C. R. Acad. Sci. Paris, Ser. I 356 (2018) 571–574

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Statistique

Test d'hétéroscédasticité quand les covariables sont fonctionnelles

Heteroscedasticity test when the covariables are functionals

Aicha Henien [a], Larbi Ait-Hennani [b], Jacques Demongeot [d], Ali Laksaci [c], Mustapha Rachdi [d]

[a] *Laboratoire de statistique et processus stochastiques, Université Djillali-Liabès, BP 89, Sidi Bel-Abbès 22000, Algeria*
[b] *Université Lille-2, Droit et Santé, IUT C, Roubaix, France*
[c] *Department of Mathematics, College of Science, King Khalid University, Abha, Saudi Arabia*
[d] *Université Grenoble Alpes, laboratoire AGEIS EA 7407, France*

INFO ARTICLE

Historique de l'article :
Reçu le 6 février 2017
Accepté le 23 février 2018
Disponible sur Internet le 5 mars 2018

Présenté par Paul Deheuvels

RÉSUMÉ

Dans cette note, nous construisons et étudions un test non paramétrique de détection de l'hétéroscédasticité quand les covariables sont fonctionnelles. Ce dernier est construit en évaluant la différence entre la variance conditionnelle et la variance inconditionnelle. Nous montrons la normalité asymptotique de cette statistique de test sous l'hypothèse nulle. En outre, nous prouvons que ce test est également robuste contre toutes les déviations possibles à l'homoscédasticité.

ABSTRACT

We present in this paper a consistent nonparametric test for heteroscedasticity when data are of functional kind. The latter is constructed by evaluating the difference between the conditional and unconditional variances. We show the asymptotic normality of the statistical test under the null hypothesis. In addition, we prove that this test is consistent against all deviations from homoscedasticity condition.

1. Introduction

L'inférence statistique pour l'analyse des données fonctionnelles (FDA) a été le centre de plusieurs études (voir, par exemple, Ramsay et Silverman [6] et Ferraty et Vieu [3] pour des références de base et Hsing et Eubank [5] ou Goia et Vieu [4] pour des avancées récentes). Dans ce contexte, la prédiction d'une réponse scalaire Y étant donné une variable aléatoire

Adresses e-mail : aichahenien@yahoo.com (A. Henien), larbi.aithennani@univ-lille2.fr (L. Ait-Hennani), jacques.demongeot@yahoo.fr (J. Demongeot), alilak@yahoo.fr (A. Laksaci), mustapha.rachdi@univ-grenoble-alpes.fr (M. Rachdi).

https://doi.org/10.1016/j.crma.2018.02.010
1631-073X/

fonctionnelle X, est une question très importante. Le modèle de régression est le plus utilisé pour résoudre ce genre de problème. Cependant, ce modèle n'est pas efficace dans le cas où les données sont héteroscédastiques. Il est donc nécessaire de vérifier l'homoscédasticité des données afin de préciser le modèle qui convient. En dimension finie, où les données peuvent être visualisées, ce phénomène est facilement détectable. Mais ceci n'est point possible lorsque la variable explicative est fonctionnelle (de dimension éventuellement finie), car il n'est pas possible de faire des représentations graphiques montrant le lien entre celle-ci et la réponse scalaire. Malgré l'importance de l'hétéroscédasticité des résidus, beaucoup moins d'attention lui a été portée dans le cadre de la FDA, et même, à notre connaissance, ce problème n'a pas été traité jusqu'à présent. Cependant, la littérature sur le sujet est très abondante dans le cas vectoriel (en dimension finie). Nous citons, à titre d'exemple, Breusch et Pagan [1] Dette et al. [2] pour les anciens travaux et Zheng [7] pour des références assez récentes.

Dans cette note, nous proposons un test non paramétrique consistant pour la détection de l'hétéroscédasticité. Ce test est basé sur l'estimation à noyau de l'opérateur de régression. Ensuite, nous établissons la normalité asymptotique de la statistique de test construite sous l'hypothèse nulle (voir (5))et nous montrons que nous pouvons détecter des variantes locales distinctes de zéro. Mentionnons que les résultats de cette note sont obtenus sous des conditions standards en FDA non paramétrique et permettent d'éviter le fléau de la dimension du cas multivarié.

Dans la section 2, nous présentons le modèle fonctionnel. Nous construisons la statistique de test dans la section 3. Enfin, dans la section 4, nous exposons les résultats sur le comportement asymptotique de cette statistique de test. Notons qu'une étude sur des données simulées a été réalisée et a montré l'efficacité du test proposé : celle-ci peut être obtenue sur simple demande.

2. Position du problème

Soit (X_i, Y_i) pour $i = 1, \ldots, n$, des couples de variables aléatoires indépendantes et identiquement distribuées comme $(X, Y) \in \mathcal{F} \times \mathbb{R}$, où $\mathcal{F}$ est un espace semi-métrique. Dans ce qui suit, nous noterons d une semi-métrique sur l'espace $\mathcal{F}$, x est une courbe fixée dans $\mathcal{F}$, N_x est un voisinage fixé de x et on note $B(x, \alpha) = \{y \in \mathcal{F} \text{ tel que } d(y, x) \leq \alpha\}$ la boule fermée de centre x et de rayon α. De plus, nous supposons que les variables aléatoires X et Y sont reliées par la relation :

$$Y = r(X) + \varepsilon, \tag{1}$$

où r est un opérateur défini de $\mathcal{F}$ vers $\mathbb{R}$ et ε est une variable aléatoire d'erreur, telle que : $\mathbb{E}[\varepsilon|X] = 0$.

Dans ce qui suit, nous supposons que l'opérateur r satisfait la condition suivante :

$$\text{pour tout } x_1, x_2 \in N_x,\ |r(x_1) - r(x_2)| \leq C d^{\beta}(x_1, x_2),\ \text{où } C \text{ et } \beta > 0. \tag{2}$$

Dans le but de tester l'hypothèse d'hétéroscédasticité de ce modèle, nous supposons qu'il existe deux fonctions, ϕ et f, positives et continues et telles que : $\mathbb{P}(X \in B(x, \alpha)) = \phi(\alpha).f(x) + o(\phi(\alpha))$. De plus, nous supposons que la fonction ϕ a une dérivée bornée au voisinage de 0 et que :

$$\text{pour tout } s \in [0, 1],\ \lim_{\alpha \to 0} \frac{\phi(s\alpha)}{\phi(\alpha)} = \tau(s). \tag{3}$$

Concernant la fonction f, nous supposons que la ε-entropy de Kolmogorov[1] $\psi_{\mathcal{S}}$ du support $\mathcal{S}$ de f est telle que :

$$\sum_{n=1}^{\infty} \exp\left\{(1-\eta)\psi_{\mathcal{S}}\left(\frac{\log n}{n}\right)\right\} < \infty,\ \text{pour } \eta > 1. \tag{4}$$

3. Construction de la statistique de test

Nous souhaitons tester l'hétéroscédasticité du modèle (1). Typiquement, nous testons :

$$\mathcal{H}_0 :\ \text{var}[\varepsilon|X] = \sigma^2 \quad \text{versus} \quad \mathcal{H}_1 :\ var[\varepsilon|X] \neq \sigma^2. \tag{5}$$

Dans le but de construire une statistique de test, nous supposons que la fonction f est telle que :

$$f(X) > 0 \text{ presque sûrement, et } \mathbb{E}[f(X)] < \infty. \tag{6}$$

Notons que l'hypothèse (6) permet de montrer que l'hypothèse $\mathcal{H}_0$ est équivalente à :

$$S_1 := \mathbb{E}[(\varepsilon^2 - \sigma^2)\mathbb{E}[(\varepsilon^2 - \sigma^2)|X]f(X)] = \mathbb{E}[\mathbb{E}^2[(\varepsilon^2 - \sigma^2)|X]f(X)] = 0,$$

[1] Soit $\varepsilon > 0$. Un ensemble fini de points $x_1, x_2, \ldots, x_N$ dans $\mathcal{F}$ est dit un ε-net pour $\mathcal{S}$ si $\mathcal{S} \subset \bigcup_{k=1}^{N} B(x_k, \varepsilon)$. La quantité $\psi_{\mathcal{S}}(\varepsilon) = \log(N_\varepsilon(\mathcal{S}))$, où $N_\varepsilon(\mathcal{S})$ est le nombre minimal de boules ouvertes dans $\mathcal{F}$ de rayon ε qui est nécessaire pour couvrir $\mathcal{S}$, et est appelé la ε-entropie de Kolmogorov de l'ensemble $\mathcal{S}$.

tandis que l'hypothèse $\mathcal{H}_1$ est équivalente à $S_1 > 0$. Donc, notre problème de test (voir (5), est équivalent au problème suivant :

$$\mathcal{H}_0 : S_1 = 0 \quad \text{versus} \quad \mathcal{H}_1 : S_1 > 0.$$

L'estimateur naturel de S_1 est :

$$W_n = \frac{1}{n(n-1)\phi(h)} \sum_{i=1}^{n} \sum_{j\neq i,=1}^{n} (\widehat{\varepsilon_i}^2 - \widehat{\sigma}^2) K\left(\frac{d(X_i, X_j)}{h}\right) (\widehat{\varepsilon_j}^2 - \widehat{\sigma}^2),$$

où

$$\widehat{\varepsilon_i} = Y_i - \widehat{r}(X_i) \quad \text{et} \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \widehat{\varepsilon_i}^2 \text{ avec } \widehat{r}(x) = \frac{\sum_{i=1}^{n} K\left(d(x, X_i)/h\right) Y_i}{\sum_{i=1}^{n} K\left(d(x, X_i)/h\right)},$$

et K est un noyau devant vérifier quelques conditions (voir (H3)).

Afin d'établir la distribution asymptotique de la statistique W_n, nous avons besoin des hypothèses suivantes :

(H1) il existe $m \geq 2$ tel que $\mathbb{E}[Y^m | X = x] < \delta_m(x) < C < \infty$ avec $\delta_m(.)$ est continue sur $\mathcal{S}$;
(H2) $\mathbb{E}[\varepsilon^8 | X = x] \leq b(x)$ où $b(\cdot)$ est un opérateur continu sur $\mathcal{S}$, tel que $\mathbb{E}[b^2(X)] < \infty$;
(H3) le noyau K est de classe $\mathcal{C}^1$ sur son support $[0, 1]$ et vérifie :

$$K^2(1) - \int_0^1 (K^2(s))' \tau(s)\, \mathrm{d}s > 0 \quad \text{et} \quad K(1) - \int_0^1 (K(s))' \tau(s)\, \mathrm{d}s \neq 0 ;$$

(H4) la largeur de fenêtre $h := h(n)$ est strictement positive et telle que :

$$h \to 0, \ n\phi(h) \to \infty, \ n\sqrt{\phi(h)} \max\left(h^{4\beta}, \frac{1}{\log^2 n}\right) \to 0 \ \text{ et } \ \frac{(\log n)^2}{n\phi(h)} < \psi_{\mathcal{S}}\left(\frac{\log n}{n}\right) < \frac{n\phi(h)}{\log n}$$

quand n tend vers l'infini.

4. Principaux résultats

On a les résultats suivants :

Théorème 4.1. *Si les hypothèses (H1)–(H4) et (2)–(6) sont satisfaites, alors :*

– *Sous l'hypothèse nulle $\mathcal{H}_0$:*

$$n\sqrt{\phi(h)}\, W_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, s^2) \ \textit{quand} \ n \to \infty$$

où $s^2 = 2\left(K^2(1) - \int_0^1 (K^2(s))' \tau(s)\, \mathrm{d}s\right) \mathbb{E}\left[f(X) var^2\left[\varepsilon^2 | X\right]\right]$ *et* $\xrightarrow{\mathcal{D}}$ *désigne la convergence en loi.*

De plus,

$$T_n = n\sqrt{\phi(h)}\, \frac{W_n}{\widehat{s}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \ \textit{quand} \ n \to \infty$$

où

$$\widehat{s}^2 = \frac{1}{n(n-1)\phi(h)} \sum_{i=1}^{n} \sum_{j\neq i,=1}^{n} K\left(\frac{d(X_i, X_j)}{h}\right) (\widehat{\varepsilon_j}^2 - \widehat{\sigma}^2)^2 (\widehat{\varepsilon_i}^2 - \widehat{\sigma}^2)^2.$$

– *Sous l'hypothèse alternative* $\mathcal{H}_1$ *:*

$$\frac{T_n}{n\sqrt{\phi(h)}} \longrightarrow \mathbb{E}[(var[\varepsilon|X]-\sigma^2)^2 f(X)]/s_1, \text{ en probabilité,}$$

où

$$s_1^2 = \frac{\left(K^2(1)-\int_0^1 (K^2(s))'\tau(s)\,\mathrm{d}s\right)}{\left(K(1)-\int_0^1 (K(s))'\tau(s)\,\mathrm{d}s\right)}\mathbb{E}[(var[\varepsilon^2|X]+(var[\varepsilon|X]-\sigma^2)^2)f(X)].$$

Maintenant, nous examinons la robustesse du test contre toutes les déviations possibles à l'homoscédasticité. Pour ce faire, nous introduisons la séquence d'hypothèses alternatives locales suivantes :

$$\mathcal{H}_{1n}: \quad var[\varepsilon|x]-\sigma^2 = \delta_n g(x)$$

où la fonction $g(\cdot)$ est connue, continue sur $\mathcal{S}$ et telle que $\mathbb{E}[g^2(X)] < \infty$. Donc, nous obtenons le corollaire suivant.

Corollaire 4.2. *Sous les hypothèses (H1)–(H4) et (2)–(6), nous avons, sous* $\mathcal{H}_{1n}$ *avec* $\delta_n = n^{-1/2}\phi^{-1/4}(h)$ *:*

$$T_n \xrightarrow{\mathcal{D}} \mathcal{N}(\mu,1) \text{ quand } n\to\infty, \quad \text{où} \quad \mu = \left(K(1)-\int_0^1 (K(s))'\tau(s)\,\mathrm{d}s\right)\mathbb{E}[g^2(X)f(X)]/s.$$

Remerciements

Les auteurs souhaitent remercier le Professeur P. Deheuvels, pour ses remarques et conseils avisés. Le second auteur voudrait, également, exprimer sa gratitude à l'Université King Khalid (Arabie Saoudite) pour son soutien administratif et technique.

Références

[1] T.S. Breusch, A.R. Pagan, A simple test for heteroscedasticity and random coefficient variation, Econometrica 47 (1979) 1287–1294.
[2] H. Dette, A. Munk, Testing heteroscedasticity in nonparametric regression, J. R. Stat. Soc. B 60 (1998) 693–708.
[3] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis. Theory and Practice, Springer-Verlag, New York, 2006.
[4] A. Goia, P. Vieu, An introduction to recent advances in high/infinite dimensional statistics, J. Multivar. Anal. 146 (2016) 1–6.
[5] T. Hsing, R. Eubank, Theoretical foundations of functional data analysis, with an introduction to linear operators, Wiley Series in Probability and Statistics, John Wiley & Sons, Chichester, UK, 2015.
[6] J. Ramsay, B. Silverman, Functional Data Analysis, Springer Series in Statistics, Springer, New York, 2005.
[7] J.X. Zheng, Testing heteroscedasticity in nonlinear and nonparametric regressions, Can. J. Stat. 37 (2009) 282–300.