

Institut de Statistiques de l'Université de Paris

Series Temporelles : projet

Modélisation de la série *fret*

par

Romain FRANÇOIS francoisromain@free.fr
Gwladys TOULEMONDE gwladystoulemonde@yahoo.fr

16 mars 2005

Table des matières

1 Etude de la stationnarité	2
1.1 Test DFA	2
1.2 Test de Ljung–Box	5
1.3 Test de Phillips-Perron	5
1.4 Conclusion	6
2 Modélisation de la série stationnaire	7
2.1 Modélisation AR	7
2.1.1 Construction du modèle	7
2.1.2 Etude des résidus	8
2.1.3 Conclusion	11
2.2 Modélisation MA	11
2.2.1 Construction du modèle	11
2.2.2 Etude des résidus	14
2.2.3 Conclusion	14
2.3 Modèle ARMA	14
2.4 Conclusion	16
3 Prévision	17
3.1 .. normale	17
3.2 .. dynamique	17
4 Robustesse	19
Bibliographie	21

Chapitre 1

Etude de la stationnarité

Nos données concernent le transport de marchandises, 117 mesures ont été relevées ($T=117$). Au vu du graphe 1.1 page suivante, il n'est pas déraisonnable de penser que nous sommes en présence d'une structure autorégressive. En effet, la série semble garder un souvenir de son passé. En d'autres termes, une forte valeur au temps t influencera la valeur au temps $t + 1$ qui aura donc tendance à être également élevée.

Un autre élément distinctif de ce graphe est la présence d'un choc¹ entre les temps 46 et 47. Les valeurs restent élevées jusqu'au temps 53, ceci étant certainement dû à la structure autorégressive de la série. Remarquons que ce souvenir s'efface à moyen terme, les autres valeurs n'étant plus affectées par ce choc.

Nous travaillerons par la suite sur la série centrée en la moyenne. Le graphe de l'auto-corrélogramme partiel 1.2 page suivante nous confirme la structure autorégressive de la série que nous étudierons plus en détail dans la partie 2.1 page 7. La série étudiée semble stationnaire, cette impression de stationnarité va se vérifier en effectuant les tests de racines unitaires de DICKEY-FULLER^[Bou03] augmenté² (DFA) et de PHILLIPS-PERRON^[BT04] (PP). Les tests DFA et PP évaluent l'hypothèse H_0 de présence d'une racine unitaire dans la série $(Y_t, t \in \mathbb{Z})$, la présence d'une racine unitaire impliquant la non-stationnarité de la série.

1.1 Test DFA

Soit Y_t notre série initiale centrée. Le test de DICKEY-FULLER augmenté consiste à effectuer la régression 1.1 page 4 et tester la significativité du coefficient $\rho - 1$ à l'aide des seuils de MACKINNON^[Mac94]. $D_t = \mu + \alpha t$ est la partie déterministe de la

¹La valeur du fret passe de 83.63 au temps 46 à 105.67 au temps 47.

²Le test de DICKEY-FULLER augmenté est ici préféré au test de DICKEY-FULLER simple afin de tenir compte des autocorrélations de la série *fret*.

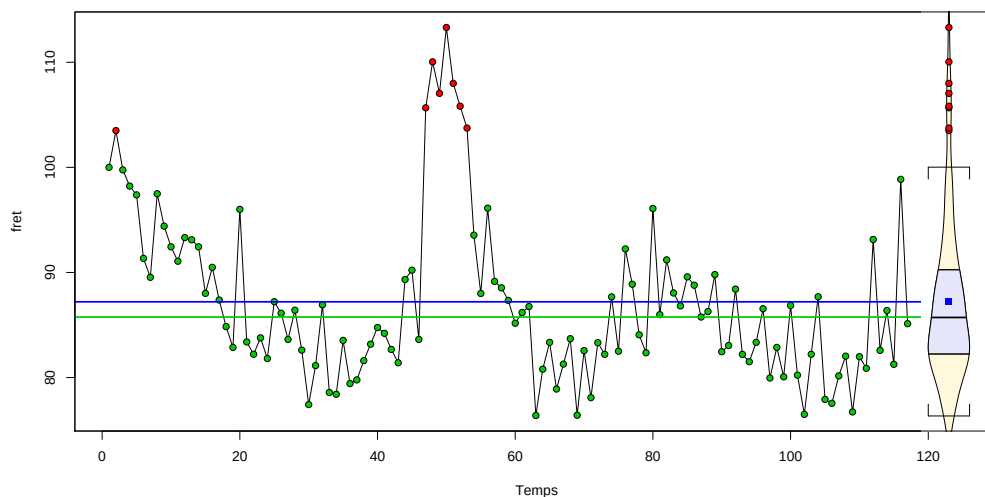


FIG. 1.1 – Représentation de la série *fret*. La moyenne sur les 117 valeurs est représentée par une ligne verte. Un violonplot^[Ver04] est représenté sur la droite de la série, ce graphique confirme le caractère singulier (Une valeur est considérée comme aberrante pour une variable X si elle n'appartient pas à l'intervalle $[Q_1^X - 1.5 \times (Q_3^X - Q_1^X); Q_3^X + 1.5 \times (Q_3^X - Q_1^X)]$, où Q_1^X et Q_3^X sont les quantiles empiriques de X à 25% et 75%.) des valeurs du *fret* entre les temps 47 et 53.

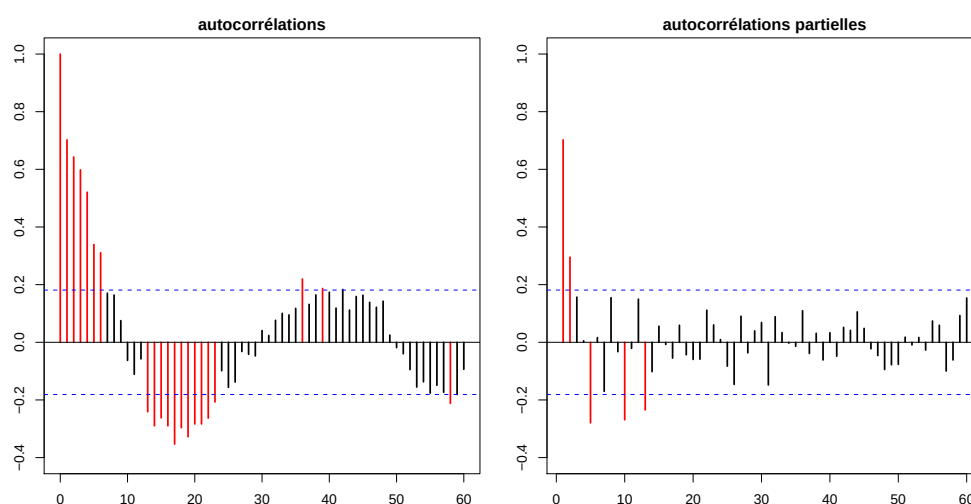


FIG. 1.2 – Autocorrélations et autocorrélations partielles de la série *fretC*. La bande en pointillé correspond à l'intervalle $\pm \frac{2}{\sqrt{T}}$.

série, α et μ pouvant être nuls. Selon les valeurs de α et μ , on obtient les différentes versions du test DFA du tableau 1.1.

$$\Delta Y_t = D_t + (\rho - 1)Y_{t-1} + \sum_{i=1}^{p-1} \psi_i \Delta Y_{t-i} + \epsilon_t \quad (1.1)$$

DICKEY-FULLER augmenté : test de racine unitaire

- $H_0 : \rho - 1 = 0$ dans la régression 1.1. Racine unitaire, non stationnarité.
- $H_1 : \rho - 1 < 0$, stationnarité de la série.
- statistique de test : statistique de STUDENT du coefficient $(\rho - 1)$.
- rejet de H_0 : seuils de MACKINNON, tableau 1.1.

version du test		valeur critique à 5%	résultat
Tendance + constante	$\alpha \neq 0, \mu \neq 0$	-3.454	-3.739
Constante	$\alpha = 0, \mu \neq 0$	-2.890	-3.622
ϕ	$\alpha = 0, \mu = 0$	-1.944	-3.615

TAB. 1.1 – Différentes versions du test DFA. Valeurs critiques de MACKINNON^[Mac94] pour le rejet de la non stationnarité au risque de première espèce 5% et résultats des tests.

La détermination du lag p est essentielle, elle est effectuée à l'aide d'un critère automatique comme le critère d'AKAIKE. On choisit l'entier p^* qui minimise l'AIC (équation 1.4) pour p variant pour de 0 à p_{max} . Notons cependant que le critère AIC a tendance^[NP01] à choisir des valeurs trop petites pour p , ce qui a pour conséquence de distordre la taille du test. p_{max} est déterminé par la formule 1.2 de SCHWERT. En arrondissant à l'entier supérieur, nous obtenons $p_{max} = 13$.

$$p_{max} = 12 \left(\frac{T}{100} \right)^{2/9} \quad (1.2)$$

$$p^* = \arg \min_{0 \leq p \leq 13} AIC(p) \quad (1.3)$$

$$AIC(p) = \ln \hat{\sigma}_{\hat{\epsilon}(p)}^2 - 2 \frac{p}{T} \quad (1.4)$$

Pour la série *fretC*, l'AIC est minimum pour $p=13$ (voir figure 1.3 page suivante), le lag retenu pour effectuer le test de DFA est 13. Les statistiques obtenues pour les trois versions du test DFA (tableau 1.1) conduisent à rejeter au risque de première espèce 5% la présence d'une racine unitaire dans la série *fretC*. La série *fretC* est stationnaire au sens du test de DICKEY-FULLER augmenté.

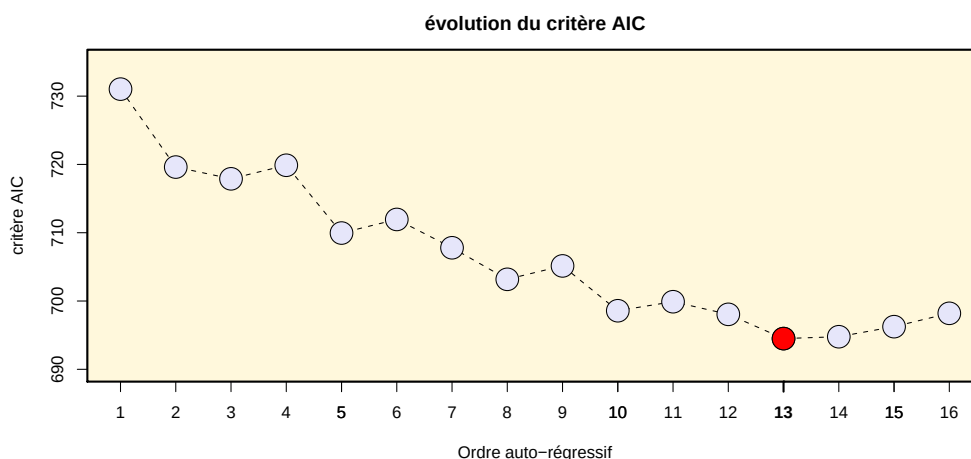


FIG. 1.3 – evolution du critère aic selon l'ordre auto-régressif. Le minimum est atteint pour $p=13$.

1.2 Test de Ljung–Box

Le test de LJUNG–BOX permet d'évaluer les autocorrélations des résidus d'une modélisation $ARMA(p, q)$ de la série d'intérêt.

LJUNG–BOX : test de bruit blanc

- $H_0 : \rho_1 = \rho_2 = \dots = \rho_K = 0$
- $H_1 : \exists i \in 1, \dots, K : \rho_i \neq 0$
- statistique de test : $Q' = n(n+2) \sum_{k=1}^K \frac{\hat{\rho}^2(k)}{n-k}$
- rejet de $H_0 : Q' > \chi_{K-(p+q)}^2(1-\alpha)$

Pour effectuer ce test, il est usuellement préconisé^[BT04] de choisir K proche du tiers du nombre d'observations. Dans la suite de l'étude, nous prendrons $K = 30$. Avec $p = q = 0$, on évalue les autocorrélations de la série initiale, nous obtenons $Q' = 357$ à comparer avec $\chi_{30}^2(0.95) = 44$. Il y a rejet de l'hypothèse nulle et donc présence d'autocorrélations.

1.3 Test de Phillips-Perron

Le test de PHILLIPS-PERRON est une adaptation^[Bou03] du test de DICKEY-FULLER de racine unitaire à utiliser en présence d'erreurs hétéroscédastiques et/ou autocorrélées. Or les résultats du test de LJUNG-BOX en partie 1.2 montrent la

présence d'autocorrélations significatives. La correction non-paramétrique des statistiques de DICKEY-FULLER s'effectue de la façon suivante.

1. Estimation par la méthode des moindres carrés ordinaires des trois modèles de DICKEY-FULLER.

$$\Delta Y_t = D_t + (\rho - 1)Y_{t-1} \quad (1.5)$$

La partie déterministe D_t est identique à celle utilisée dans le cadre des tests de DICKEY-FULLER augmentés. Notons e_t le résidu estimé.

2. Estimation de la variance de *court terme* $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T e_t^2$.
3. Estimation de la variance de *long terme* s_t^2 à partir de la structure des covariances des modèles 1.5 de telle sorte que les transformations réalisées conduisent à des distributions identiques à celles du test standard de DICKEY-FULLER :

$$s_t^2 = \hat{\sigma}^2 + 2 \sum_{i=1}^l \left(1 - \frac{i}{l+1}\right) \frac{1}{T} \sum_{t=i+1}^T e_t e_{t-i} \quad (1.6)$$

L'estimation de cette variance de long terme passe par le choix du nombre de retards l appelé troncature de NEWEY-WEST : $l \approx 4(T/100)^{2/9}$

4. Calcul de la statistique 1.7 de PHILIPPS-PERRON

$$PP = \sqrt{k} \frac{\hat{\rho} - 1}{\hat{\sigma}_{\hat{\rho}}} + \frac{T(k-1)\hat{\sigma}_{\hat{\rho}}}{\sqrt{k}}, \text{ avec : } k = \frac{\hat{\sigma}^2}{s_t^2} \quad (1.7)$$

5. Comparaison de la statistique de PHILIPPS-PERRON aux différents seuils de MACKINNON du tableau 1.1 page 4

PHILLIPS-PERRON : test de racine unitaire

- $H_0 : \rho - 1 = 0$ dans la régression 1.6. Racine unitaire.
- $H_1 : \rho - 1 < 0$, stationnarité de la série.
- statistique de test : $PP = \sqrt{k} \frac{\hat{\rho}-1}{\hat{\sigma}_{\hat{\rho}}} + \frac{T(k-1)\hat{\sigma}_{\hat{\rho}}}{\sqrt{k}}$
- rejet de H_0 : seuils de MACKINNON^[Mac94].

Les résultats du test de PHILLIPS-PERRON coïncident avec ceux des tests DFA et amènent donc à rejeter l'hypothèse de non-stationnarité de la série *fretC*.

1.4 Conclusion

Les tests DICKEY-FULLER augmentés et PHILLIPS-PERRON conduisent au rejet de l'hypothèse de non stationnarité de la série *fretC*. Cette série est donc stationnaire et ne sera pas différenciée. De plus, une différenciation aurait entraîné des problèmes de surdifférenciation.

Chapitre 2

Modélisation de la série stationnaire

La série *fretC* est stationnaire sur la période $t = 1, \dots, 117$. L'observation des autocorrélogrammes et des autocorrélogrammes partiels (graphique 1.2 page 3) ainsi que les résultats du test de LJUNG–BOX (partie 1.2 page 5) montrent que la série ne peut toutefois pas être considérée comme un bruit blanc. Nous allons modéliser la série *fretC* dans la classe des processus *ARMA*. Nous procéderons à une modélisation auto-régressive (AR), à une modélisation moyenne mobile (MA), puis enfin à une modélisation combinée (ARMA).

2.1 Modélisation AR

2.1.1 Construction du modèle

Le graphe de l'auto-corrélogramme partiel 1.2 page 3 montre que la série est autorégressive d'ordre 13. En effet, l'autocorrélation partielle d'ordre 13 est la dernière à sortir de la bande délimitée par $\pm 2/\sqrt{T}$. Nous retrouvons le résultat obtenu lors de l'étude de la stationnarité car c'est à l'ordre 13 que le critère d'AKAIKE est minimum (figure 1.3 page 5). Nous allons donc construire un modèle autorégressif d'ordre 13 (équation 2.1). Les coefficients $\varphi_1, \dots, \varphi_{13}$ sont estimés par la méthode du maximum de vraisemblance, ils suivent asymptotiquement une loi normale, leur significativité pourra donc être testée à l'aide du test de STUDENT.

$$\Phi(L)Y_t = \epsilon_t, \text{ avec : } \Phi(L) = 1 - \varphi_1 L - \varphi_2 L^2 - \dots - \varphi_{13} L^{13} \quad (2.1)$$

Le tableau 2.1 page suivante présente les estimations des coefficients $\varphi_1, \dots, \varphi_{13}$. Il présente également les p-values associées aux tests de STUDENT. Certains coefficients ne sont manifestement pas significativement différents de 0. Nous sommes

en présence d'un modèle auto-régressif à trous. Une sélection descendante¹ va alors être mise en oeuvre jusqu'à l'obtention d'un modèle minimisant le critère AIC.

	Estimate	Std. Error	t value	Pr(> t)
ar1	0.508	0.089	5.687	0.000
ar2	0.294	0.099	2.968	0.004
ar3	0.173	0.105	1.656	0.101
ar4	0.119	0.103	1.162	0.248
ar5	-0.304	0.104	-2.910	0.004
ar6	0.014	0.106	0.131	0.896
ar7	-0.138	0.107	-1.292	0.199
ar8	0.235	0.108	2.169	0.032
ar9	0.079	0.107	0.736	0.463
ar10	-0.224	0.108	-2.086	0.039
ar11	-0.103	0.107	-0.959	0.340
ar12	0.304	0.104	2.921	0.004
ar13	-0.231	0.096	-2.404	0.018

TAB. 2.1 – estimation et significativité des coefficients du modèle AR(13). La p-value associée au coefficient ar6 est la plus forte (0.896), ce coefficient sera donc le premier à être retiré.

Le graphique 2.1 page suivante résume la sélection. Le tableau se lit en ligne, chaque ligne correspondant à un modèle de la sélection. Le modèle initial du tableau 2.1 se retrouve à la première ligne, le coefficient le moins significatif (ar6) est retiré et ensuite la sélection se poursuit pas à pas. L'estimation des coefficients est donnée au centre de chaque case et représentée par la couleur² de la case. La significativité des coefficients est donnée par les triangles en bas à droite de chaque case. L'AIC est donné pour chaque modèle et est représenté à sa gauche par un point bleu.

On retient le modèle de l'étape 4, correspondant à l'AIC minimum³. Enfin, une étude des résidus du modèle retenu (étape 4) est présentée à la droite du graphique. L'autocorrélogramme (acf) et l'autocorrélogramme partiel (pacf) montrent que les résidus ne présentent plus de structure ARMA. De plus, le QQ-plot indique qu'il n'est pas déraisonnable de considérer les résidus gaussiens.

2.1.2 Etude des résidus

Dans un premier temps, il convient d'étudier l'hétéroscédasticité des résidus. Pour ce faire, nous allons regarder si les résidus présentent ou non une structure

¹A chaque étape de la sélection, le coefficient le moins significatif est retiré du modèle.

²L'échelle des valeurs des coefficients est présentée dans la légende en bas à gauche du graphique.

³Marqué d'un gros point rouge.

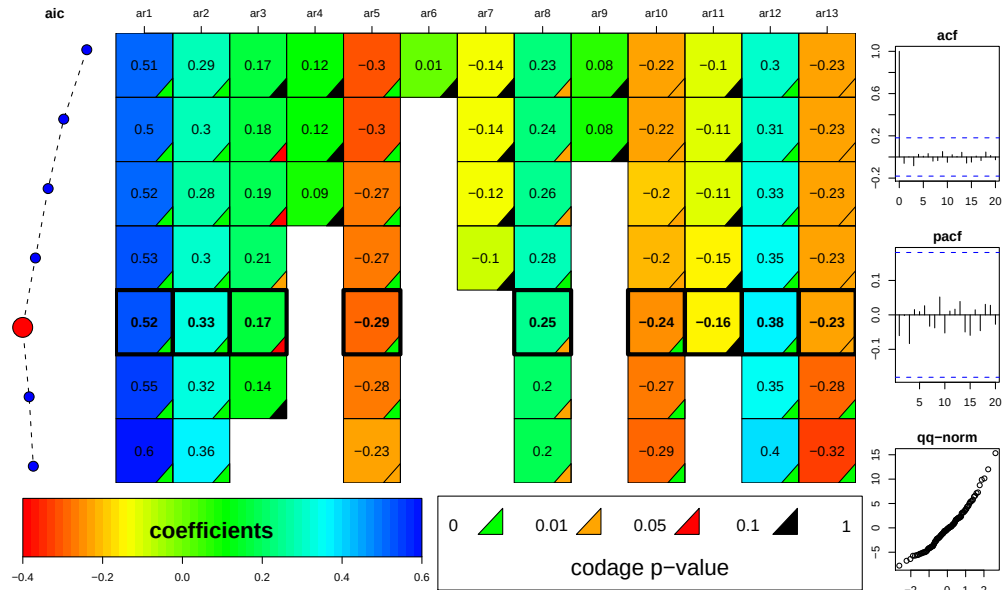


FIG. 2.1 – Sélection des coefficients significatifs du modèle auto-régressif.

GARCH. Après avoir étudié différents modèles, le plus adapté est un ARCH(1). Les résidus du modèle sélectionné dans la partie 2.1.1 page 7 vérifient donc l'équation 2.2 où u_t est un bruit blanc fort.

$$\epsilon_t^2 = c + \alpha \epsilon_{t-1}^2 + u_t \quad (2.2)$$

Les résultats du modèle ARCH(1) sont présentés dans le tableau 2.3 page suivante. Nous allons pouvoir étudier les résidus du modèle ARCH(1). Tout d'abord, ces résidus doivent être gaussiens. C'est pourquoi le test de JARQUE et BERA va être mis en oeuvre.

JARQUE-BERA : test de normalité

- H_0 : normalité.
- H_1 : écart à la normalité.
- statistique de test : $JB = \beta_1 \left(\frac{n}{6}\right) + (\beta_2 - 3)^2 \left(\frac{n}{24}\right)$ où β_1 et β_3 correspondent aux *kurtosis* et au *skewness*.
- rejet de H_0 : $JB > \chi_2^2(1 - \alpha)$

Nous rejetons l'hypothèse H_0 de normalité car la statistique de test JB vaut 9.75. Cependant le nombre d'observations dont nous disposons est très faible et l'hypothèse de normalité est rarement parfaitement vérifiée. Le qqplot présenté à la figure 2.2 page suivante n'est pas extrêmement éloigné de celui d'une loi normale et c'est pourquoi nous pouvons supposer que nos résidus sont normaux.

	Estimate	Std. Error	t value	Pr(> t)
ar1	0.524	0.082	6.404	<.001
ar2	0.327	0.086	3.787	<.001
ar3	0.171	0.091	1.887	0.062
ar5	-0.286	0.095	-3.018	0.003
ar8	0.249	0.097	2.574	0.011
ar10	-0.242	0.091	-2.664	0.009
ar11	-0.155	0.095	-1.630	0.106
ar12	0.376	0.094	4.008	<.001
ar13	-0.234	0.096	-2.434	0.017

TAB. 2.2 – estimation et significativité des coefficients du meilleur modèle AR(13) au sens du critère AIC.

	Estimate	Std. Error	t value	Pr(> t)
a0	12.078	2.306	5.237	<.001
a1	0.268	0.135	1.978	0.048

TAB. 2.3 – estimation et significativité des coefficients du modèle ARCH(1)

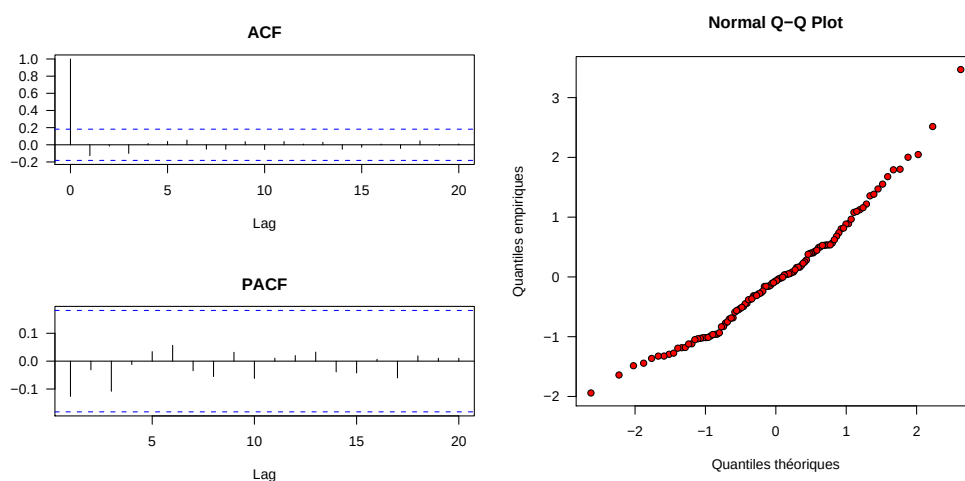


FIG. 2.2 – Etude des résidus du modèle ARCH(1). Autocorrélogramme et auto-corrélogramme partiel, QQ-plot.

Le test de LJUNG et BOX va nous permettre de vérifier que nos résidus sont un bruit blanc en vérifiant l'absence d'autocorrélation entre ces derniers. Le test a été effectué en prenant $K = 30$. La statistique de test vaut 17.3 ce qui correspond à une p-value supérieure à 0.95. On ne rejette pas l'hypothèse de bruit blanc.

2.1.3 Conclusion

Le modèle AR-ARCH retenu est le suivant :

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varphi_3 Y_{t-3} + \varphi_5 Y_{t-5} + \varphi_7 Y_{t-7} + \varphi_8 Y_{t-8} \quad (2.3)$$

$$+ \varphi_{10} Y_{t-10} + \varphi_{11} Y_{t-11} + \varphi_{12} Y_{t-12} + \varphi_{13} Y_{t-13} + \epsilon_t$$

$$\text{avec } \epsilon_t^2 = c + \alpha \epsilon_{t-1}^2 + u_t \text{ où } u_t \text{ est un bruit blanc fort} \quad (2.4)$$

Les estimations des coefficients φ_j , c et α sont donnés dans les tableaux 2.2 page précédente et 2.3 page précédente.

2.2 Modélisation MA

2.2.1 Construction du modèle

En premier lieu, il convient de déterminer l'ordre du modèle moyenne mobile. Cet ordre peut être calculé à l'aide du critère d'AKAIKE. On choisit l'entier q^* qui minimise l'AIC pour q variant pour de 0 à q_{max} ¹. On trouve $q^*=12$ (voir figure 2.3 page suivante).

Nous allons donc construire un modèle moyenne mobile d'ordre 12 (equation 2.5). Les coefficients $\vartheta_1, \dots, \vartheta_{12}$ sont estimés par la méthode du maximum de vraisemblance. Comme ils suivent asymptotiquement une loi normale, nous pourrions à l'aide du test de STUDENT tester la nullité des coefficients.

$$Y_t = \Theta(L)\epsilon_t \text{ avec } \Theta(L) = 1 + \vartheta_1 L + \vartheta_2 L^2 + \dots + \vartheta_{12} L^{12} \quad (2.5)$$

Les estimations des coefficients $\vartheta_1, \dots, \vartheta_{12}$ ainsi que les p-values associées sont présentées dans le tableau 2.4 page 13. Certains coefficients n'étant pas significatifs, une sélection descendante va être mise en oeuvre². Cette sélection est résumée sur la figure 2.4 page suivante qui se lit de la même manière que précédemment. Nous retenons le modèle de l'étape 4. Il s'agit du modèle minimisant l'AIC et c'est de plus le premier modèle ne contenant que des coefficients significatifs à 5%. Le modèle retenu pour la modélisation MA est un modèle moyenne mobile à trous.

¹ q_{max} est défini par la formule de SCWHERT (voir partie 1.1 page 2) et est égal à 13

²A chaque étape de la sélection, le coefficient le moins significatif est retiré du modèle.

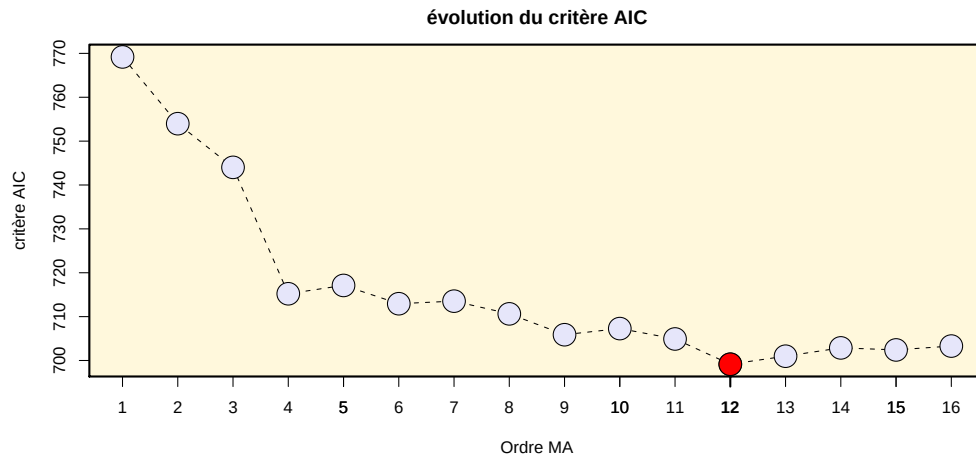


FIG. 2.3 – evolution du critère aic selon l'ordre moyenne mobile. Le minimum est atteint pour $q=12$.

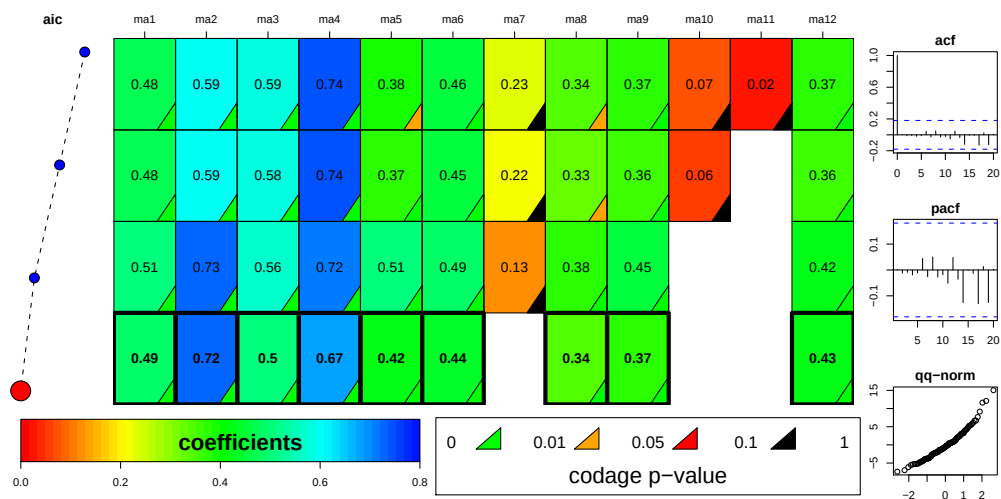


FIG. 2.4 – Sélection des coefficients significatifs du modèle MA.

	Estimate	Std. Error	t value	Pr(> t)
ma1	0.477	0.094	5.091	0.000
ma2	0.593	0.105	5.629	0.000
ma3	0.585	0.124	4.705	0.000
ma4	0.742	0.140	5.295	0.000
ma5	0.385	0.152	2.533	0.013
ma6	0.455	0.147	3.101	0.002
ma7	0.228	0.141	1.616	0.109
ma8	0.340	0.144	2.366	0.020
ma9	0.370	0.124	2.985	0.004
ma10	0.072	0.107	0.667	0.506
ma11	0.023	0.116	0.197	0.844
ma12	0.374	0.102	3.654	0.000

TAB. 2.4 – estimation et significativité des coefficients du modèle MA(12) avant toute sélection.

L'autocorrélogramme (acf) et l'autocorrélogramme partiel (pacf) présentés sur cette même figure 2.4 page précédente, et correspondant aux résidus du modèle minimisant l'AIC, montrent que les résidus ne présentent plus de structure ARMA. Les estimations correspondant au modèle retenu sont présentées au tableau 2.5

	Estimate	Std. Error	t value	Pr(> t)
ma1	0.486	0.091	5.342	<.001
ma2	0.722	0.119	6.061	<.001
ma3	0.501	0.104	4.816	<.001
ma4	0.674	0.126	5.347	<.001
ma5	0.424	0.119	3.558	0.001
ma6	0.444	0.113	3.935	<.001
ma8	0.344	0.123	2.789	0.006
ma9	0.375	0.110	3.417	0.001
ma12	0.428	0.090	4.764	<.001

TAB. 2.5 – Estimation et significativité des coefficients du meilleur modèle MA(12) au sens du critère AIC.

2.2.2 Etude des résidus

Les tentatives de modélisation GARCH des résidus n'ont pas été couronnées de succès. Les résidus du modèle moyenne mobile peuvent ainsi être considérés comme homoscedastiques. Le test de JARQUE BERA va ensuite nous permettre de tester la normalité des résidus. La statistique de test vaut 37 ce qui correspond au rejet de l'hypothèse de normalité. Cependant le qqplot de la figure 2.4 page 12 nous indique que l'écart à la normalité est faible. Pour que le modèle soit tout à fait valide, il reste à vérifier que les résidus sont un bruit blanc. C'est pourquoi nous utilisons le test de LJUNG et BOX. La statistique de test est égale à 19 et il convient de la comparer au quantile à 95 % du χ^2 à 21 degrés de liberté ce qui correspond à une p-valeur de 0.58. Nous rejetons l'autocorrélation des résidus. Ces derniers sont donc un bruit blanc pouvant être considéré comme gaussien et de variance constante.

2.2.3 Conclusion

Le modèle MA retenu est le suivant :

$$\begin{aligned} Y_t = & \vartheta_1 \epsilon_{t-1} + \vartheta_2 \epsilon_{t-2} + \vartheta_3 \epsilon_{t-3} + \vartheta_4 \epsilon_{t-4} + \vartheta_5 \epsilon_{t-5} \\ & + \vartheta_6 \epsilon_{t-6} + \vartheta_8 \epsilon_{t-8} + \vartheta_{12} \epsilon_{t-12} \end{aligned} \quad (2.6)$$

Les estimations des coefficients ϑ_j sont donnés dans le tableau 2.5 page précédente.

2.3 Modèle ARMA

Afin de déterminer les ordres p et q d'un éventuel modèle ARMA, nous allons utiliser la méthode basée sur les coefficients de corrélations canoniques^[GM95]. Cette méthode repose sur un tableau analogue à celui de la méthode du coin. Le tableau 2.5 page suivante présente les p-values associées au test dont l'hypothèse nulle est la nullité de la plus petite valeur propre de la matrice $A(i, j)$ définie dans l'équation 2.7. Le modèle retenu est un modèle $ARMA(7, 3)$, après sélection descendante (présentée en figure 2.6 page suivante) selon le critère AIC, on retient un modèle $ARMA(7, 3)$ à trous dont les coefficients sont présentés dans le tableau 2.6 page 16.

$$A(i, j) = D(i, -1)^{-1} D(i, j)^t D(i, -1)^{-1} D(i, j) \quad (2.7)$$

$$\text{avec } D(i, j) = \begin{bmatrix} \rho_X(j+1) & \dots & \rho_X(j+i+1) \\ \vdots & \ddots & \vdots \\ \rho_X(j+1-i) & \dots & \rho_X(j+1) \end{bmatrix} \quad (2.8)$$

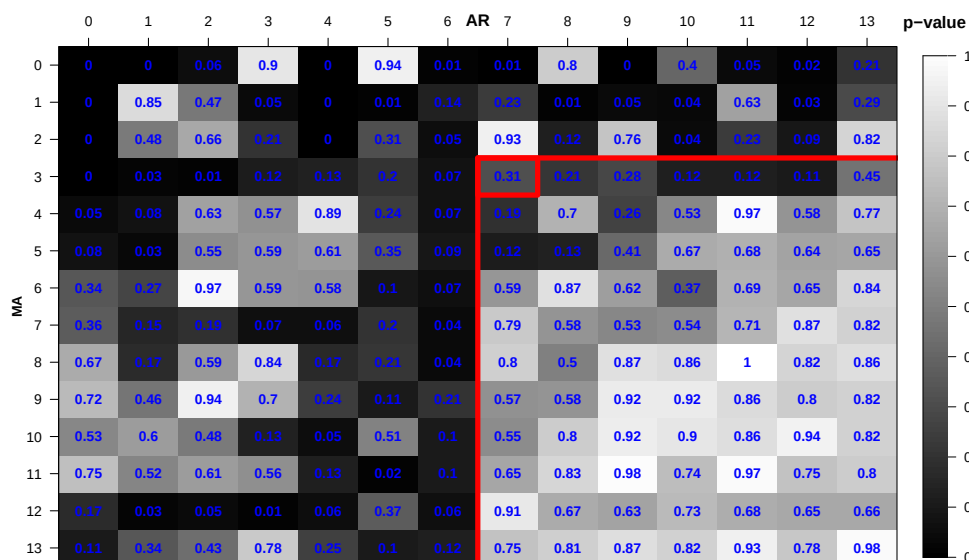


FIG. 2.5 – Méthode du coin. Représentation des p-values. Le modèle retenu, $ARMA(7,3)$, correspond au coin supérieur gauche délimité par les segments rouge.

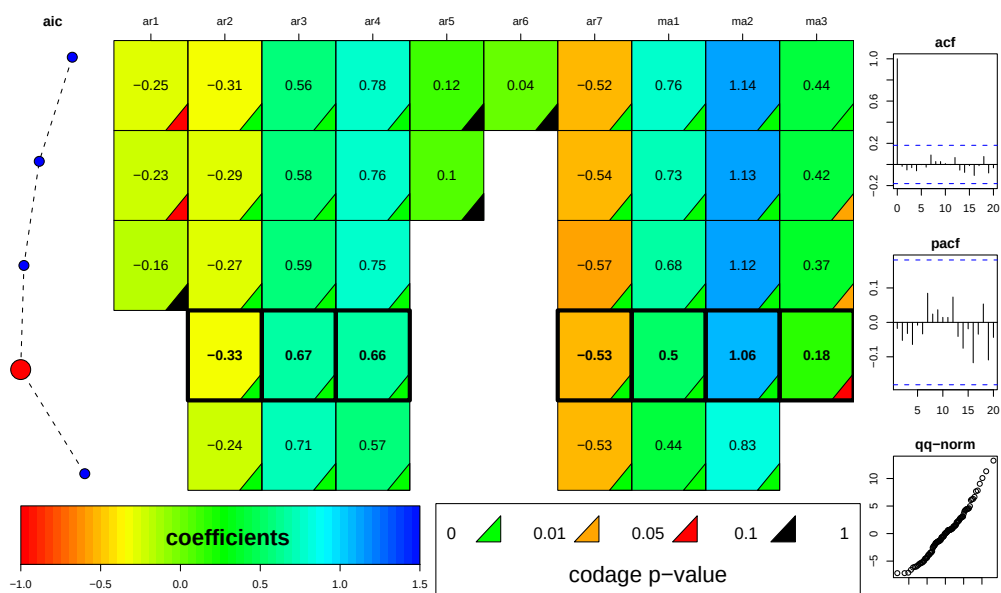


FIG. 2.6 – Sélection des coefficients significatif du modèle ARMA.

	Estimate	Std. Error	t value	Pr(> t)
ar2	−0.327	0.063	−5.172	<.001
ar3	0.672	0.069	9.714	<.001
ar4	0.657	0.069	9.505	<.001
ar7	−0.534	0.071	−7.524	<.001
ma1	0.504	0.094	5.344	<.001
ma2	1.058	0.053	19.797	<.001
ma3	0.181	0.101	1.794	0.076

TAB. 2.6 – Estimation et significativité des coefficients du modèle $ARMA(7, 3)$ après sélection des coefficients par le critère AIC.

Les tentatives de modélisation GARCH des résidus n’aboutissent à aucun résultat concluant. Les perturbations (ϵ_t) du modèle ARMA sont donc considérés comme homoscedastiques. Le test de JARQUE-BERA¹ conduit au rejet de la normalité, cependant le QQ-plot présenté en figure 2.6 page précédente témoigne d’un faible écart à la normalité. Enfin, les tests de LJUNG-BOX² confirment l’absence d’autocorrélations observable sur les graphiques acf et pacf de la figure 2.6 page précédente.

2.4 Conclusion

Les modèles AR–ARCH, MA et ARMA permettent une modélisation correcte de la série *fret*. Le modèle ARMA minimise le critère AIC parmi les 3 modèles étudiés. On retiendra ce modèle pour effectuer la prévision et l’étude de robustesse.

¹ $df = 2, JB = 8.08 > 5.99 = \chi^2_2(0.95), p = 0.018.$

² $df = 30 - 7 = 23, LB = 19.01 < 35.17 = \chi^2_{23}(0.95), p = 0.70.$

Chapitre 3

Prévision

3.1 Prévision normale

La prévision normale consiste à tronquer la série *fretC* en deux parties. La première partie ($t = 1, \dots, 87$) est utilisée pour recalculer les paramètres du modèle ARMA(7,3) en conservant la structure 2.6 page précédente. Les observations restantes ($t = 88, \dots, 117$) sont prédites en utilisant l'écriture du modèle avec les paramètres précédemment estimés.

3.2 Prévision dynamique

La prévision *dynamique* consiste à effectuer la prévision d'une valeur au temps t en utilisant toutes les observations précédentes. Autrement dit, pour chaque temps $t \in 88, \dots, 117$:

1. Estimer les coefficients du modèle en utilisant les observations de la série de 1 à $t - 1$.
2. Prédire la valeur au temps t en utilisant l'écriture du modèle.

La prévision dynamique est évidemment plus précise que la prévision normale car plus d'information est utilisée pour le calcul de chaque prévision. Les valeurs prédites sont plus proches des valeurs observées, une prévision de cette forme est plus réactive aux chocs subis par la série au cours du temps. La bande de confiance évolue également selon les fluctuations de la série.

Seules les valeurs de la série aux temps 115 et 116 se retrouvent en dehors de la bande de confiance. Ce taux d'erreur ($2/30 = 6.66\%$) correspond au risque de première espèce que l'on s'est fixé (5%).

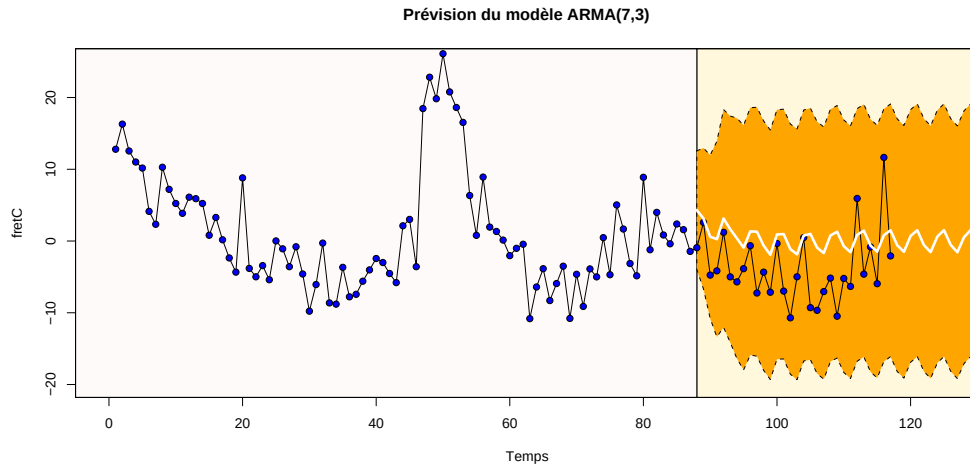


FIG. 3.1 – Prédiction réalisée par le modèle ARMA(7,3). Les valeurs prédites sont représentées par la courbe blanche. La bande orange correspond à un intervalle de confiance à 95% sous des hypothèses de normalité des résidus.

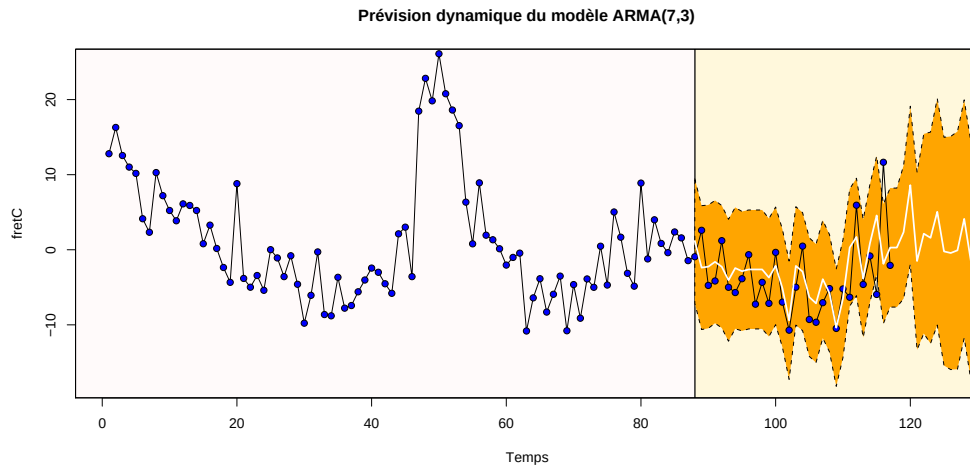


FIG. 3.2 – Prédiction dynamique réalisée par le modèle ARMA(7,3) . Les valeurs prédites sont représentées par la courbe blanche. La bande orange correspond à un intervalle de confiance à 95% sous des hypothèses de normalité des résidus.

Chapitre 4

Robustesse

Etant donné les liaisons temporelles entre les observations de la série, une étude de robustesse de type bootstrap ne peut pas être menée. Il est tout de même intéressant d'étudier la stabilité des coefficients du modèle. Nous adopterons alors la démarche suivante : pour chaque entier $i = 1, \dots, 30$ nous estimerons les coefficients du modèle à partir des observations allant du temps i au temps $87 + i$. De cette façon, nous obtiendrons 30 estimations pour chaque coefficient.

Les résultats sont représentés en figure 4.1 page suivante sous forme de violinplots. Les violinplots^[Ver04] sont des diagrammes à moustache (boxplot) qui ont la particularité de montrer la densité d'une variable telle qu'elle est estimée par la méthode des noyaux^[Sim96, Tsy03]. Nous avons utilisé le noyau gaussien et un choix de la fenêtre par validation croisée.

Certains coefficients peuvent être considérés comme atypiques¹ parmi les 30 valeurs dont on dispose. Il s'agit des coefficients associés aux modèles calculés en démarrant des temps 1, 2, 9, 10, 11, 12.

On peut constater que les supports des densités estimées ne rencontrent jamais la valeur nulle, ce qui conforte les résultats de significativité des coefficients du modèle. Le modèle choisi semble stable avec le glissement de la série.

¹Une valeur est considérée atypique si elle n'appartient pas à l'intervalle $[Q_1 - 1.5(Q_3 - Q_1), Q_3 + 1.5(Q_3 - Q_1)]$ où Q_1 et Q_3 sont les quantiles empiriques à 25% et 75%.

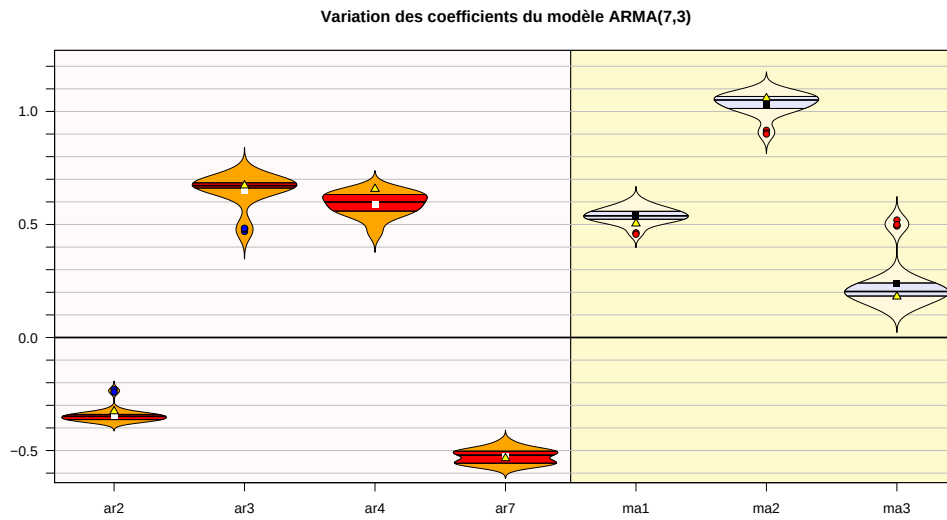


FIG. 4.1 – Robustesse des coefficients AR et MA du modèle ARMA(7,3). Violin-plots sur 30 modèles. Les valeurs des coefficients du modèle complet (avec les 117 observations) sont présentées sous forme de triangle jaune.

Bibliographie

- [Bou03] Régis Bourbonnais. *Econométrie*. Dunod, 2003.
- [BT04] Régis Bourbonnais and Michel Terraza. *Analyse des séries temporelles*. Dunod, 2004.
- [Des03] Bernard Desgraupes. *LaTeX - Apprentissage, guide et référence*. Vuibert, deuxième édition, 2003.
- [GM95] Gouriéroux and Monfort. *Séries temporelles et modèles dynamiques*. Economica, 1995.
- [Mac94] James G. MacKinnon. Approximate asymptotic distribution functions for unit-root and cointegration tests. *Journal of Business & Economic Statistics*, 12 :167–176, avril 1994.
- [NP01] Serena Ng and Pierre Perron. A note on the selection of time series models. Technical report, Boston College Department of Economics, Juin 2001.
- [OPHE04] Tobias Oetiker, Hubert Partl, Irene Hyna, and Schlegl Elisabeth. *The not so short introduction to LaTeX*, avril 2004.
- [Sim96] Jeffrey S. Simonoff. *Smoothing Methods in Statistics*. Springer, 1996.
- [Tea05] R Development Core Team. *R : A language and environment for statistical computing*, 2005.
- [Tra04] Adrian Trapletti. *tseries : Time series analysis and computational finance*, 2004. R package version 0.9-24.
- [Tsy03] Alexandre B. Tsybakov. *Introduction à l'estimation non paramétrique*. Springer, 2003.
- [Ver04] John Verzani. *Using R for Introductory Statistics*. Chapman & Hall, 2004.
- [VR02] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer, quatrième édition, 2002.
- [Wue04] Diethelm Wuertz. *fSeries : Financial Software Collection - fSeries*, 2004. R package version 200.10058.