Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 346 (2008) 451–455
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistics/Probability Theory

The marked empirical process to test a general AR-ARCH against an other general AR-ARCH when the random vectors are nonstationary and absolutely regular

Michel Harel [a,b], Echarif Elharfaoui [a,b]

[a] *LSP U.M.R. 5219 CNRS, 31062 Toulouse cedex 9, France*
[b] *IUFM du Limousin, 209, boulevard de Vanteaux, 87036 Limoges cedex, France*

Received 2 July 2007; accepted after revision 19 February 2008

Available online 24 March 2008

Presented by Paul Deheuvels

Abstract

In this Note, we study a procedure on goodness-of-fit testing for nonlinear time-series models against a large class of alternatives under nonstationarity and absolute regularity. For that, we define a marked empirical process based on residuals which converges in distribution to a Gaussian process with respect to the Skorohod topology. This method was first introduced by Stute (1997) and then widely developed by Ngatchou-Wandji (2002, 2005, 2008) [1–3] under more general conditions. Applications to general AR-ARCH models are given. ***To cite this article: M. Harel, E. Elharfaoui, C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

Résumé

Le processus empirique marqué pour tester un modèle AR-ARCH général contre un autre AR-ARCH général lorsque les vecteurs aléatoires sont non stationnaires et absolument réguliers. Nous étudions une procédure pour tester des modèles de régression non stationnaires et absolument réguliers contre une large classe d'alternatives. Notre idée est d'utiliser un processus empirique marqué basé sur les résidus qui converge en loi vers un processus gaussien. ***Pour citer cet article : M. Harel, E. Elharfaoui, C. R. Acad. Sci. Paris, Ser. I 346 (2008).***

Version française abrégée

Notre but est de tester un modèle de régression hétéroscédastique de la forme

$$Y_i = m(X_{i-1}, \dots, X_{i-d}; \theta) + v(X_{i-1}, \dots, X_{i-d})\epsilon_i, \quad i \geqslant 1 + d$$

en utilisant une approche non paramétrique, et en prenant en compte l'estimation de θ sous l'hypothèse nulle H_0 d'appartenance de la fonction m à un modèle paramétrique $\mathcal{H} = m(\cdot; \theta)\colon\ \theta \in \Theta\}$. La fonction v est inconnue et les bruits ϵ_i sont absolument réguliers.

E-mail addresses: michel.harel@limousin.iufm.fr (M. Harel), eelharfaoui@yahoo.fr (E. Elharfaoui).

1631-073X/$ – see front matter
doi:10.1016/j.crma.2008.02.018

La suite $\{\mathbf{Z}_i = (\mathbf{X}_i = (X_{i-1}, \ldots, X_{i-d})', Y_i)\}$ est non stationnaire et absolument régulière. Nos statistiques de test sont construites selon le processus défini par :

$$R_n^*(\mathbf{x}) = n^{-1/2} \sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{X}_i \leqslant \mathbf{x}\}} \big(Y_i - m(\mathbf{X}_i; \tilde{\theta}_n)\big), \quad \mathbf{x} \in \mathbb{R}^d$$

$\tilde{\theta}_n$ désignant un estimateur $n^{-1/2}$-convergeant du vrai paramètre θ_0, et vérifiant la Condition 1 ci-dessous.

Nous pouvons en déduire plusieurs statistiques de test possibles : en particulier, un test de type Cramér–von Mises basé sur

$$\mathcal{T}_n = \int \big(R_n^*(\mathbf{x})\big)^2 w\big(\widehat{F}_n(\mathbf{x})\big)\, \mathrm{d}\widehat{F}_n(\mathbf{x})$$

($w(\cdot)$ désigne une fonction poids et $\widehat{F}_n$ est la f.r. empirique de l'échantillon). D'après le Théorème 1 ci-dessous, le processus R_n^* converge en distribution vers un processus R_∞^*. Par conséquent, sous H_0, $\mathcal{T}_n$ converge en loi vers le processus $\mathcal{T}$ défini ci-dessous.

1. Introduction

The purpose of this Note is to study a general method on goodness-of-fit testing for a nonlinear parametric regression model. Now, we define our model.

Let $\{\mathbf{Z}_i = (\mathbf{X}_i, Y_i);\ i \geqslant 1\}$ be a sequence of random vectors with continuous distribution functions $H_i(\mathbf{z})$, $i \geqslant 1$, $\mathbf{z}$ is in $\mathbb{R}^{d+1}$ and we assume that $H_i(\mathbf{z})$ admits a strictly positive density and H_i has the two marginals F_i and G_i.

In this paper, we will suppose that the sequence $\{\mathbf{Z}_i\}_{i \geqslant 1}$ is absolutely regular with the rate

$$\beta(n) = \mathrm{O}(\tau^n), \quad 0 < \tau < 1. \tag{1}$$

Suppose that H_i converges to the distribution function H (for the norm of total variation noted $\|\cdot\|$) which admits a strictly positive density and H has the two marginals F and G. Put $H_{i,j}$ the distribution function of $(\mathbf{Z}_i, \mathbf{Z}_j)$. Furthermore, assume that for any $l > 1$, there exists a continuous distribution function $\widetilde{H}_l$ on $\mathbb{R}^{2d+2}$ admitting a strictly positive density with marginals $\widetilde{F}_l$ on $\mathbb{R}^{2d}$, $\widetilde{G}_l$ on $\mathbb{R}^2$ such that

$$\|H_{i,j} - \widetilde{H}_{j-i}\| = \mathrm{O}(\rho_0^i), \quad 1 \leqslant i < j \leqslant n,\ n \geqslant 1,\ 0 < \rho_0 < 1 \tag{2}$$

for which there exists a sequence $\{\widetilde{\mathbf{Z}}_i = (\widetilde{\mathbf{X}}_i, \widetilde{Y}_i),\ i \geqslant 1\}$ of stationary random vectors absolutely regular with rate (1) and $(\widetilde{\mathbf{Z}}_i, \widetilde{\mathbf{Z}}_j)$ has $\widetilde{H}_{j-i}$ as distribution function $(i < j + 1)$.

Suppose also that there exists a random vector $(\mathbf{X}, Y)$ in $\mathbb{R}^{d+1}$ with finite expectation $E|Y|$ and which admits H as distribution function, so that the regression function $m(\mathbf{x}) = E(Y \mid \mathbf{X} = \mathbf{x})$ of Y on $\mathbf{X}$ is well defined, where $\mathbf{x}$ is in $\mathbb{R}^d$, is a.s. in $\mathbf{x}$ uniquely defined in view of the equation

$$m(\mathbf{X}) = E(Y \mid \mathbf{X}). \tag{3}$$

Some literature is concerned with parametric modeling in that m is assumed to belong to a given family

$$\mathcal{H} = \big\{m(\cdot;\theta) \colon \theta \in \Theta\big\} \tag{4}$$

of functions, where $\Theta \subset \mathbb{R}^p$ is a proper parameter set.

Consider the general hypothesis testing the null hypothesis that H_0 is a parametric regression model and belongs to a family given: $m \in \mathcal{H}$ versus the local alternatives $H_{1,n}$: $m \equiv m_n \in \mathcal{H}_{1,n}$: $\mathcal{H}_{1,n} = \{m = m(\cdot, \theta) + n^{-1/2} r \colon \theta \in \Theta\}$ where r is a function satisfying $E(r(\widetilde{X}_1)) \neq 0$.

For that, we consider an empirical process such that under H_0 this process depends of a parameter θ_0. First, we start by estimating the parameter and we prove that the empirical process converges in distribution to a certain centered Gaussian process when the parameter is replaced by its estimator $\tilde{\theta}_n$. Under $H_{1,n}$, the empirical process converges in distribution to a noncentered Gaussian process which has the same limit covariance function. Put

$$R_n^*(\mathbf{x}) = n^{-1/2} \sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{X}_i \leqslant \mathbf{x}\}} \big(Y_i - m(\mathbf{X}_i; \tilde{\theta}_n)\big), \quad \mathbf{x} \in \mathbb{R}^d \tag{5}$$

a marked empirical process.

The main results will be to prove the weak convergence of the process R_n^* with respect to the Skorohod topology under some reasonable conditions and to investigate the power of tests based on R_n^*.

2. Conditions and weak convergence of the marked empirical process

For simplicity, we now suppose $d = 1$.

We know that the process defined in (5) takes its values in the Skorohod space $D(-\infty, \infty)$ and the convergence in this space is equivalent to the weak convergence on compacts. This excludes the possibility of handling goodness-of-fit statistics such as $\sup_{x\in\mathbb{R}} |R_n^*(x)|$.

To also deal with such statistics, we continuously extend R_n^* to $-\infty$ and ∞ by setting: $R_n^*(-\infty) = 0$, $R_n^*(x)$ is defined by (5) for $x \in \mathbb{R}$ and $R_n^*(\infty) = n^{-1/2} \sum_{i=1}^n (Y_i - m(\mathbf{X}_i; \tilde{\theta}_n))$.

Then R_n^* becomes a process in $D[-\infty, \infty]$, which, modulo a continuous transformation, is the same as $D[0, 1]$.

Consider the sequence of distribution functions $\{\overline{F}_n\}_{n\geqslant 1}$ defined by

$$\overline{F}_n = n^{-1} \sum_{i=1}^n F_i.$$

For the behavior of the process R_n^* defined in (5), some regularity assumptions on the estimator $\tilde{\theta}_n$ will be needed. These conditions are similar to those of Stute [4] but our sequence $\{\mathbf{Z}_i\}_{i\geqslant 1}$ is nonstationary and geometrically absolutely regular, rather than being iid.

Condition 1. Under H_0, that is $m = m(\cdot; \theta_0)$ for some unknown θ_0 in Θ, $\tilde{\theta}_n$ admits an expansion: $n^{1/2}(\tilde{\theta}_n - \theta_0) = n^{-1/2} \sum_{i=1}^n \mathbf{l}(\mathbf{Z}_i; \theta_0) + o_p(1)$ for some vector-valued function $\mathbf{l}$ such that

(i) $E[\mathbf{l}(\mathbf{Z}_i; \theta_0)] = 0$ for any $i \geqslant 1$;

(ii) $L_{i,j}(\theta_0) = E[\mathbf{l}(\mathbf{Z}_i; \theta_0)\mathbf{l}'(\mathbf{Z}_j; \theta_0)]$ exists for all $i, j \geqslant 1$.

Condition 2. (i) $m(x; \theta)$ is continuously differentiable at each θ in the interior set Θ^0 of Θ. Put

$$\mathbf{g}(x; \theta) = \frac{\partial m(x; \theta)}{\partial \theta} = \big(g_1(x; \theta), \ldots, g_p(x; \theta)\big)' \tag{6}$$

(ii) there exists an $\{F_i\}_{i\geqslant 1}$ and F-integrable function $M(x)$ such that

$$\big|g_j(x; \theta)\big| \leqslant M(x), \quad \text{for all } \boldsymbol{\theta} \in \Theta \text{ and } 1 \leqslant j \leqslant p. \tag{7}$$

Theorem 1. *Assume that for any $u \in [0, 1]$,*

$$\sup_{i\geqslant 1} E\Big(\big|Y_i - m(X_i)\big|^{2+\gamma_0} \mid U_i = u\Big) < CE\Big(\big|Y - m(X)\big|^{2+\gamma_0} \mid U = u\Big) < \infty,$$

where $\gamma_0 > 0$, $U_i = \overline{F}_n(X_i)$, $1 \leqslant i \leqslant n$, $U = F(X)$, C is some positive constant and the conditions (1) *and* (2) *hold and let Conditions* 1 *and* 2 *be satisfied, then $R_n^* \to R_\infty^*$ in distribution in the space $D[-\infty, \infty]$ where R_∞^* is a centered Gaussian process with covariance function $K^*(x, y)$ where*

$$\begin{aligned} K^*(x, y) = {} & K(x, y) + \mathbf{G}'(x; \theta_0)\left(L_{1,1}(\theta_0) + 2\sum_{k=1}^{\infty} L_{1,k}(\theta_0)\right)\mathbf{G}(y; \theta_0) \\ & - \mathbf{G}'(x; \theta_0) \sum_{k=0}^{\infty} E\big[\mathbb{1}_{\{\widetilde{X}_1 \leqslant x\}}\big(\widetilde{Y}_1 - m(\widetilde{X}_1; \theta_0)\big)\mathbf{l}(\widetilde{Z}_{1+k}; \theta_0)\big] \\ & - \mathbf{G}'(y; \theta_0) \sum_{k=0}^{\infty} E\big[\mathbb{1}_{\{\widetilde{X}_1 \leqslant y\}}\big(\widetilde{Y}_1 - m(\widetilde{X}_1; \theta_0)\big)\mathbf{l}(\widetilde{Z}_{1+k}; \theta_0)\big], \end{aligned} \tag{8}$$

$$\mathbf{G}(x;\theta)=\big(G_1(x;\theta),\dots,G_p(x;\theta)\big)',\quad G_j(x;\theta)=\int_{-\infty}^{x} g_j(u;\theta)\,\mathrm{d}F(u),\quad 1\leqslant j\leqslant p,$$

and

$$K(x,y)=\int_{-\infty}^{x\wedge y}\mathrm{Var}(Y\mid X=u)\,\mathrm{d}F(u)+2\sum_{k=1}^{\infty}\int_{-\infty}^{x}\int_{-\infty}^{y}\mathrm{Cov}(\widetilde{Y}_1,\widetilde{Y}_{1+k}\mid \widetilde{X}_1=u,\widetilde{X}_{1+k}=v)\,\mathrm{d}\widetilde{F}_k(u,v). \tag{9}$$

Corollary 1. *Under $H_{1,n}$, and the conditions of Theorem* 1, *$R_n^*\to R_\infty^*$ in distribution in the space $\boldsymbol{D}_d[-\infty,\infty]$ where R_∞^* is a Gaussian process with mean $s(\mathbf{x})$ and covariance function $K^*(\mathbf{x},\mathbf{y})$ defined in* (8)*, where*

$$s(\mathbf{x})=\int_{\mathbf{u}\leqslant\mathbf{x}} r(\mathbf{u})\,\mathrm{d}F(\mathbf{u})-\mathbf{G}(\mathbf{x};\theta_0)\int_{\mathbf{u}\leqslant\mathbf{x}}\int_{\mathbb{R}}\frac{r(\mathbf{u})}{v(\mathbf{u})}\mathbf{l}(\mathbf{u},y;\theta_0)\,\mathrm{d}H(\mathbf{u},y)$$

and $v(\cdot)$ continuous are unknown.

3. The testing procedure

From the results obtained in Theorem 1, some testing procedure can be derived.

We can consider the Cramér–von Mises type test defined by

$$\mathcal{T}_n=\int\big(R_n^*(\mathbf{x})\big)^2 w\big(\widehat{F}_n(\mathbf{x})\big)\,\mathrm{d}\widehat{F}_n(\mathbf{x}), \tag{10}$$

where w is a weight function and $\widehat{F}_n$ is the empirical distribution function of the random vectors $\mathbf{X}_1,\dots,\mathbf{X}_n$.

We easily deduce that under the conditions of Corollary 1, $\mathcal{T}_n$ converges in law to $\mathcal{T}=\int(R_\infty^*(F^{-1}(\mathbf{u})))^2w(\mathbf{u})\,\mathrm{d}\mathbf{u}$. We remark that $\mathcal{T}_n$ can be also written as

$$\mathcal{T}_n=\frac{1}{n}\sum_{i=1}^{n}w\big(\widehat{F}_n(\mathbf{X}_i)\big)\left[\sum_{j=1}^{n}\mathbb{1}_{[\mathbf{X}_j\leqslant\mathbf{X}_i]}\big(Y_j-m(\mathbf{X}_j,\tilde{\theta}_n)\big)\right]^2.$$

The tails probability of the limiting distribution of the Cramer–von Mises test statistics would be very difficult to compute. That is why it is necessary to proceed to a discretization of $\mathcal{T}$ like in Ngatchou-Wandji [1].

As in Ngatchou-Wandji [1], the discretization that we can propose, follows from the Karhunen–Loève expansion of the processes $\mathcal{T}$.

Denote by $W(\cdot)=R_\infty^*(F^{-1}(\cdot))$ the process defined on $[0,1]^d$. Its Karhunen–Loève expansion can be written as

$$W=\sum_{j=1}^{\infty}\lambda_j^{1/2}W_jf_j, \tag{11}$$

where $\lambda_1\geqslant\lambda_2\geqslant\cdots$ are the eigenvalues of the covariance operator $B(\cdot)=K^*(F^{-1}(\cdot),F^{-1}(\cdot))$ which are supposed strictly positive, the sequence of functions $f_1,f_2,\dots$ is a complete orthonormal base for $L^2[0,1]^d$ of eigenvectors of the operator B and the random variables $W_j=\lambda_j^{-1/2}\int_{[0,1]^d}W(\mathbf{v})f_j(\mathbf{v})\,\mathrm{d}(\mathbf{v})$ are independent $\mathcal{N}(0,1)$ under H_0.

Then it is possible to choose a test statistic on the form $\mathcal{T}_n^J=\sum_{j=1}^{J}W_{n,j}^2$, where $J>1$ is the number of the more informative terms in the development (9) and for any $j\geqslant 1$

$$W_{n,j}=\lambda_j^{-1}n^{-1}\sum_{i=1}^{n}R_n^*(\mathbf{X}_i)w\big(\widehat{F}_n(\mathbf{X}_i)\big)f_j\big(\widehat{F}_n(\mathbf{X}_i)\big).$$

Under H_0, $\mathcal{T}_n^J$ converges is law to $\mathcal{T}^J=\sum_{j=1}^{J}W_j^2$ which has asymptotically a chi-square distribution with J degrees of freedom. However, the $\lambda_j's$ and $f_j's$ are difficult to compute in practice. A way to overcome this difficulty was suggested by Ngatchou-Wandji [1,3] by approximating the integrals by discretization.

4. Applications to the AR-ARCH model

Now we apply the results of Section 3 to test an AR-ARCH model against an other AR-ARCH model. Consider a model which can be written in the form

$$X_i = m(X_{i-1}, \dots, X_{i-d}; \theta) + v(X_{i-1}, \dots, X_{i-d})\epsilon_i, \quad i \geqslant 1+d, \tag{12}$$

where $\theta \in \Theta \subset \mathbb{R}^p$ a proper parameter set, $m(\cdot)$ satisfying Condition 2 and $v(\cdot)$ continuous are unknown.

Let $\{\mathbf{Z}_i = (\mathbf{X}_i, Y_i);\ i \geqslant 1+d\}$ denotes the random sequence of vectors in $\mathbb{R}^{d+1}$ defined by

$$Y_i = X_i \quad \text{and} \quad \mathbf{X}_i = (X_{i-1}, \dots, X_{i-d})', \quad i \geqslant 1+d.$$

We suppose that the sequence $\{\mathbf{Z}_i\}_{i\geqslant 1+d}$ satisfies the conditions (1) and (2) in the introduction and $\{\epsilon_i\}_{i\geqslant 1+d}$ is a sequence of absolutely regular random variables satisfying (1).

We will use the results of Section 3 to test H_0: $m(\cdot;\theta) \in \mathcal{H}$ versus the sequence of alternatives $H_{1,n}$: $m(\cdot;\theta) \equiv m_n \in \mathcal{H}_{1,n}$.

Theorem 2. *Assume that* $\sup_{i\geqslant 1+d} E(|v(\mathbf{X}_i)\epsilon_i|^{2+\gamma_0}) < \infty$ *and* $E(|v(\widetilde{\mathbf{X}}_{1+d})\epsilon_{1+d}|^{2+\gamma_0}) < \infty$ *hold and that Conditions* 1 *and* 2 *also hold. Then under* $H_{1,n}$, $\mathcal{T}_n^J$ *converges in law to* $\mathcal{T}^J = \sum_{j=1}^J W_j^2$ *which has asymptotically a chi-square distribution with* J *degrees of freedom and noncentrality parameter* $\Delta_J = \sum_{j=1}^J (\delta_j)^2$ *where*

$$\delta_j = \lambda_j^{-1/2}\Biggl\{ \int\limits_{[0,1]^d} \Biggl[\int\limits_{\mathbf{v}\leqslant\mathbf{u}} \mathrm{ro}\, F^{-1}(\mathbf{v})\, \mathrm{d}F(\mathbf{v}) - \mathbf{G}\bigl(F^{-1}(\mathbf{u}); \theta_0\bigr) \Biggr] \times \Biggl[\int\limits_{\mathbf{v}\leqslant\mathbf{u}} \int\limits_{\mathbb{R}} \frac{\mathrm{ro}\, F^{-1}(\mathbf{v})}{\mathrm{vo}\, F^{-1}(\mathbf{v})} \mathbf{l}\bigl(F^{-1}(\mathbf{v}), y; \theta_0\bigr)\, \mathrm{d}\widetilde{F}(\mathbf{v}, y) \Biggr] \mathrm{d}\mathbf{u} \Biggr\} f_j,$$

$\widetilde{F}$ *is the distribution function of* $(\widetilde{\mathbf{U}}_1, Y_1)$ *and* $\widetilde{\mathbf{U}}_1 = F^{-1}(\widetilde{\mathbf{X}}_1)$.

References

[1] J. Ngatchou-Wandji, Weak convergence of some marked empirical processes: Application to testing heteroscedasticity, J. Nonparametr. Statist. 14 (2002) 325–339.
[2] J. Ngatchou-Wandji, Checking nonlinear heteroscedastic time series models, J. Statist. Plann. Inference 133 (2005) 33–68.
[3] J. Ngatchou-Wandji, N. Laîb, Local power of a Cramer–von Mises type test for parametric autoregressive models of order one, Comput. Math. Apll. (2008), in press.
[4] W. Stute, Nonparametric model checks for regression, Ann. Statist. 25 (1997) 613–641.