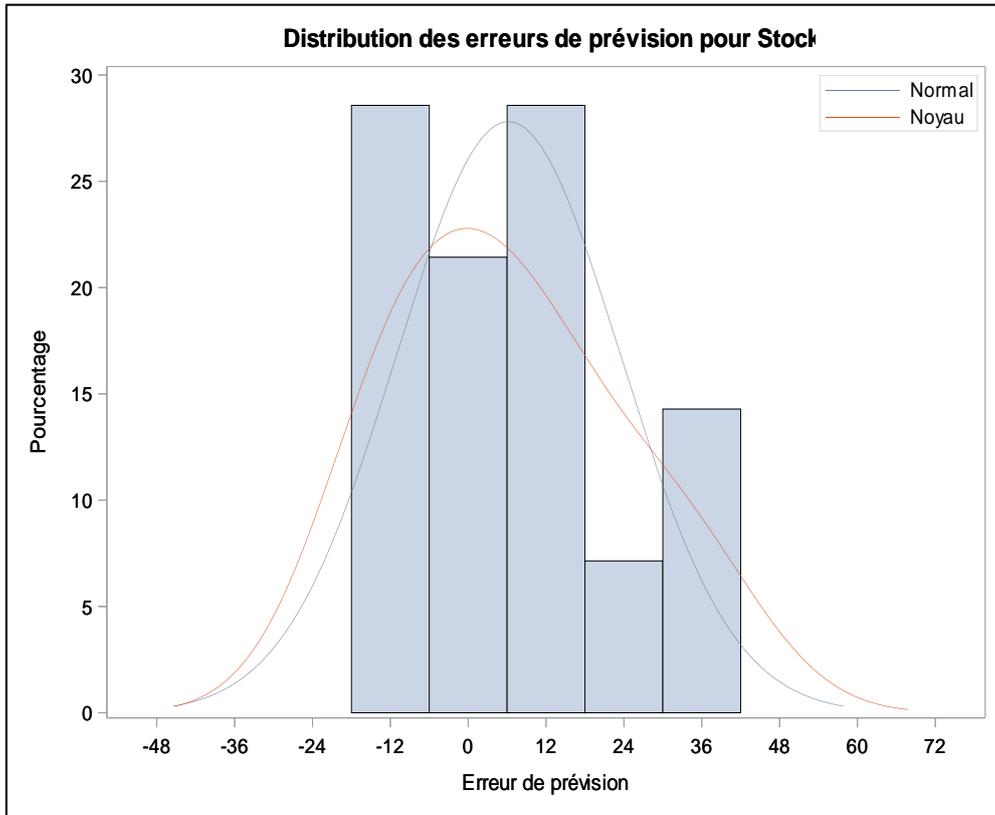
## Méthode de prévision : Document 5

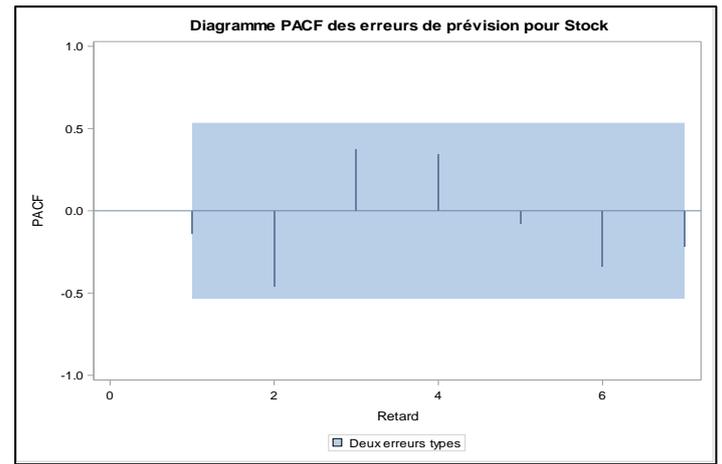
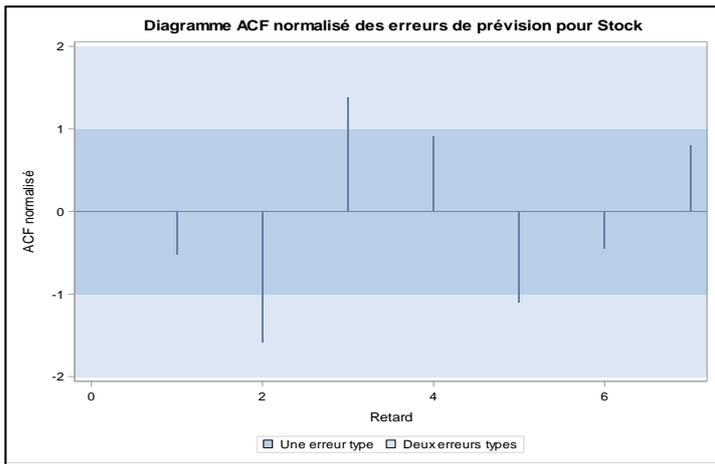
### Q1/ Effectuer un lissage exponentiel avec une constante de lissage $\alpha=0.4$ et une valeur initiale 143



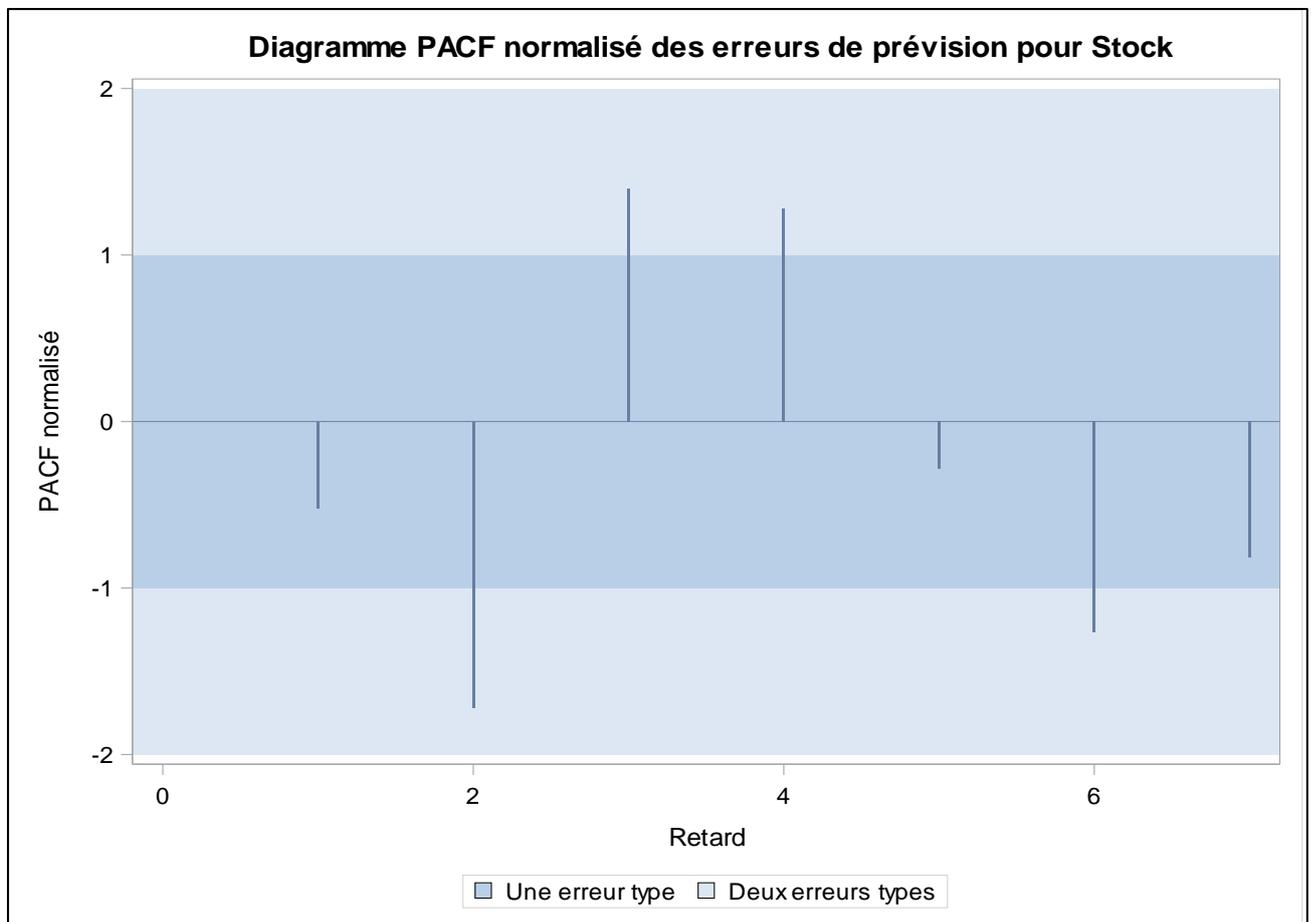
- On remarque la présence d'un trend haussière dans la série.
- On sépare notre échantillon.
- On ne peut pas mettre la valeur initiale dans la proc esm donc on utilise la proc forecast même si elle est obsolète elle convient parfaitement dans ce cas.
- On peut utiliser la proc forecast ou utiliser la proc esm qui donne les prévisions en utilisant les valeurs optimales des paramètres de lissage.
- Ici, on a utilisé outest=est\_LES, on peut donc y récupérer notre coef de lissage optimal.
- On sait d'avance que la méthode de lissage par exponentielle simple n'est pas adaptée en présence d'un trend.

## Prévisions avec la méthode de lissage exponentiel simple Alpha = 0.4





On peut regarder le Diagramme ACF/PACF normalisé des erreurs de prévisions pour stock on voit qu'il y a des autocorrélation ou autocorrélation partielle des erreurs qui sont significatifs. Sachant que pour la méthode de prévision soit retenu il faut que les erreurs de prévisions issu de cette méthode ne possède pas d'autocorrélation ou d'autocorrélation partielle significatif. Or c'est bien le cas ici, donc la méthode de prévision de lissage exponentielle simple n'est pas adaptée.



## Q2/ Effectuer un lissage exponentiel avec une constante de lissage $\alpha=0.8$ et une valeur initiale 143

### Prévisions avec la méthode de lissage exponentiel simple Alpha = 0.8

- On a exactement les mêmes prévisions pour les deux modèles, seuls les intervalles de confiances varient différemment.

On peut regarder le Diagramme ACF/PACF normalisé des erreurs de prévisions pour stock

- On voit qu'il y a des autocorrélation ou autocorrélation partielle des erreurs sont toujours significatifs
- Or on sait que la méthode de prévision soit retenue il faut que les erreurs de prévision, issues de cette méthode, ne possèdent pas d'autocorrélation ou d'autocorrélation partielle significative.
- Or c'est bien le cas ici, donc la méthode de prévision de lissage exponentielle simple n'est pas adaptée.

## Q3/ Quelle constante de lissage donne les meilleures prévisions

### Procédure ESM :

### Prévisions avec la méthode de lissage exponentiel simple

Obs.	_EST_
1	0.48135

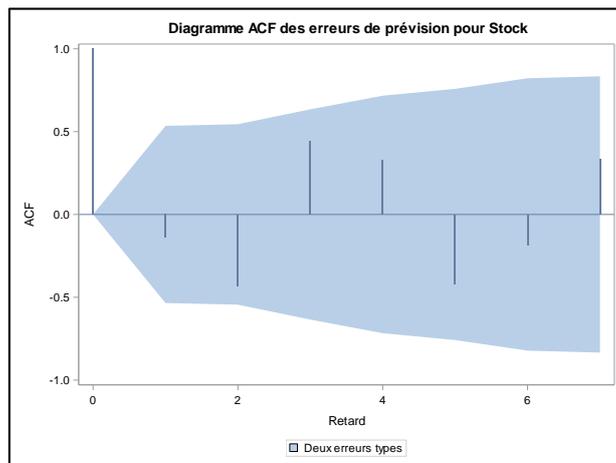
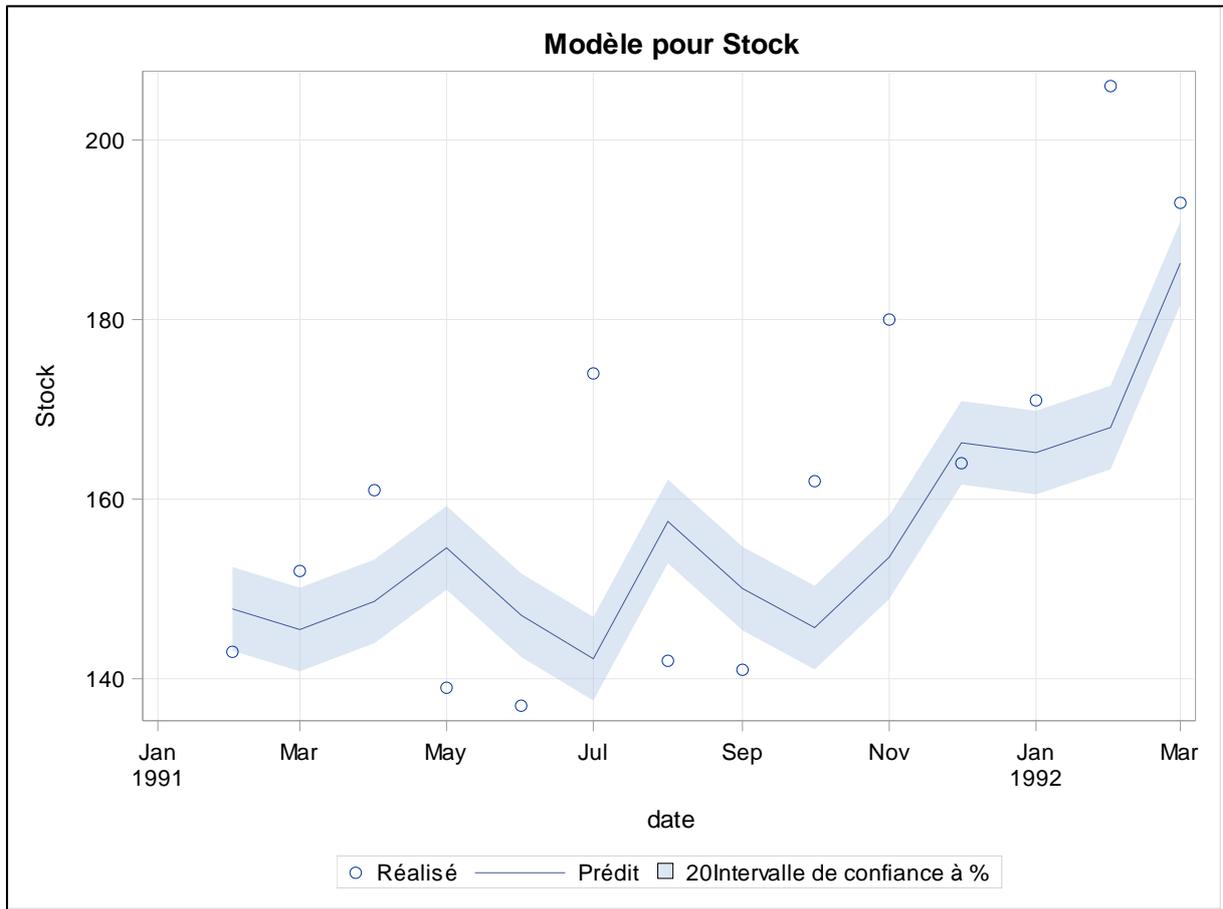
- Afin d'obtenir le meilleur modèle on utilise la constante de lissage optimal égale à 0.48135 trouvé à l'aide de la proc esm
- La série de prévision n'intègre pas la tendance et la saisonnalité de la série 'stocks' car la méthode de prévision de lissage exponentiel ne considère pas la tendance.

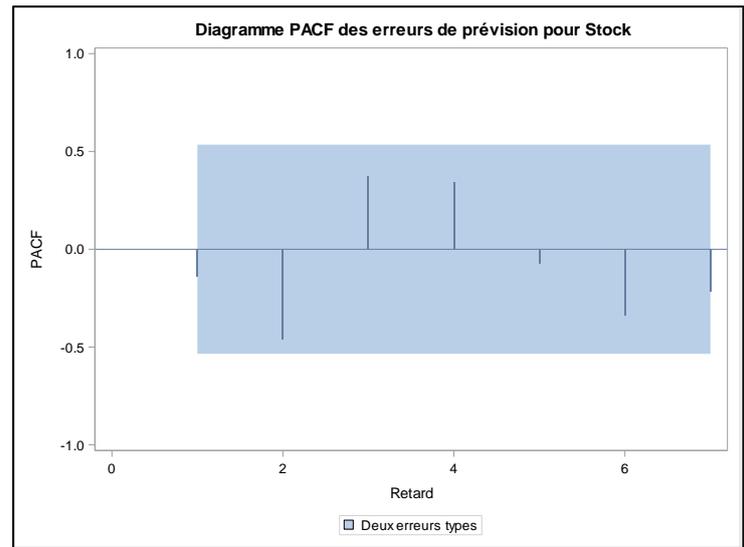
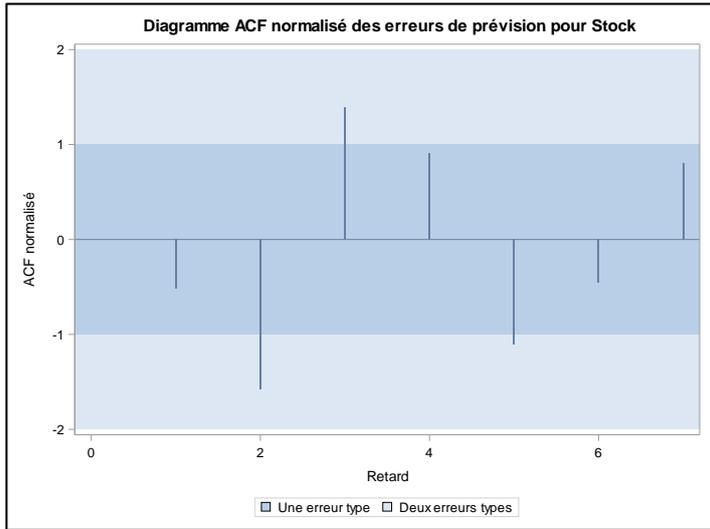
### Prévisions avec la méthode de lissage exponentiel simple alpha =0.48

- Afin d'obtenir le meilleur modèle on utilise la constante de lissage optimal égale à 0.48135 trouvé à l'aide de la proc ESM.
- Lissage exponentielle alpha optimal =0.48135

## Procédure ESM :

### Prévisions avec la méthode de lissage exponentiel simple alpha =0.8





- On peut regarder le Diagramme ACF/PACF normalisé des erreurs de prévisions pour stock
- On voit qu'il y a des autocorrélation ou autocorrélation partielle des erreurs qui sont significatifs,
  - ➔ Donc on sait que pour la méthode de prévision soit retenu il faut que les erreurs de prévision, issues de cette méthode, ne possèdent pas d'autocorrélation ou d'autocorrélation partielle significative.
  - ➔ Or c'est le cas ici, donc la méthode de prévision de lissage exponentielle simple n'est pas adaptée.

## Méthode de prévision : Document 6

### Q1/ La variable x étant la variable dépendante, déterminer la meilleure combinaison de variables explicatives parmi x1, x2 et x3 qui donne les meilleures prévisions ?

- Problème d'importation : les données ne sont pas toutes sous format numérique : il y a des 'variables caractères', on utilise la fonction INPUT.
- Il existe trois types de méthode de sélection automatique :
  - FORWARD
  - BACKWARD
  - STEPWISE (combinant FORWARD & BACKWARD)

On a du renommer X1=Y X2=Z X3=W

### STEPWISE :

Le modèle sélectionné est le modèle à la dernière étape (Etape 3).

Effets : Z W Z\*W

On a Z et Z\*W qui sont les meilleurs estimateurs de X.

Analyse de variance				
Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F
Modèle	3	2779.89563	926.63188	665.29
Erreur	97	135.10437	1.39283	
Total sommes non corrigées	100	2915.00000		

Résultats estimés des paramètres				
Paramètre	DDL	Estimation	Erreur type	Valeur du test t
Z	1	2.631167	0.277761	9.47
W	1	0.236393	0.039538	5.98
Z*W	1	-0.121795	0.024332	-5.01

## **FORWARD :**

Le modèle sélectionné est le modèle à la dernière étape (Etape 3).

Effets : Z W Z\*W

On a Z et Z\*W qui sont les meilleurs estimateurs de X.

Analyse de variance				
Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F
Modèle	3	2779.89563	926.63188	665.29
Erreur	97	135.10437	1.39283	
Total sommes non corrigées	100	2915.00000		

Racine MSE	1.18018
Moyenne dépendante	5.19000
R carré	0.9537
R car. ajust.	0.9522
AIC	138.08774
AICC	138.50879
SBC	43.90325

Résultats estimés des paramètres				
Paramètre	DDL	Estimation	Erreur type	Valeur du test t
Z	1	2.631167	0.277761	9.47
W	1	0.236393	0.039538	5.98
Z*W	1	-0.121795	0.024332	-5.01

## **BACKWARD :**

Le modèle sélectionné est le modèle à la dernière étape (Etape 4).

Effets : Z W Z\*W

On a Z et Z\*W qui sont les meilleurs estimateurs de X.

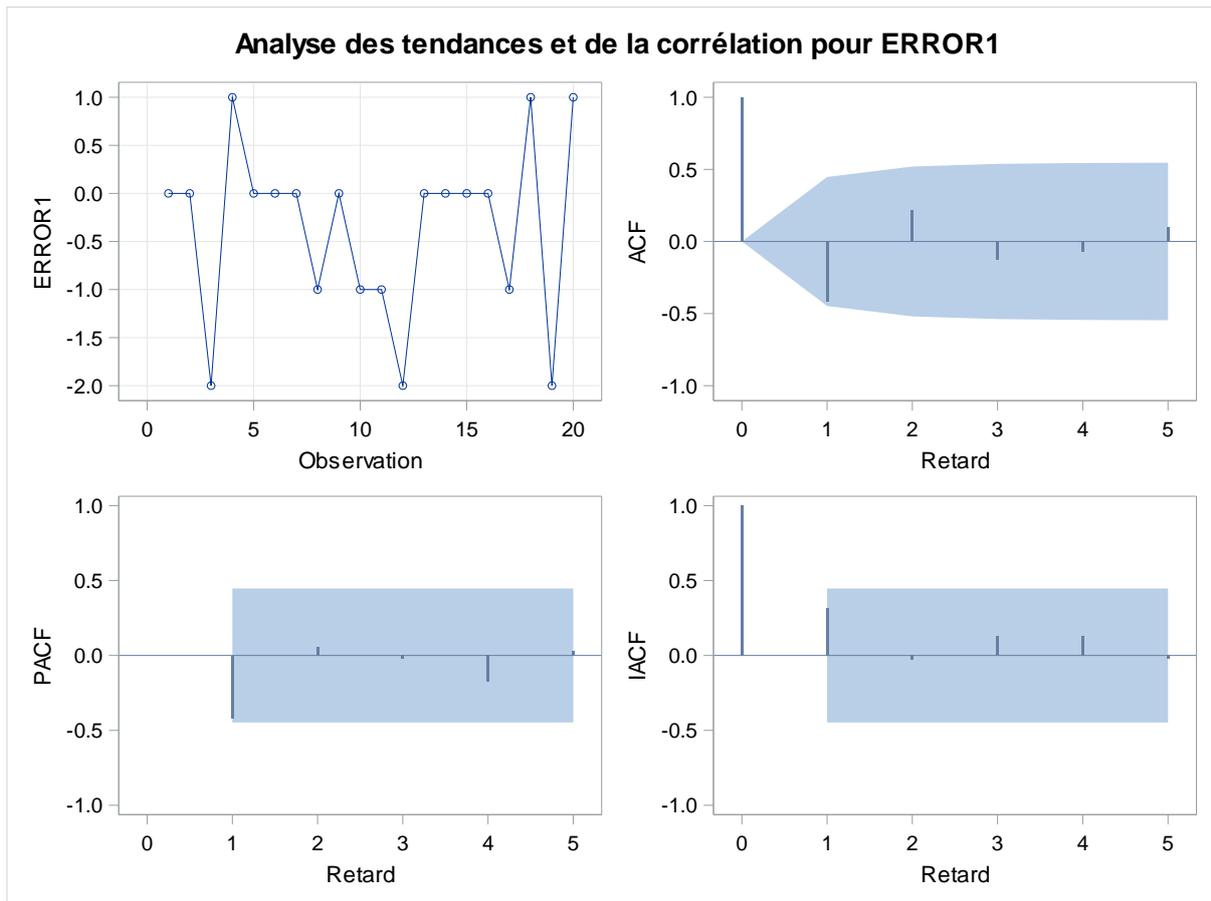
Analyse de variance				
Source	DDL	Somme des carrés	Moyenne quadratique	Valeur F
Modèle	3	2779.89563	926.63188	665.29
Erreur	97	135.10437	1.39283	
Total sommes non corrigées	100	2915.00000		

Racine MSE	1.18018
Moyenne dépendante	5.19000
R carré	0.9537
R car. ajust.	0.9522
AIC	138.08774
AICC	138.50879
SBC	43.90325

Résultats estimés des paramètres				
Paramètre	DDL	Estimation	Erreur type	Valeur du test t
Z	1	2.631167	0.277761	9.47
W	1	0.236393	0.039538	5.98
Z*W	1	-0.121795	0.024332	-5.01

**Q4/ Utiliser la statistique de Diebold et Mariano pour comparer la qualité prévisionnelle des deux modèles précédents.**

Estimation des moindres carrés conditionnels					
Paramètre	Estimation	Erreur type	Valeur du test t	Approx Pr >  t	Retard
MU	-0.35000	0.20869	-1.68	0.1099	0



On a pas de problème dans les autocorrélations et les autocorrélations partielles du modèle on peut donc juger la bonne estimation pour la prédiction.

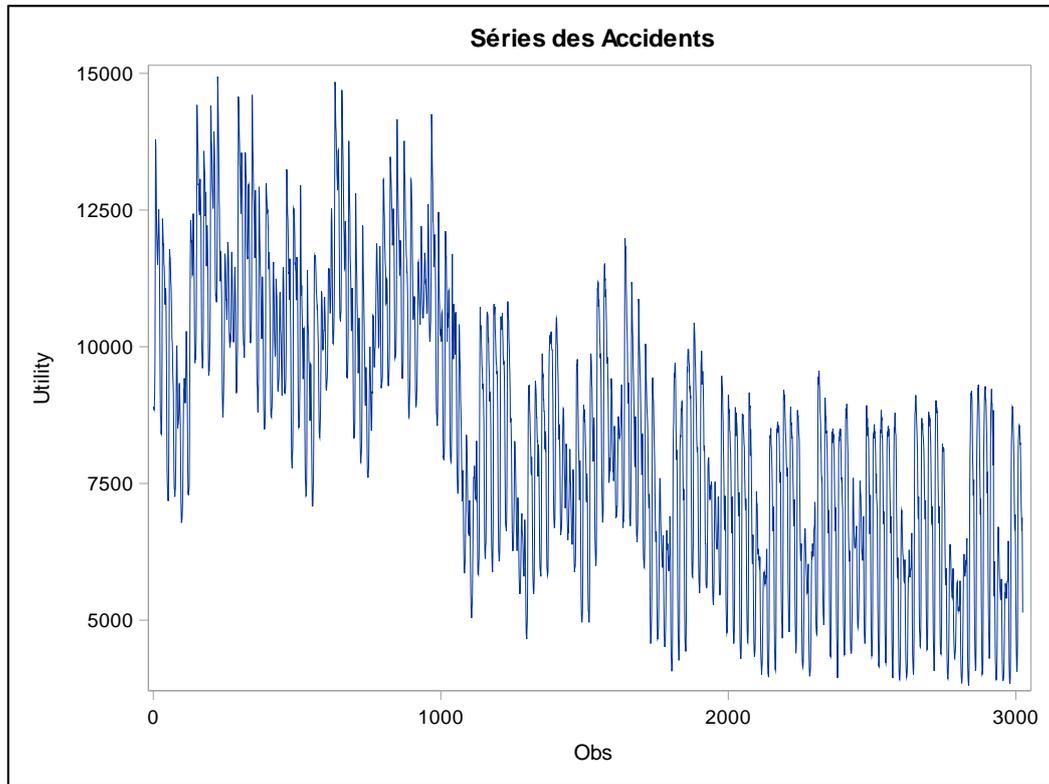
Modèle pour la variable ERROR1	
Moyenne estimée	-0.35

- ➔ Rejet  $H_0$ , le modèle non restreint donne de meilleures prévisions
- ➔ Puisque la moyenne des résidus n'est pas significativement inférieure à zéro, nous concluons que la MAE de population de la première méthode de prévision est supérieure à la MAE de la population de la seconde méthode de prévision.

## Méthode de prévision : Document 7

### Q1/ Générer 100 prévisions avec la méthode du Lissage exponentiel simple

On souhaite à l'aide des données effectuer une prévision avec différentes méthodes de lissage exponentiel.



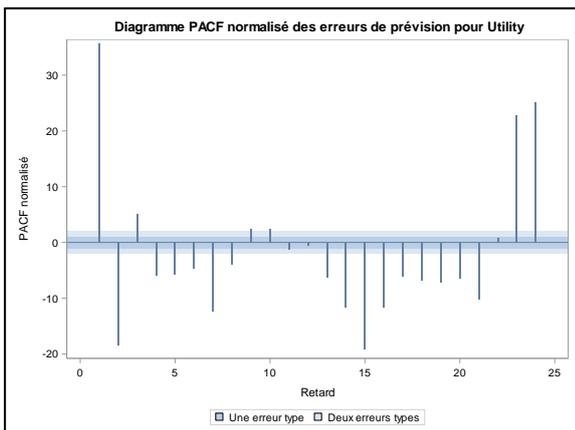
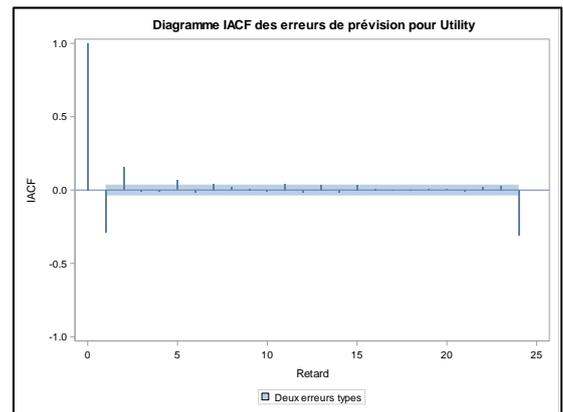
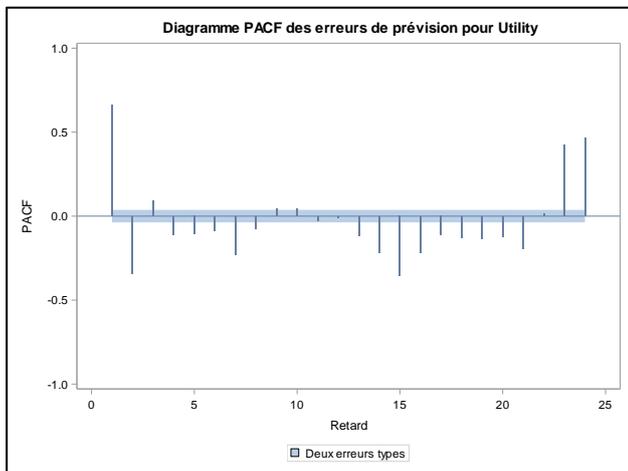
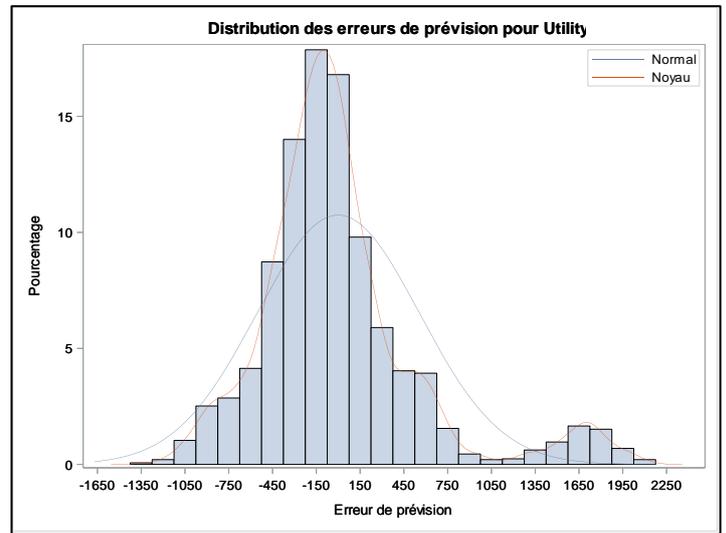
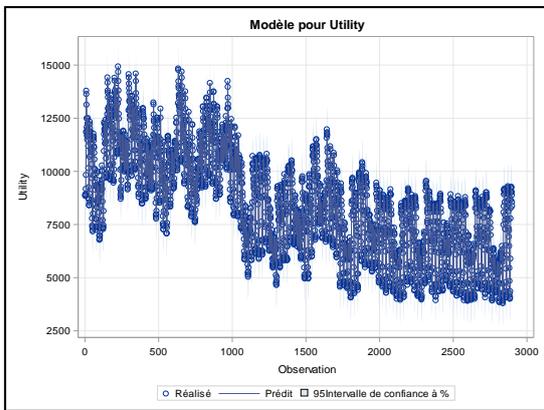
- Ce graphique est difficile à analyser afin de savoir si il y a d'avance un trend et une saisonnalité
- Application de la méthode de lissage exponentiel simple.
- On souhaite connaître la longueur de la saisonnalité à l'aide de la proc timeseries
- De plus, on remarque qu'à 1000 observations on a un changement du trend, en effet le modèle continue avec un trend inférieur à celui de départ.

- Table d'entrée	
Nom	WORK.TABLE
Libellé	
Variable ID temps	Obs
Intervalle de temps	MONTH
Longueur du cycle saisonnier	12

On remarque qu'il y a une saisonnalité d'une année, pour longueur du cycle, donc le modèle de lissage exponentielle simple n'est pas adapté.

## Procédure ESM :

### Prévisions avec la méthode de lissage exponentiel simple



Obs.	_EST_
1	0.999

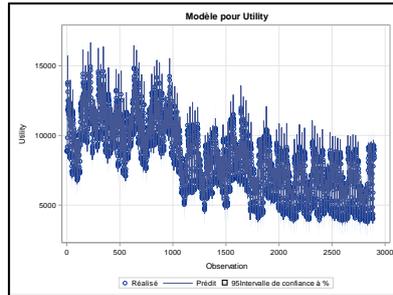
- La valeur optimale du paramètre de lissage de la LES on aura 0.999.

Modèle pas accepté, pour quelle soit accepté elle ne doit pas avoir d'autocorrélation significatives dans les résidus.

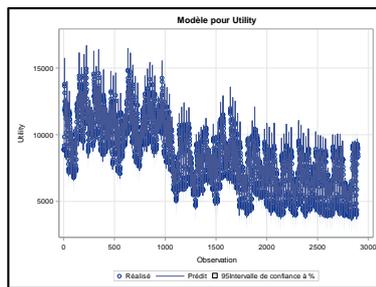
## Q2/ Générer 100 prévisions avec la méthode du Lissage exponentiel double

### Procédure ESM :

#### Prévisions avec la méthode de lissage exponentiel double



## Q3/ Générer 100 prévisions avec la méthode du Lissage exponentiel de Holt

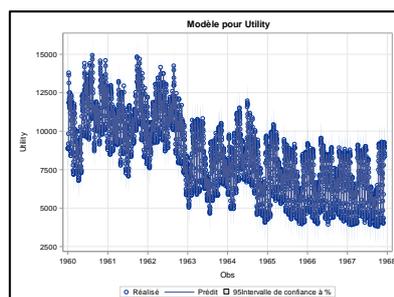


## Q4/ Générer 100 prévisions avec la méthode du Lissage exponentiel triple

- La 'proc esm' ne me permet pas de choisir les paramètres de lissage mais elle permet d'utiliser les paramètres optimaux.

### Procédure ESM :

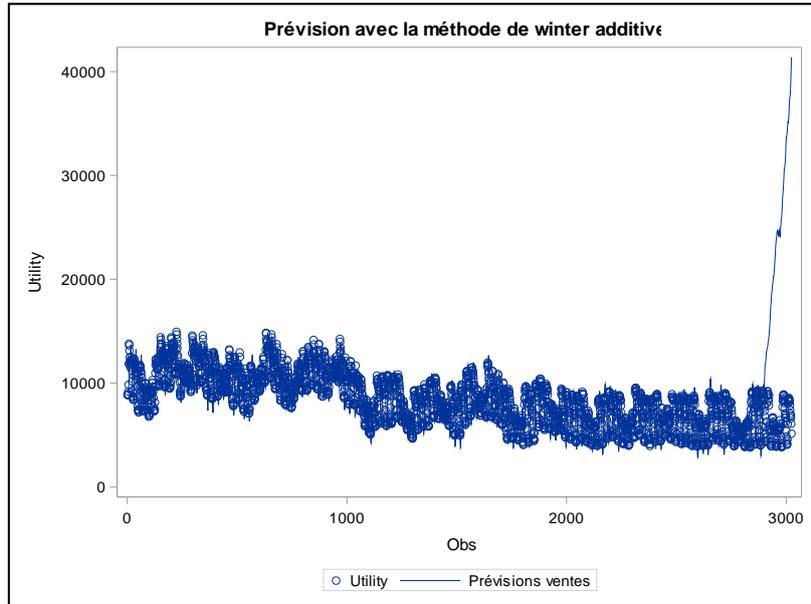
- Prévisions avec la méthode de lissage exponentiel Triple



## Q5/ Générer 100 prévisions avec la méthode du lissage exponentiel de Winters additif

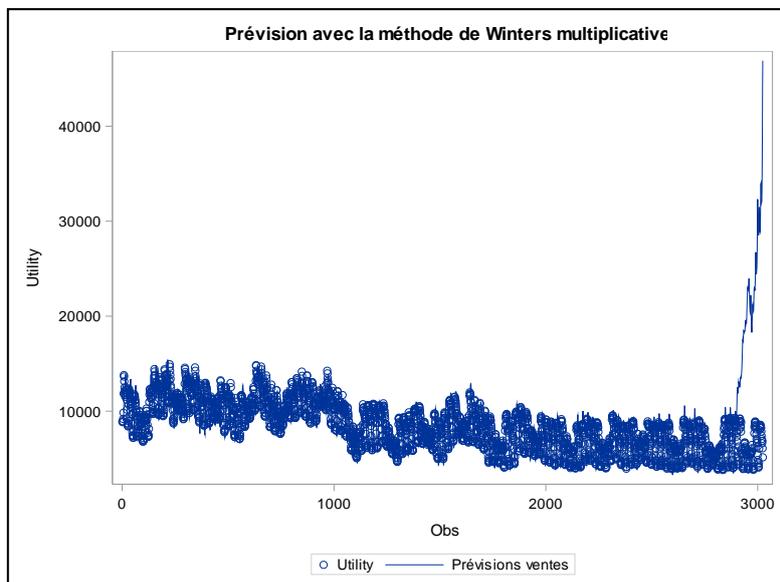
- L'utilisateur ne peut pas changer la valeur des paramètres avec proc esm donc on utilise la proc forecast

### Prévisions avec la méthode de lissage exponentiel de Winters additif

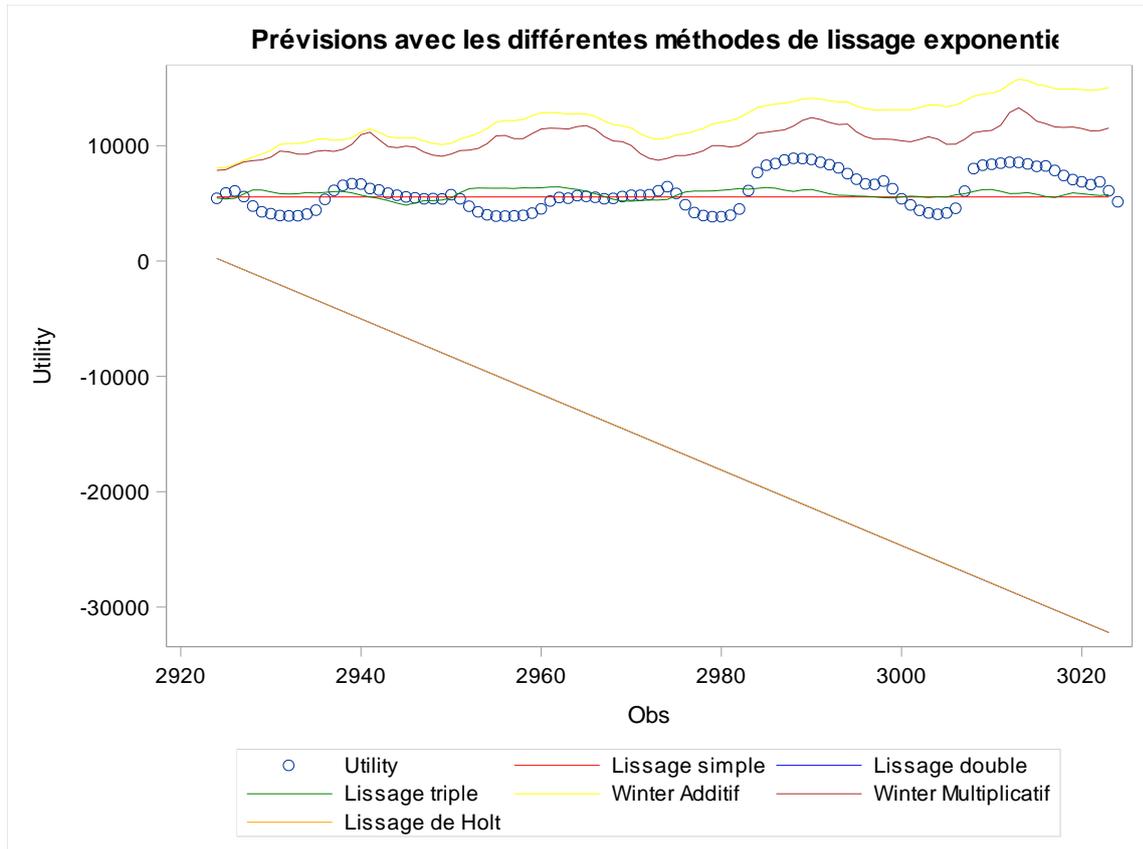


## Q6/ Générer 100 prévisions avec la méthode du lissage exponentiel de Winters multiplicatif

### Prévisions avec la méthode de lissage exponentiel de Winters multiplicatif



**Q7 / Représenter graphiquement la série 'utilité' et les prévisions**  
**obtenues avec les méthodes précédentes.**  
**Représentation graphique des observations**  
**et des prévisions sur un même graphique**



## Q8 / Comparer les méthodes précédentes à l'aide des critères RMSE, MAE et MAPE.

### Prévisions avec les différentes méthodes de lissage exponentiel

	RMSE	MAE	MPE
Lissage exponentiel simple	1580.952	1249.186	0.403995
Lissage exponentiel double	191328.9	165863.3	2771.154
Lissage exponentiel de HOLT	191240.2	165785.2	2769.842
Lissage exponentiel triple	1584.226	1250.483	0.624252
winter Additif	3096.587	2788.315	-55.7561
winter Multiplicatif	3096.587	2788.315	-55.7561

On remarque que les RMSE, MAE, MPE des modèles de lissages exponentiel simple et triple sont les meilleurs.

Cependant, nous allons garder le modèle de lissage exponentiel triple car il incorpore à la fois un trend et une saisonnalité.

La meilleure méthode afin d'estimer ce modèle serait de séparer les données en deux dès lors où on a un changement de trend (1000obs) et de réeffectuer des estimations afin que les paramètres soit le plus adéquat possible.