Test pseudo-gaussien optimal pour les modèles de régression à coefficient aléatoire

Mohamed FIHRI[1] & Abdelhadi AKHARIF[2]

Laboratoire de Mathématiques et Applications, Faculté des Sciences et Techniques de Tanger, ancienne Route de l'Aéroport, Km 10, Ziaten. BP : 416. Tanger - Maroc.
[1]*m.fihri@gmail.com* & [2]*aakharif@gmail.com*

Résumé. On s'intéresse au problème de détection du caractère aléatoire dans un modèle de régression. Il s'agit du problème de tester une régression classique contre une alternative d'une régression à coefficient aléatoire(*RCR model*) . Nous proposons une procédure de test paramétrique localement et asymptotiquement optimale au sens de Le Cam en donnant les hypothèses de base nécessaires. Une version non standard de la normalité locale asymptotique (LAN) pour un *RCR model* au voisinage d'une régression classique est établie par rapport à la constante μ, le coefficient de régression β, le paramètre d'échelle σ^2 et le paramètre d'intérêt σ_b^2, avec une densité fixée f_1. Nous construisons le test localement et asymptotiquement optimal et nous traitons le cas particulier de test pseudo-gaussien. On signale que notre problème de test est unilatéral et que l'estimation de la variance des résidus ne peut pas se faire sans perdre d'efficacité.

Mots-clés. LAN, optimal test, pseudo Gaussian test, *RCR model*, linéarité locale asymptotique.

Abstract. We consider the problem of detection of randomness in a regression model. This is the problem of testing a classical regression model against an alternative with random coefficient regression model (*RCR model*). We propose a parametric test procedure locally and asymptotically optimal in the Le Cam sense with giving necessary assumptions. A non-standard version of the local asymptotic normality (LAN) for the *RCR model* in the neighbourhood of the classical regression is established with respect to constant, regression coefficient, scale parameter σ^2, and the parameter of interest σ_b^2, for fixed density f_1. Locally asymptotically optimal test and the particular case of pseudo Gaussian test are treated. Note that our test problem is unilateral and the estimation of the variance of the residual cannot be done without losing efficiency.

Keywords. LAN, test optimal, test pseudo-gaussien, *RCR model*, local asymptotic linearity.

1 Introduction

Un modèle de régression linéaire simple est un modèle de régression où la variable dépendante (Y) est continue, s'explique par une seule variable exogène (X), et linéaire par rapport aux paramètres μ et β :

$$y_i = \mu + \beta x_i + \varepsilon_i, \qquad i = 1, \ldots, n \tag{1}$$

Une version générale de ce modèle est le modèle de régression à coefficient aléatoire :

$$y_i = \mu + (\beta + b_i)x_i + \varepsilon_i, \qquad i = 1, \ldots, n \tag{2}$$

Avec

- i désigne l'individu, y_i est la réponse observée pour l'individu i, x_i est la variable explicative exogène (l'observation),
- μ et β sont des paramètres inconnus de régression,
- $(\varepsilon_i)_{1\leq i\leq n}$ est une suite de variables aléatoires indépendantes et identiquement distribuées i.i.d. avec $\mathbb{E}(\varepsilon_i) = 0$, $0 < \mathbb{E}(\varepsilon_i^2) = \sigma^2 < \infty$ et de densité $\varepsilon \mapsto f(\varepsilon) := (1/\sigma)f_1(\varepsilon/\sigma)$,
- $(b_i)_{1\leq i\leq n}$ est une suite de variables aléatoires i.i.d. $(0, \sigma_b^2)$ de densité $b \mapsto h(b) := (1/\sigma_b)h_1(b/\sigma_b)$, et indépendante de $(\varepsilon_i)_{1\leq i\leq n}$.

Parmis les travaux qui se sont intéressés à cette problématique, on cite : Newbold et Bos (1985) qui donnent le test du multiplicateur de Lagrange, Ramanathan et Rajarshi (1994) qui proposent le test de rang signé basé sur les scores de Wilcoxon, Gumpertz et Pantula (1989) qui établissent une version simple d'estimateur généralisé des moindres carrés. Et en se basant sur les travaux de Akharif et Hallin (2003) et Bennala et *al* (2012), nous proposons une procédure de test paramétrique localement et asymptotiquement optimale au sens de Le Cam en donnant les hypothèses de base nécessaires.

2 Plan de la présentation

Nous montrons que le modèle présenté jouit de la propriété de normalité locale asymptotique (LAN), ce qui nous amène à construire des tests localement et asymptotiquement optimal sous une densité f_1 spécifiée, et pseudo gaussien sous une densité g quelconque (plus largement des tests basés sur les rangs au sens de Hájek).

On établit, pour ce modèle, une propriété de normalité locale asymptotique uniforme par rapport à $(\mu, \beta, \sigma^2, \sigma_b^2)'$ en $\vartheta := (\mu, \beta, \sigma^2, 0)'$, avec une densité fixée f_1 de la famille des distributions :

$$\mathcal{P}^{(n)}_{f_1,h_1} := \left\{ \mathbb{P}^{(n)}_{\mu,\beta,\sigma^2,\sigma_b^2;f_1,h_1} : (\mu, \beta)' \in \mathbb{R}^2,\ \sigma^2 > 0 \text{ and } \sigma_b^2 \geq 0 \right\}$$

Nous considérons une sequence d'hypothèses alternatives locales de forme $\vartheta+n^{-1/2}\nu^{(n)}\tau^{(n)}$ avec $\tau^{(n)} := \left(\tau_1^{(n)}, \tau_2^{(n)}, \tau_3^{(n)}, \tau_4^{(n)}\right)' \in \mathbb{R}^3 \times \mathbb{R}^+$ est tel que : $\sup_n \left(\tau_i^{(n)}\right)^2 < \infty, \forall i = 1, 2, 3, 4$.

Pour tout $\vartheta^{(n)} := (\mu^{(n)}, \beta^{(n)}, \sigma^{2(n)}, 0)'$ tel que $\mu^{(n)}-\mu = O\left(n^{-1/2}\right)$, $\left(K_1^{(n)}\right)^{-1}\left(\beta^{(n)} - \beta\right) = O\left(n^{-1/2}\right)$ et $\sigma^{2(n)}-\sigma^2 = O\left(n^{-1/2}\right)$ et pour toute sequence bornée $\tau^{(n)} := \left(\tau_1^{(n)}, \tau_2^{(n)}, \tau_3^{(n)}, \tau_4^{(n)}\right)' \in \mathbb{R}^3 \times \mathbb{R}^+$,on a, sous $\mathbb{P}^{(n)}_{\vartheta^{(n)};f_1}$, quand $n \to \infty$,

$$\begin{aligned}\Lambda^{(n)}_{\vartheta^{(n)}+n^{-1/2}\nu^{(n)}\tau^{(n)}/\vartheta^{(n)};f_1,h_1} :=& \log\left(\frac{d\mathbb{P}^{(n)}_{\vartheta^{(n)}+n^{-1/2}\nu^{(n)}\tau^{(n)};f_1,h_1}}{d\mathbb{P}^{(n)}_{\vartheta^{(n)};f_1}}\right)\\ =&\tau^{(n)\prime}\Delta^{(n)}_{f_1}(\vartheta^{(n)}) - \frac{1}{2}\tau^{(n)\prime}\Gamma_{f_1}(\vartheta)\tau^{(n)} + o_{\mathbb{P}}(1)\end{aligned}$$

avec $\Gamma_{f_1}(\vartheta) := \left(\Gamma_{f_1;ij}(\vartheta)\right)_{1\le i,j\le 4}$ est la matrice variance-covariance et qui prend la forme :

$$\Gamma_{f_1}(\vartheta) = \begin{pmatrix}\Gamma_{f_1;11}(\vartheta) & 0 & \Gamma_{f_1;13}(\vartheta) & \Gamma_{f_1;14}(\vartheta)\\ 0 & \Gamma_{f_1;22}(\vartheta) & 0 & \Gamma_{f_1;24}(\vartheta)\\ \Gamma_{f_1;13}(\vartheta) & 0 & \Gamma_{f_1;33}(\vartheta) & \Gamma_{f_1;34}(\vartheta)\\ \Gamma_{f_1;14}(\vartheta) & \Gamma_{f_1;24}(\vartheta) & \Gamma_{f_1;34}(\vartheta) & \Gamma_{f_1;44}(\vartheta)\end{pmatrix}$$

Et la suite centrale $\Delta^{(n)}_{f_1}(\vartheta^{(n)}) := \left(\Delta^{(n)}_{f_1;1}(\vartheta^{(n)}), \Delta^{(n)}_{f_1;2}(\vartheta^{(n)}), \Delta^{(n)}_{f_1;3}(\vartheta^{(n)}), \Delta^{(n)}_{f_1;4}(\vartheta^{(n)})\right)'$ suit asymptotiquement une loi normale de moyenne nulle et de matrice variance-covariance $\Gamma_{f_1}(\vartheta)$.

On remarquera que h_1 sert seulement à la définition de la vraisemblance et que la forme non-diagonale de la matrice $\Gamma_{f_1}(\vartheta)$ entraine la non orthogonalité entre μ, β, σ^2 et σ_b^2 mutuellement.

Partant du fait que le logarithme du rapport de vraisemblance local en ϑ est asymptotiquement équivalent au logarithme du rapport de vraisemblance dans un modèle de position gaussien classique (famille de lois $\mathcal{N}\left(\Gamma_{f_1}(\vartheta)\tau, \Gamma_{f_1}(\vartheta)\right)$) et du troisième lemme de Le Cam, on construit le test localement et asymptotiquement optimal le plus puissant basé sur le résidu de la régression de $\Delta^{(n)}_{f_1;4}(\vartheta^{(n)})$ par rapport à $\left(\Delta^{(n)}_{f_1;1}(\vartheta^{(n)}), \Delta^{(n)}_{f_1;2}(\vartheta^{(n)}), \Delta^{(n)}_{f_1;3}(\vartheta^{(n)})\right)'$ et prend la forme suivante :

$$\begin{aligned}\Delta^{*(n)}_{f_1;4}(\vartheta) :=&\Delta^{(n)}_{f_1;4}(\vartheta) - \left(\Gamma_{f_1;14}(\vartheta), \Gamma_{f_1;24}(\vartheta), \Gamma_{f_1;34}(\vartheta)\right)\\ &\times\begin{pmatrix}\Gamma_{f_1;11}(\vartheta) & 0 & \Gamma_{f_1;13}(\vartheta)\\ 0 & \Gamma_{f_1;22}(\vartheta) & 0\\ \Gamma_{f_1;13}(\vartheta) & 0 & \Gamma_{f_1;33}(\vartheta)\end{pmatrix}^{-1}\begin{pmatrix}\Delta^{(n)}_{f_1;1}(\vartheta)\\ \Delta^{(n)}_{f_1;2}(\vartheta)\\ \Delta^{(n)}_{f_1;3}(\vartheta)\end{pmatrix}\\ =&\Delta^{(n)}_{f_1;4}(\vartheta) - \Gamma^*_{f_1;1}(\vartheta)\Delta^{(n)}_{f_1;1}(\vartheta) - \Gamma^*_{f_1;2}(\vartheta)\Delta^{(n)}_{f_1;2}(\vartheta) - \Gamma^*_{f_1;3}(\vartheta)\Delta^{(n)}_{f_1;3}(\vartheta)\end{aligned}$$

qui suit asymptotiquement, sous $\mathbb{P}^{(n)}_{\vartheta^{(n)};f_1}$, une loi normal de moyenne zero et de variance

$$
\begin{aligned}
\Gamma^{*(n)}_{f_1;44}(\vartheta) :=& \Gamma^{(n)}_{f_1;44}(\vartheta) - (\Gamma_{f_1;14}(\vartheta), \Gamma_{f_1;24}(\vartheta), \Gamma_{f_1;34}(\vartheta)) \\
& \times \begin{pmatrix} \Gamma_{f_1;11}(\vartheta) & 0 & \Gamma_{f_1;13}(\vartheta) \\ 0 & \Gamma_{f_1;22}(\vartheta) & 0 \\ \Gamma_{f_1;13}(\vartheta) & 0 & \Gamma_{f_1;33}(\vartheta) \end{pmatrix}^{-1} \begin{pmatrix} \Gamma^{(n)}_{f_1;14}(\vartheta) \\ \Gamma^{(n)}_{f_1;24}(\vartheta) \\ \Gamma^{(n)}_{f_1;34}(\vartheta) \end{pmatrix} \\
=& \Gamma^{(n)}_{f_1;44}(\vartheta) - \Gamma^{*}_{f_1;1}(\vartheta)\Gamma^{(n)}_{f_1;14}(\vartheta) - \Gamma^{*}_{f_1;2}(\vartheta)\Gamma^{(n)}_{f_1;24}(\vartheta) - \Gamma^{*}_{f_1;3}(\vartheta)\Gamma^{(n)}_{f_1;34}(\vartheta)
\end{aligned}
$$

avec

$$
\begin{aligned}
\Gamma^{*}_{f_1;1}(\vartheta) :=& \frac{\Gamma_{f_1;14}(\vartheta)\Gamma_{f_1;33}(\vartheta) - \Gamma_{f_1;34}(\vartheta)\Gamma_{f_1;13}(\vartheta)}{\Gamma_{f_1;11}(\vartheta)\Gamma_{f_1;33}(\vartheta) - \Gamma^{2}_{f_1;13}(\vartheta)}, \\
\Gamma^{*}_{f_1;2}(\vartheta) :=& \frac{\Gamma_{f_1;24}(\vartheta)}{\Gamma_{f_1;22}(\vartheta)} \text{ et} \\
\Gamma^{*}_{f_1;3}(\vartheta) :=& \frac{\Gamma_{f_1;11}(\vartheta)\Gamma_{f_1;34}(\vartheta) - \Gamma_{f_1;13}(\vartheta)\Gamma_{f_1;14}(\vartheta)}{\Gamma_{f_1;11}(\vartheta)\Gamma_{f_1;33}(\vartheta) - \Gamma^{2}_{f_1;13}(\vartheta)}.
\end{aligned}
$$

Le test qui rejette l'hypothèse nulle si :

$$
T^{*(n)}_{f_1}(\vartheta) := \left(\Gamma^{*}_{f_1;44}(\vartheta)\right)^{-1/2} \Delta^{*(n)}_{f_1;4}(\vartheta) > z_{1-\alpha}
$$

est asymptotiquement le plus puissant au niveau α. La linéarité locale asymptotique nous permet de remplacer le paramètre inconnu ϑ par un estimateur $\sqrt{n}$-convergent $(\widehat{\vartheta})$. La statistique de test $T^{*(n)}_{f_1}(\widehat{\vartheta})$ est asymptotiquement normale de moyenne nulle et de variance un sous l'hypothèse nulle $\mathbb{P}^{(n)}_{\vartheta^{(n)};f_1}$; de moyenne $\left(\Gamma^{*(n)}_{f_1;44}(\vartheta)\right)^{1/2} \tau_4$ et de variance un sous sous la contre hypothèse locale $\mathbb{P}^{(n)}_{\vartheta^{(n)}+n^{-1/2}\nu^{(n)}\tau^{(n)};f_1,h_1}$.

Ensuite, on propose un test pseudo gaussien asymptotiquement le plus puissant qui est valide sous une densité g quelconque. Ce test correspond au cas où $f_1(z) = \phi_1(z) := \sqrt{a/2}\exp(-az^2/2)$ tel que $\mathbb{V}ar(Z) = 1/a$, $a \approx 0.4549$. On donne ci-dessous la version gaussienne de la suite centrale, sa matrice variance-covariance et la statistique de test :

$$
\Delta^{(n)}_{\mathcal{N}}(\vartheta) = \begin{pmatrix} \frac{a}{\sigma\sqrt{n}} \sum_{i=1}^{n} Z_i \\ \frac{a}{\sigma\sqrt{n}} \sum_{i=1}^{n} Z_i K^{(n)}_1 x_i \\ \frac{a}{2\sigma^2\sqrt{n}} \sum_{i=1}^{n} (Z_i^2 - 1/a) \\ \frac{a^2}{\sigma^2\sqrt{n}} \sum_{i=1}^{n} (Z_i^2 - 1/a) K^{(n)}_2 x_i^2 \end{pmatrix}
$$

$$\Gamma_{\mathcal{N}}(\vartheta) = \begin{pmatrix} \frac{a}{\sigma^2} & 0 & 0 & 0 \\ 0 & \frac{a}{n\sigma^2} & 0 & 0 \\ 0 & 0 & \frac{1}{2\sigma^4} & \frac{a}{n\sigma^4}\frac{K_2^{(n)}}{{K_1^{(n)}}^2} \\ 0 & 0 & \frac{a}{n\sigma^4}\frac{K_2^{(n)}}{{K_1^{(n)}}^2} & \frac{2a^2}{n\sigma^4} \end{pmatrix}$$

$$T^{*(n)}_{\mathcal{N};4}(\vartheta) = \frac{a{K_1^{(n)}}^2 K_2^{(n)}}{\sqrt{2\left({K_1^{(n)}}^4 - \frac{1}{n}{K_2^{(n)}}^2\right)}} \sum_{i=1}^{n} Z_i^2 \left(x_i^2 - \overline{x^2}\right)$$

avec $K_1^{(n)} = \left(\sum_{i=1}^n x_i^2\right)^{-1/2}$ et $K_2^{(n)} = \left(\sum_{i=1}^n x_i^4\right)^{-1/2}$.

On signale que l'estimation de la variance des résidus dans ce cas ne peut pas se faire sans perdre d'efficacité.

Bibliographie

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

[2] Bennala, N., Hallin, M., Paindaveine, D. (2012) *Pseudo-Gaussian and rank-based optimal tests for random individual effects in large n small T panels*, Journal of Econometrics **170**, p. 50-67.

[3] Beran, R. (1993) *semiparametric random coefficient regression models*, Ann. Inst. Statist. Math **45**, p. 639-654.

[4] Droesbeke, J.J., Jeanne, J. (1996) *Inférence non paramétrique : Les statistiques de rangs*, Association pour la statistique et ses utilisations.

[5] Gumpertz, M., Pantula, S.G. (1989) *A Simple Approach to Inference in Random Coefficient Models*, The American Statistician **43**, p. 203-210.

[6] Hallin, M., Werker, B. (1999) *Optimal testing for semiparametric AR models : From Gaussian Lagrange multipliers to autoregression rank scores and adaptive tests. In*, Asymptotics, Nonparametrics and Time Series (S. Ghosh, ed.) 295-350. Dekker, New York.

[7] Le Cam, L.M. (1986) *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New York.

[8] Newbold, P., Bos, T. (1986) *Stochastic Stochastic parameter regression models*, Sage University Paper Series on Quantitative Applications in the Social Sciences **51**.

[9] Ramanathan, T.V., Rajarshi, M.B. (1994) *Efficient detection of random coefficients in autoregressive models*, Statist. Probab. Lett. **21**, p. 115-120.

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

[11] van der Vaart, A.W. (1998) *Asymptotic Statistic*, Cambridge University Press, Cambridge.