ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 20/21 — 1991

Une nouvelle forme du test de la matrice d'information

Russell DAVIDSON, James G. MacKINNON *

RÉSUMÉ. – Nous développons une nouvelle forme du test de la matrice d'information pour une large classe de modèles statistiques, en particulier nous détaillons complètement le cas de la régression linéaire univariée. Chesher a montré que l'hypothèse alternative implicite pour le test de la matrice d'information est un modèle où les paramètres varient aléatoirement. Nous utilisons cette propriété pour construire un test contre une alternative explicite de ce type. Le nouveau test est calculé grâce à une régression artificielle dite « Double-Longueur », au lieu de la régression classique sur le produit extérieur du gradient qui, tout en étant facile à utiliser, admet une distribution exacte différente de la distribution asymptotique même pour un échantillon de taille importante. La nouvelle forme se comporte parfaitement au moins dans le contexte des modèles de régression. Des distributions à distance finies approximées sont calculées. Elles justifient l'emploi de la nouvelle forme du test.

A New Form of a Matrix Test.

ABSTRACT. – We develop a new form of the information matrix test for a wide variety of statistical models, and present full details for the special case of univariate nonlinear regression models. Chesher (1984) showed that the implicit alternative of the information matrix test is a model with random parameter variation. We exploit this fact by constructing the test against an explicit alternative of this type. The new test is computed using a double-length artificial regression, instead of the more conventional outer product of the gradient regression, which although easy to use, is known to give test statistics with distributions very far from the asymptotic nominal distribution even in rather large samples. The new form on the other hand performs remarkably well, at least in the context of regressions models. Some approximate finite-sample distribution are calculated and lend support to the use of the new form of the test.

R. DAVIDSON : GREQE-EHESS, Hospice de la Vieille Charité, 2, rue de la Charité, 13002 Marseille; J. G. MACKINNON : Dept of Economics, Queen's University, Kingston, Ontario, K7L3N6, Canada. Ce travail a été réalisé en partie grâce au concours du Social Sciences and Humanities Research Council of Canada. Nous avons discuté une version antérieure de cet article avec Andrew Chesher, et nous avons bénéficié des remarques des participants des séminaires de l'Université d'Oxford et du Econometric Study Group. Cet article a été traduit en français par G. LANOT.

1 Introduction

L'égalité de la Matrice d'Information entre d'une part, l'espérance du produit extérieur des gradients, et moins l'espérance du Hessien d'autre part, est utilisée depuis longtemps dans la théorie du maximum de vraisemblance. Récemment WHITE [1982] a suggéré son utilisation comme base pour un test de spécification. Le test proposé par White n'est pas facile à calculer. CHESHER [1983] et LANCASTER [1984] ont montré comment calculer une version asymptotiquement équivalente à l'aide d'une régression « artificielle ». Cette régression est une variante de la régression sur le produit extérieur des gradients (régression Outer Product of Gradient) popularisé par GODFREY et WICKENS [1981] comme une forme du test du multiplicateur de Lagrange. Malheureusement toutes les études disponibles montrent que le test basé sur une régression OPG tend à rejeter trop souvent l'hypothèse nulle pour les échantillons finis. Plus récemment KENNAN et NEUMANN [1988] ainsi que CHESHER et SPADY [1988] ont montré que la forme OPG du test de la Matrice d'Information se comporte particulièrement mal. Ces derniers proposent une nouvelle forme du test basée sur des corrections jusqu'à l'ordre n^{-1} (n = taille de l'échantillon) d'un développement de type Edgeworth de la fonction de répartition de la statistique de test. Puisque cette forme du test est inévitablement plus difficile à calculer que le test basé sur une seule régression artificielle, on est amené à se demander s'il existe une méthode alternative facilement calculable du test de la Matrice d'Information (pour la suite : test IM), et si cette méthode alternative a des propriétés à distance finies supérieures aux tests existant basés sur la régression OPG.

Dans cet article nous développons une forme alternative du test IM, qui est applicable presque aussi généralement que le test basé sur la régression OPG. Nous explicitons cette nouvelle forme dans le cas du modèle de régression univariée. Ce test est basé sur une régression artificielle « Double-Longueur » (DLR) du test du multiplicateur de Lagrange (LM) proposé par DAVIDSON et MACKINNON [1984 *a*]. Nous commençons par construire une hypothèse alternative explicite pour laquelle les paramètres du modèle sont aléatoires. CHESHER [1984] a montré que cette alternative est l'hypothèse alternative implicite du test IM. Pour créer un test automatiquement équivalent au test IM il suffit de dériver classiquement la forme DLR du test LM contre cette alternative.

Dans la section 2 nous décrivons la forme OPG traditionnelle du test IM et son application au modèle de régression. La section 3 contient les principaux résultats théoriques de l'article. Ensuite nous présentons dans la section 4 des simulations qui suggèrent que la nouvelle forme DLR du test IM se comporte beaucoup mieux que la forme OPG; en particulier quand le nombre de paramètres est suffisamment grand. Ces résultats corroborent ceux de KENNAN et NEUMANN [1983] et de CHESHER et SPADY [1988] : même quand la taille de l'échantillon est grande les performances du test IM sous la forme OPG sont faibles, contrairement aux performances du test IM sous la forme DLR. Enfin, dans la section 5, nous réalisons des développements de tests à un degré de liberté liés au test IM mais dans leur forme $\mathcal{N}(0, 1)$. Les développements considérés par Chesher et Spady étaient réalisés à partir de statistiques sous leur forme conventionnelle de χ^2; on note plus clairement la cause des différences de comportements à distance finie des deux formes de tests.

2 La forme OPG du test IM pour les modèles de régression

Considérons le modèle de régression nonlinéaire univarié :

$$(1) \qquad y_t = \xi_t(\beta) + u_t, \qquad u_t \sim \mathcal{N}(0, \sigma^2),$$

où $\xi_t(\beta)$ est une fonction deux fois différentiable qui dépend de β un vecteur de paramètres de dimension k; l'indice t indique que cette fonction peut aussi dépendre de variables exogènes (généralement ce sera le cas) qui varient suivant les observations. Pour ce modèle la contribution à la fonction de log-vraisemblance de l'observation t s'écrit :

$$l_t(\beta, \sigma) = \frac{1}{2}\log(2\pi) - \log(\sigma) - \frac{(y_t - \xi_t(\beta))^2}{2\sigma^2}.$$

Ainsi la contribution de l'observation t au gradient pour l'élément i de β est :

$$(2) \qquad G_{ti}(\beta, \sigma) = (y_t - \xi_t(\beta))\, Z_{ti}(\beta)\frac{1}{\sigma^2},$$

où

$$Z_{ti}(\beta) = \frac{\partial \xi_t(\beta)}{\partial \beta_i}.$$

Quand (1) est un modèle de régression linéaire simple, $Z(\beta)$ est simplement la matrice des régresseurs. Nous supposerons que la matrice $n^{-1} Z'(\beta) Z(\beta)$ tend vers une matrice positive quand le nombre d'observations, n, tend vers l'infini.

La contribution de l'observation t au gradient pour σ s'écrit :

$$(3) \qquad G_{t,k+1}(\beta, \sigma) = -\frac{1}{\sigma} + (y_t - \xi_t(\beta))^2 \frac{1}{\sigma^3}.$$

Les éléments généraux de la matrice Hessienne de $l_t(\beta, \sigma)$ s'écrivent :

$$\frac{\partial^2 l_t(\beta, \sigma)}{\partial \sigma^2} = \frac{1}{\sigma^3} - 3(y_t - \xi_t(\beta))^2 \frac{1}{\sigma^4}. \tag{4}$$

$$\frac{\partial^2 l_t(\beta, \sigma)}{\partial \sigma \, \partial \beta_i} = -2(y_t - \xi_t(\beta)) Z_{ti}(\beta) \frac{1}{\sigma^3}. \tag{5}$$

$$\frac{\partial^2 l_t(\beta, \sigma)}{\partial \beta_i \, \partial \beta_j} = -Z_{ti}(\beta) Z_{tj}(\beta) \frac{1}{\sigma^2} + (y_t - \xi_t(\beta)) Z^*_{tij} \frac{1}{\sigma^2}, \tag{6}$$

où $Z^*_{tij} = \dfrac{\partial Z_{ti}(\beta)}{\partial \beta_j}$.

La forme OPG du test IM se calcule à l'aide de la somme des carrés expliqués de la régression artificielle d'un vecteur de 1 sur les k régresseurs correspondant aux β_i, un régresseur correspondant à σ et au plus $\frac{1}{2}(k+2)$ $(k+1)$ régresseurs spécifiques du test dont la forme générale est :

$$\frac{\partial^2 l_t(\theta)}{\partial \theta_i \, \partial \theta_j} + \frac{\partial l_t(\theta)}{\partial \theta_i} \frac{\partial l_t(\theta)}{\partial \theta_i}, \tag{7}$$

$i = 1, \ldots, k+1$; $j = 1, \ldots, i$. θ désigne le vecteur $(\beta', \sigma)'$. Ces éléments sont évalués en $\tilde{\theta}$ l'estimateur des moindres carrés de θ. Le test statistique suit asymptotiquement un χ^2 ayant autant de degrés de liberté que de régresseurs tests, au maximum $\frac{1}{2}(k+2)(k+1)$.

A partir des expressions (2) à (6), il est facile d'exprimer les régresseurs de la version OPG du test IM. Définissons :

$$\tilde{e}_t = \frac{(y_t - \xi_t(\tilde{\beta}))}{\tilde{\sigma}}, \qquad \tilde{Z}_{ti} = Z_{ti}(\tilde{\beta}).$$

Remarquons de plus quel que soit le régresseur un changement d'échelle n'aura aucune influence sur l'ajustement de la régression. Les régresseurs peuvent alors s'exprimer de la façon suivante :

(8) pour β_i. $\tilde{e}_t \tilde{Z}_{ti}$.

(9) pour σ : $\tilde{e}_t^2 - 1$.

(10) pour $\beta_i \times \beta_j$: $(\tilde{e}_t^2 - 1) \tilde{Z}_{ti} \tilde{Z}_{tj} + \tilde{e}_t \tilde{Z}^*_{tij}$.

(11) pour $\sigma \times \beta_i$: $(\tilde{e}_t^3 - 3\tilde{e}_t) \tilde{Z}_{ti}$.

(12) pour $\sigma \times \sigma$: $\tilde{e}_t^4 - 5\tilde{e}_t^2 + 2$.

Si le modèle de départ contient un terme constant, (10) est parfaitement colinéaire avec (9) quand i et j indicent la constante. Dans ce dernier cas le terme (9) doit être ignoré. Le nombre de degré de liberté pour le test devient $\frac{1}{2}(k+2)(k+1) - 1$.

Les expressions (8) à (12) montrent clairement quelles formes de mauvaises spécifications sont testées par le test IM. HALL [1987]. En considérant le test appliqué au modèle linéaire, ses résultats s'étendent à une exception près au cas étudié ici. (11) montre clairement que les régresseurs associés aux couples (σ, β_i) correspondent à l'effet croisé de la dissymétrie et des $\tilde{Z}_{ti}$. Si l'on soustrait (9) de (12) on obtient $\tilde{e}_t^4 - 6\,\tilde{e}_t^2 + 3$. Ainsi la partie linéairement indépendante du régresseur associé au couple (σ, σ) teste l'applatissement. Si $\xi_t(\beta)$ est linéaire, le régresseur associé au couple (β_i, β_j) (cf. : (10)) teste l'hétéroscédasticité contre laquelle le test de WHITE [1980] est construit (c'est-à-dire l'hétéroscédasticité qui affecte l'estimation convergente de la matrice de variance covariance des estimateurs des moindres carrés). Dans le cas du modèle de régression non linéaire étudié ici, ces régresseurs testent contre la mauvaise spécification de la fonction de régression, puisque $1/n$ fois le second terme de (10), sommé sur tous les t, n'admet pas zéro comme limite en probabilité si la fonction n'est pas correctement spécifiée.

Le test IM dans le cadre du modèle de régression peut s'interpréter comme testant l'hétéroscédasticité, l'asymétrie, l'applatissement et la forme fonctionnelle (dans le cas de la régression non linéaire).

3 La forme DLR du test IM

Nous dérivons tout d'abord une forme explicite de l'alternative locale implicite du test IM pour tout modèle pouvant s'exprimer sous une forme compatible avec la construction de la régression artificielle « Double-Longueur » de DAVIDSON et MACKINNON [1984*a*]. Comme l'a établit CHESHER [1984], l'alternative implicite du test IM correspond à l'hétérogénéité due à des paramètres aléatoires. Une expression explicite de « l'alternative implicite » nous permet de construire la régression DLR pour mener le test IM. Nous décrivons en détail la procédure dans le cas du modèle de régression non linéaire (1).

Tout d'abord nous calculons le vecteur du score pour les directions testées par le test IM pour les modèles de la forme :

$$(13) \qquad f_t(y_t, \theta) = \varepsilon_t \qquad t = 1 \ldots n \qquad \varepsilon \sim \mathcal{N}(0, \mathrm{I}_n),$$

où :

θ : paramètres du modèle tels que $\theta \in \Theta \subset \mathbb{R}$.

$f_t(.,.)$ est deux fois continuement différentiable par rapport à θ.

$f_t(.,.)$ est au moins une fois continuement différentiable par rapport à y_t.

$\dfrac{\partial f_t(.,.)}{\partial y_t}$ est deux fois continuement différentiable par rapport à θ.

La contribution de l'observation t à la log-vraisemblance s'écrit :

$$l_t = -\frac{1}{2}\log(2\pi) - \frac{1}{2}f_t^2 + k_t,$$

où

(14) $$k_t \equiv \log|f_t'|,$$

est le jacobien de la transformation f.

Les éléments génériques du gradient et du hessien s'écrivent :

(15) $$l_{ti} = -f_t \mathrm{F}_{ti} + \mathrm{K}_{ti}; \qquad l^*_{tij} = -\mathrm{F}_{ti}\mathrm{F}_{tj} - f_t \mathrm{F}^*_{tij} + \mathrm{K}^*_{tij},$$

où

$$\mathrm{F}_{ti} = \frac{\partial f_t(.,.)}{\partial \theta_i}; \qquad \mathrm{K}_{ti} = \frac{\partial k_t(.,.)}{\partial \theta_i};$$

$$l^*_{tij} = \frac{\partial^2 l_t}{\partial\theta_i \partial\theta_j}; \qquad \mathrm{F}^*_{tij} = \frac{\partial^2 f_t}{\partial\theta_i \partial\theta_j}; \qquad \mathrm{K}^*_{tij} = \frac{\partial^2 k_t}{\partial\theta_i \partial\theta_j}.$$

On déduit de (7) qu'une direction générale testée par le test IM est donnée par $l^*_{tij} + l_{ti} l_{tj}$. En utilisant (15), cette quantité vaut :

(16) $$-\mathrm{F}_{ti}\mathrm{F}_{tj} - f_t \mathrm{F}^*_{tij} + \mathrm{K}^*_{tij} + f_t^2 \mathrm{F}_{ti}\mathrm{F}_{tj} - f_t(\mathrm{F}_{ti}\mathrm{K}_{tj} + \mathrm{F}_{tj}\mathrm{K}_{ti}) + \mathrm{K}_{ti}\mathrm{K}_{tj}.$$

Ces directions sont celles induites par l'hétérogénéité due aux paramètres aléatoires. Ce qui suggère le modèle suivant :

(17) $$f_t(y_t, \theta + \eta_t) = \varepsilon_t; \qquad \varepsilon_t \sim \mathcal{N}(0, 1); \qquad \eta_t \sim \mathcal{N}(0, \bar{\Omega})$$

où η_t est un vecteur aléatoire de dimension k distribué indépendamment de ε. Un développement de Taylor de (17) au second ordre en η_t amène :

(18) $$f_t(y_t, \theta) + \mathrm{F}_t \eta + \frac{1}{2}\eta' \mathrm{F}^*_t \eta \simeq \varepsilon_t.$$

F_t est un vecteur ligne dont l'élément générique est F_{ti} et F^*_t est une matrice $k \times k$ dont l'élément générique est F^*_{tij}.

On a de plus :

$$\mathbb{E}\left[\mathrm{F}_t \eta + \frac{1}{2}\eta' F^*_t \eta / y_t\right] = \frac{1}{2}\mathrm{Tr}\,\bar{\Omega}\,\mathrm{F}^*_t.$$

$$\mathbb{V}\left[\mathrm{F}_t \eta + \frac{1}{2}\eta' \mathrm{F}^*_t \eta / y_t\right] = \mathrm{Tr}\,\bar{\Omega}\,\mathrm{F}_t' \mathrm{F}_t.$$

Localement autour de $\bar{\Omega} = 0$ le modèle est équivalent à :

(19) $$q(y, \theta, \Omega) \equiv \frac{f_t(y_t, \theta) - \mathrm{Tr}\,\Omega\,\mathrm{F}^*_t}{(1 + 2\,\mathrm{Tr}\,\Omega\,\mathrm{F}_t' \mathrm{F}_t)^{1/2}} = \nu_t,$$

pour un vecteur de perturbation v, normalement distribué, d'espérance nulle et de variance I_n, et pour $\Omega \equiv \frac{\bar{\Omega}}{2}$.

A première vue, le modèle (17) n'est pas approprié pour tester l'hypothèse $\Omega=0$, puisque celle-ci se situe sur la frontière des valeurs possibles pour Ω, et ainsi une des hypothèses habituelles de régularité n'est pas vérifiée. Comme la matrice Ω est proportionnelle à une matrice de variance-covariance, elle doit être, *a priori*, semi-definie positive. Ce n'est pas un problème pour deux raisons :

– Tout d'abord, si les tests de Wald et du rapport de vraisemblance ne sont pas applicables pour des valeurs des paramètres sur la frontiére de l'ensemble admissible, le test du Multiplicateur de Lagrange peut l'être (CHANT [1974]). Dans ces circonstances, le test du Multiplicateur de Lagrange perd généralement ses propriétés d'optimalité.

– De plus, même si le modèle (17) n'a aucun sens si $\bar{\Omega}$ n'est pas semi définie positive, le modèle approché (19), quant à lui, est défini pour toute matrice Ω suffisamment proche de la matrice 0, le test du Multiplicateur de Lagrange conserve dans ce cas ces propriétés d'optimalité. Il est donc possible d'utiliser la théorie usuelle pour tester $\Omega=0$.

La régression DLR va être construite à partir de la fonction $q_t(.)$ définie par (19) pour tester l'hypothèse $\Omega=0$. Nous devons calculer pour (19) les termes analogues à F_{ti} et K_{ti} en $\Omega=0$.

Le premier de ces termes s'obtient aisément; l'analogue de F_{ti} s'écrit :

$$Q_{tij} \equiv \frac{\partial q_t}{\partial \Omega_{ij}} = F^*_{tij} - f_t F_{ti} F_{tj}, \qquad i=1, \ldots, k, \qquad j=1, \ldots, i.$$

Il y a deux indices car Ω est une matrice $k \times k$, mais seul le triangle inférieur nous intéresse. L'analogue de K_{ti}, $R_{tij} \equiv \frac{\partial q'_t}{\partial \Omega_{ij}}$, est plus difficile à obtenir, les détails du calculs se trouvent dans l'appendice 1. On a :

$$R_{tij} = K^*_{tij} + K_{ti} K_{tj} - f_t (F_{ti} K_{tj} + F_{tj} K_{ti}) - F_{ti} F_{tj}.$$

De (15) nous déduisons que les vecteurs du score associés à la direction Ω, c'est-à-dire $-q_t Q_{tij} + R_{tij}$, sont de la forme :

$$-F_{ti} F_{tj} - f_t F^*_{tij} + K^*_{tij} + f_t^2 F_{ti} F_{tj} - f_t (F_{ti} K_{tj} + F_{tj} K_{ti}) + K_{ti} K_{tj}.$$

On remarque qu'ils sont identiques aux directions (16), qui sont inspectées par le test IM.

Les résultats de DAVIDSON et MACKINNON [1984*a*] permettent de déduire facilement la régression DLR pour tester $\Omega=0$. Cette régression comporte $2n$ observations. Les endogènes sont f_t pour les observations t et 1 pour les observations $t+n$. Les régresseurs associés à θ sont F_t pour les observations t et K_t pour les observations $t+n$. Le régresseur correspondant à Ω_{ij} est :

$$Q_{tij} = f_t F_{ti} F_{tj} - F^*_{tij}$$

pour les observations t et

$$R_{tij} = K^*_{tij} + K_{ti} K_{tj} - f_t (F_{ti} K_{tj} + F_{tj} K_{ti}) - F_{ti} F_{tj}$$

pour les observations $t+n$.

Toutes ces quantités sont évaluées sous l'hypothèse nulle $\Omega=0$. Le test statistique est constitué par la somme des carrés expliqués de cette régression artificielle. Cette statistique constitue l'écriture complète du test IM sous sa forme DLR pour les modèles du type (13). Quand l'hypothèse nulle est le modèle de régression linéaire (1), certaines simplifications sont possibles. Notons à nouveau $\tilde{\beta}$ et $\tilde{\sigma}$ les estimateurs du maximum de vraisemblance, et

$$\tilde{e}_t \equiv \frac{(y_t - \xi_t(\tilde{\beta}))}{\tilde{\sigma}}.$$

Nous obtenons ainsi la régression artificielle décrite par les éléments suivants :

(21) variable dépendante : $\{\tilde{e}_t;\ 1\}$

régresseurs :

(22)	pour β_i :	$\{\tilde{Z}_{ti};\ 0\}$.
(23)	pour σ :	$\{\tilde{e}_t;\ -1\}$.
(24)	pour $\beta_i x \beta_j$:	$\{\tilde{e}_t \tilde{Z}_{ti} \tilde{Z}_{tj} - \tilde{Z}^*_{tij};\ -\tilde{Z}_{ti}\tilde{Z}_{tj}\}$.
(25)	pour $\sigma x \beta_i$:	$\{(\tilde{e}_t^2 - 1)\tilde{Z}_{ti};\ -2\tilde{e}_t \tilde{Z}_{ti}\}$.
(26)	pour $\sigma x \sigma$:	$\{\tilde{e}_t - 3\tilde{e}_t;\ 3 - 3\tilde{e}_t\}$.

Le test statistique est égal à $2n$ moins la somme des carrés des résidus de cette régression artificielle. Pour simplifier certaines des expressions ci-dessus, nous avons utilisé le fait que prendre des combinaisons linéaires des régresseurs n'a aucun effet sur le pouvoir explicatif d'une régression. Notons que lorsque $\xi_t(\beta)$ contient un terme constant, le terme (24) qui correspond au terme constante × constante, est colinéaire avec le terme (23). Il doit donc être ôté de la régression.

On remarque aisément que la régression artificielle DLR teste les mêmes directions que la régression OPG décrite par les équations (8) à (12). (Il faut considérer pour cela le produit scalaire de (21) avec chacun des régresseurs.) Les $\frac{1}{2}k(k+1)$ $\left(\text{ou les } \frac{1}{2}k(k+1)-1\right)$ régresseurs (24) testent l'hétéroscedasticité et la forme fonctionnelle de la même façon que le test proposé par White [1980].Les k régresseurs (25) quant à eux testent l'interaction de l'applatissement et des $\tilde{Z}_{ti}$ et enfin, le régresseur (26) teste l'asymétrie.

4 Performance du test à distance finie

Dans cette section nous présentons une expérience de Monte-Carlo relativement simple, destinée à montrer s'il existe des différences systématiques entre les formes DLR et OPG du test IM, pour le modèle de régression simple. La différence entre les deux tests est frappante : à la différence de la forme DLR, la forme OPG rejette trop souvent l'hypothèse nulle correcte. Nous avons examiné les performances des tests uniquement sous l'hypothèse nulle. La forme OPG rejette si souvent qu'aucune comparaison de puissance n'a de sens.

Nous considérons les modèles linéaires de la forme :

$$y_t = \beta_0 + \sum_{i=1}^{k-1} \beta_i X_{ti} + u_t, \qquad u_t \sim \mathcal{N}(0, \sigma).$$

Dans notre expérience k varie de 2 à 6. Les X_{ti} sont des variables aléatoires distribuées normalement indépendamment entre les observations et équicorrélées, de corrélation égale à un demi. Nous avons généré seulement un ensemble de X_{ti} pour les cent premières observations; quand la taille dépasse cent, ces cent observations ont été répétées autant de fois que nécessaire. Ceci permet d'augmenter la taille de l'échantillon sans que $\dfrac{X'X}{n}$ en soit affectée. (Nous avons fait quelques essais pour lesquels toutes les observations étaient générées à partir du même processus aléatoire, les résultats obtenus étaient comparables à ceux présentés dans le tableau 1.) On voit que les formes OPG et DLR du test IM sont invariants à la fois à β et σ, ces paramètres ont donc été choisis arbitrairement.

Les résultats pour des échantillons de taille 100, 200, 500 et 1 000, basés sur 2 000 répliques sont présentés dans le tableau 1. Les résultats sont étonnants. Le test OPG rejette l'hypothèse nulle beaucoup trop souvent : l'hypothèse nulle n'est jamais rejetée moins de quinze fois sur cent ($k=2$, $n=1\,000$) pour une taille de 5 %, elle est même rejetée dans 96 % des cas pour $k=6$ et $n=100$. Le test ne rejette plus l'hypothèse nulle à mesure que la taille de l'échantillon augmente. Cette dernière propriété est prédite par la théorie asymptotique. La première propriété est beaucoup plus importante pour les résultats de travaux appliqués, où le nombre de régresseurs est généralement supérieur à 6.

Bien que le nombre de régresseurs pour cette expérience soit plutôt faible, le nombre de degrés de liberté du test IM peut être quelquefois assez grand : 27 par exemple pour $k=6$. Ce nombre est bien plus grand que ce qui a déjà été étudié dans des expériences précédentes s'intéressant au test OPG. Ceci explique certainement que ce test fonctionne beaucoup moins bien que ce que d'autres auteurs ont rapporté. La détérioration des résultats à mesure que k augmente, suggère que la forme OPG du test IM fonctionne, en pratique, encore moins bien que ce que nous montre les résultats du tableau 1.

Ce dernier phénomène n'affecte pourtant pas autant la forme DLR du test IM. La fréquence de rejet est toujours raisonnablement près de la taille du test, sauf pour les échantillons de taille 100 et bien que l'on suspecte aussi une divergence par rapport à la distribution asymptotique du χ^2 pour $n=200$. Pour $n=100$ le test DLR ne rejette pas assez pour $k=3$, alors qu'il rejette trop souvent pour $k=5$ et $k=6$.

TABLEAU 1

Performance des formes OPG et DLR du test IM sous l'hypothèse nulle

Fréquence de rejet suivant la taille :

k	d.f.	n	Test	Mean	S.D.	10 %	5 %	1 %
2	5	100	OPG	13.82**	10.05**	59.6**	51.1**	35.1**
			DLR	5.13	3.09	10.3	5.1	0.9
		200	OPG	10.85**	9.07**	43.8**	36.1**	23.4**
			DLR	4.89	3.06*	9.1	4.3	0.7
		500	OPG	8.30**	7.00**	32.1**	24.0**	13.6**
			DLR	5.05	3.24	10.6	5.2	1.3
		1 000	OPG	6.80**	5.78**	22.2**	15.0**	6.9**
			DLR	4.90	3.20	9.2	5.2	1.0
3	9	100	OPG	23.43**	11.23**	76.4**	68.1**	50.5**
			DLR	8.84	3.83**	7.4**	3.1**	0.7
		200	OPG	19.77**	11.03**	62.2**	53.1**	35.8**
			DLR	9.01	4.04*	9.9	4.5	0.6*
		500	OPG	15.41**	9.15**	43.3**	35.1**	21.5**
			DLR	8.99	4.15	9.2	4.1	0.7
		1 000	OPG	12.63**	7.40**	30.9**	21.7**	11.0**
			DLR	8.88	4.16	8.7	4.4	0.9
4	14	100	OPG	34.68**	12.37**	87.5**	80.7**	63.6**
			DLR	14.25*	5.10*	9.5	4.7	0.9
		200	OPG	31.47**	13.29**	77.8**	69.5**	50.3**
			DLR	13.99	4.97**	9.2	3.7*	0.6*
		500	OPG	24.64**	11.78**	55.0**	45.2**	28.3**
			DLR	13.83	5.04*	9.2	3.9*	0.5*
		1 000	OPG	20.24**	9.56**	38.4**	28.9**	15.2**
			DLR	13.83	5.19	9.7	4.6	1.1
5	20	100	OPG	47.43**	12.10**	95.2**	91.6**	77.5**
			DLR	21.29**	6.40	12.75**	7.2**	1.8**
		200	OPG	45.90**	15.63**	89.2**	83.1**	67.2**
			DLR	20.69**	6.22	10.2	5.2	1.5*
		500	OPG	36.50**	14.00**	69.2**	59.0**	39.9**
			DLR	20.04	6.03*	9.0	4.4	0.7
		1 000	OPG	29.92**	11.64**	48.7**	38.7**	22.6**
			DLR	19.93	6.22	10.2	5.0	1.0
6	27	100	OPG	59.68**	10.92**	98.3**	96.3**	86.2**
			DLR	29.40**	7.15	15.2**	7.8**	1.4
		200	OPG	60.85**	16.58**	94.8**	91.3**	77.8**
			DLR	27.84**	7.20	11.1	5.4	1.0
		500	OPG	51.44**	16.69**	81.0**	72.8**	55.6**
			DLR	27.15	7.16	10.0	4.9	0.9
		1 000	OPG	42.90**	14.53**	60.7**	51.9**	33.8**
			DLR	27.10	7.37	10.8	5.1	1.2

- Tous les résultats sont déduits de 2 000 répliques.
- * et ** indiquent que la quantité est significativement différente de sa valeur asymptotique au niveau de 5 % et de 1 %.

Si l'on suit les résultats issus du tableau 1, la forme OPG du test IM rejette si souvent qu'il ne faut certainement jamais l'utiliser à moins que la taille de l'échantillon soit réellement importante. D'autre part la forme DLR du test se comporte remarquablement bien et peut certainement être utilisée dès que la taille atteint quelques centaines. Bien entendu, ces résultats sont sujets aux réserves habituelles formulées pour ce type d'expériences. Bien qu'il soit possible que d'autres expériences montrent que la forme DLR n'est pas aussi performante, il est difficile d'imaginer que les performances de la forme OPG du test IM s'améliorent.

La faiblesse des performances du test OPG à distance finie a été remarquée récemment par CHESHER et SPADY [1988], KENNAN et NEUMANN [1988] et ORME [1990 *a*]. En particulier, KENNAN et NEUMANN [1988] ont obtenu des résultats à partir d'une expérience de Monte-Carlo, pour une autre forme du test IM, où le test rejette d'autant plus que le nombre de régresseurs augmente, moins nettement pourtant que la forme OPG dans notre expérience. Ils montrent que la valeur de la statistique de la forme OPG du test IM est toujours supérieure à la valeur de la statistique pour leur forme du test. Ces résultats sont cohérents avec les nôtres. Orme propose une manière de calculer le test IM qui améliore les propriétés à distance finie. Cette approche est plus générale que celle que nous avons utilisée ici, mais est aussi plus difficile à appliquer. Orme a discuté le calcul du test IM pour des modèles dichotomiques (ORME [1988]), et récemment (ORME [1990 *b*]) il a proposé des régressions doubles et triples pour calculer une version du test, pour un ensemble plus restreint de modèles que celui qui est considéré ici.

5 Développements asymptotiques de quelques statistiques de test

Les résultats de la section précédente sont extrêmement clairs, mais n'expliquent pas les raisons des comportements différents à distance finie des formes OPG et DLR du test IM. Dans cette section nous tentons d'expliquer ces raisons. Afin de simplifier les calculs, nous ne considèrerons qu'un modèle très simple, et nous limiterons notre attention à l'une des directions qui est testée par le test IM : l'applatissement (kurtosis). Cette seconde restriction permet d'étudier la statistique de test sous une forme qui admet une loi normale centrée réduite comme distribution asymptotique sous l'hypothèse nulle de spécification correcte du modèle. Cette forme est plus facile à étudier qu'un test du Khi deux à plusieurs degrés de liberté. De plus le rôle de l'asymétrie pour une statistique asymptotiquement normale apparaît clairement. Cette propriété pouvant être confondue avec un biais quand la statistique est élevée au carré pour obtenir une forme suivant un Khi deux.

Le modèle que nous considérons est :

$$y_t = \mu + u_t, \qquad u_t \sim \mathcal{N}(0, \sigma), \qquad \mathbb{E}[u_t u_s] = 0,\ t \neq s.$$

Le fait que la fonction de régression n'est ici qu'une constante est une simplification importante pour ce qui suit. CHESHER et SPADY [1988] ont montré que le test contre l'applatissement que nous considérons est invariant a des variations de la fonction de régression à l'ordre n^{-1}. Puisque nos résultats finals ne sont vrais que pour cet ordre, nous ne perdrons pas de généralité à nous restreindre à cette forme simple. Pour construire la forme DLR du test, nous réécrivons le modèle sous la forme :

$$f_t(y_t, \mu, \sigma) \equiv \frac{y_t - \mu}{\sigma} = \varepsilon_t \qquad \text{avec} \quad \varepsilon_t \sim \mathcal{N}(0, 1). \tag{27}$$

Nous en déduisons que $k_t = \log |f_t'| = -\log \sigma$. La régression DL pour tester dans la direction (σ, σ) peut s'écrire de la forme suivante :
variable dépendante :

$$(\tilde{f}_t, 1);$$

régresseurs :

$$\begin{gathered}(1, 0) \text{ pour } \mu,\\ (\tilde{f}_t, -1) \text{ pour } \sigma,\\ (\tilde{f}_t^3 - 3\tilde{f}_t, 3 - 3\tilde{f}_t^2) \text{ pour la direction du test.}\end{gathered} \tag{28}$$

Le modèle (27) est d'abord estimé par maximum de vraisemblance, ce qui donne les mêmes estimateurs que les moindres carrés ordinaires mis à part le dénominateur de l'estimateur de σ^2. Ainsi sous l'hypothèse nulle,

$$\tilde{\mu} = \frac{1}{n} \sum_{t=1}^{n} y_t = \mu + \frac{1}{n} \sum_{t=1}^{n} f_t, \tag{29}$$

et

$$\tilde{\sigma} = \frac{1}{n} \sum_{t=1}^{n} (y_t - \tilde{\mu})^2 = \frac{1}{n} \sum_{t=1}^{n} \left(f_t - \frac{1}{n} \sum_{s=1}^{n} f_s\right)^2. \tag{30}$$

Nous introduisons maintenant les notations qui seront utilisées dans le reste de la section. Soit :

$$h_t \equiv n^{-1/2} \sum_{t=1}^{n} \mathrm{He}_i(f_t). \tag{31}$$

$\mathrm{He}_i(.)$ dénote le polynôme d'Hermite de degré i. Ces polynômes sont définis par :

$$\mathrm{He}_{i+1}(x) = x\,\mathrm{He}_i(x) - \mathrm{He}_i'(x); \qquad \mathrm{He}_i(x) = x.$$

avec $\mathrm{He}_i'(x) = \dfrac{d\mathrm{He}_i(x)}{dx}$.

Ainsi les premiers éléments de la suite sont :

$$\mathrm{He}_1(x) = x, \qquad \mathrm{He}_2(x) = x^2 - 1, \qquad \mathrm{He}_3(x) = x^3 - 3x$$
$$\mathrm{He}_4(x) = x^4 - 6x^2 + 3.$$

Les polynômes d'Hermites sont orthogonaux, dans le sens suivant :

$$\mathbb{E}[\mathrm{He}_i(x)\,\mathrm{He}_j(x)] = 0 \qquad \text{si} \quad x \sim \mathcal{N}(0, 1),$$

on a de plus la normalisation suivante :

$$\mathbb{E}[(\mathrm{He}_i(x))^2] = i!.$$

Enfin, si $x \sim \mathcal{N}(0, 1)$ alors $\mathbb{E}[\mathrm{He}_i(x)] = 0$, $\forall i$. Puisque sous l'hypothèse nulle $f_t \sim \mathcal{N}(0, 1)$, les variables aléatoires définies en (31) admettent les propriétés suivantes :

$$\mathbb{E}[h_i] = 0; \qquad \mathbb{V}[h_i] = i!; \qquad \mathrm{Cov}[h_i, h_j] = 0,\ i \neq j. \tag{32}$$

D'après leur définition, les h_i sont évidemment asymptotiquement normales. Les équations (29) et (30) peuvent s'écrire en fonction des h_i de la façon suivante :

$$\tilde{\mu} = \mu + n^{-1/2} h_1 \sigma;$$
$$\tilde{\sigma} = \sigma(1 + n^{-1/2} h_2 - n^{-1} h_1^2).$$

Les $\tilde{f}_t$ qui sont utilisés pour la régression DL peuvent ainsi se réécrire jusqu'à l'ordre $n^{-3/2}$ de la façon suivante :

$$\begin{aligned} \tilde{f}_t = \frac{y_t - \tilde{\mu}}{\tilde{\sigma}} &= f_t - n^{-1/2}\left(h_1 + \frac{1}{2} h_2 f_t\right) \\ &+ \frac{1}{8} n^{-1}(4 h_1^2 f_t + 3 h_2^2 f_t + 4 h_1 h_2) \\ &- \frac{1}{16} n^{-3/2}(8 h_1^3 + 6 h_1 h_2^2 + 12 h_1^2 h_2 f_t + 5 h_2^3 f_t). \end{aligned} \tag{33}$$

Quelle que soit la version du test d'applatissement considérée, le numérateur s'écrit : (comme on peut le voir à partir de la spécification (28))

$$\sum_{t=1}^{n} (\tilde{f}_t^4 - 3).$$

Si l'on remplace $\tilde{f}_t$ dans l'expression ci-dessus par la dernière égalité de (33), et si l'on divise le résultat par $n^{1/2}$ afin d'obtenir une expression dont l'ordre du terme principal est d'ordre 1 en probabilité, nous obtenons :

$$\begin{aligned} n^{-1/2} \sum_{t=1}^{n} (\tilde{f}_t^4 - 3) &= h_4 - n^{-1/2}(4 h_1 h_3 + 2 h_2 h_4 + 3 h_2^2) \\ &+ n^{-1}(2 h_1^2 h_4 + 12 h_1^2 h_2 + 3 h_2^2 h_4 + 6 h_2^3 + 8 h_1 h_2 h_3) + 0(n^{-3/2}). \end{aligned} \tag{34}$$

Suspendons un instant notre analyse de la statistique de test DLR afin de considérer une statistique plus simple qui s'impose immédiatement. Cette dernière statistique est simplement la statistique du test du score efficient :

$$(24n)^{-1/2} \sum_{t=1}^{n} (\tilde{f}_t^4 - 3), \tag{35}$$

pour laquelle l'estimateur de la variance du numérateur $\sum_{t=1}^{n} (\tilde{f}_t^4 - 3)$ est obtenu par la matrice d'information pour le modèle (27) étendu pour admettre la possibilité d'un applatissement non normal, et évaluée à l'aide des estimateurs du maximum de vraisemblance sous l'hypothèse nulle. La matrice information pertinente est bloc diagonale entre les paramètres associés à la moyenne et à la variance de la variable dépendante et entre les paramètres associés aux moments d'ordre plus élevés. Ainsi seul un élément de la matrice d'information est nécessaire, de plus cet élément est indépendant des paramètres du modèle.

Le test (35) est très comparable à ceux proposés par plusieurs auteurs pour tester l'applatissement de la distribution des résidus dans un modèle de régression (voir par exemple KIEFER et SALMON [1983]). En étudiant tout d'abord ce simple test, nous pouvons distinguer aisément ce qui caractérise l'approximation à distance finie de chaque test que nous allons considérer.

Nous allons utiliser la caractéristique associée à la statistique (35). Rappelons tout d'abord que pour une variable aléatoire quelconque x, la fonction caractéristique est définie par : $\varphi(t) \equiv \mathbb{E}[\exp(itx)]$. Cette fonction existe toujours si x est réelle. Pour une variable aléatoire suivant une loi normale centrée et réduite la fonction caractéristique est : $\exp(-t^2/2)$. Bien que le premier terme dans (34), h_4, soit d'espérance nulle et de variance unitaire, sa distribution asymptotique seule est normale. Afin d'obtenir sa fonction caractéristique jusqu'à l'ordre n^{-1} (nous ne travaillerons que jusqu'à cet ordre), nous suivrons les preuves élémentaires du théorème central limite. Supposons que x_t dénote une variable aléatoire identiquement et indépendamment distribuée selon une $\mathcal{N}(0, 1)$. Nous avons :

$$
\begin{aligned}
\mathbb{E}[\exp\{ith_4\}] &= \mathbb{E}\left[\exp\left\{itn^{-1/2}\sum_{t=1}^{n}\mathrm{He}_4(x_t)\right\}\right] \\
&= \mathbb{E}\left[\prod_{t=1}^{n}\exp\{itn^{-1/2}\mathrm{He}_4(x_t)\}\right] \\
&= \prod_{t=1}^{n}\mathbb{E}[\exp\{itn^{-1/2}\mathrm{He}_4(x_t)\}] \\
&= (\mathbb{E}[\exp\{itn^{-1/2}\mathrm{He}_4(x_t)\}])^n.
\end{aligned}
$$

Le logarithme de la fonction caractéristique s'appelle la fonction génératrice des moments. Dans notre cas elle s'écrit :

$$
\begin{aligned}
& n\log(\mathbb{E}[\exp\{itn^{-1/2}\mathrm{He}_4(x_t)\}]) \\
&= n\log\mathbb{E}[1+itn^{-1/2}\mathrm{He}_4(x_t)-\frac{t}{2}n^{-1}(\mathrm{He}_4(x_t))^2 \\
&\quad -\frac{it^3}{6}n^{-3/2}(\mathrm{He}_4(x_t))^3+\frac{t^4}{24}n^{-2}(\mathrm{He}_4(x_t))^4+0(n^{-5/2})] \qquad (36)\\
&= n\log\mathbb{E}[1-12n^{-1}t^2-288n^{-3/2}it^3+15\,264n^{-2}t^4+0(n^{-5/2})] \\
&= -12t^2-288n^{-1/2}it^3+15\,264n^{-1}t^4+0(n^{-3/2}).
\end{aligned}
$$

La ligne précédent (36) utilise les espérances des puissances de $\mathrm{He}_4(x_t)$; celles-ci peuvent être calculées en les exprimant comme une combinaison linéaire de polynômes d'Hermite plus une constante, ce qui amène le résultat désiré puisque les espérances de tous les polynômes d'Hermite sont nulles. La dernière ligne résulte d'un développement de Taylor du logarithme.

En prenant l'exponentielle de (36) on obtient l'approximation désirée de la fonction caractéristique :

$$\exp(-12t^2)(1-288n^{-1/2}it^3+15\,264n^{-1}t^4-41\,472n^{-1}t^6+0(n^{-3/2})).$$

Pour les calculs suivants il est nécessaire de remplacer t dans cette dernière expression par $\dfrac{t}{\sqrt{24}}$. On a :

$$(37) \qquad \varphi(t)=\exp(-t^2/2)(1-\sqrt{6}n^{-1/2}it^3+(53/2)n^{-1}t^4-3n^{-1}t^6),$$

où nous avons omis le symbole, $0(n^{-3/2})$, pour alléger les notations. A partir de maintenant ce symbole est sous entendu. Le logarithme de cette fonction caractéristique est du même ordre que :

$$(38) \qquad \log\varphi(t)=-\frac{t^2}{2}-\frac{it^3}{6}(3\sqrt{24}n^{-1/2})+\frac{t^4}{24}(636n^{-1}).$$

A l'aide de la fonction caractéristique (37), nous pouvons étudier la fonction caractéristique de l'expression approximée du membre de droite de (34). Si l'on note A le terme proportionnel à $n^{-1/2}$ et B le terme proportionnel à n^{-1} dans (34), l'équation devient :

$$h_4+n^{-1/2}\mathrm{A}+n^{-1}\mathrm{B}.$$

La fonction caractéristique peut se développer de la façon suivante :

$$\begin{aligned}
&\mathbb{E}[\exp\{it(h_4+n^{-1/2}\mathrm{A}+n^{-1}\mathrm{B})\}]\\
&\quad=\mathbb{E}[\exp\{ith_4\}\exp\{itn^{-1/2}\mathrm{A}+itn^{-1}\mathrm{B}\}]\\
&\quad=\mathbb{E}[\exp\{ith_4\}(1+itn^{-1/2}\mathrm{A}+itn^{-1}\mathrm{B}-t^2n^{-1}\mathrm{A}^2)]\\
&\quad=\varphi(\sqrt{24}\,t)+(itn^{-1/2}\,\mathbb{E}[\mathrm{A}\exp\{ith_4\}]+itn^{-1}\,\mathbb{E}[\mathrm{B}\exp\{ith_4\}]\\
&\qquad-t^2n^{-1}\,\mathbb{E}[\mathrm{A}^2\,exp\{ith_4\}]).
\end{aligned}$$

Ce qui peut être calculé en évaluant les espérances du produit de $\exp\{ith_4\}$ avec un produit de h_i. Nous détaillons maintenant le calcul dans le cas le plus simple, pour lequel il y a un seul h_i, puis nous donnerons simplement le résultat pour les cas plus compliqués.

Considérons l'espérance $\mathbb{E}[h_i \exp\{ith_4\}]$, avec $i \neq 4$. De la définition des h_i nous déduisons :

$$\begin{aligned}
&\mathbb{E}[h_1 \exp\{ith_4\}] \\
&\quad = \mathbb{E}\left[n^{-1/2} \sum_{t=1}^{n} \mathrm{He}_i(x_t) \prod_{s=1}^{n} \exp\{itn^{-1/2}\,\mathrm{He}_4(x_s)\}\right] \\
&\quad = n^{-1/2} \sum_{t=1}^{n} \mathbb{E}\left[\mathrm{He}_i(x_t) \prod_{s=1}^{n} \exp\{itn^{-1/2}\,\mathrm{He}_4(x_s)\}\right] \\
&\quad = n^{-1/2} \sum_{t=1}^{n} \mathbb{E}\left[\mathrm{He}_i(x_t) \exp\{itn^{-1/2}\,\mathrm{He}_4(x_t)\} \prod_{\substack{s=1 \\ s\neq t}}^{n} \exp\{itn^{-1/2}\,\mathrm{He}_4(x_s)\}\right].
\end{aligned}$$

A l'intérieur du dernier produit tous les facteurs sont identiques, et il n'est pas difficile de voir qu'en inclure un autre, pour $s=t$, induit un changement d'ordre n^{-1} uniquement. Puisque, comme nous allons le voir, l'expression entière est d'ordre $n^{-1/2}$, nous pouvons remplacer le dernier produit par $\varphi(\sqrt{24}\,t)$. Enfin en développant $\exp\{itn^{-1/2}\,\mathrm{He}_4(x_s)\}$ et en utilisant les propriétés des polynômes d'Hermite, nous obtenons le résultat :

$$\mathbb{E}[h_i \exp\{ith_4\}] = -n^{-1/2}\,\varphi(\sqrt{24}\,t)\frac{t^2}{2}\,\mathbb{E}[\mathrm{He}_i(x)(\mathrm{He}_4(x)])^2]. \tag{39}$$

Des manipulations identiques permettent d'obtenir les résultats suivant :

$$\begin{aligned}
\mathbb{E}[h_1 h_j \exp\{ith_4\}] = \varphi(\sqrt{24}\,t)\big(&-(24)^2 t^2 \delta_{i4}\delta_{j4} \\
&- 12\,n^{-1/2}\, it^3\, \mathbb{E}[\mathrm{He}_i(x)\,\mathrm{He}_j(x)\,\mathrm{He}_4(x)](\delta_{i4}+\delta_{j4}) \\
&+ \mathbb{E}[\mathrm{He}_i(x)\,\mathrm{He}_j(x)]\,\delta_{ij} \\
&+ it\,n^{-1/2}\,\mathbb{E}[\mathrm{He}_i(x)\,\mathrm{He}_j(x)\,\mathrm{He}_4(x)]\big),
\end{aligned} \tag{40}$$

$$\begin{aligned}
\mathbb{E}[h_i h_j h_k \exp\{ith_4\}] = \varphi(\sqrt{24}\,t)\big((24)^3&(-it^3)\,\delta_{i4}\delta_{j4}\delta_{k4} \\
&+ 24\,it\,\mathbb{E}[\mathrm{He}_i(x)\,\mathrm{He}_j(x)]\,\delta_{ij}\,\delta k4 \\
&+ 2 \text{ autres termes en permutant } (i, j, k)\big);
\end{aligned} \tag{41}$$

$$\begin{aligned}
\mathbb{E}[h_i h_j h_k h_l \exp\{ith_4\}] = \varphi(\sqrt{24}\,t)\big((24)^4&\, t^4\,\delta_{i4}\delta_{j4}\delta_{k4}\delta_{l4} \\
&+ (24)^2(it^2)\,\mathbb{E}[\mathrm{He}_i(x)\,\mathrm{He}_j(x)]\,\delta_{ij}\delta_{k4}\delta_{l4} \\
&+ 5 \text{ autres termes} \\
&+ \mathbb{E}[\mathrm{He}_i(x)\,\mathrm{He}_j(x)]\,\mathbb{E}[\mathrm{He}_k(x)\,\mathrm{He}_l(x)]\,\delta_{ij}\delta_{kl} \\
&+ 2 \text{ autres termes}\big).
\end{aligned} \tag{42}$$

Armés de ces résultats, nous pouvons calculer l'approximation de la fonction caractéristique de (35). Nous remarquons à partir de (39), (40), (41) et (42) que $\varphi(\sqrt{24}\,t)$ est un facteur commun à toute l'expression. Le reste du calcul est long et technique, mais peut être réalisé heureusement à l'aide d'un ordinateur. La réponse est donnée en fonction de la fonction génératrice des cumulants (f.g.c.) plutôt qu'en terme de la fonction caractéristique, car non seulement elle est plus simple, mais son interprétation est plus immédiate. La f.g.m. s'écrit :

$$(43) \qquad it\left(-\frac{1}{4}\sqrt{24}\,n^{-1/2}\right)-\frac{1}{2}t^2(1-15n^{-1})-\frac{1}{6}it^3(3\sqrt{24}\,n^{-1/2}) \\ +\frac{t^4}{24}(540(n^{-1})).$$

Rappelons que la f.g.c. est appelée ainsi car dans son développement de Taylor dans les puissances de t, le coefficient de $(it)^k/k!$ est le k-ème cumulant de la distribution. D'après (43) seuls les quatre premiers cumulants de la distribution de la statistique (35) sont différents des valeurs associées à la distribution normale centrée réduite à l'ordre n^{-1}. Ainsi tous les moments d'ordre supérieur sont nuls pour cet ordre d'approximation. De plus les moments d'ordre impair sont perturbés à l'ordre $n^{-1/2}$ et non à l'ordre n^{-1}, alors que les moments d'ordre pair sont perturbés à l'ordre n^{-1}. C'est une caractéristique partagée par tous les développements asymptotiques semblables à ceux considérés ici.

Le biais de la statistique, qui est son espérance, est déduit de (43). Il est égal à : $-\frac{1}{4}\sqrt{24}\,n^{-1/2}$. L'utilisation d'une statistique admettant asymptotiquement une distribution normale centrée réduite, nous a permis de trouver le signe du biais. Nous remarquons que le test est biaisé vers un applatissement négatif. La variance de la statistique est, à l'ordre que nous avons considéré ici, $1-15n^{-1}$. *Ceteris paribus*, une variance inférieure à 1 fait que le test rejette trop peu l'hypothèse nulle. Le coefficient d'asymétrie, le troisième moment, est égal à $3\sqrt{24}\,n^{-1/2}$ et est positif. Tandis que le coefficient d'applatissement $540(n^{-1})$ montre que la distribution sera leptokurtique, c'est-à-dire aura des queues de distribution plus épaisses que la distribution normale. *Ceteris paribus*, le test aura tendance à trop rejeter l'hypothèse nulle.

Le comportement de la statistique dépend ainsi de la taille de l'échantillon et de la taille du test choisie. Nous n'allons pas ici détailler plus le problème, mais nous allons décrire les sources des corrections de la distribution d'ordre $n^{-1/2}$ et n^{-1}. L'examen de la f.g.c. de h_4, (38), montre que les contributions non normales de $\mathrm{He}_4(x)$ à la distribution de h_4 induisent une asymétrie de $3\sqrt{24}\,n^{-1/2}$ et un applatissement de $636n^{-1}$. Ainsi l'utilisation des résidus plutôt que les « vraies » erreurs dans la statistique (35) n'amène aucun changement pour l'asymétrie et baisse l'applatissement. Le biais et la réduction de la variance sont bien entendu attribuables à l'utilisation des résidus.

Nous pouvons conclure notre étude de la statistique DLR. Il serait fastidieux de dériver l'expression qui correspond à (34) pour la statistique

DLR. $\sqrt{24}$ fois la statistique DLR s'écrit :

$$\begin{aligned}
h_4 - n^{-1/2} & \left(4 h_1 h_3 + 3 h_2^2 + 2 h_2 h_4 + \frac{3}{8} h_4^2 + \frac{1}{48} h_4 h_6\right) \\
& + n^{-1} \left(12 h_1^2 h_2 + 2 h_1^2 h_4 + 8 h_1 h_2 h_3 + 3 h_1 h_3 h_4\right. \\
& + \frac{1}{12} h_1 h_3 h_6 + \frac{1}{8} h_1 h_4 h_5 + 6 h_2^3 \\
& + \frac{21}{4} h_2^2 h_4 + \frac{1}{16} h_2^2 h_6 + \frac{29}{16} h_2 h_4^2 \\
& + \frac{5}{48} h_2 h_4 h_6 + \frac{1}{48} h_3^2 h_4 + \frac{85}{384} h_4^3 \\
& \left. + \frac{3}{128} h_4^2 h_6 + \frac{1}{1\,536} h_4 h_6^2\right).
\end{aligned}$$

Cette dernière expression est nettement plus complexe que l'expression (34)! Heureusement le calcul peut être fait par un logiciel de calcul, comme celui amenant la f.g.c. correspondante :

$$\begin{aligned}
& it\left(-\frac{5}{8}\sqrt{24}\, n^{-1/2}\right) - \frac{1}{2} t^2 \left(1 - \frac{319}{8} n^{-1}\right) \\
& - \frac{1}{6} it^3 \left(\frac{3}{4}\sqrt{24}\, n^{-1/2}\right) + \frac{t^4}{24}\left(\frac{261}{4} n^{-1}\right).
\end{aligned} \tag{44}$$

Le biais est légèrement plus grand que celui de la statistique directe (35), mais l'asymétrie et l'applatissement ont été réduits, comme on peut le voir dans la formule (43). La variance est inférieure à la variance de la statistique directe. Tout cela explique le bon comportement de la statistique DLR, puisque une faible variance à tendance à rendre le test trop conservateur. Il faut noter que la statistique (35), qui utilise comme estimateur de la variance l'espérance du terme approprié de la matrice d'information (l'estimateur est donc certain), n'est pas nécessairement supérieure à la statistique DLR. Cette dernière corrige une partie de la non normalité due à h_4.

Il reste à considérer le cas désespérant de la statistique OPG. On voit facilement que la régression artificielle peut être décrite de la façon suivante :

variable dépendante :

$$1;$$

régresseurs :

$$\begin{aligned}
1 & \quad \text{pour } \mu; \\
\mathcal{J}_t^2 - 1 & \quad \text{pour } \sigma^2; \\
\mathcal{J}_t^4 - 3 & \quad \text{pour la direction du test.}
\end{aligned}$$

Le calcul de la f.g.c. pour la statistique OPG donne :

$$(45)\qquad \begin{aligned} &it\left(-\frac{7}{4}\sqrt{24}\,n^{-1/2}\right)-\frac{1}{2}t^2\,(1+519\,n^{-1}) \\ &-\frac{1}{6}it^3\,(6\sqrt{24}\,n^{-1/2})+\frac{t^4}{24}(1\,698\,n^{-1}). \end{aligned}$$

Tout est plus mauvais que dans la f.g.c. pour la statistique DLR. Le biais est plus important, la variance est nettement plus grande que un, l'asymétrie est multipliée par huit environ, et l'applatissement est énorme. Tout cela est dû à l'utilisation de l'estimateur OPG de la matrice d'information qui est inefficace. Les détails du calcul de (45) montrent que cela vient de l'apparition de variables aléatoires à l'intérieur de cet estimateur jusqu'à h_8, cette variable ayant une variance égale à 8! est particulièrement variable. Contrairement à la statistique DLR, la statistique OPG n'utilise pas toute l'information sur le modèle. En particulier, cette dernière utilise le moment empirique d'ordre huit pour estimer le moment d'ordre huit de f_t. Comme l'explique Kennan et Neumann [1988], cette procédure a le désavantage d'amener des estimateurs très imprécis.

6 Conclusion

Nous avons développé une nouvelle manière de calculer le test de la matrice d'information pour des modèles de régression univariée linéaires et non linéaires. La nouvelle forme du test est basée sur une régression artificielle Double-Longueur, qui est une manière de calculer le test du Multiplicateur de Lagrange contre une alternative explicite de variation aléatoire des paramètres. Bien que la forme OPG et la forme DLR du test IM soient asymptotiquement équivalentes, les propriétés à distance finie sont très différentes. La forme OPG habituelle rejette bien trop souvent (dans certain cas tout le temps!), à des niveaux de significativité courants, tandis que la forme DLR se comporte particulièrement bien.

A cause de la complexité des calculs nécessaires, l'étude de la section 5 basée sur des développements limités en probabilité est peu générale. Pourtant cette étude clarifie les caractéristiques des tests qui déterminent les résultats des expériences de Monte-Carlo. Par exemple, même si il n'y a pas de paramètre de nuisance à estimer (c'est-à-dire la moyenne μ et la variance σ^2), on observe alors f_t directement dans l'équation (27), la statistique de test contre l'applatissement (35) est affectée par l'asymétrie et l'applatissement. Ce dernier étant positif, le test aura tendance à trop rejeter. L'utilisation des résidus réduit cet applatissement, alors que l'on introduit un biais et que l'on réduit la variance.

On remarque que la forme DLR du test semble corriger, au moins jusqu'à un certain point, plusieurs de ces caractéristiques indésirables. La baisse significative de la valeur de la variance, associée à une faible valeur de l'asymétrie, semble être la cause principale du bon comportement de cette forme du test. Pour un nombre de degrés de liberté faible, cela fait plus que compenser le biais et l'asymétrie de la statistique de test, de plus le test est plus conservateur. Néanmoins, après l'observation des résultats de l'expérience de Monte-Carlo, il pourrait apparaître que l'arbitrage se fasse dans l'autre sens quand le nombre de degrés de liberté augmente.

Les résultats théoriques expliquent aussi le faible comportement de la forme OPG du test IM. Il semble que cela soit dû principalement à l'utilisation d'un estimateur de la matrice de variance covariance qui est très variable, et ainsi induit une variance trop importante, augmente l'applatissement de façon significative, et accroît largement le biais et l'asymétrie, par rapport à la forme DLR.

Dans le contexte des tests à un degré de liberté, il a été malheureusement impossible de découvrir les raisons qui font que les performances de la forme OPG se détériorent énormément quand le nombre de paramètres augmente, et donc quand le nombre de degrés de liberté augmente, ni les raisons qui font que les performances de la forme DLR du test se détériorent moins dans les mêmes conditions. Ces deux questions seront l'objet d'une recherche future.

Dans cet appendice nous dérivons R_{tij}, l'analogue de K_{ti} pour le modèle (19). La première étape est d'obtenir l'expression (14) pour ce modèle. On a $\log|q_t'|$ qui est égal a :

$$\log\left|\frac{f_t'+\operatorname{tr}(\Omega F_t^{*\prime})}{(1+2\operatorname{tr}(\Omega F_t^{T} F_t))^{1/2}}-\frac{f_t'(f_t+\operatorname{tr}(\Omega F_t^{*}))\operatorname{tr}(\Omega(K_t^{T} F_t+F_t^{T} K_t))}{(1+2\operatorname{tr}(\Omega F_t^{T} F_t))^{3/2}}\right|.$$

D'où :

$$R_{tij}=\frac{1}{f_t'}(F_{tij}^{*}-f_t' F_{ti} F_{tj}-f_t(F_{ti}' F_{tj}+F_{ti} F_{tj}')). \tag{46}$$

Ainsi, puisqu'en général dans le contexte de la forme DLR on a $k_t=\log|f_t'|$, on obtient :

$$K_{ti}=\frac{1}{f_t'}\frac{\partial f_t'}{\partial\theta_i}=\frac{F_t'}{f_t'},$$

d'où on tire $F_{ti}'=f_t' K_{ti}$. De la même façon :

$$F_{tij}^{*}=f_t'(K_{tij}^{*}+K_{ti} K_{tj}).$$

(46) devient alors :

$$R_{tij}=K_{tij}^{*}+K_{ti}K_{tj}-F_{ti}F_{tj}-f_t'(K_{ti}F_{tj}-F_{ti}K_{tj}),$$

ce qui est l'expression (20) dans le texte.

• Références bibliographiques

BERA, A. K. et MCKENZIE, C. R. (1986). – « Alternative Forms and Properties of the Score Test », *Journal of Applied Statistics*, 13, pp. 13-25.

CHANT, D. (1974). – « On Asymptotic Tests of Composite Hypotheses in Nonstandard Conditions », *Biometrika*, 61, pp. 291-298.

CHESHER, A. (1983). – « The Information Matrix Test: Simplified Calculation Via a Score Test Interpretation », *Economics Letters*, 13, pp. 45-48.

CHESHER, A. (1984). – « Testing for Neglected Heterogeneity », *Econometrica*, 52, pp. 865-872.

CHESHER, A. et SPADY, R. (1988). – « Asymptotic Expansions of the Information Matrix Test Statistic », paper presented at June 1988 meeting of *Econometric Study Group*, Bristol.

DAVIDSON, R. et MACKINNON, J. G. (1983). – « Small Sample Properties of Alternative Forms of the Lagrange Multiplier Test », *Economics Letters*, 12, pp. 269-275.

DAVIDSON, R. et MACKINNON, J. G. (1984*a*). – « Model Specification Tests Based on Artificial Linear Regressions », *International Economic Review*, 25, pp. 485-502.

Davidson, R. et MacKinnon, J. G. (1984*b*). – « Convenient Specification Tests for Logit and Probit Models », *Journal of Econometrics*, 25, pp. 241-262.

Davidson, R. et MacKinnon, J. G. (1985). – « Testing Linear and Log-Linear Regressions Against Box-Cox Alternatives », *Canadian Journal of Economics*, 25, pp. 499-517.

Godfrey, L. G., McAleer, M. et McKenzie, C. R. (1988). – « Variable Addition and Lagrange Multiplier Tests for Linear and Logarithmic Regression Models », *Review of Economics and Statistics*, 70, pp. 492-503.

Godfrey, L. G. and Wickens, M. R. (1981). – « Testing Linear and Log-Linear Regressions for Functional Form », *Review of Economic Studies*, 48, pp. 487-496.

Hall, A. (1987). – « The Information Matrix Test for the Linear Model », *Review of Economic Studies*, 54, pp. 257-263.

Keñnan, J. et Neumann, G. R. (1988). – « Why Does the Information Matrix Test Reject too Often? A Diagnosis of Some Monte Carlo Symptoms », Hoover Institution, Stanford University, *Working Papers in Economics E*-88-10.

Kiefer, N. M. et Salmon, M. (1983). – « Testing Normality in Econometric Models », *Economics Letters*, 11, pp. 123-127.

Lancaster, T. (1984). – « The Covariance Matrix of the Information Matrix Test », *Econometrica*, 52, pp. 1051-1053.

Orme, C. (1990*a*). – « The Small Sample Performance of the Information Matrix Test », *Journal of Econometrics*, forthcoming.

Orme, C. (1990*b*). – « Double and Triple Length Regressions for the Information Matrix Test and Other Conditional Moment Tests », University of York, *mimeo*.

White, H. (1980). – « A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity », *Econometrica*, 48, pp. 817-838.

White, H. (1982). – « Maximum Likelihood Estimation of Misspecified Models », *Econometrica*, 50, pp. 1-25.