Ann. sc. math. Québec, 1982, vol. 6, no 2, pp. 151-161

EFFICACITÉ RELATIVE ASYMPTOTIQUE DE CERTAINS TESTS DE TENDANCE MONOTONE

Denis Labelle

1. Introduction et test paramétrique

Soit $\{Y_1,Y_2,\ldots,Y_T\}$ une série chronologique de la forme

$$Y_t = \delta + \theta g(\tfrac{t}{T}) + X_t$$

où g est une fonction monotone, définie sur l'intervalle $[0,1]$, telle que $g(0) = 0$, $g(1) = 1$, $\int_0^1 g^2(x)\,dx - (\int_0^1 g(x)\,dx)^2 > 0$ et où $X_1,\ldots,X_T$ sont des variations aléatoires, indépendantes et équidistribuées, avec $E(X_t) = 0$ et $\mathrm{Var}(X_t) = \sigma_x^2 > 0$.

On veut tester l'hypothèse $H : \theta = 0$ contre l'alternative $K : \theta \neq 0$.

Si la fonction g est connue, on peut estimer directement θ par la méthode classique des moindres carrés. On trouve

$$\hat{\theta} = \frac{T \sum Y_t g(\frac{t}{T}) - (\sum Y_t)(\sum g(\frac{t}{T}))}{T \sum g^2(\frac{t}{T}) - (\sum g(\frac{t}{T}))^2} \tag{1}$$

qui est sans biais pour θ. De plus, si les X_t sont de loi normale,

$$\hat{\theta} \text{ est } N(\theta\,, \frac{T\sigma_x^2}{T \sum g^2(\frac{t}{T}) - (\sum g(\frac{t}{T}))^2})\,. \tag{2}$$

Si les X_t ne sont pas de loi normale, l'expression (2) ne sera valide qu'asymptotiquement et on obtient

$$\frac{\sqrt{T}\sigma_g}{\sigma_x}(\hat{\theta}-\theta) \xrightarrow{\text{loi}} N(0,1) \tag{3}$$

où

$$\sigma_g^2 = \int_0^1 g^2(x)\,dx - \left(\int_0^1 g(x)\,dx\right)^2 .$$

Pour tester H , on estime σ_x^2 par $S_x^2 = \frac{1}{T-2}\sum_{t=1}^{T}(Y_t - \hat{\delta} - \hat{\theta}g(\frac{t}{T}))^2$ et on utilise la loi de Student. La convergence de S_x^2 vers σ_x^2 (et de la loi de Student vers la loi normale) permet de déterminer la taille de la série qui donnera un test de niveau α et de puissance $1 - \beta$ (avec $\alpha > 0$, $\beta > 0$ et $\alpha + \beta < 1$).

Pour un test bilatéral, on trouve

$$T = T(\alpha,\beta;\theta) \sim \left(\frac{k(\alpha,\beta)\sigma_x}{\sigma_g\theta}\right)^2 \tag{4}$$

où $k(\alpha,\beta)$ est tel que $P(|N(k(\alpha,\beta),1)| < \gamma_{1-(\alpha/2)}) = \beta$, où

$$\int_{-\infty}^{\gamma_x} \frac{1}{\sqrt{2\pi}} e^{\frac{-y^2}{2}}\,dy = x$$

et où le symbole " $\sim$ " indique que le quotient asymptotique (quand $\theta \to 0$) des deux termes qu'il sépare est un. Si α et β sont petits, $k(\alpha,\beta)$ est convenablement approximé par $\gamma_{1-(\alpha/2)} + \gamma_{1-\beta}$.

La formule (4) n'est utilisable que pour les petites valeurs de $|\theta|$, ce qui est suffisant pour le calcul de l'efficacité relative asymptotique des différents tests que nous considérerons.

Lorsque la véritable fonction g est inconnue, on peut quand même tester $H : \theta = 0$ en effectuant une régression où on utilise une fonction h arbitraire, soumise aux mêmes restrictions que g . L'estimateur $\hat{\theta}$ qu'on obtient alors de la formule (1) où g est remplacé par h est alors biaisé et on trouve

$$E(\hat{\theta}) = \frac{T\sum((\delta+\theta g(\frac{t}{T}))h(\frac{t}{T})) - (\sum(\delta+\theta g(\frac{t}{T})))(\sum h(\frac{t}{T}))}{T\sum h^2(\frac{t}{T}) - (\sum h(\frac{t}{T}))^2} \sim \frac{\theta\sigma_{gh}}{\sigma_h^2} \tag{5}$$

où

$$\sigma_{gh} = \int_0^1 g(x)h(x)\ dx - (\int_0^1 g(x)\ dx)(\int_0^1 h(x)\ dx)\ .$$

La valeur qu'on doit alors utiliser pour

$$S_x^2 = \frac{1}{T-2} \sum_{t=1}^{T} (Y_t - \hat{\delta} - \hat{\theta}h(\frac{t}{T}))^2$$

a tendance à surestimer légèrement le véritable σ_x^2 mais ce biais est de l'ordre de θ^2 et, pour les petites valeurs de $|\theta|$, il peut être négligé. De l'expression (5) on trouve alors, pour un test bilatéral de niveau et de puissance $1 - \beta$,

$$T = T(\alpha,\beta;\theta,g,h) \sim (\frac{k(\alpha,\beta)\sigma_x\sigma_h}{\sigma_{gh}\theta})^2 \ . \tag{6}$$

2. <u>Deux tests non paramétriques</u>

Plusieurs tests non paramétriques existent qui sont excellents pour détecter les tendances monotones. Pour les deux tests les plus usuels (le test des rangs et le test des médianes), nous obtiendrons une expression donnant, en fonction de θ (petit), de g et de la loi des X_t, la taille T requise pour que le test soit de niveau $\alpha > 0$ et de puissance $1 - \beta < 1$ désirés (avec $\alpha + \beta < 1$).

a) <u>Le test des rangs</u>. Soit R_t le rang obtenu par Y_t en ordonnant, de la plus petite à la plus grande, les T observations $Y_1,Y_2,\ldots,Y_T$. On pose alors

$$R = \sum_{t=1}^{T} (t-R_t)^2 \quad \text{et} \quad S = \sum_{t=1}^{T} tR_t\ .$$

Les statistiques R et S, liées linéairement, donnent le même test. On obtient aisément

$$E(S|H) = \frac{T(T+1)^2}{4} \quad \text{et} \quad Var(S|H) = \frac{T^2(T^2-1)(T+1)}{144} \tag{7}$$

(voir Hájek & Šidák [1], p. 114) et la normalité asymptotique est très rapide.

Si les Y_t ont tendance à croître, S aura tendance à être supérieur à sa valeur espérée sous H. Inversement, une petite valeur de S est statistiquement indicatrice d'une tendance décroissante. Le test bilatéral de niveau α s'obtient en convenant de rejeter H si $|S - \mu_s| > \sigma_s \gamma_{1-(\alpha/2)}$.

Sous l'alternative,

$$E(S) = \sum_{t=1}^{T} t(1 + \sum_{\substack{s=1\\ s\neq t}}^{T} P(Y_s < Y_t))$$

$$= \sum_{t=1}^{T} t(1 + \sum_{\substack{s=1\\ s\neq t}}^{T} P(X_s < X_t + \theta(g(\tfrac{t}{T}) - g(\tfrac{s}{T})))) \ . \tag{8}$$

Or, pour $s \neq t$ et pour $|\theta|$ très petit, $P(X_s < X_t + \theta(g(\frac{t}{T}) - g(\frac{s}{T})))$ se réduit essentiellement à $\frac{1}{2} + \theta(g(\frac{t}{T}) - g(\frac{s}{T})) \int_{-\infty}^{\infty} f_x^2(x)\, dx$.

De l'équation (8) on obtient donc

$$E(S) - E(S|H) \sim \theta \int_{-\infty}^{\infty} f_x^2(x)\, dx \sum_{t=1}^{T} (t \sum_{s=1}^{T} (g(\tfrac{t}{T}) - g(\tfrac{s}{T}))$$

$$\sim \theta T^3 A(f_x) B(g) \tag{9}$$

où

$$A(f_x) = \int_{-\infty}^{\infty} f_x^2(x)\, dx \quad \text{et} \quad B(g) = \int_0^1 xg(x)\, dx - \tfrac{1}{2} \int_0^1 g(x)\, dx = \sigma_{xg} \ .$$

La non-nullité de θ affecte aussi la variance de S mais, pour les petites valeurs de $|\theta|$, la formule donnée en (7) est satisfaisante et la normalité asymptotique n'est pas menacée. De (9) on obtient donc, pour le test des rangs,

$$T(\alpha,\beta;\theta) \sim \left(\frac{k(\alpha,\beta)}{12\theta A(f_x) B(g)}\right)^2 \ . \tag{10}$$

b) Le test des médianes. Dans le test des médianes, on compte simplement le nombre d'observations Y_t qui, dans la seconde moitié de l'échantillon, sont supérieures à la médiane générale de ces observations. Si T est pair,

$$Z = \sum_{t=\frac{T}{2}+1}^{T} V_t \quad \text{où} \quad V_t = \begin{cases} 1 & \text{si} \quad Y_t > \text{méd}(Y_{t'}) \\ 0 & \text{si} \quad Y_t \leq \text{méd}(Y_{t'}) \end{cases} .$$

Sous l'hypothèse $H : \theta = 0$, on trouve

$$E(Z) = \frac{T}{4} \quad \text{et} \quad \text{Var}(Z) = \frac{T^2}{16(T-1)} . \tag{11}$$

Si T est impair, on permet à Z de prendre des valeurs fractionnaires, ce qui a un effet négligeable sur les formules (11). Encore une fois, la normalité de Z est très rapide.

Si $\theta \neq 0$, $E(V_t)$ dépendra de t . Il faut d'abord déterminer la valeur limite de $\text{méd}(Y_t)$, pour les petites valeurs de θ . Il est ici plus commode de considérer une variable aléatoire Y , obtenue sous la forme $Y = \delta + \theta g(U) + X$ où U est distribué uniformément sur l'intervalle $(0,1)$.

Pour tout $a \in R$, on a alors

$$P(Y \leq a) = E(F_x(a - \delta - \theta g(U)) \mid U) .$$

Si F_x est dérivable en $a - \delta$ et si $|\theta|$ est suffisamment petit, on obtient

$$P(Y \leq a) - F_x(a - \delta) \sim \theta \mu_g f_x(a - \delta) .$$

En particulier, si f_x est continue et positive au point $\text{Méd}(X) = F_x^{-1}(\frac{1}{2})$ et si $|\theta|$ est petit, on trouve

$$\text{Méd}(Y) - \text{Méd}(X) - \delta \sim \theta \mu_g .$$

Pour les grandes valeurs de T on obtient alors

$$\begin{aligned} E(V_t) = P(Y_t > \text{méd}(Y_{t'})) &\simeq P(Y_t > \text{Méd}(Y)) \\ &\simeq P(X_t > \text{Méd}(X) - \theta(g(\tfrac{t}{T}) - \mu_g)) \\ &\simeq \tfrac{1}{2} + \theta f_x(\text{Méd}(X))(g(\tfrac{t}{T}) - \mu_g) \end{aligned}$$

et

$$E(Z) \simeq T\int_{\frac{1}{2}}^{1} \tfrac{1}{2} + \theta f_X(\text{Méd}(X))(g(x) - \mu_g)\ dx$$
$$= \frac{T}{4} + \frac{\theta T C(f_X) D(g)}{4}$$

où

$$C(f_X) = f_X(F_X^{-1}(\tfrac{1}{2})) \quad \text{et où} \quad D(g) = 2\left(\int_{\frac{1}{2}}^{1} g(x)\ dx - \int_0^{\frac{1}{2}} g(x)\ dx\right)$$

est simplement la différence entre les valeurs moyennes prises par g de part et d'autre du point $\frac{1}{2}$.

Comme dans le cas précédent, l'effet de la non-nullité de θ sur la variance de Z peut, si $|\theta|$ est petit, être négligé. Utilisant (11), on obtient donc, pour le test des médianes,

$$T(\alpha,\beta;\theta) \sim \left(\frac{k(\alpha,\beta)}{\theta C(f_X) D(g)}\right)^2 . \qquad (12)$$

3. Efficacité relative asymptotique et cas particuliers

L'efficacité relative asymptotique d'un test τ_1 , par rapport à un autre test τ_2 est le quotient limite, quand $\theta \to 0$, des tailles échantillonnales requises par chacun de ces tests. En général ce quotient limite, s'il existe, ne dépend pas de α et β (pourvu que $\alpha + \beta < 1$) .

$$\text{ERA}(\tau_1;\tau_2) = \lim_{\theta \to 0} \frac{T_2(\alpha,\beta;\theta)}{T_1(\alpha,\beta;\theta)} \qquad (13)$$

(voir Lehmann [3], p. 239).

Selon le choix de g , h et de f_X , les test qu'on vient de présenter sont d'efficacité relative asymptotique très variable.

En ce qui concerne la fonction g (ou h) monotone, définie sur l'intervalle $[0,1]$, avec $g(0) = 0$ et $g(1) = 1$, les choix les plus naturels sont:

□ Tendance linéaire: $g_1(x) = x$.

□ Tendance à deux niveaux: $g_2(x) = \begin{cases} 0 & \text{si } 0 \le x < c \\ 1 & \text{si } c \le x \le 1 \end{cases}$ où $0 < c < 1$.

□ Tendance parabolique: $g_3(x) = ax^2 + (1-a)x$ ou $|a| \le 1$.

Cette tendance est convexe si $0 < a \le 1$, concave si $-1 \le a < 0$ et se réduit à la tendance linéaire si $a = 0$.

Dans le tableau qui suit on trouvera, pour chacune de ces trois tendances, les valeurs de $B(g)$ et de $D(g)$ ainsi que celles de σ_g^2 et des σ_{gh_i} où $h_i = g_i$.

	g_1	g_2	g_3
$B(g)$	$\frac{1}{12}$	$\frac{c(1-c)}{2}$	$\frac{1}{12}$
$D(g)$	$\frac{1}{2}$	$1 - \|2c-1\|$	$\frac{1}{2}$
σ_g^2	$\frac{1}{12}$	$c(1-c)$	$\frac{a^2+15}{180}$
σ_{gh_1}	$\frac{1}{12}$	$\frac{c_g(1-c_g)}{2}$	$\frac{1}{12}$
σ_{gh_2}	$\frac{c_h(1-c_h)}{2}$	$\min\{c_g,c_h\}(1-\max\{c_g,c_h\})$	$\frac{c_h(1-c_h)}{6}(3+a_g(2c_h-1))$
σ_{gh_3}	$\frac{1}{12}$	$\frac{c_g(1-c_g)}{6}(3+a_h(2c_g-1))$	$\frac{a_g a_h+15}{180}$

Tableau 1. Contribution de g *dans le calcul de* T .

En ce qui concerne la loi des X_t , les choix les plus naturels sont:

□ Loi normale: $f_1(x) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma_x^2}}$.

□ Loi uniforme: $f_2(x) = \begin{cases} \frac{1}{2\sqrt{3}\sigma_x} & \text{si } |x| < \sqrt{3}\sigma_x \\ 0 & \text{sinon} \end{cases}$.

□ Loi triangulaire isocèle: $$f_3(x) = \begin{cases} \dfrac{\sqrt{6}\sigma_x - |x|}{6\sigma_x^2} & \text{si } |x| < \sqrt{6}\sigma_x \\ 0 & \text{sinon .} \end{cases}$$

□ Loi double-exponentielle: $$f_4(x) = \frac{1}{\sqrt{2}\sigma_x} e^{-\frac{\sqrt{2}|x|}{\sigma_x}} .$$

Pour chacune de ces lois, le tableau suivant donne les valeurs de $A(f)$ et $C(f)$.

	f_1	f_2	f_3	f_4
$A(f)$	$\frac{1}{2\sqrt{\pi}\sigma_x}$	$\frac{1}{2\sqrt{3}\sigma_x}$	$\frac{2}{3\sqrt{6}\sigma_x}$	$\frac{1}{2\sqrt{2}\sigma_x}$
$C(f)$	$\frac{1}{\sqrt{2\pi}\sigma_x}$	$\frac{1}{2\sqrt{3}\sigma_x}$	$\frac{1}{\sqrt{6}\sigma_x}$	$\frac{1}{\sqrt{2}\sigma_x}$

Tableau 2. Contribution de f_x *dans le calcul de* T .

En introduisant, dans les formules 6, 10 et 12, les valeurs obtenues des Tableaux 1 et 2, on compare aisément l'efficacité asymptotique des quelques tests que nous avons présentés.

Dans le Tableau 3, pour faciliter les comparaisons, on a préféré la représentation décimale à la représentation algébrique exacte. Selon la loi des X_t et selon la tendance réelle g , l'un ou l'autre des tests est asymptotiquement supérieur.

REMARQUE 1. Si les X_t sont de loi uniforme (f_2) alors, quelle que soit la tendance réelle g , le test des rangs est asymptotiquement équivalent en efficacité au test de régression linéaire (h_1) . En effet, les formules (6) et (10) donnent toutes deux

$$T(\alpha,\beta;\theta) \sim \frac{1}{12} \left(\frac{k(\alpha,\beta)\sigma_x}{\sigma_{gx}\theta}\right)^2 .$$

		TENDANCE RÉELLE g					
		$g = g_1$	$g = g_2$, $c_g = \frac{1}{2}$	$g = g_2$, $c_g = \frac{1}{3}$	$g = g_2$, $c_g = \frac{2}{3}$	$g = g_3$, $a_g = -1$	$g = g_3$, $a_g = 1$
TEST DES RANGS	$f_x = f_1$	0,9549	0,7162	0,6366	0,6366	0,8952	0,8952
	$f_x = f_2$	1,0000	0,7500	0,6667	0,6667	0,9375	0,9375
	$f_x = f_3$	0,8889	0,6667	0,5926	0,5926	0,8333	0,8333
	$f_x = f_4$	1,5000	1,1250	1,0000	1,0000	1,4063	1,4063
TEST DES MÉDIANES	$f_x = f_1$	0,4775	0,6366	0,3183	0,3183	0,4476	0,4476
	$f_x = f_2$	0,2500	0,3333	0,1667	0,1667	0,2344	0,2344
	$f_x = f_3$	0,5000	0,6667	0,3333	0,3333	0,4688	0,4688
	$f_x = f_4$	1,5000	2,0000	1,0000	1,0000	1,4063	1,4063
TEST DE RÉGRESSION	$h = h_1$	1,0000	0,7500	0,6667	0,6667	0,9375	0,9375
	$h = h_2$, $c_h = \frac{1}{2}$	0,7500	1,0000	0,5000	0,5000	0,7031	0,7031
	$h = h_2$, $c_h \equiv \frac{1}{3}$	0,6667	0,5000	1,0000	0,2500	0,7716	0,4938
	$h = h_2$, $c_h = \frac{2}{3}$	0,6667	0,5000	0,2500	1,0000	0,4938	0,7716
	$h = h_3$, $a_h = -1$	0,9375	0,7031	0,7716	0,4938	1,0000	0,7656
	$h = h_3$, $a_h = 1$	0,9375	0,7031	0,4938	0,7716	0,7656	1,0000

Tableau 3. E.R.A. des tests par rapport au test de régression où $h = g$.

REMARQUE 2. Si les X_t sont de loi double-exponentielle (f_4) et si $g = g_2$ (avec $c_g = \frac{1}{3}$ ou $\frac{2}{3}$), alors les *trois* tests (où $h = g$ dans le test de régression) sont asymptotiquement également efficaces. Chaque test donne

$$T(\alpha,\beta;\theta) \sim \frac{9}{2}\left(\frac{k(\alpha,\beta)\sigma_x}{\theta}\right)^2 .$$

REMARQUE 3. Si la tendance réelle est à deux niveaux (g_2) et si la loi des X_t est normale ou uniforme, alors le test des rangs est asymptotiquement plus efficace que le test des médianes. Si les X_t sont de loi triangulaire, le test des rangs l'emporte encore sur le test des médianes sauf si $c_g = \frac{1}{2}$ où les deux tests sont asymptotiquement également efficaces. Si les X_t sont de loi double-exponentielle, le test des médianes l'emporte sur le test des rangs pour $\frac{1}{3} < c_g < \frac{2}{3}$.

REMARQUE 4. Si la tendance réelle est parabolique (g_3, incluant le cas linéaire), alors le test des rangs est asymptotiquement plus efficace que le test des médianes pour des X_t de loi normale, uniforme ou triangulaire. Si les X_t sont de loi double-exponentielle, alors ces deux tests sont asymptotiquement également efficaces.

4. Note sur la paramétrisation

Il y a souvent plus d'une façon de paramétriser une tendance. Par exemple, si elle est linéaire, il est plus naturel de considérer le saut unitaire θ' plutôt que le saut total $\theta = T\theta'$. L'efficacité relative asymptotique des tests s'en trouve modifiée. Les formules (6), (10) et (12) deviennent alors

$$T \sim \left(\frac{2\sqrt{3}k(\alpha,\beta)\sigma_x}{\theta'}\right)^{2/3} \quad \text{pour le test de régression linéaire,}$$

$$T \sim \left(\frac{k(\alpha,\beta)}{A(f_x)\theta'}\right)^{2/3} \quad \text{pour le test des rangs,}$$

$$T \sim \left(\frac{2k(\alpha,\beta)}{C(f_x)\theta'}\right)^{2/3} \quad \text{pour le test des médianes.}$$

Par exemple, si les X_t sont de loi normale, on obtient que l'efficacité relative asymptotique du test des rangs par rapport au test de régression linéaire est $\sqrt[3]{3/\pi} \simeq 0{,}9847$, résultat donné dans Kendall & Stuart [2], p. 358.

De même, sous les mêmes conditions, l'efficacité relative asymptotique du test des médianes par rapport au test de rangs est $\sqrt[3]{\frac{1}{2}} \simeq 0{,}7937$.

Le choix d'une paramétrisation plutôt qu'une autre n'a d'autre effet que de comprimer (ou dilater) les efficacités relatives sans que l'ordre des tests n'en soit affecté.

Dans le cas linéaire que nous venons de considérer, l'utilisation de θ' donne des efficacités relatives asymptotiques qui sont les racines cubiques de celles qu'on obtient en paramétrisant plutôt le saut total θ .

Bibliographie

[1] HÁJEK, J. et ŠIDÁK, Z., *Theory of Rank Tests*, Academic Press (1967).

[2] KENDALL, M.G. et STUART, A., *The Advanced Theory of Statistics*, Vol. 3, Seconde édition, Griffin (1968).

[3] LEHMANN, E.L., *Testing Statistical Hypotheses*, Wiley (1959).

Département de mathématiques
Université du Québec à Montréal
C.P. 8888, Succ. "A"
Montréal, Québec
H3C 3P8

Manuscrit reçu le 9 septembre 1980.
Revisé le 23 mars 1982.