Test pseudo-gaussien du modèle $AR(1)$ contre l' $EXPAR(1)$ en présence d'une nuisance non identifiable sous l'hypothèse nulle

Nabil AZOUAGH [1] & Said El MELHAOUI [2]

[1] *Université Mohammed Premier, Faculté des Sciences, Département Mathématiques et Informatique.60000 BP 717 Oujda Maroc. E-mail: n.azouagh@ump.ac.ma*
[2] *Université Mohammed Premier, Faculté de Droit, Département d'économie, Oujda 60 000 B.P. 724 Maroc. E-mail: safir.time@gmx.fr*

Résumé. Dans ce travail, nous considérons le problème de la détection de l'éventuelle existence d'une composante exponentielle dans les modèles autorégressifs d'ordre 1 $AR(1)$. Ce problème revient à tester une dépendance linéaire $AR(1)$ contre une non linéaire du modèle exponentiel autorégressif $EXPAR(1)$. En utilisant l'approche de Le Cam, nous avons extrait un test pseudo-gaussien localement et asymptotiquement valide pour n'importe quelle densité d'innovation. Cependant, la statistique de test dépend d'une nuisance β non identifiable sous l'hypothèse nulle. Pour répondre à ce problème, nous suggérons de prendre le maximum de la statistique de test sur toute la gamme de la nuisance β, puis utiliser la procédure AR-sieve bootstrap pour approximer sa distribution asymptotique.

Mots-clés. Test de non linéarité, modèle exponentiel autorégressif, Méthodologie de Le Cam, test pseudo-gaussien, AR-sieve bootstrap.

Abstract. In this work, we consider the problem of detecting the eventual existence of an exponential component in autoregressive models of order 1 $AR(1)$. This problem is equivalent to test a linear dependence $AR(1)$ against a nonlinear one of exponential autoregressive model $EXPAR(1)$. Using Le Cam approach, We extracted a locally asymptotically valid pseudo-gaussian test, whatever the underlying innovation density. However, the test statistic depends on a nuisance parameter β not identified under the null hypothesis. To deal with this problem, we suggest to use the Maximum of the test statistic over the whole range of the nuisance β, then we apply the AR-sieve bootstrap procedure to approximate its asymptotic distribution .

Keywords. Nonlinearity test, exponential autoregressive models, Le Cam approach, pseudo-gaussian test, AR-sieve bootstrap.

1 Introduction

Dans le domaine des séries temporelles, plusieurs modèles non linéaires ont été développés dans les dernières décennies, afin de reproduire les caractéristiques non linéaires qui ne sont pas prises en compte par les modèles linéaires autorégressifs usuels $ARMA$ ou VAR. Mais avant d'adopter un modéle non linéaire, il faut s'assurer qu'un tel choix sophistiqué est bien justifié. Ainsi, le problème revient à tester l'hypothèse de la dépendance linéaire contre l'hypothése d'une dépendance non linéaire. Dans ce travail on s'interésse au modèle non linéaire exponentiel autorégressif $EXPAR$ introduit par Haggan et Ozaki (1981), dans le cas de l'ordre 1. Un modèle exponentiel autorégressif d'ordre 1 $EXPAR(1)$ est la solution de l'équation aux différences stochastiques de la forme :

$$X_t = \left(\alpha + \beta \exp(-\varphi X_{t-1}^2)\right) X_{t-1} + \varepsilon_t, \tag{1}$$

avec $(\alpha, \beta, \varphi)'$ est le vecteur des paramètres autorégressifs et de la composante exponentielle $\varphi \geq 0$, et $\{\varepsilon_t;\ t \in \mathbb{Z}\}$ est une suite i.i.d.
On cherche à construire un test pseudo-gaussien du modèle $AR(1)$ contre un $EXPAR(1)$ valide pour n'importe quelle densité d'innovation. Afin d'atteindre cet objectif et en utilisant l'approche de Le Cam, nous avons établi la Normalité Asymptotique Locale [LAN] du modèle $EXPAR(1)$ au voisinage de AR(1). Puis, on a extrait le test pseudo-gaussien localement et asymptotiquement valide, qui atteint l'optimalité sous la densité gaussienne. Toutefois, la statistique de test dépend d'une nuisance β présent uniquement sous l'alternative. Ce problème a été étudié dans plusieurs papiers, Chernoff Et Zacks (1964) pour un test de sup-Lagrange Multiplier [LM], Davies (1977, 1987) pour un test Sup-Likelihood Ratio [LR], d'autre contributions incluent Andrews et Ploberger (1994, 1995), Hansen (1996), Andrews et Cheng (2012, 2013, 2014), Di et Liang (2014)...
Inspiré par Davies (1977), on propose le maximum de la statistique de test pris sur toute la gamme de la nuisance β, on le note par la suite $MaxT$. Cependant, la distribution du maximum de test n'est pas connue en général, c'est pourquoi nous suggérons d'approximer cette distribution par une méthode qui utilise l'algorithme AR-sieve bootstrap proposé par Berg et al (2010).

2 Résultats principaux

2.1 Normalité asymptotique locale

La propriété LAN et la linéarité locale requièrent des conditions de régularité. Ces conditions sont A1 pour la densité d'innovation et A2 pour les paramètres d'autorégression :

- **A1. Conditions sur la densité d'innovation** f

(i) $\forall x \in \mathbb{R},\ f(x) > 0,\ \int x f(x) dx = 0$; et $(m_i)_f = \int x^i f(x) dx < \infty,\ i = 1, \ldots, 6$;

(ii) f absolument continue, dérivable presque partout et sa dérivée est notée $\dot{f}$;

(iii) L'information de Fisher $I(f) = \displaystyle\int \phi_f^2(x) f(x) dx < \infty$, avec $\phi_f = -\dfrac{\dot{f}}{f}$;

- **A2. Conditions sur les paramètres d'autorégression**

  (i) $\mid \alpha + \beta \mid < 1$; (ii) $\beta \neq 0$; (iii) $\varphi \geq 0$.

Supposons que $X_0^{(n)}$ est observée, et soit f la densité de probabilité de ε_t.

Soient les résidus

$$Z_t^{(n)} = X_t^{(n)} - \theta X_{t-1}^{(n)}; \qquad t = 1, \ldots, n, \text{ et } \theta = \alpha + \beta \; .$$

Considérons la suite locale $(\alpha + \frac{\gamma^{(n)}}{\sqrt{n}}, \beta, \frac{\delta^{(n)}}{\sqrt{n}})$, où $\tau^{(n)} = (\gamma^{(n)}, \delta^{(n)})' \in \mathbb{R} * \mathbb{R}^{+}$ est une suite de vecteurs vérifiant: $\sup\limits_{n \in \mathbb{N}} (\tau^{(n)\prime} \tau^{(n)}) < \infty$. On signale que notre problème de test est unilatéral.

On s'intéresse à :

l'hypothèse nulle : $H_f^{(n)}(\alpha, \beta, 0)$, contre l'alternative: $H_f^{(n)}(\alpha + \frac{\gamma^{(n)}}{\sqrt{n}}, \beta, \frac{\delta^{(n)}}{\sqrt{n}})$

Notons $\Lambda_f^{(n)}(\theta, \beta, \tau^{(n)})$ le logarithme du rapport de vraisemblance (conditionnelle à $X_0^{(n)}$) de $H_f^{(n)}(\alpha + \frac{\gamma^{(n)}}{\sqrt{n}}, \beta, \frac{\delta^{(n)}}{\sqrt{n}})$; par rapport à $H_f^{(n)}(\alpha, \beta, 0)$:
$\Lambda_f^{(n)}(\theta, \beta, \tau^{(n)}) = \sum_{t=1}^{n} \log \left[f \left(Z_t - (\frac{\gamma^{(n)}}{\sqrt{n}} - \beta + \beta \exp[-\frac{\delta^{(n)}}{\sqrt{n}} X_{t-1}^2] X_{t-1} \right) \right] - \log[f(Z_t)] + o_p(1).$

Ainsi on a le résultat de la normalité locale asymptotique suivante.

Proposition 2.1.
Supposons que les conditions A1 et A2 sont satisfaites. Alors pour toute suite $\tau^{(n)}$ vérifiant $\sup\limits_{n \in \mathbb{N}} (\tau^{(n)\prime} \tau^{(n)}) < \infty$, sous $H_f^{(n)}(\alpha, \beta, 0)$, quand $n \to \infty$, on a :

(i) $\Lambda_f^{(n)}(\theta, \beta, \tau^{(n)}) = \tau^{(n)\prime} \Delta_f^{(n)}(\theta, \beta) - \frac{1}{2} \tau^{(n)\prime} \Gamma_f(\theta, \beta) \tau^{(n)} + o_p(1)$,

avec,

$$\Delta_f^{(n)}(\theta, \beta) = \frac{1}{\sqrt{n}} \begin{pmatrix} \displaystyle\sum_{t=1}^{n} X_{t-1} \phi_f(Z_t) \\ \displaystyle\sum_{t=1}^{n} -\beta X_{t-1}^3 \phi_f(Z_t) \end{pmatrix},$$

et

$$\Gamma_f(\theta,\beta) = I(f)\begin{pmatrix} E(X_0^2) & -\beta E(X_0^4) \\ -\beta E(X_0^4) & \beta^2 E(X_0^6) \end{pmatrix};$$

(ii) $\Delta_f^{(n)}(\theta,\beta) \xrightarrow{\mathcal{L}} N\big(0, \Gamma_f(\theta,\beta)\big)$.

2.2 Extraction des tests pseudo-gaussiens

Le modèle présent jouit de la propriété de normalité locale asymptotique (LAN), ce qui nous amène à construire des tests localement et asymptotiquement optimaux sous une densité d'innovation spécifiée, et pseudo-gaussien sous une densité f quelconque. Ainsi d'après Le Cam (1986), on déduit la statistique de test de l'hypothèse nulle $H_f^{(n)}(\alpha,\beta,0)$, contre l'alternative: $H_f^{(n)}(\alpha + \frac{\gamma^{(n)}}{\sqrt{n}}, \beta, \frac{\delta^{(n)}}{\sqrt{n}})$:

$$T\left(\Delta_f^{(n)}(\theta,\beta)\right) = \Delta_f^{(n)\prime}(\theta,\beta)\frac{\Gamma_f^{-1}(\theta,\beta)e_1}{\sqrt{e_1'\Gamma_f^{-1}(\theta,\beta)e_1}} \ ; \ e_1 = (0,1)'$$

La statistique de test dépend de deux nuisances β et θ, on remarque que θ figure sous l'hypothèse nulle contrairement à β. Supposons pour le moment que β est connu, ainsi il ne nous reste plus qu'à estimer θ. Afin de contrôler les effets de la substitution d'une valeur estimée $\hat{\theta}^{(n)}$ à la place de la valeur exacte θ dans la suite centrale $\Delta_f^{(n)}(\theta,\beta)$, on a établit la linéarité locale asymptotique. Puis on s'est penché sur les propriétés de notre test, présentées dans la proposition 2.2.

Proposition 2.2.
Supposons que les conditions A1 et A2 sont satisfaites. Considérons la suite des tests $\hat{\phi}^{(n)}$ qui rejettent l'hypothèse nulle si la statistique $\hat{T}^{(n)}$ dépasse le $(1-\nu)$ quantile z_ν de la loi normale centré réduite, alors :

(i) les tests $\hat{\phi}^{(n)}$ sont de niveau ν;

(ii) les tests $\hat{\phi}^{(n)}$ sont localement et asymptotiquement les plus puissants, au niveau de probabilité ν, contre l'alternative gaussienne de la forme :

$$\bigcup_{\sigma>0}\bigcup_{\varphi>0} H_{\Phi_\sigma}^{(n)}(\alpha,\beta,\varphi),$$

avec Φ_σ est la fonction de répartition gaussienne centré;

(iii) sous $H_f^{(n)}(\alpha,\beta,\frac{\delta}{\sqrt{n}})$, les tests $\hat{\phi}^{(n)}$ sont de puissance asymptotique, lorsque $n\to\infty$, $1-\Phi(z_\nu - \rho(\theta,\beta,\delta,f))$,
où $\quad \rho(\theta,\beta,\varphi,f) = \left(\frac{\Gamma_f^{-1}(\theta,\beta)e_1}{\sqrt{e_1'\Gamma_f^{-1}(\theta,\beta)e_1}}\right)'\Gamma_f(\theta,\beta)e_1\delta.$

3 Test pseudo-gaussien quand le paramètre de nuisance est présent uniquement sous l'alternative

Dans cette section nous supposons que β est inconnu, donc il doit être estimé. Cependant, la nuisance β n'est pas identifiable sous l'hypothèse nulle, elle ne se manifeste que sous l'alternative. Face à ce problème, on suggère de considérer une statistique de test transformée, en prenant le maximum de la statistique de test sur toute la gamme de la nuisance β, on le note $MaxT$. Étant donnée que la distribution asymptotique de notre nouveau test est en général inconnue, on propose donc d'approximer cette distribution en utilisant l'algorithme AR-sieve bootstrap. Les étapes et la preuve de la consistance asymptotique de cette procédure (à la fois sous l'hypothèse nulle et l'alternative) peuvent être trouvées dans l'étude de Berg et al (2010). Cet algorithme génère un échantillon bootstrap noté $X^* = (X_1^*, \dots, X_n^*)$ à partir des observations réelles $X = (X_1, \dots, X_n)$, l'idée est de calculer $MaxT$ pour l'échantillon bootstrap X^*, la procédure est répétée pour un grand nombre de fois (par exemple B). En conséquence, on dispose de B bootstrap pseudo-statistiques $MaxT_1^*, \dots, MaxT_B^*$. La distribution empirique des B bootstrap pseudo-statistiques peut être ensuite utilisée pour approximer la distribution réelle de $MaxT$ sous l'hypothèse nulle, de sorte que le test devienne réalisable. Une étude de simulation du niveau et de la puissance empirique de ce test a été réalisée en utilisant l'ensemble de cette procédure. Nous présentons ci dessous un aperçu de quelques résultats de cette étude.

3.1 Résultats des simulations

Soit le modèle AR(1) avec $|\theta| < 1$ condition de régularité (causalité du modèle sous H_0) :

$$X_t = \theta X_{t-1} + \varepsilon_t. \tag{2}$$

Soit le modèle $EXPAR(1)$:

$$X_t = \left(-2 + 2.5 \exp(-0.1 X_{t-1}^2)\right) X_{t-1} + \varepsilon_t. \tag{3}$$

. Le tableau 1 donnes les niveaux empiriques des tests pour les données générées par le modèle (2) pour différents paramètres θ au niveau $\nu = 0.05$, on a simulé 1000 pseudo échantillons de taille 100, la densité d'innovation utilisée est gaussienne centré réduite.

θ	Niveau empirique
0.3	0.047
0.5	0.05
0.7	0.052

Table 1. Niveau empirique sous une dépendance $AR(1)$.

. Le tableau 2 donne les puissances empiriques des tests pour les données générées par le modèle (3) au niveau $\nu = 0.05$, on a simulé 1000 pseudo échantillons, la densité d'innovation utilisée est gaussienne centré réduite.

taille de l'échantillon	Puissance empirique
30	0.151
100	0.498

Table 2. Puissance empirique sous une dépendance $EXPAR(1)$.

4 Conclusion

En utilisant l'approche de Le Cam, la propriété (LAN) des modèles $EXPAR(1)$ au voisinage de $AR(1)$ a été établie. Ensuite, un test pseudo-gaussien est proposé pour la détection de l'éventuelle existence d'une composante exponentielle dans les modèles $AR(1)$. Ce test est localement et asymptotiquement valide pour n'importe quelle densité d'innovation (vérifiant les conditions de régularité) et optimal pour les densités gaussiennes lorsque la nuisance β est supposée connue. Lorsque cette dernière est supposée inconnue, nous avons montré qu'une alternative est possible. En effet, nous avons pris le maximum de la statistique de test sur toute la gamme de la nuisance β, puis, utilisé l'algorithme AR-sieve bootstrap pour approximer la distribution asymptotique de cette nouvelle statistique $MaxT$ sous l'hypothèse nulle. Finalement, nous avons vérifié les bonnes propriétés de test $MaxT$ par des simulations sur des échantillons réels de tailles finies.

Bibliographie

[1] A. Akharif, and M. Hallin, (2003), Efficient detection of random coefficients in autoregressive models, The Annals of Statistics, Series 31(2), 675-704.

[2] A. R. Swensen, (1985), The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend, Journal of Multivariate Analysis, Series 16(1), 54-70.

[3] Berg, A., Paparoditis, E., & Politis, D. N. (2010). A bootstrap test for time series linearity. Journal of Statistical Planning and Inference, 140(12), 3841-3857.

[4] Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative, Biometrika, 64(2), 247-254.

[5] L. Le Cam, (1986),*Asymptotic methods in statistical decision theory*, Springer-verlag, New York.

[6] Tong, H. (1990). *Non-linear time series: a dynamical system approach*, Oxford University Press, Oxford.

[7] V. Haggan, and T. Ozaki, (1981), Modelling nonlinear random vibrations using an amplitude-dependent autoregressive time series model, Biometrika, Series 68(1), 189-196.