JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

MARC HALLIN

# Séquences généralisées : un outil pour l'analyse des séries hétéroscédastiques ?

*Journal de la société statistique de Paris*, tome 135, n° 3 (1994), p. 7-19

<http://www.numdam.org/item?id=JSFS_1994__135_3_7_0>

II

COMMUNICATION

# SÉQUENCES GÉNÉRALISÉES : UN OUTIL POUR L'ANALYSE DES SÉRIES HÉTÉROSCÉDASTIQUES ?

par Marc HALLIN
Institut de Statistique et Département de Mathématique
Université Libre de Bruxelles*

## 1. Bruits blancs, tests de permutation, de rangs, de signes

### 1.1. Bruits blancs

Le concept de *bruit blanc* constitue le fondement de la quasi-totalité des modèles statistiques couramment utilisés. Soit $X_t$, $t = 1, ..., n$ une série observée : la plupart des modèles statistiques spécifient qu'une certaine transformation (elle-même dépendant des paramètres du modèle) de ces observations est un *bruit blanc :*

– modèle *de position*

$$X_t - \mu = e_t, \qquad t = 1, ..., n, \qquad \mu \in \mathbb{R}$$

est une réalisation (de longueur $n$) du *bruit blanc* $\{e_t;\ t \in \mathbb{N}\}$ ;

– modèle *linéaire général*

$$X_t - \sum_{k=1}^{K} c_{tk} \beta_k = e_t \quad t = 1, ..., n, \quad \beta = (\beta_1, ..., \beta_K)' \in \mathbb{R}^K$$

(les $c_{tk}$ sont des constantes fixées) est une réalisation (de longueur $n$) du *bruit blanc* $\{e_t;\ t \in \mathbb{N}\}$ ;

– modèles *autorégressifs*

$$X_t = \sum_{k=1}^{K} a_k X_{t-k} = e_t \quad t = 1, ..., n$$

* Conférence du 20 avril 1994 prononcée à l'occasion de la remise du prix du statisticien d'expression française.

avec $\mathbf{a} = (a_1, ..., a_K)' \in \mathbb{R}^K$ tel que

$$1 - \sum_k a_k z^k = 0, \; z \in \mathbb{C}$$

ne possède pas de racines à l'intérieur du disque unité, est une réalisation (de longueur $n$) du *bruit blanc* $\{e_t; t \in \mathbb{N}\}$ (on suppose $X_{-K+1}, ..., X_0$ observables) ;

- modèles *bilinéaires*, modèles *fractionnaires*, modèles *à erreurs ARCH*, modèles *à seuil*, etc.

Le sens précis donné à cette notion de *bruit blanc* détermine, dans chaque modèle, la nature des outils statistiques à prendre en considération – même si, dans la pratique, l'attitude perverse qui consiste, tout à l'inverse, à choisir en fonction des méthodes utilisées les hypothèses techniques à imposer aux modèles sous-jacents s'observe le plus fréquemment.

La conséquence immédiate de ce rôle fondamental du *concept* de bruit blanc dans la construction des modèles statistiques est l'importance considérable du problème du test de l'*hypothèse* de bruit blanc. Tout test d'une hypothèse fixant la valeur du paramètre dans l'un des modèles décrits ci-dessus se ramène en effet à un test de bruit blanc :

- modèles de *position :*

$$\mu = \mu_0 \quad \text{ssi} \quad Z_t(\mu_0) = X_t - \mu_0$$

est la réalisation finie d'un *bruit blanc ;*

- modèle *linéaire général :*

$$\beta = \beta^0 \quad \text{ssi} \quad Z_t(\beta^0) = X_t - \sum_k c_{tk} \beta_k^0$$

est la réalisation finie d'un *bruit blanc ;*

- modèle *autorégressif :*

$$\mathbf{a} = \mathbf{a}^0 \quad \text{ssi} \quad Z_t(\mathbf{a}^0) = X_t - \sum_k z_k^0 X_{t-k}$$

est la réalisation finie d'un *bruit blanc ;* etc. Ce rôle central de l'hypothèse de bruit blanc n'est pas propre aux problèmes de tests, et s'étend immédiatement (via la relation tests-zones de confiance) aux problèmes d'estimation par zones de confiance, ou (via la $M$-estimation, la fonction $M(\theta)$ utilisée dans ce cadre constituant le plus souvent une mesure de l'écart entre $Z_t(\theta)$ et un bruit blanc) aux problèmes d'estimation ponctuelle.

Le sens le plus restrictif pouvant être donné à la notion de *bruit blanc* est celui de *bruit blanc gaussien :*

$$\mathbf{e} = (e_1, ..., e_n)' \sim N(\mathbf{0}, \sigma^2 \mathbf{I}).$$

Cette définition débouche sur l'ensemble des méthodes de la statistique dite classique :

- modèle de *position* : le test de $H_0 : \mu = \mu_0$ se fonde sur la statistique $n^{1/2}(\overline{X} - \mu_0)/\sigma$ (si $\sigma$ est spécifié), sur la statistique $(n-1)^{1/2}(\overline{X} - \mu_0)/s$ (test de *Student*, si $\sigma$ est non spécifié) – les notations utilisées sont les notations classiques

$$\overline{X} = n^{-1} \Sigma_t X_t \quad \text{et} \quad s^2 = n^{-1} \Sigma_t (X_t - \overline{X})^2;$$

- modèle *linéaire général* : le test de $H_0 : \beta = \beta^0$ débouche sur une procédure de type chi-deux si $\sigma$ est spécifié, de type Fisher-Snedecor si $\sigma$ n'est pas spécifié ;
- modèle *autorégressif* : le test de $H_0 : \mathbf{a} = \mathbf{a}^0$ se fonde sur la *blancheur* du *corrélogramme résiduel*, ensemble des autocorrélations des résidus

$$Z_t(\mathbf{a}^0) = X_t - \Sigma_k a_k^0 X_{t-k} ;$$

- etc.

Dans un grand nombre de situations cependant l'hypothèse gaussienne est, au mieux, douteuse. On peut donc préférer l'abandonner au profit d'une autre définition, moins restrictive, de la notion de *bruit blanc*, celle de *bruit blanc homogène symétrique* :

*$e_1, ..., e_n$ indépendantes, de même densité f partout positive et symétrique par rapport à l'origine.*

Cette notion à son tour peut être affaiblie en celle de *bruit blanc homogène centré* :

*$e_1, ..., e_n$ indépendantes, de même densité f partout positive et de médiane nulle,*

de *bruit blanc non homogène symétrique*

*$e_1, ..., e_n$ indépendantes, de densités $f_i$ partout positives et symétriques par rapport à l'origine,*

de *bruit blanc non homogène centré*

*$e_1, ..., e_n$ indépendantes, de densités $f_i$ partout positives, toutes de médiane nulle*

ou de *bruit blanc échangeable*

*$e_1, ..., e_n$ échangeables,*

etc. (toutes les densités ci-dessus sont prises par rapport à la mesure de Lebesgue). Les techniques gaussiennes traditionnelles décrites plus haut doivent alors, selon le cas, faire place à l'une ou l'autre catégorie de techniques *non paramétriques*. Ces techniques s'obtiennent généralement à partir de deux types d'arguments : les *principes de non-biais*, et les *principes d'invariance*.

### 1.2. Non-biais et invariance

Considérons, à titre d'exemple, le cas des modèles fondés sur la notion de *bruit blanc homogène symétrique*. Le principe de *non-biais* consiste, rappelons-le, à ne prendre en considération que les tests dont la puissance est uniformément supérieure au risque de première espèce. On sait (cf. par exemple Lehmann (1986), chapitre 4.3) que, s'il existe une statistique **T** exhaustive et complète pour le sous-modèle

associé à la frontière commune entre l'hypothèse et la contre-hypothèse, les tests sans biais (en fait, les tests *semblables*) s'obtiennent en conditionnant par rapport à T (tests à *α-structure de Neyman* en T). Dans l'exemple qui nous occupe, notons $Z_1, ..., Z_n$ les résidus (notés $Z_t(\mu_0)$, $Z_t(\beta^0)$ et $Z_t(\mathbf{a}^0)$, respectivement, dans les trois cas décrits plus hauts) associés à la valeur spécifiée sous l'hypothèse nulle $H_0$ ($\mu = \mu_0, \beta = \beta^0$ ou $\mathbf{a} = \mathbf{a}^0$ dans les cas considérés plus hauts) du paramètre du modèle. L'hypothèse nulle est satisfaite ssi $Z_1, ..., Z_n$ est une réalisation de longueur $n$ d'un bruit blanc homogène symétrique, de densité non spécifiée. Une statistique exhaustive et complète pour cette hypothèse est la statistique d'ordre

$$|\mathbf{Z}|_{(\cdot)} = (|Z|_{(1)}, ..., |Z|_{(n)})$$

des valeurs absolues des résidus (caractérisée par

$$|Z|_{(1)} \leq |Z|_{(2)} \leq ... \leq |Z|_{(n)}).$$

La loi conditionnelle (sous $H_0$) de $Z_1, ..., Z_n$, sachant $|\mathbf{Z}|_{(\cdot)}$, est uniforme discrète sur les $2^n(n!)$ permutations de $|Z|_{(1)}, ..., |Z|_{(n)}$ affectées de signes quelconques. Les tests (conditionnels) ainsi obtenus sont les tests dits de *permutation* (en réalité, des tests de *permutation* et *affectation de signes*). L'exemple le plus connu en est celui du *test de Student permutationnel* (pour les modèles de position ou de régression : cf. Dufour et Hallin, 1991). Ce test est basé sur la statistique de Student traditionnelle (dont la loi inconditionnelle dépend malheureusement de la densité non spécifiée des $Z_t$), mais les valeurs critiques considérées sont prises dans la loi conditionnelle (permutationnelle) décrite plus haut. Ces lois permutationnelles peuvent également être prises en considération dans le cas d'un coefficient de corrélation (pour les modèles autorégressifs, par exemple : Dufour et Hallin, 1993, 1994a).

Le principal (et très puissant) argument théorique en faveur de la prise en considération de ces tests de permutation est que tout test qui n'appartient pas à cette catégorie est automatiquement *biaisé*. Les points critiques permutationnels sont indépendants des densités sous-jacentes, et peuvent aisément être obtenus par simulation.

La seconde méthode de construction de tests dans ce contexte de modèles à bruit blanc homogène symétrique est fondée sur le *principe d'invariance*. Considérons en effet le groupe $\mathcal{G}$, $\circ$ des transformations $\mathbf{g}$ de $\mathbb{R}^n$ telles que

$$\mathbf{g}(\mathbf{z}) = (g(z_1), ..., g(z_n)) \qquad \mathbf{z} = (z_1, ..., z_n) \in \mathbb{R}^n,$$

où $g : \mathbb{R} \to \mathbb{R}$ est une fonction continue, monotone croissante, impaire ($g(-z) = -g(z)$) et telle que $\lim_{z \to \pm \infty} g(z) = \pm \infty$. Un tel groupe, appliqué au vecteur $\mathbf{Z}$ des résidus associés à $H_0$, est un *groupe générateur* de $H_0$ (cf. par exemple Lehmann (1986) chapitre 6). L'application du principe d'invariance pour le test de l'hypothèse $H_0$ conduit à ne prendre en considération que les tests invariants par $\mathcal{G}$, $\circ$. Si T constitue un *invariant maximal* pour ce groupe, les tests invariants sont les tests T-mesurables (même référence).

Un invariant maximal dans le cas qui nous occupe est le couple $(\mathbf{s}, \mathbf{R}_+^{(n)})$ constitué

- du vecteur $\mathbf{s} = (s_1, ..., s_n)$ des signes de $Z_1, ..., Z_n$ $\left(s_t = I\,[Z_t \geq 0] - I\,[Z_t \leq 0]\right)$;
- du vecteur $\mathbf{R}_+^{(n)} = \left(R_{+;1}^{(n)}, ..., R_{+;n}^{(n)}\right)$ des rangs des valeurs absolues des $Z_t$ $\left(R_{+;t}^{(n)}\right.$ est le rang de $|Z_t|$ parmi $|Z_1|, ..., |Z_n|\left.\right)$.

Sous $H_0$, $\left(\mathbf{s}, \mathbf{R}_+^{(n)}\right)$ est uniformément distribué sur les $2^n n!$ combinaisons possibles d'un $n$-uple de signes et d'une permutation des entiers $\{1, ..., n\}$.

Les tests ainsi obtenus (tests mesurables en $\left(\mathbf{s}, \mathbf{R}_+^{(n)}\right)$) sont les tests dits de *rangs signés* dont les plus connus sont, pour les problèmes de position, le test de rangs signés de Wilcoxon et celui de Fisher-Yates, pour les problèmes d'analyse de la variance les tests de Kruskal-Wallis et de Friedman, etc. Des corrélogrammes de rangs signés peuvent également être construits pour l'étude des modèles AR et ARMA (Hallin et Puri 1993 ; 1994). Ces tests possèdent tous les avantages découlant de leur invariance : ils sont libres, naturellement robustes, et plus puissants souvent que les tests gaussiens correspondants (cf. Chernoff et Savage, 1958 ; Hallin ; 1994).

### 1.3. Le cas des bruits blancs non homogènes

Qu'obtient-on en appliquant les mêmes principes de non-biais et d'invariance dans le cadre des modèles construits à partir d'un bruit blanc *non homogène ?*

La réponse à cette question est d'autant plus intéressante que les procédures gaussiennes – qui, sous des bruits blancs homogènes (centrés ou non, symétriques ou non), et moyennant quelques hypothèses techniques additionnelles, demeurent en général asymptotiquement ou approximativement valides – perdent ici toute fiabilité. Or la présence de bruits blancs non homogènes (au moins, hétéroscédastiques) est à soupçonner dans un grand nombre de domaines d'applications (séries économétriques, financières, biomédicales, ...).

Notons, comme précédemment, $Z_1, ..., Z_n$ la série des résidus associés à une hypothèse fixant les paramètres du modèle. Sous cette hypothèse, et dans le cas d'un modèle construit à partir d'un bruit blanc non homogène, les $Z_t$ sont donc indépendants, pas nécessairement identiquement distribués, mais de médiane nulle (bruit blanc *centré*), ou de loi symétrique par rapport à zéro (bruit blanc *symétrique*).

Pour l'application du principe de non-biais, on se placera dans le cas du bruit blanc *symétrique*. Une statistique exhaustive complète dans ce cas est le vecteur $|\mathbf{Z}| = \left(|Z_1|, ..., |Z_n|\right)$ des valeurs absolues des résidus. Conditionnellement à $|\mathbf{Z}|$, $\mathbf{Z}$ est uniformément distribuée sur les $2^n$ points résultant des $2^n$ façons d'attribuer un signe aux $|Z_t|$. Les tests fondés sur ces lois conditionnelles portent toujours le nom de *tests de permutation* mais seraient plus exactement définis comme des tests *d'attribution de signes*. On remarquera que, dans le modèle de position comme dans le modèle de régression, la loi permutationnelle de la statistique de Student obtenue en conditionnant par rapport à $|\mathbf{Z}|$ est la même que celle précédemment obtenue en conditionnant par rapport à $|\mathbf{Z}|_{(\cdot)}$. Le cas des coefficients d'autocorrélation est étudié dans Dufour et Hallin (1994b).

Quant au principe d'invariance, il s'applique de la façon suivante aux cas de bruit blanc non-homogène symétrique. Un groupe générateur de l'hypothèse est le groupe $\mathcal{G}$, $\circ$ des transformations g de $\mathbb{R}^n$ définies par

$$\mathbf{g}(\mathbf{z}) = \big(g_1(z_1),\ ...,\ g_n(z_n)\big), \qquad \mathbf{z} = (z_1,\ ...,\ z_n) \in \mathbb{R}^n,$$

où $(g_1, ..., g_n)$ est un $n$-uple de fonctions continues de $\mathbb{R}$ dans $\mathbb{R}$, monotones croissantes et telles que $g_i(0) = 0$, $\lim_{z \to \pm\infty} g_i(z) = \pm\infty$. L'invariant maximal pour ce groupe est le vecteur $\mathbf{s} = (s_1, ..., s_n)$ des signes des $z_t$. Le principe d'invariance conduit donc à ne considérer que les tests mesurables en les signes $s_t$ des résidus $Z_t$. Les tests ainsi obtenus constituent une sous-classe de l'ensemble des tests de permutations fournis par le principe de non-biais (ces derniers peuvent dépendre également des valeurs prises par $|\mathbf{Z}|$).

Le très classique *test du signe* (cf. Lehmann (1986) pp. 106-197) est celui des tests ainsi obtenus qui s'applique dans le cadre du modèle de position. Il s'agit d'un test exact et particulièrement simple à mettre en œuvre, puisque reposant sur une loi binomiale. On sait en outre (même référence) qu'il est à puissance uniformément maximum (dans la classe de tous les tests de même niveau) dans ce contexte. Il n'est donc pas possible de faire mieux.

Une question qui vient immédiatement à l'esprit est : ce résultat peut-il être généralisé aux modèles plus généraux évoqués plus haut ? Peut-on construire des *tests de signes exacts* pour les problèmes de modèle linéaire général ? pour les problèmes d'autorégression ?

C'est ce dernier cas qui fera plus particulièrement l'objet de notre attention : comme nous le verrons, ce sont les tests *de séquences* et leurs généralisations à des délais d'ordre supérieur qui fourniront, dans le contexte sériel, les extensions recherchées du test du signe.

## 2. Tests de séquences et tests de séquences généralisés

### 2.1. Séquences prises par rapport à l'origine ; séquences prises par rapport à la médiane empirique

Les tests de séquences figurent parmi les outils d'inférence statistique les plus anciens, puisqu'on les trouve décrits et étudiés de façon approfondie dans Mood (1940) ; pour une bibliographie commentée, voir Hallin et Puri (1992). Deux types de *séquences* seront considérés ici : les *séquences prises par rapport à l'origine*, et les *séquences prises par rapport à la médiane empirique*. Revenons à la série de résidus $Z_1, ..., Z_n$ déjà envisagée au paragraphe précédent. Les signes de ces résidus forment un vecteur de la forme

$$\mathbf{s}^{(n)} = (\underline{s_1},\ ...,\ \underline{1,1},\ \underline{-1},\ \underline{1},\ \underline{-1, -1, -1},\ \underline{1,1},\ \underline{-1},\ ...,\ \underline{s_n});$$

on appelle *séquence* (prise par rapport à l'origine) toute succession de signes identiques dans $s^{(n)}$ (ces *séquences* sont soulignées dans l'exemple ci-dessus). Leur nombre est $\sum_{t=2}^{n} I[Z_1 Z_{t-1} < 0] + 1$ (avec probabilité 1, les $Z_t$ sont tous non nuls). La quantité

$$S_{+;1}^{(n)} = (n-1)^{-1} \sum_{t=2}^{n} I[Z_1 Z_{t-1} < 0] + 1$$

est appelée *statistique des séquences prises par rapport à l'origine.*

Les séquences prises par rapport à la médiane empirique se calculent de façon entièrement analogue, mais à partir des signes des différences $Z_t - Z_{1/2}^{(n)}$, où $Z_{1/2}^{(n)}$ représente la médiane empirique des résidus $Z_1, ..., Z_n$. La quantité

$$S_1^{(n)} = (n-1)^{-1} \sum_{t=2}^{n} I\left[(Z_t - Z_{1/2}^{(n)})(Z_{t-1} - Z_{1/2}^{(n)}) < 0\right]$$

sera appelée *statistique des séquences prises par rapport à la médiane empirique.*

Clairement, $S_{+;1}^{(n)}$ est mesurable en les signes $s_t$ des résidus $Z_t$, donc invariante pour le groupe générateur $\mathcal{G}$, $\circ$ décrit plus haut, donc libre. Sa loi exacte (sous l'hypothèse que $Z_1, ..., Z_n$ est la réalisation de longueur $n$ d'un bruit blanc *centré*) est de type binomial.

La statistique $S_1^{(n)}$, elle, n'est pas mesurable en les signes $s_t$. Dans la mesure où $Z_{1/2}^{(n)}$ constitue une estimation de la médiane « vraie » des $Z_t$, $S_1^{(n)}$ peut être considérée comme une version *alignée* de $S_{+;1}^{(n)}$, et sera donc particulièrement utile dans les situations où les modèles de régression et les modèles autorégressifs définis plus haut sont de la forme

$$X_t - \sum_{k=1}^{K} c_{tk} \beta_k = e_t + \mu$$

(hypothèse $H_0$ à tester : $\beta = \beta^0$, $\mu$ non spécifié ; résidus $Z_t = X_t - \sum_{k=1}^{K} c_{tk} \beta_k^0$)

et

$$X_t - \sum_{k=1}^{K} a_k X_{t-k} = e_t + \mu$$

(hypothèse $H_0$ à tester : $\mathbf{a} = \mathbf{a}^0$, $\mu$ non spécifié ; résidus $Z_t - X_t - \sum_{k=1}^{K} a_k X_{t-k}$), respectivement. Les résidus $Z_t$, sous $H_0$, sont alors un bruit blanc non homogène, centré en une médiane $\mu$ non spécifiée, que l'on peut, de façon très naturelle, estimer par $Z_{1/2}^{(n)}$. Malheureusement, cette estimation détruit les propriétés d'invariance des signes : les signes des différences $Z_t - Z_{1/2}^{(n)}$ ne possèdent pas les mêmes propriétés

que ceux des différences $Z_t - \mu$, et, la statistique $S_1^{(n)}$ n'étant pas libre, sa loi exacte, qui dépend de la loi non spécifiée des $e_t$, ne peut être calculée.

Si $S_1^{(n)}$ devait pouvoir être considérée comme *libre* sous l'hypothèse de bruit blanc, celle-ci devrait être renforcée en une hypothèse de bruit blanc *homogène*. On peut en effet exprimer $S_1^{(n)}$ en fonction des rangs $R_t^{(n)}$ des résidus $Z_t$ :

$$S_1^{(n)} = (n-1)^{-1} \sum_{t=2}^{n} I\left[\left(R_t^{(n)} - \frac{n}{2}\right)\left(R_{t-1}^{(n)} - \frac{n}{2}\right) < 0\right].$$

Toute statistique mesurable en les rangs d'un bruit homogène étant *libre*, $S_1^{(n)}$ l'est également. Mais les tests de séquences, dans un contexte de bruit blanc homogène, sont peu intéressants, d'autres statistiques fournissant des tests plus puissants (autocorrélations de van der Waerden, Wilcoxon, etc. : cf. Hallin et Puri, 1994).

Dans les limites imposées par les conditions de validité ci-dessus, les tests de séquences peuvent cependant être utilisés en vue de détecter la présence de dépendances sérielles dans la série des résidus. En fait,

$$S_{+;1}^{(n)} = \frac{1}{2}\left(1 - \frac{1}{n-1} \sum_{t=2}^{n} s_t s_{t-1}\right),$$

où $(n-1)^{-1} \Sigma s_t s_{t-1}$ s'interprète (à des constantes multiplicatives près) comme une autocorrélation calculée à partir de la série des *signes* des $Z_t$ (cf. Dufour, 1981). Le test des séquences (prises par rapport à l'origine) est donc au test fondé sur le coefficient d'autocorrélation d'ordre un ce que le test du signe est au test de Student.

### 2.2. Séquences généralisées

Les arguments d'invariance développés plus haut conduisent à ne prendre en considération, dans le cadre de modèles définis à partir de bruits blancs non homogènes, que les statistiques de test mesurables en les signes des résidus. Le test des séquences appartient bien à cette catégorie. Malheureusement, son utilisation se heurte à trois problèmes.

(i) Typiquement, le test des séquences classique n'est sensible qu'à des dépendances sérielles de délai un : autorégressions ou moyennes mobiles d'ordre un, par exemple. Ceci en limite considérablement le champ d'application.

(ii) Dans la majorité des applications, la médiane $\mu$ des résidus reste non spécifiée. On serait tenté, donc, de considérer plutôt les séquences prises par rapport à la médiane empirique. Malheureusement, on l'a vu, celles-ci ne sont pas *libres* dans le contexte qui nous intéresse, qui est celui des modèles fondés sur un bruit blanc non homogène.

(iii) Les statistiques classiquement utilisées pour la détection des dépendances sérielles sont les autocorrélations résiduelles. Évidemment, leur loi sous une hypothèse de bruit blanc non homogène n'est pas connue avec exactitude,

non plus que leur loi asymptotique. Mais la seule prise en considération des signes constitue, à première vue, une perte d'information si considérable qu'on peut être tenté de prendre quelques risques du côté de la validité des tests (risques de première espèce) afin de ne pas consentir une perte rédhibitoire sur la plan de la puissance (risques de seconde espèce).

Les réponses à ces trois problèmes sont données ci-après dans les paragraphes 2.2., 2.3. et 2.4., respectivement.

L'introduction du concept de séquences généralisées (Dufour et Hallin, 1994b) a pour but de répondre à la première des trois objections faites ci-dessus. Soit $Z_1, ..., Z_n$ la série de résidus étudiée ; considérons la partition de $(Z_1, ..., Z_n)$ en les deux sous-séries

$$(Z_1, Z_3, Z_5, ..., Z_{n \text{ ou } n-1}) \quad \text{et} \quad (Z_2, Z_4, Z_6, ..., Z_{n-1 \text{ ou } n}).$$

À chacune de ces deux sous-séries correspond un nombre de séquences (prises par rapport à l'origine ou par rapport à la médiane empirique $Z^{(n)}_{1/2}$). Leur somme fournit pour la série d'origine un concept de séquences de délai 2. Pour le délai 3, trois sous-séries seront constituées, et $k$ pour le délai $k$. Ceci justifie l'appellation *statistique de séquences de délai k* pour les statistiques

$$S^{(n)}_{+;k} = (n-k)^{-1} \sum_{t=k+1}^{n} I\,[Z_t Z_{t-k} < 0]$$

et

$$S^{(n)}_{k} = (n-k)^{-1} \sum_{t=k+1}^{n} I\,[(Z_t - Z^{(n)}_{1/2})(Z_{t-k} - Z^{(n)}_{1/2}) < 0]$$

respectivement. De même que $S^{(n)}_{+;1}$ et $S^{(n)}_{1}$ constituent des mesures de dépendance sérielle à l'horizon 1, $S^{(n)}_{+;k}$ et $S^{(n)}_{k}$ mesurent les dépendances sérielles à l'horizon $k$, et sont liées aux autocorrélations de délai $k$ des séries de signes correspondantes. Les lois exactes de $S^{(n)}_{+;k}$ (sous une hypothèse de bruit blanc non homogène centré) et de $S^{(n)}_{k}$ (sous une hypothèse de bruit blanc homogène) peuvent être obtenues sous forme explicite : cf. Dufour et Hallin (1994b). Des corrélogrammes généralisés, fondés sur $S^{(n)}_{+;k}$ et $S^{(n)}_{k}$, et accompagnés de $p$-valeurs exactes, peuvent ainsi être établis. Des tests de type *portemanteau* peuvent également être construits (même référence).

Si une réponse satisfaisante pouvait être donnée aux objections (ii) et (iii) formulées plus haut, les séquences généralisées de délai $k$ (les statistiques $S^{(n)}_{k}$) pourraient donc fournir un outil parfaitement adapté à l'étude des dépendances sérielles dans les séries non homogènes (les séries hétéroscédastiques, par exemple).

### 2.3. Le problème de l'alignement

Les versions, généralisées à l'ordre $k$ qu'on vient de donner des statistiques de séquences traditionnelles présentent malheureusement les mêmes graves inconvénients que les statistiques de séquences au délai un :

- $S^{(n)}_{+;k}$, comme $S^{(n)}_{+;1}$ n'est invariante et libre que si la médiane exact des $Z_t$ est nulle, ce qui n'est pas le cas dans la plupart des situations pratiques ;
- $S^{(n)}_k$, comme $S^{(n)}_1$, n'est invariante et libre que sous une hypothèse de bruit blanc *homogène* – hypothèse sous laquelle on préférera avoir recours à d'autres outils plus puissants (les tests de rangs, par exemple).

La portée pratique de ces extensions aux délais $k$ supérieurs à un des traditionnelles statistiques de séquences de délai un semble donc, à première vue, des plus limitée.

L'intérêt des tests fondés sur les statistiques de séquences de délai $k$ ainsi définies demeurerait toutefois (presque) intact si l'effet de l'alignement (estimation de la médiane vraie inconnue par la médiane empirique $Z^{(n)}_{1/2}$) était négligeable ou asymptotiquement négligeable. Si, pour être plus précis, il pouvait être montré que la différence entre les statistiques de séquences *« exactes »* $S^{(n)}_{+;k}$ et les statistiques de séquences *« alignées »* $S^{(n)}_k$ était $o_p(n^{-1/2})$ sous l'hypothèse que les $Z_t$ constituent un bruit blanc non homogène centré (de médiane nulle), la statistique alignée $S^{(n)}_k$ bénéficierait, sous forme asymptotique, des propriétés d'invariance et de liberté qui sont celles de $S^{(n)}_{+;k}$.

Une telle propriété, qui restaurerait tout l'intérêt d'une inférence fondée sur les $S^{(n)}_k$, est hélas peu vraisemblable à première vue. À moins de renforcer considérablement les hypothèses, la convergence vers la médiane « vraie » commune $\text{med}\left(Z^{(n)}_t\right) = \zeta_{1/2}$ de la médiane empirique $Z^{(n)}_{1/2}$ n'est, en effet, nullement acquise. Une condition nécessaire et suffisante pour une telle convergence a été établie par Sen (1970), qui montre que si $Z_t\,(t \in \mathbb{N})$ admet une densité $f_t$ de médiane 0, et que $Z_1, \ldots, Z_n$ sont indépendantes, leur médiane empirique $Z^{(n)}_{1/2}$ converge vers 0 en probabilité si et seulement si

$$\lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} f_t(0) = \infty.$$

Ce genre d'hypothèse technique est particulièrement désagréable à faire, car tout à fait invérifiable : il serait donc souhaitable de pouvoir établir l'équivalence asymptotique entre $S^{(n)}_{+;k}$ et $S^{(n)}_k$ sans y avoir recours. L'absence de cette hypothèse rend cependant l'entreprise singulièrement difficile. Toutes les techniques de démonstration classiques dans une telle situation reposent en effet sur l'établissement d'une propriété de *continuité en probabilité* de la variable considérée par rapport au paramètre inconnu (ici, la médiane théorique), dans le voisinage de sa valeur « exacte » (non alignée). Si l'hypothèse de convergence de $Z^{(n)}_{1/2}$ vers la médiane « vraie » $\zeta_{1/2}$ ne peut être faite, une telle méthode de preuve est inapplicable, puisque aussi bien la *continuité en probabilité* de $S^{(n)}_{+;k}$ dans le voisinage de $\zeta_{1/2}$ ne suffirait pas à établir l'équivalence asymptotique souhaitée.

De façon inattendue, cependant, et à partir de techniques de nature purement combinatoire, il est possible (Dufour et Hallin, 1994b) de montrer que

$(n-k)^{1/2}\left(S^{(n)}_{+;k} - S^{(n)}_k\right)$ tend vers zéro, en probabilité, sous l'hypothèse de bruit blanc non homogène centré, lorsque $n$ tend vers l'infini. La preuve procède par conditionnements successifs : sous l'hypothèse non restrictive que $\text{med}(Z_t) = \xi_{1/2} = 0$, les lois conditionnelles de $S^{(n)}_k$ sachant :

- le vecteur $\left(|Z_1|, \ldots, |Z_n|\right)$ des valeurs absolues ;
- le vecteur $R^{(n)}_+ = \left(R^{(n)}_{+;1}, \ldots, R^{(n)}_{+;n}\right)$ des rangs des $|Z_t|$ ;
- $\tau^{(n)}_* = \left\{t \leq n \mid Z^{(n)}_{1/2} \leq Z_t \leq 0 \quad \text{ou} \quad 0 \leq Z_t \leq Z^{(n)}_{1/2}\right\}$ ;
- $\max_{t \in \tau^{(n)}_*} R^{(n)}_{+;t}$ ;
- $\#\tau^{(n)}_*$ ;

sont de type hypergéométrique ou hypergéométrique négatif, et permettent d'obtenir l'équivalence asymptotique désirée.

Un corrélogramme, accompagné de zones de confiance asymptotiquement valides, peut donc être construit à partir des statistiques $S^{(n)}_k$. Ce corrélogramme peut être utilisé pour le test de l'hypothèse de bruit blanc non homogène de la même façon qu'un corrélogramme classique pour l'hypothèse de bruit blanc homogène du second ordre. Les limites de confiance peuvent être calculées soit à partir des lois exactes des $S^{(n)}_{+;k}$ (cf. Dufour et Hallin, 1994b), soit à partir de leurs approximations asymptotiques normales. Des tests de type portemanteau peuvent également être mis en œuvre (même référence).

### 2.4. Efficacité asymptotique relative

Grâce à l'équivalence asymptotique de $S^{(n)}_k$ et $S^{(n)}_{+;k}$, les statistiques de séquences généralisées fournissent donc bien, pour le test d'une hypothèse de bruit blanc non homogène, un outil comparable à celui que constituent les corrélogrammes usuels dans le cas homogène. Les praticiens cependant répugneront à abandonner le corrélogramme classique en faveur des séquences généralisées si ce changement de méthode et d'habitudes correspond à une perte trop grande dans le domaine de l'efficacité.

Cette attitude est discutable, car le recours à des procédures de test dont le risque de première espèce n'est pas contrôlé (pas même approximativement) est difficilement justifiable. Une comparaison des puissances, d'autre part, ne rend pas vraiment justice aux tests de séquences. Il n'est en effet pas légitime de mettre en compétition deux méthodes dont les conditions de validité sont aussi différentes : la comparaison ne peut être établie que sous les conditions techniques les plus restrictive et, les méthodes de corrélogramme classiques requérant des conditions de bruit blanc homogène, ce n'est que sous ces conditions, évidemment défavorables aux tests de séquences généralisés, que peuvent s'effectuer des comparaisons de puissances.

Ces dernières sont néanmoins fort satisfaisantes. Il peut sembler, à première vue, que l'abandon de toute information concernant l'amplitude des écarts à la médiane

des observations doive, irrémédiablement, conduire à un effondrement de la fonction de puissance. Comme l'indiquent les efficacités asymptotiques relatives (AREs) rapportées dans le tableau ci-dessous (Hallin et Puri, 1991), il n'en est rien, puisque les valeurs prises par ces AREs varient entre 0,40 et 0,50. Ceci indique, rappelons-le que le nombre d'observations nécessaire pour que les tests traditionnels atteignent à la même puissance (localement et asymptotiquement) que les tests de séquences généralisés ne représente que 40 % ou 50 % des effectifs requis par ces derniers. Rappelons cependant que ces tests traditionnels, en revanche, ne sont plus valides dans le cadre non homogène qui nous intéresse ici.

*Tableau 1.* Efficacités asymptotiques relatives (AREs) des tests fondés sur les statistiques de séquences généralisées par rapport aux tests classiques correspondants fondés sur les autocorrélations usuelles, sous les densités gaussienne, logistique et double-exponentielle, respectivement. Ces efficacités sont calculées sous les hypothèses d'homogénéité indispensables à la prise en considération de ces autocorrélations.

| densités | *gaussiennes* | *logistiques* | *double-exponentielles* |
|---|---|---|---|
| AREs | 0,405 | 0,438 | 0,500 |

Le complémentaire de ces efficacités asymptotiques relatives – de l'ordre de 50 % – constitue en quelque sorte la prime de risque du contrat d'assurance contre l'hétéroscédasticité ou l'hétérogénéité que constitue le recours aux séquences généralisées. Il appartient à l'utilisateur de décider si cette prime est suffisamment compensée par la garantie correspondante de non-dérapage du risque de première espèce.

## 3. Un outil pour l'analyse des séries hétéroscédastiques ?

Les séquences généralisées fournissent, on l'a vu, un outil adéquat pour le test de l'hypothèse de bruit blanc non homogène, et constituent sans doute, grâce à leurs propriétés d'invariance, l'outil le plus approprié dans le traitement des séries à innovations indépendantes mais non stationnaires. Le problème majeur qui subsiste est celui du calcul le plus approprié des résidus estimés – celui qui, par exemple, généraliserait au cas de l'estimation d'un modèle ARMA l'utilisation de la médiane empirique comme estimateur (éventuellement, non convergent) de la médiane vraie. Même si des expériences numériques (Campbell et Dufour, 1993, 1994 ; Campbell et Galbraith, 1994) suggèrent un certain optimisme, beaucoup de travail reste encore à réaliser avant que l'interrogation du titre puisse faire place à une affirmation. Bien que jalonnée de multiples questions, une voie semble toutefois tracée.

## RÉFÉRENCES

CAMPBELL B. and DUFOUR J.-M. (1993) "Exact Nonparametric Orthogonality and Random Walk Tests", *R. Economics and Statistics*, à paraître.

CAMPBELL B. and DUFOUR J.-M. (1994) *Exact Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter*. Discussion paper CRDE, Université de Montréal, Montréal.

CAMPBELL B. and GALBRAITH J.W. (1994) "Inference in Expectations Models of the Term Structure" *in New Developments in Time Series Econometrics*, J.-M. Dufour and B. Raj, Eds, Springer-Verlag, New York, pp. 67-82.

CHERNOFF H. and SAVAGE I.R. (1958) "Asymptotic Normality and Efficiency of Certain Nonparametric Tests". *Annals of Mathematical Statistics*, pp. 972-994.

DUFOUR J.-M. (1981) "Rank Tests for Serial Dependence". *J. of Time Series Analysis*, pp. 117-228.

DUFOUR J.-M. and HALLIN M. (1991) "Nonumiform Bounds for Nonparametric *t*-Tests". *Econometric Theory*, pp. 253-263.

DUFOUR J.-M. and HALLIN M. (1993) "Improved Eaton Bounds for Linear Combinations of Bounded Random Variables, with Statistical Applications", *J. of the American Statistical Association*, pp. 1026-1033.

DUFOUR J.-M. and HALLIN M. (1994a). *Distribution-free Bounds for Serial Correlation Coefficients*, Unpublished manuscript, CRDE, Université de Montréal, Montréal.

DUFOUR J.-M. and HALLIN M. (1994b) *Exact Inference for Time Series Analysis Based on Generalized Runs*, Unpublished manuscript, Département de Mathématique, U.L.B.

HALLIN M. (1994) "On the Pitman Nonadmissibility of Correlogram-based Methods", *J. Time Series Analysis*, à paraître.

HALLIN M. and PURI M.L. (1991) "Time Series Analysis via Rank-order Theory : Signed-rank Tests for ARMA Models", *J. Multivariate Analysis*, pp. 1-29.

HALLIN M. and PURI M.L. (1992) "Rank Tests for Time-series Analysis : a Survey", *in New Directions in Time Series Analysis*, D. Brillinger, E. Parzen and M. Rosenblatt, Eds, Springer-Verlag, New York, pp. 111-154.

HALLIN M. and PURI M.L. (1994). "Aligned Rank Tests for Linear Models with Autocorrelated Error Terms", *J. of Multivariate Analysis*, pp. 175-237.

LEHMANN E.L. (1986) *Testing Statistical Hypotheses*, 2nd Edition, Wiley, New York.

MOOD A.M. (1940) "The Distribution Theory of Runs", *Annals of Mathematical Statistics*, pp. 367-392.

SEN P.K. (1970) "A Note on Order Statistics for Heterogeneous Distributions", *Annals of Mathematical Statistics*, pp. 2137-2139.