NON-LINEARITE ET RESEAUX NEURONAUX

Vêlayoudom MARIMOUTOU
Laboratoire d'Analyse et de Recherche Economiques
Université de Bordeaux IV
Avenue. Leon Duguit, 33608 PESSAC, France
tel. 05 56 84 85 77
e-mail : marimout@montesquieu.u-bordeaux.fr

et

Thierry RONCALLI
Laboratoire d'Analyse et de Recherche Economiques
Université de Bordeaux IV
Avenue. Leon Duguit, 33608 PESSAC, France
e-mail : roncalli@montesquieu.u-bordeaux.fr

Résumé
Dans cette étude, nous montrons l'influence du choix d'un modèle (conditionnel ou non conditionnel) pour tester la linéarité dans un réseau de neurones. La détection de la non-linéarité dans ce cadre de référence pose des problèmes d'interprétation statistique, qui limitent la validité de ces méthodes.

1 Introduction

Dans les années récentes, plusieurs développements importants ont eu lieu dans l'analyse des séries chronologiques non-linéaires ; les tests de linéarité, ou plus précisément la détection de la linéarité contre la non-linéarité, sont devenus plus populaires. L'objet est d'éviter la construction de modèle non-linéaire complexe lorsque cela n'est pas nécessaire et de détecter si un modèle non-linéaire pourrait capturer tous les faits "non-linéaires" dans les données. La question naturelle est de savoir quel type de tests utiliser pour détecter la non-linéarité en moyenne conditionnelle ou en variance conditionnelle lorsqu'on a considéré une certaine chronique.

On se concentrera ici sur la linéarité en moyenne conditionnelle comme définie par Lee, White et Granger [1993] (LWG par la suite), c'est-à-dire que le processus y_t est linéaire en moyenne conditionnelle sur X_t si

$$\Pr\left\{E^{\theta}\left(y_t \mid X_t\right) = X_t^{\top}\theta\right\} = 1 \quad \forall\, \theta \in \mathbb{R}^K$$

où y_t et X_t sont des vecteurs de dimension G et K. L'alternative est que cette probabilité est inférieure à 1 (y_t n'est pas linéaire en moyenne conditionnelle). Plusieurs tests sont admissibles (Granger et Teräsvirta [1993]). On distingue deux catégories importantes :

- les tests sans alternative non-linéaire spécifique comme le test du RESET.
- les tests contre un modèle non linéaire spécifique tel que le modèle STR (Smooth Transition Regression).

La plupart de ces tests peuvent être formulés comme des test de type multiplicateur de Lagrange (LM), ce qui est intéressant car le modèle n'a ainsi pas besoin d'être estimé sous l'alternative. D'après Davidson et McKinnon (1993), ce type de test peut être obtenu à partir d'une régression artificielle de Gauss-Newton.

Cet article porte essentiellement sur les particularités du test de réseau neuronal (procédure développée par des cogniticiens) dans un cadre conditionnel et non conditionnel avec comparaison des performances en faisant des études de Monte Carlo. La stratégie de test LWG se développe dans le cadre d'un modèle conditionnel et la statistique de test porte sur l'existence de non linéarité en moyenne, l'alternative intéressante étant la non linéarité de la moyenne conditionnelle. Le problème que nous traitons est le suivant : Que se passe t-il si au lieu de traiter le modèle conditionnel, nous nous focalisons sur le modèle non conditionnel, c'est à dire si nous utilisons la série filtrée de sa structure linéaire ?

2 Test de non-linéarité dans un réseau neuronal

2.1 Le réseau

Le réseau considéré est constitué de K entrées, de J unités cachées et de G sorties. Il est défini par les relations suivantes

$$\begin{aligned} \zeta_{j,n} &= a_{0,j} + \sum_{k=1}^{K} a_{k,j} x_{k,n} \\ \xi_{j,n} &= f\left(\zeta_{j,n}\right) \\ \kappa_{g,n} &= \sum_{j=1}^{J} b_{j,g} \xi_{j,n} \end{aligned} \tag{1}$$

La fonction f est la fonction d'activation du réseau. On considère généralement une fonction non linéaire bornée, comme la fonction logistique. C'est pourquoi les relations (1) sont perçues comme celles définissant un modèle non linéaire.

Le modèle statistique associé au réseau neuronal est

$$y_{g,n} = \kappa_{g,n} + \varepsilon_{g,n}$$

avec

$$\varepsilon_n = \begin{bmatrix} \varepsilon_{1,n} \\ \vdots \\ \varepsilon_{G,n} \end{bmatrix}$$

et $\varepsilon_n \sim \mathcal{N}\left(0, \Sigma\right)$.

2.2 Linéarité et Non-linéarité

Considérons deux séries temporelles y et x de dimensions respectives G et K. Si la série temporelle y est linéaire par rapport à celle de x, alors le processus de génération de données est de la forme suivante

$$y_{g,n} = x_n^\top \beta_g + u_{g,n}$$

avec $x_n = \left[\begin{array}{ccc} x_{1,n} & \cdots & x_{K,n} \end{array}\right]^\top$. Considérons que le bruit sur la série est gaussien de moyenne nulle. En reprenant la terminologie de LWG, l'hypothèse de linéarité est

$$\Pr\left\{E\left[y_{g,n}|\, x_n\right] = x_n^\top \beta_g\right\} = 1$$

Cela veut dire que si nous considérons le modèle statistique suivant

$$y_{g,n} = x_n^\top \beta_g + \kappa_{g,n} + \varepsilon_{g,n} \tag{2}$$

alors nous devons vérifier que

$$\Pr\left\{E\left[\kappa_{g,n}|\, x_n\right] = 0\right\} = 1$$

Imposer l'hypothèse de linéarité revient donc à poser $b_{j,g} = 0$. L'étude de la linéarité peut alors être menée selon une approche non conditionnelle ou conditionnelle :

1. Dans la première approche, le modèle (2) est estimé par maximum de vraisemblance. La log-vraisemblance pour l'observation n est donnée par l'expression suivante

$$-\frac{NG}{2}\ln 2\pi - \frac{G}{2}\ln\left|\det\Sigma\right| - \sum_{n=1}^{N}\varepsilon_n^\top \Sigma^{-1}\varepsilon_n$$

2. Dans la seconde approche, nous considérerons le modèle conditionnel suivant

$$y_{g,n} - x_n^\top \beta_g = \kappa_{g,n} + \varepsilon_{g,n} \tag{3}$$

Dans ce modèle, la linéarité de y par rapport à x est imposée a priori. Soit $u_{g,n} = y_{g,n} - x_n^\top \beta_g$ la partie non prise en compte par la linéarité. Nous cherchons alors à extraire une structure non linéaire par le réseau.

Le vecteur des paramètres d'intérêt est

$$\theta = \mathrm{vec}\left[\begin{array}{cccc} a & b & \beta & \Sigma \end{array}\right]$$

Dans le modèle non conditionnel, θ est estimé par maximum de vraisemblance. Dans le modèle conditionnel, β sous la condition de linéarité est approchée de façon efficace par un estimateur des moindres carrés. Les paramètres a, b et Σ sont ensuite estimés par maximum de vraisemblance.

Si $u_{g,n}$ est orthogonal à $\kappa_{g,n}$ l'hypothèse de linéarité est acceptée. Dans le cas contraire, deux problèmes se posent :

- la fonction $\kappa_{g,n}$ est elle réellement engendrée par une fonction non linéaire ?
- la modification de la distribution sous-jacente par le conditionnement joue-t-elle un rôle dans la détection des non-linéarités ?

3 Résultats

3.1 La construction de la procédure de test n'est pas indifférente au modèle considéré

On considére le modèle AR(1) suivant

$$y_t = 0.95 y_{t-1} + u_t$$

En utilisant une seule unité cachée, nous devons vérifier que

$$a_0 = a_1 = b_1 = \beta_0 = 0$$

et que

$$\beta_1 = 0.95$$

Nous avons effectué 250 simulations pour une taille d'échantillon de 1000 points et une variance σ_u^2 égale à 0.25. Nous avons estimé les fonctions de densité des différents estimateurs par la méthode du noyau gaussien (Härdle et Linton [1994]).

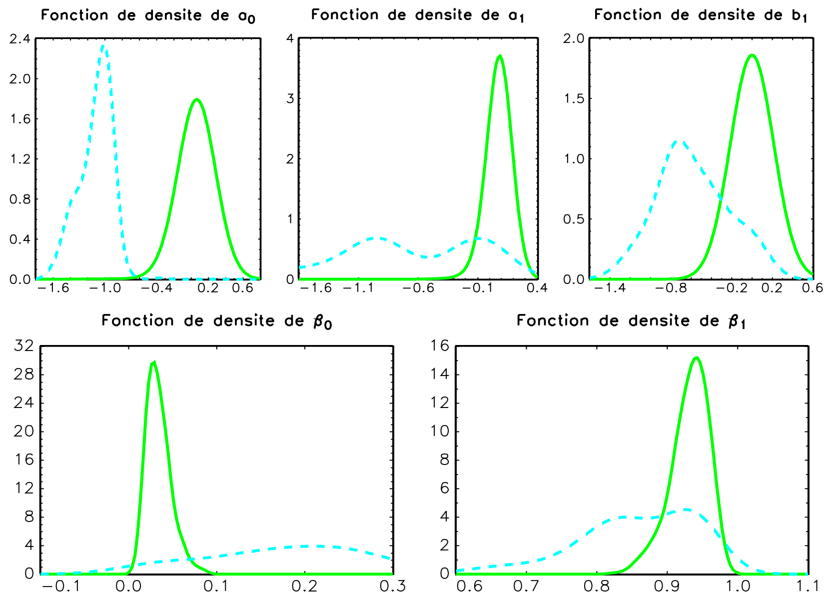

Figure 1:

Il est clair que le choix d'une modélisation n'est pas neutre sur les résultats (voir graphique[1] 1). En effet, avec le modèle non conditionnel, il existe un biais important concernant les coefficients. De plus, les fonctions de densité des estimateurs ne sont pas symétriques par rapport au mode, lorsqu'il en existe un. Pour le modèle statistique conditionnel, nous observons un biais seulement pour le coefficient β_0. Cependant, ce biais est relativement faible.

3.2 Linéarité et réseau neuronal

3.2.1 Processus linéaires et réseau neuronal

Pour tester l'hypothèse de linéarité, il ne suffit pas de tester l'hypothèse $b_{j,g} = 0$. En effet, cela n'indique pas forcément que le processus sous-jacent est non linéaire. La non linéarité est introduite dans le réseau en spécifiant les unités cachées telles que nous avons $\xi_{j,n} = f\left(\zeta_{j,n}\right)$. Dans ce cas, la relation entre $\xi_{j,n}$ et $\zeta_{j,n}$ est non linéaire si la fonction f est non linéaire. Cependant,

[1] Les courbes en traits pleins (respectivement, en tirets) correspondent aux fonctions de densité pour le modèle conditionnel (respectivement, non conditionnel).

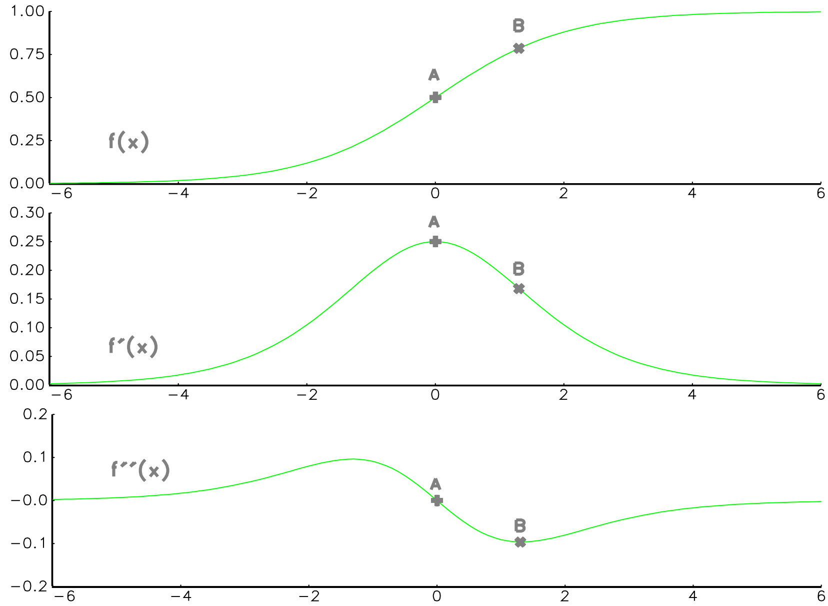

Figure 2:

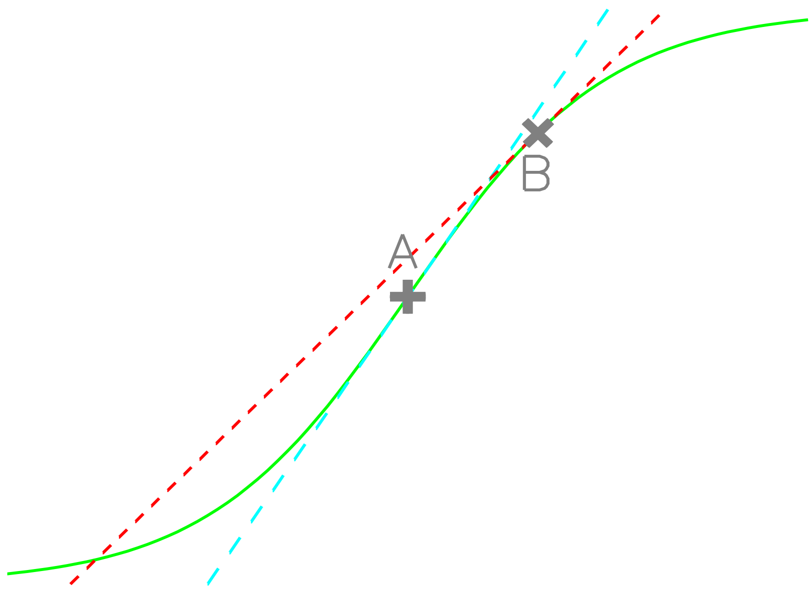

Figure 3:

la fonction f peut parfois être approximée localement par un fonction linéaire. De plus, nous pourrions utiliser une fonction linéaire. **Il faut donc être très prudent et ne pas assimiler de façon systématique non linéarité et réseau neuronal**.

Considérons la fonction logistique

$$f(x) = \frac{1}{1 + \exp(-x)}$$

Considérons les points A et B (graphiques 2 et 3). Au voisinage de A, la fonction logistique présente une courbure nulle. Au point B, la courbure est maximale (voir graphique 2). Dans le premier cas, la fonction logistique peut être raisonnablement approximée par une droite (graphique 3). Si les valeurs prises par $\xi_{j,n}$ sont proches de 0.5, alors les relations (1) ne peuvent plus être considérées comme celles définissant un modèle non linéaire (graphique 3).

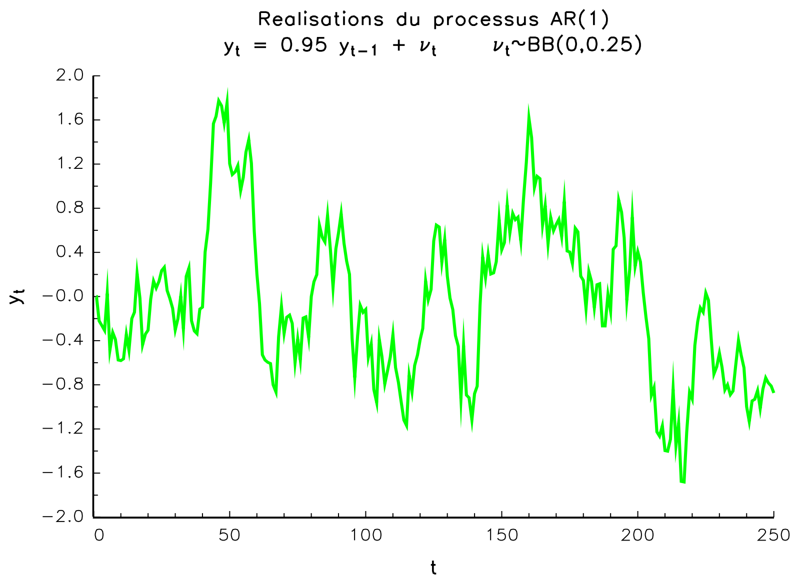

Figure 4:

Considérons une réalisation du processus AR(1) précédent. Le graphique 4 montre la trajectoire de cette série. Supposons que nous voulons modéliser ce processus avec un réseau de neurones. La valeur de sortie du réseau correspond à $\kappa_{1,n}$ (nous n'utilisons pas de lien direct), c'est-à-dire que le modèle statistique est

$$y_n = b_1 f\left(a_0 + a_1 y_{n-1}\right) + \varepsilon_n$$

avec f la fonction logistique. Le coefficient b_1 apparait significativement différent de zéro. Si nous analysons les valeurs prises par $\xi_{1,n}$, la répartition du nuage de points ne se fait pas sur l'ensemble du segment $[0, 1]$, mais seulement sur une portion beaucoup plus faible $[0.75, 0.87]$ (voir graphique 5). Même si la fonction est logistique et donc non linéaire, la relation entre $\zeta_{1,n}$ et $\xi_{1,n}$ peut être considérée comme linéaire. En effet, le nuage des points peut être raisonnablement approchée par une droite.

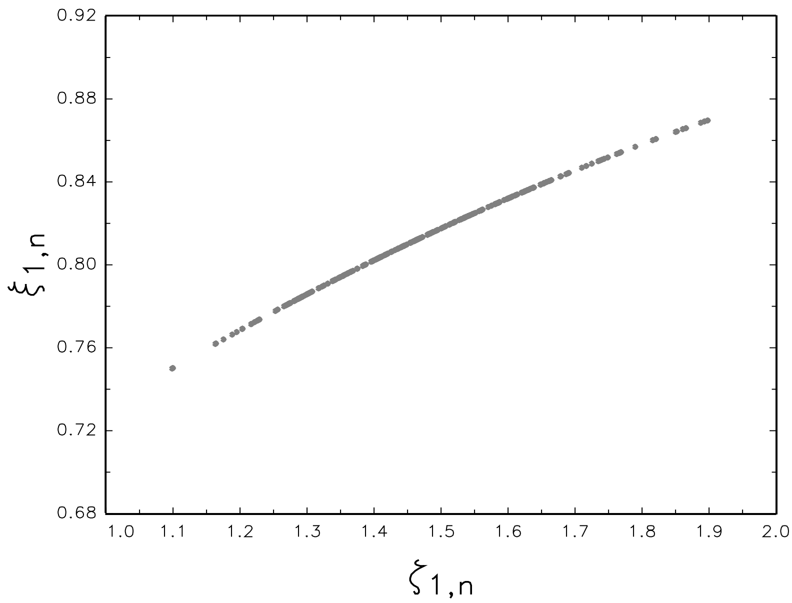

Figure 5:

3.2.2 Processus non linéaires et réseau neuronal

Considérons maintenant le processus logistique bruité suivant

$$y_n = \frac{1}{1 + \exp\left(-x_n\right)} + \varepsilon_n$$

avec $\varepsilon_n \sim \mathcal{N}(0, 0.01)$. C'est un processus non linéaire par rapport à x_n. La question posée est de savoir si le test de réseau neuronal permet de rejeter l'hypothèse de linéarité. Car, si nous utilisons le modèle conditionnel, alors les cibles du réseau ne sont plus y_n, mais la projection orthogonale de y_n sur x_n. La distribution sous-jacente est donc modifiée puisque nous avons appliqué un filtre linéaire. La fonction de densité empirique de b_1 (graphique 6) montre clairement que celle-ci est symmétrique autour du point 0. Dans la plupart des cas (en prenant une approximation gaussienne), l'hypothèse de nullité du coefficient ne peut être rejeté. Cela indique que le réseau n'a pas pu modéliser la série filtrée. Une implication importante pour la modélisation est que l'imposition a priori de la linéarité peut favoriser cette hypothèse de linéarité, car le processus transformé n'a plus les mêmes caractéristiques que le processus original.

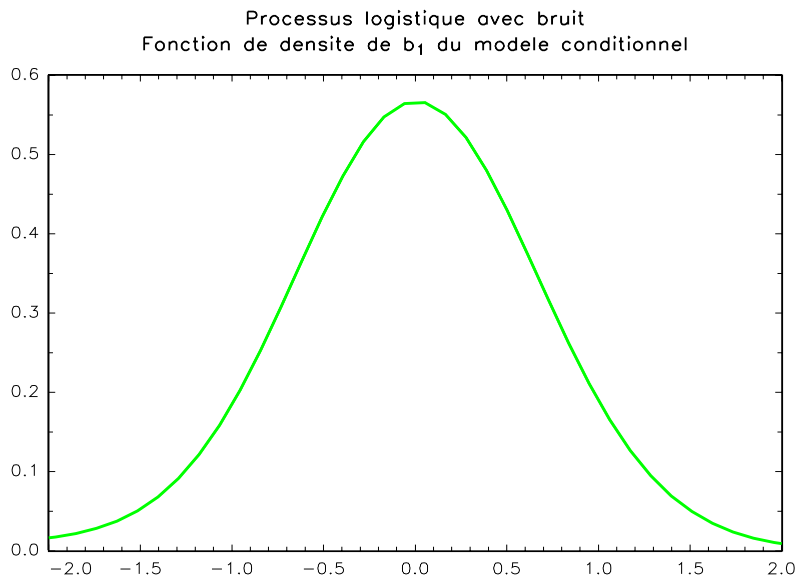

Figure 6:

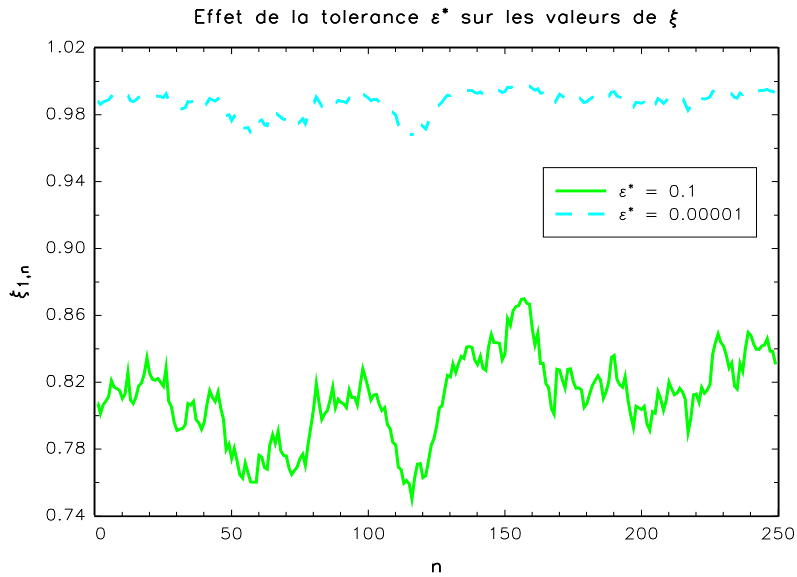

Figure 7:

3.3 Le test de réseau neuronal se ramène-t-il seulement à un problème de test statistique ?

On considère le processus AR(1) précédent. On estime le modèle non conditionnel associé par maximum de vraisemblance. Le critère d'arrêt pour l'optimisation porte sur le vecteur gradient de la fonction de vraisemblance. Nous arrêtons l'algorithme de maximisation lorsque l'ensemble des valeurs du vecteur gradient sont inférieures à la tolérance ε^*. Nous envisageons différentes valeurs prises par ε^*. Nous reportons dans le tableau suivant les estimations des paramètres.

	$\varepsilon^* = 0.1$	$\varepsilon^* = 0.001$	$\varepsilon^* = 0.00001$
a_0	0.9999	0.9688	2.9361
a_1	1.0008	1.1272	3.3950
b_1	0.0637	0.7204	0.7745
β_0	-0.0258	-0.5082	-0.7279
β_1	0.9430	0.8316	0.9215
σ	0.0499	0.0498	0.0498

Nous remarquons que les valeurs des paramètres dépendent significativement de ce critère d'optimisation. En effet, la valeur prise par β_0 varie de -0.02 à -0.72 pour des valeurs de ε^* égales à 0.1 et 0.00001. Si nous nous intéressons au coefficient b_1, on peut supposer que le modèle est linéaire pour ε^* égal à 0.1, puisque b_1 prend une valeur proche de 0, alors que pour ε^* égal à 0.00001, cette hypothèse est difficilement tenable. La modélisation par le réseau dépend donc du critère de convergence et ce fait observé influe dramatiquement sur l'objet à tester.

Le modèle non conditionnel que nous avons utilisé est le suivant

$$y_n = \beta_0 + \beta_1 y_{n-1} + b_1 f\left(a_0 + a_1 y_{n-1}\right) + \varepsilon_n$$

Soit c_n la composant linéaire pure. Le modèle précédent peut s'écrire

$$y_n = d_n + c_n + \varepsilon_n$$

avec $c_n = \beta_1 y_{n-1}$ et $d_n = \beta_0 + b_1 f\left(a_0 + a_1 y_{n-1}\right)$. Nous remarquons que les valeurs prises par $\hat{d}_n$ appartiennent à l'intervalle[2] $[0.022, 0.029]$ pour ε^* égal à 0.1, et à l'intervalle $[0.016, 0.038]$ pour ε^* égal à 0.00001. Cela veut dire que même si b_1 n'est pas nul pour ε^* égal à 0.00001, la partie considérée comme non linéaire est négligeable. Ceci vient du fait qu'imposer un critère de tolérance plus contraignant (de 1 à 10000) nous amène à modéliser la partie irrégulière du processus, ce qui explique par ailleurs l'élargissement du segment des valeurs prises par $\hat{d}_n$.

4 Conclusion

La conclusion principale est qu'il est préférable d'utiliser un modèle conditionnel dans une perspective d'un test de linéarité. Cependant, des problèmes peuvent se poser si le processus sous-jacent est non linéaire du fait de la modification de la distribution sous-jacente.

Dans une perspective d'une modélisation non linéaire par un réseau neuronal, deux points doivent être vérifiés :

1. Le modèle statistique associé au réseau est-il non linéaire ?
2. Le réseau ne modèlise-t-il pas la partie irrégulière du processus ?

[2] Nous avons $\hat{d}_n = \hat{\beta}_0 + \hat{b}_1 \xi_{1,n}$.

5 Reférences

DAVIDSON, R. et J.G. MACKINNON [1993], Estimation and Inference in Econometrics, Oxford University Press, Oxford

GRANGER, C.W.J. et T. TERÄSVIRTA [1993], Modelling Nonlinear Economic Relations, Oxford University Press, Oxford

HÄRDLE, W.et O. LINTON [1994], Applied nonparametric methods, in Engle et MacFadden (eds.), Handbook of Econometrics, vol. IV, Elsevier Science B.V.

KEENAN, D.M. [1985], A Tukey nonadditivity-type test for time series nonlinearity, *Biometrika*, **72**, 39-44

LEE, T-H., H. WHITE et C.W.J. GRANGER [1993], Testing for neglected nonlinearity in time series models: A comparaison of neural network methods and alternative tests, *Journal of Econometrics*, **56**, 269-290

RAMSEY, J. [1969], Tests for specification errors in classical linear least squares regression analysis, *Journal of the Royal Statistical Society, B*, 350-371

SMITH, M. [1994], Neural Networks for Statistical Modelling, Van Nostrand Reinhold, New York

TSAY, R.[1986], Nonlinearity tests for time series, *Biometrika*, **73**, 461-466