Propriétés en échantillon fini des tests robustes à l'hétéroscédasticité de forme inconnue

Emmanuel Flachaire

Propriétés en échantillon fini des tests robustes à l'hétéroscédasticité de forme inconnue

Emmanuel Flachaire

EUREQua, Université Paris I Panthéon-Sorbonne

Septembre 2003

Résumé

Dans cet article, nous étudions les propriétés en échantillon fini des tests robustes à l'hétéroscédasticité de forme inconnue, basés sur l'estimateur de la matrice de covariance proposé par Eicker (1963) et White (1980). Nous montrons tout d'abord que, tels qu'ils sont habituellement utilisés dans la pratique, ces tests ne sont pas fiables et pas efficaces, même si la taille de l'échantillon est grande. Nous montrons alors qu'une inférence fiable et beaucoup plus efficace peut être obtenue en petit échantillon, si ces tests sont construits sur la base des résidus contraints et si une version appropriée du bootstrap est utilisée.

Mots-Clés : Hétéroscédasticité, HCCME, Bootstrap, Méthodes de Monte Carlo

Classification JEL : C1

1 Introduction

Dans le cadre d'un modèle de régression linéaire où les régresseurs sont supposés exogènes, l'inférence statistique doit se faire avec précaution si les aléas sont hétéroscédastiques, c'est à dire s'ils n'ont pas tous la même variance. En effet, même si l'estimateur par Moindres Carrés Ordinaires (MCO) des paramètres reste convergent, celui de la matrice de covariance des paramètres estimés ne l'est généralement plus, et les tests basés dessus ne sont plus valides. Le problème a été résolu par Eicker (1963) et White (1980) qui ont proposé un estimateur de la matrice de covariance des paramètres estimés valide asymptotiquement, appelé le HCCME ou *Heteroskedasticity Consistent Covariance Matrix Estimator*. Ce dernier permet d'obtenir une inférence valide en présence d'hétéroscédasticité de forme inconnue dans le modèle. Toutefois, MacKinnon et White (1985) montrent que la distorsion du niveau des tests basés sur cet estimateur peut être importante en échantillon fini, notamment lorsque les régresseurs contiennent des observations à fort effet de levier, voir aussi Chesher et Jewitt (1987).

Les mauvaises propriétés de tests robustes à l'hétéroscédasticité ont naturellement conduit d'autres auteurs à suggérer l'utilisation des méthodes du bootstrap pour corriger la distorsion du niveau en échantillon fini. Les méthodes du bootstrap permettent d'obtenir une autre approximation de la vraie loi d'une statistique de test, que celle donnée par la loi asymptotique. On parle alors d'inférence ou de test bootstrap. Des développements théoriques montrent que souvent, l'approximation donnée par les méthodes du bootstrap est de meilleure qualité que celle donnée par la loi asymptotique, voir Beran (1988) et Davidson et MacKinnon (1999). Il en résulte que la distorsion du niveau des tests peut être largement diminuée et les tests peuvent être de bien meilleure qualité. Davidson et Flachaire (2001) et Godfrey et Orme (2001) montrent que la distorsion du niveau des tests robustes à l'hétéroscédasticité peut être largement réduite et les tests beaucoup plus fiables, si la méthode du wild bootstrap est utilisée. Néanmoins, ces travaux portent seulement sur l'étude de la distorsion du niveau des tests, sans s'intéresser à la puissance.

Un test est d'autant plus performant en échantillon fini qu'il peut discriminer entre l'hypothèse nulle et l'hypothèse alternative. D'une part, si l'hypothèse nulle est vraie, un test est *fiable* s'il rejette rarement cette hypothèse, à un niveau nominal donné (ex : dans 1% ou 5% des cas). D'autre part, si l'hypothèse nulle n'est pas vraie, un test est *efficace* s'il la rejette très fortement. Pour être utilisable en pratique, un test doit avant tout être fiable, c'est la raison pour laquelle l'étude de la distorsion du niveau est fréquente. Néanmoins, lorsqu'on dispose de tests fiables, l'étude de la puissance peut s'avérer très utile, elle permet de comparer l'efficacité entre différents tests.

Dans cet article, nous étudions le niveau et la puissance des tests robustes à l'hétéroscédasticité de forme inconnue. Nous montrons que l'étude de la puissance apporte des résultats nouveaux, intéressants pour l'utilisation de ces tests en pratique. Dans la section 2, nous présentons les tests robustes à l'hétéroscédasticité de forme inconnue. Dans la section 3, nous décrivons le modèle utilisé dans les simulations. Les résultats expérimentaux sur le niveau sont présentés en section 4, ceux sur la puissance en section 5. La section 6 conclut.

2 Inférence asymptotique et bootstrap

Soit le modèle de régression linéaire hétéroscédastique suivant :

$$y_t = X_t \beta + \sigma_t \varepsilon_t \qquad \varepsilon_t \sim \text{IID}(0,1) \tag{1}$$

où y_t est une variable dépendante et X_t un vecteur ligne contenant k régresseurs supposés exogènes, dans le sens où, pour tout t, ces éléments sont indépendants du terme d'erreur $u_t = \sigma_t \varepsilon_t$. Le vecteur des paramètres β, de dimension k, est inconnu ainsi que les variances des aléas σ_t^2, pour $t = 1, \ldots, n$ où n est la taille de l'échantillon. En présence d'hétéroscédasticité, l'estimateur MCO de la matrice de covariance des paramètres estimés est en général biaisé et non-convergent. De ce fait, les tests basés dessus ne sont plus valides.

Test asymptotique

Eicker (1963) et White (1980) ont proposé un estimateur robuste à l'hétéroscédasticité de forme inconnue, ou HCCME (*Heteroskedasticity Consistent Covariance Matrix Estimator*), qui permet de faire de l'inférence asymptotiquement valide sur les paramètres :

$$(X^\top X)^{-1} X^\top \widehat{\Omega} X (X^\top X)^{-1} \tag{2}$$

où $\widehat{\Omega}$ est une matrice diagonale, de taille $n \times n$, qui a pour élément type $a_t^2 \widehat{u}_t^2$, où $\widehat{u}_t$ est le résidu d'une estimation par MCO du modèle (1). MacKinnon et White (1985) considèrent plusieurs formes différentes du HCCME et montrent qu'en échantillon fini les tests basés dessus peuvent être très peu fiables, même lorsqu'on dispose d'un nombre important de données. Chesher and Jewitt (1987) montrent que ce problème est largement dû à la structure des régresseurs, notamment à la présence d'observations influentes ou à fort effets de levier (Davidson et MacKinnon 1993, chapitre 1, ou 2003, chapitre 2). Nous noterons la forme simple du HCCME, telle qu'elle a été proposée par Eicker (1963) et White (1980), HC_0, et les formes alternatives proposées par MacKinnon et White (1985), HC_1, HC_2 et HC_3 :

$$HC_0 : a_t = 1 \qquad HC_1 : a_t = \sqrt{\frac{n}{n-k}} \qquad HC_2 : a_t = \frac{1}{\sqrt{1-h_t}} \qquad HC_3 : a_t = \frac{1}{1-h_t}$$

où $h_t \equiv X_t (X^\top X)^{-1} X_t^\top$ est le $t^{\text{ème}}$ élément de la matrice de projection orthogonale sur l'espace engendré par les colonnes de X. MacKinnon et White (1985), Chesher et Jewitt (1987) et Long et Ervin (2000) montrent qu'en terme de distorsion du niveau, HC_0 est moins performant que HC_1, à son tour moins performant que HC_2 et HC_3. L'une de ces deux dernières versions n'est pas meilleure que l'autre dans tous les cas, même si la version HC_3 s'est montrée plus performante dans un certain nombre d'expériences Monte Carlo.

Test bootstrap

Le principe du bootstrap consiste à construire un processus générateur de données (PGD) artificiel, appelé PGD bootstrap, en remplaçant les paramètres et distributions de probabili-

tés inconnues dans le modèle, par des estimations empiriques de ces derniers. La distribution de la statistique de test sous ce PGD est appelée la loi bootstrap, elle est utilisée comme loi nominale pour calculer un seuil critique ou une P-value, voir Horowitz (1997, 2000).

Si l'hétéroscédasticité est de forme inconnue, elle ne peut pas être reproduite dans le PGD bootstrap. Le wild bootstrap, introduit par Wu (1986) et Beran (1986), permet de pallier ce problème en considèrant le PGD bootstrap suivant :

$$y_t^\star = X_t\widetilde{\beta} + a_t\widetilde{u}_t\,\varepsilon_t^\star \qquad\qquad \varepsilon_t^\star \sim G(0,1) \tag{3}$$

où $\widetilde{\beta}$ est l'estimateur des paramètres et $\widetilde{u}_t$ le résidu, issus d'une estimation par MCO du modèle de régression dans lequel on impose l'hypothèse nulle, appelé le modèle contraint. Le paramètre a_t peut prendre la forme HC_0, HC_1, HC_2 ou HC_3 même s'il est souvent présenté dans la litérature avec la forme HC_2. Pour que le bootstrap soit valide, il faut que le rééchantillonnage de $\varepsilon_t^\star$ s'effectue à partir d'une fonction de distribution quelconque G d'espérance nulle et de variance l'unité, comme par exemple la loi de Rademacher

$$G: \quad \varepsilon_t^\star = \begin{cases} +1 & \text{avec la probabilité} \quad 1/2 \\ -1 & \text{avec la probabilité} \quad 1/2. \end{cases} \tag{4}$$

Davidson et Flachaire (2001) ont montré, sur la base de développements théoriques et de simulations, que cette loi bi-atomique donne des performances toujours supérieures aux autres choix de G proposés dans la littérature. Les expériences de Godfrey et Orme (2001) et de MacKinnon (2002) confirment ce résultat.

La loi bootstrap est la distribution de la statistique de test sous le PGD bootstrap. Elle peut être approximée par simulations. Cette loi est alors utilisée comme loi nominale pour calculer un seuil critique ou une P-value bootstrap, voir Flachaire (2000) pour plus de détails.

3 Description du modèle

On considère le modèle de régression linéaire suivant

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \sigma_t\varepsilon_t \tag{5}$$

où les vraies valeurs des paramètres sont $\beta_0 = \beta_1 = \beta_2 = 0$, et ε_t est un bruit blanc Normal $N(0,1)$. La quantité qui fait l'objet de l'étude est une statistique τ de type Student, basée sur le HCCME, qui teste l'hypothèse nulle $\beta_1 = 0$,

$$\tau = \frac{x_1^\top M_2 y}{(x_1^\top M_2\widehat{\Omega}M_2 x_1)^{1/2}} \tag{6}$$

où $\widehat{\Omega} = \text{diag}\,(a_1^2\widehat{u}_1^2, \ldots, a_n^2\widehat{u}_n^2)$, avec a_t correspondant à une des transformations HC_0, HC_1, HC_2 et HC_3, et $\widehat{u}_t$ est le résidu issu d'une estimation MCO du modèle (5). Par ailleurs, $M_2 = I - X_2(X_2^\top X_2)^{-1}X_2^\top$, où $X_2 = [\iota\ x_2]$ est une matrice regroupant le vecteur constant et le régresseur x_2.

L'expression (6) est invariante par rapport à β_0, β_1, à l'échelle des variances des aléas, à l'échelle des régresseurs et à la valeur de β_1 sous l'hypothèse nulle. Par conséquent, les résultats des simulations ne seront pas sensibles aux choix de ces caractéristiques du modèle. Par contre, les choix de la forme de l'hétéroscédasticité et des régresseurs peuvent affecter sensiblement les résultats. Le modèle retenu est le plus gênant possible pour les tests robustes à l'hétéroscédasticité : l'hétéroscédasticité est une fonction des régresseurs, $\sigma_t = |x_{1t}|$, et les régresseurs contiennent des observations à fort effet de levier. Pour contrôler l'effet de levier, nous conduisons plusieurs expériences où les régresseurs sont générés de différentes manières : soit un tirage η de la loi $N(0,1)$, une réalisation des régresseurs x_{1t} ou x_{2t} est donné par la formule κ^η. Il est clair que pour $\kappa = \exp(1)$, les régresseurs sont générés selon la loi Lognormale et l'effet de levier est important. Par contre, pour κ proche de 1 les régresseurs sont homogènes. L'effet de levier est d'autant plus faible que la valeur de κ est proche de 1, il est d'autant plus fort que la valeur de κ est grande. Dans nos expériences, nous utilisons $\kappa = 1.1, 1.2, \ldots, 2.6, \exp(1), 2.8, 2.9, \ldots, 3.5$.

L'algorithme utilisé est le suivant :

1. On définit la taille de l'échantillon n, le nombre de répétitions N et les vraies valeurs des paramètres. On génère les régresseurs, puis la variance des aléas σ_t^2.
2. On génère n aléas ε_t, puis n valeurs de la variable dépendante y_t^* à partir du vrai PGD définit au préalable : $y_t^* = \sigma_t\varepsilon_t$.
3. Pour l'échantillon simulé (y_t^*, x_{1t}, x_{2t}), on estime le modèle (5), puis on calcule une réalisation τ_i^* de la statistique de test (6). On calcule une P-value asymptotique à partir de la loi de Fisher $p_i^* = 1 - F_{as}(\tau_i^{*2})$, puis une P-value bootstrap comme suit :
   (a) On génère un échantillon bootstrap $(y_t^\star, x_{1t}, x_{2t})$, à partir du PGD bootstrap, puis on calcule une réalisation $\tau_j^\star$ de la statistique de test (6).
   (b) On répète l'étape précédente B fois de manière à obtenir B réalisations bootstrap $\tau_j^\star$, pour $j = 1, \ldots, B$, de la statistique de test (6).

   Le nombre de fois où $\tau_j^{\star 2} > \tau_i^{*2}$, pour $j = 1, \ldots, B$, divisé par B, donne la valeur de la P-value bootstrap $p_i^\star$.
4. On répète N fois les étapes 2 et 3 de manière à obtenir N réalisations des P-value asymptotique p_i^* et bootstrap $p_i^\star$, pour $i = 1, \ldots, N$.

Les fonctions de distribution empiriques, ou EDF, des N réalisations des P-value asymptotique et bootstrap sont deux approximations distinctes de la vraie loi de cette P-value, d'autant plus précise que N est grand. La taille d'échantillon est égal à $n = 100$, le nombre de répétitions à $N = 10.000$ et le nombre de rééchantillonnages bootstrap $B = 999$.

4 Niveau

La probabilité de rejet d'un test (RP), appelée aussi niveau réel ou erreur de type I, est la probabilité de rejetter l'hypothèse nulle alors qu'elle est vraie. Pour un niveau nominal donné α, la probabilité de rejet d'un test est donc donné dans une expérience Monte Carlo par la proportion de rejets de l'hypothèse nulle. Pour un test asymptotique (respectivement bootstrap), c'est donc la proportion des réalisations $p_i^\star$ (resp. p_i^*) inférieure ou égale à α. Si le test est parfaitement fiable, l'écart entre la probabilité de rejet et le niveau nominal, appelé l'erreur de la probabilité de rejet (ERP) ou la distorsion du niveau, est nul.

Tests asymptotique et bootstrap

La figure 1 représente une fonction de niveau : pour un niveau nominal donné $\alpha = 0.05$ on trace les ERP en ordonnée contre différentes valeurs de κ (l'effet de levier) en abscisse, pour différentes statistiques de tests. Les courbes `HC0`, `HC1`, `HC2` et `HC3` représentent les ERP des tests *asymptotique* basés sur les transformations HC_0, HC_1, HC_2 et HC_3 du HCCME. Les courbes `BHC0`, `BHC1`, `BHC2` et `BHC3` représentent les ERP des tests *bootstrap* basés sur les transformations HC_0, HC_1, HC_2 et HC_3 pour estimer le HCCME et pour construire le PGD bootstrap.

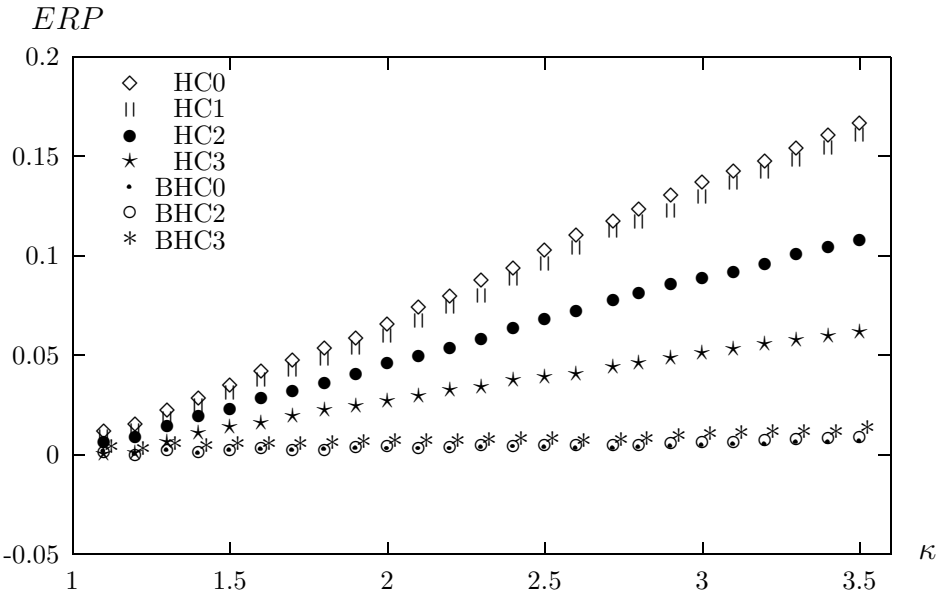

FIG. 1 – Traçé de l'ERP en fonction de l'effet de levier κ.

D'après la figure 1, on remarque que l'ERP des tests asymptotique (`HC0 HC1 HC2 HC3`) augmente fortement avec les valeurs de κ, c'est à dire lorsque l'effet de levier est de plus en plus fort. Par exemple, l'ERP est approximativement égale à 0.17 pour un test asymptotique robuste à l'hétéroscédasticité, basé sur la version HC_0 du HCCME, lorsque $\kappa = 3.5$.

Cela signifie que lorsqu'on pense faire une erreur de 5%, en réalité l'erreur que l'on fait si l'hypothèse nulle est vraie ($RP = ERP + \alpha$) est de l'ordre de 23% : le test n'est donc pas fiable. Par contre, un test basé sur la transformation HC_3 du HCCME améliore les résultats : $ERP \approx 0.06$ et donc $RP \approx 0.11$, l'erreur réellement faite est de l'ordre de 11%. La fiabilité du test basé sur la version HC_3 est meilleure, mais elle n'est pas encore de bonne qualité, malgré la taille de l'échantillon relativement importante.

Ces résultats confirment ceux de Chesher et Jewitt (1987). Les résultats des expériences sont très sensibles au choix des régresseurs et les tests ne sont pas fiables en présence d'observations à fort effet de levier. Ils suggèrent que les résultats sont moins sensibles à la taille de l'échantillon qu'à la présence de fort effet de levier. Pour vérifier cette dernière hypothèse, on refait la même expérience pour des régresseurs homogènes, tirés de la loi $N(0,1)$ et une taille d'échantillon très faible $n = 20$. On obtient les ERP suivantes,

$$ERP_{HC_0} = 0.023, \quad ERP_{HC_1} = 0.018, \quad ERP_{HC_2} = 0.014, \quad ERP_{HC_3} = 0.007$$

On constate que l'ERP de la version HC_3 du HCCME est beaucoup plus plus faible (0.007) que si on considère un grand nombre de données $n = 100$ mais où les régresseurs exhibent des fort effets de levier. Cela confirme le fait que la fiabilité des tests robustes à l'hétéroscédasticité sont beaucoup plus sensibles à la structure des régresseurs qu'à la taille de l'échantillon.

Concernant les tests bootstrap, le calcul de la P-value bootstrap étant définie comme la proportion de réalisations de la statistique bootstrap élevées au carrés $\tau_j^{\star 2}$, pour $j = 1, \ldots, B$, supérieure à la réalisation associée de la statistique asymptotique τ_i^{*2}, il est clair que la transformation HC_1, correspondant à une multiplication de la statistique par un facteur constant, n'aura aucun impact. Autrement dit, les tests bootstrap correspondant aux transformations HC_0 et HC_1 sont identiques. Dans la figure 1, la courbe de `BHC1` n'est donc pas représentée, elle est identique à celle de `BHC0`. D'après cette figure, on remarque que l'ERP des tests bootstrap `BHC0 BHC2 BHC3` est très faible, proche de l'axe des abscisse, et il n'y a pas de différence nette entre les trois versions utilisées. S'il n'y a pas de fort effet de levier ($\kappa = 1.1$), l'ERP des tests bootstrap est quasiment nulle. Par contre, en présence de très fort effet de levier ($\kappa = 3.5$), l'ERP des tests bootstrap est légèrement différente de zéro, mais elle reste très faible (≈ 0.01) comparée à celle des tests asymptotique (> 0.06).

Ces résultats montrent que les tests bootstrap, basés sur les différentes versions du HCCME, corrigent très largement la distorsion du niveau des tests asymptotique. Ils sont beaucoup plus fiables en échantillon fini, en présence ou non de forts effets de levier.

Résidus contraints ou non-contraints

Le calcul de l'estimateur HCCME avec les résidus contraints plutôt que les résidus non-contraints, a été suggéré par Davidson et MacKinnon (1985) dans le cadre des tests asymptotique. Par ailleurs, Davidson et Flachaire (2001) montrent qu'un terme supplé-

mentaire est introduit dans les développements d'Edgeworth d'un test bootstrap basé sur le HCCME, si on utilise les résidus non-contraints plutôt que les résidus contraints. Ce résultat suggère que la distorsion du niveau devrait être plus faible avec un test basé sur les résidus contraints plutôt que non-contraints. Toutefois, les expériences Monte Carlo ne montrent pas systématiquement une différence significative entre les deux approches, même si quelques expériences montrent que l'utilisation des résidus contraints plutôt que des résidus non-contraints apporte un gain de précision très sensible, voir van Giersbergen et Kiviet (2002), Li et Maddala (1993) et Nankervis et Savin (1994).

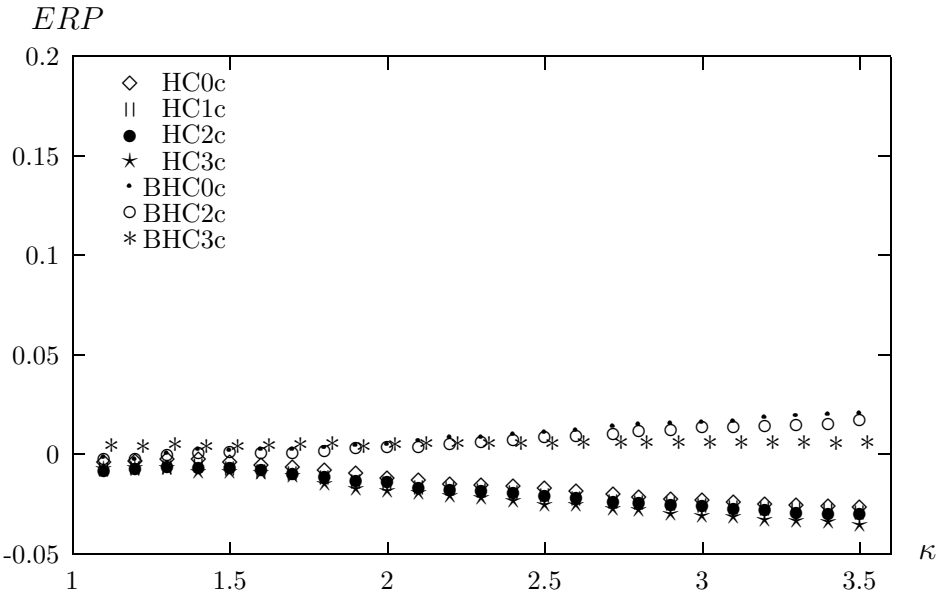

FIG. 2 – Traçé de l'ERP en fonction de l'effet de levier κ, résidus contraints.

La figure 2 représente la même fonction de niveau que la figure 1 sauf que les tests asymptotique (`HC0c HC1c HC2c HC3c`) et bootstrap (`BHC0c BHC1c BHC2c BHC3c`) sont basés sur les différentes versions du HCCME avec les résidus contraints plutôt que non-contraints. D'après ce graphique, la distorsion du niveau des tests bootstrap est faible : elle augmente légèrement pour les versions HC_0 et HC_2 lorsque l'effet de levier est fort, par contre celle de la version HC_3 reste toujours quasiment nulle. D'un autre côté, la distorsion des tests asymptotique est légèrement plus importante que celle des test bootstrap. Par ailleurs, si on compare les tests asymptotique dans les figures 1 et 2, on constate que la distorsion du niveau est largement plus faible pour les tests asymptotique basés sur les résidus contraints plutôt que non-contraints. De plus, on constate qu'il n'y a pas de différence importante entre les différentes transformations du HCCME si les résidus contraints sont utilisés.

Ces résultats montrent qu'un test bootstrap, basé sur la version HC_3 du HCCME et sur les résidus contraints, est parfaitement fiable en échantillon fini, même en présence de fort effet de levier.

5 Puissance

Pour étudier la puissance des tests robustes à l'hétéroscédasticité, on reprend le même algorithme que celui utilisé pour l'étude de la distorsion du niveau, section 3, mais on génère les données à partir d'un PGD construit sous l'hypothèse alternative. Dans nos expériences, la statistique d'intérêt teste l'hypothèse nulle $\beta_1 = 0$ contre l'hypothèse alternative $\beta_1 \neq 0$. On simule les données à partir du même PGD que celui décrit dans l'équation (5), sauf que le paramètre β_1 prend une valeur différente de zéro. L'hypothèse testée est toujours $\beta_1 = 0$, qui n'est plus vérifiée, à partir de la statistique de test (6). Cela revient donc à utiliser le même algorithme que celui décrit dans la section 3, avec le même générateur de nombres aléatoires et le même point de départ, mais en générant les données dans l'étape 2 à partir du PGD : $y_t^\diamond = \beta_1 x_{1t} + \sigma_t \varepsilon_t$. Il est clair que la puissance dépend du choix du paramètre β_1 : plus la valeur de ce paramètre est éloignée de sa valeur sous l'hypothèse nulle, plus forte est la puissance.

Dans notre cadre expérimental, la statistique de test sous l'hypothèse alternative $\tau_j^\diamond$ peut être calculée sans avoir besoin de simuler un nouvel échantillon sous un PGD qui ne respecte pas l'hypothèse nulle. Soient $\widehat{\beta}_1^\star$ l'estimateur MCO du paramètre β_1 à partir de l'échantillon simulé sous l'hypothèse nulle $(y^\star, x_1, x_2)$; et $\widehat{\beta}_1^\diamond$ l'estimateur obtenu à partir de l'échantillon simulé sous l'hypothèse alternative $(y^\diamond, x_1, x_2)$, on démontre que

$$\tau_i^\diamond \;=\; \frac{\widehat{\beta}_1^\diamond - 0}{[\widehat{V}(\widehat{\beta}_1^\diamond)]^{1/2}} \;=\; \frac{\widehat{\beta}_1^\star + \beta_1}{[\widehat{V}(\widehat{\beta}_1^\star)]^{1/2}}. \tag{7}$$

Ce résultat[1] permet de simplifier largement l'algorithme, en calculant directement la statistique bootstrap $\tau_i^\diamond$ sans avoir besoin de générer un PGD sous l'hypothèse alternative. On peut calculer différentes réalisations de cette statistique pour différentes valeurs de β_1.

Un dernier aspect important de l'étude de la puissance concerne la comparaison des puissances entre différentes statistiques de test. Dans le cas où la distorsion du niveau des tests n'est pas nulle, la comparaison de la puissance des tests pour un même niveau nominal a peu de sens, car on pourrait être amené à comparer la puissance d'un test qui sur-rejette fortement l'hypothèse nulle alors qu'elle est vrai contre la puissance d'un test qui la sous-rejette fortement. Autrement dit, il est important de comparer la puissance des tests pour un niveau réel, ou erreur de type I ou probabilité de rejet (RP), identique. On parle alors de puissance *corrigée* du niveau nominal.

Hypothèse alternative

La figure 3 représente la puissance corrigée du niveau en ordonnée, pour des tests bootstrap robustes, basés sur différentes versions du HCCME (`BHC0 BHC2 et BHC3`), en

[1] $y^\diamond = x_1\beta_1 + y^\star$ et donc, on a $\widehat{\beta}_1^\diamond = (x_1^\top M_2 x_1)^{-1} x_1^\top M_2 y^\diamond = \beta_1 + (x_1^\top M_2 x_1)^{-1} x_1^\top M_2 y^\star = \beta_1 + \widehat{\beta}_1^\star$. De plus, les variances estimées sont les mêmes car les résidus sont les même $\widehat{u}^\diamond = M_X y^\diamond = M_X y^\star = \widehat{u}^\star$.

fonction de différentes valeurs du paramètre $\beta_1 = -3, -2.9, \ldots, -0.1, 0, 0, 1, \ldots, 2.9, 3$ en abscisse, pour un niveau réel donné $RP = 0.05$ et un fort effet de levier $\kappa = 3.5$.

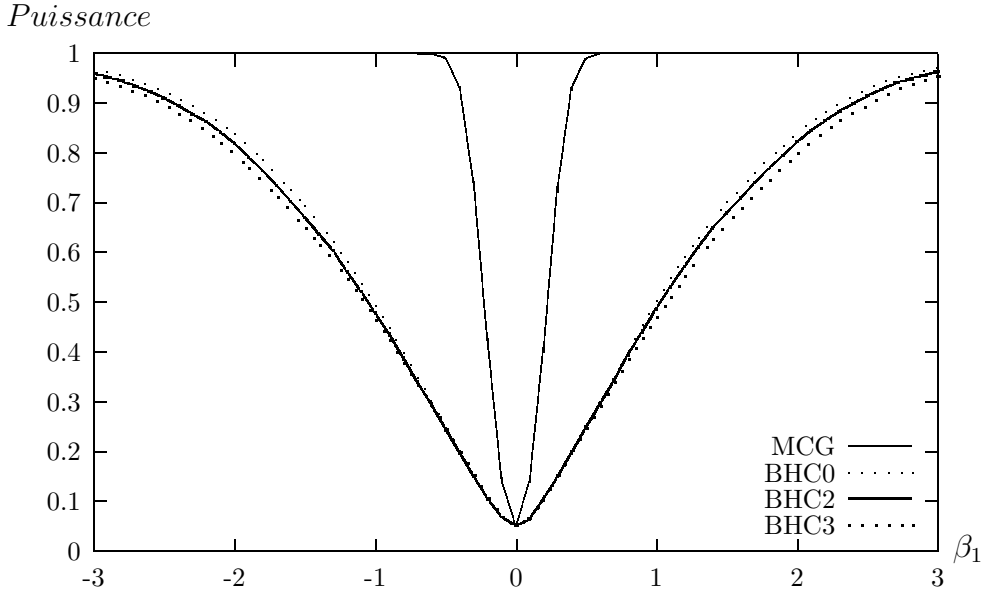

FIG. 3 – Traçé de la puissance corrigée du niveau, $RP = 0.05$ et $\kappa = 3.5$

Il est clair que la puissance augmente lorsque la valeur du paramètre β_1 s'éloigne de 0. Par contre, lorsque $\beta_1 = 0$, la puissance est équivalente à la probabilité de rejet et elle est égale à 0.05. Le test le plus puissant rejetterait l'hypothèse nulle avec certitude pour toute valeur de β_1 différente de 0, sur ce type de graphique, sa courbe prendrait alors la forme $\top$ avec le pied du segment vertical en 0.05 et le segment horizontal en 1.

D'après la figure 3, la puissance du test bootstrap avec la version HC_0 du HCCME est très légèrement supérieure à celle de la version HC_2, elle même très légèrement supérieure à celle de la version HC_3. Ce résultat suggère qu'un test bootstrap avec la version HC_0 est très légèrement plus performant que le même test avec la version HC_3 ou HC_2. Pour comparaison, on trace la puissance obtenue avec une procédure de test basée sur une estimation par Moindres Carrés Généralisés, ou Moindres Carrés Pondérés `MCG`. Cette méthode peut être mise en œuvre si on connait la forme de l'hétéroscédasticité et si on peut l'estimer de façon convergente. Si une information correcte de la forme de l'hétéroscédasticité est utilisée, cette méthode d'estimation est plus efficace que l'estimation robuste. Toutefois, celle ci n'est pas robuste à une mauvaise spécification de la forme de l'hétéroscédasticité. Il est clair que dans notre contexte où l'hétéroscédasticité est de forme inconnue, cette méthode d'estimation ne peut pas être utilisée. On l'utilise ici à titre de comparaison, comme étant la meilleure estimation sans biais que l'on puisse obtenir, si on connaissait la forme de l'hétéroscédasticité. Dans notre cadre expérimental, elle peut être mise en œuvre en transformant la variable dépendante et les régresseurs, en les divisant par la racine carré de x_{1t}^2, puis en estimant par MCO la variable dépendante et les régresseurs ainsi transformés. Il est clair, d'après

la figure 3, que la puissance du test MCG est très largement supérieure à celle des tests HCCME. La perte de puissance constatée par l'utilisation d'un test robuste est considérable. Ce résultat souligne l'arbitrage traditionnel entre robustesse et efficacité : un test robuste est moins performant mais reste valide dans le cadre d'une mauvaise spécification, alors qu'un test efficace peut être beaucoup plus performant mais il n'est pas robuste à une mauvaise spécification du modèle.

Effet de levier

On s'intéresse maintenant à l'écart de puissance constatée entre un test efficace basé sur l'estimation MCG, et un test robuste basé sur le HCCME. On se demande si la large perte de puissance observée pour le test robuste est due à la présence d'un fort effet de levier. Pour cela, on refait des expériences en diminuant l'impact de l'effet de levier, c'est-à-dire en réduisant la valeur de κ.

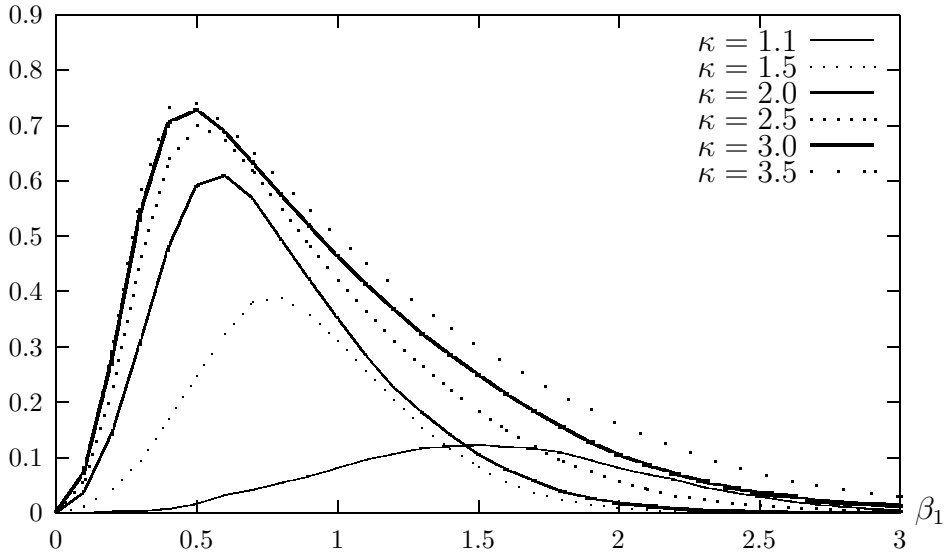

FIG. 4 – Différence de puissance entre les tests `MCG` et `BHC0`, $\beta_1 \in [0; 3]$

La figure 4 représente la différence de puissance corrigée du niveau, entre les tests `MCG` et `BHC0`. Les différentes courbes du graphique correspondent à des niveaux différents d'effet de levier $\kappa = 1.1, 1.5, 2, 2.5, 3, 3.5$. Pour $\kappa = 3.5$, la courbe correpond à la différence entre la courbe `MCG` et la courbe `BHC0` de la figure 3, sur le support $[0, 3]$. D'après ce graphique, on constate que la différence de puissance entre un test robuste et un test efficace est d'autant plus grande que l'effet de levier est fort. Ce résultat suggère que l'arbitrage entre efficacité et robustesse, évoqué ci-dessus, a moins d'impact s'il n'y a pas d'observations à fort effet de levier, c'est-à-dire si les données observées sont homogènes. Par contre, la présence d'observations à fort effet de levier a un impact important sur la puissance d'un test robuste, comparée à la puissance d'un test GLS utilisant une forme correcte de l'hétéroscédasticité.

Résidus contraints ou non-contraints

Lorsqu'on étudie la puissance, l'hypothèse nulle n'est pas vérifiée. Cela a conduit certains auteurs à avancer l'argument selon lequel l'utilisation des résidus contraints devrait conduire à une perte de puissance par rapport à l'utilisation de résidus non-contraints (van Giersbergen et Kiviet 2002), car les résidus contraints ne peuvent pas être considérés comme une meilleure estimation des aléas si H_0 n'est pas vraie. Néanmoins, les résultats de simulations ne valident pas forcément cet argument (MacKinnon 2002). Il est donc intéressant d'étudier la puissance des tests basés sur les résidus contraints et de la comparer à celle des tests basés sur les résidus non-contraints.

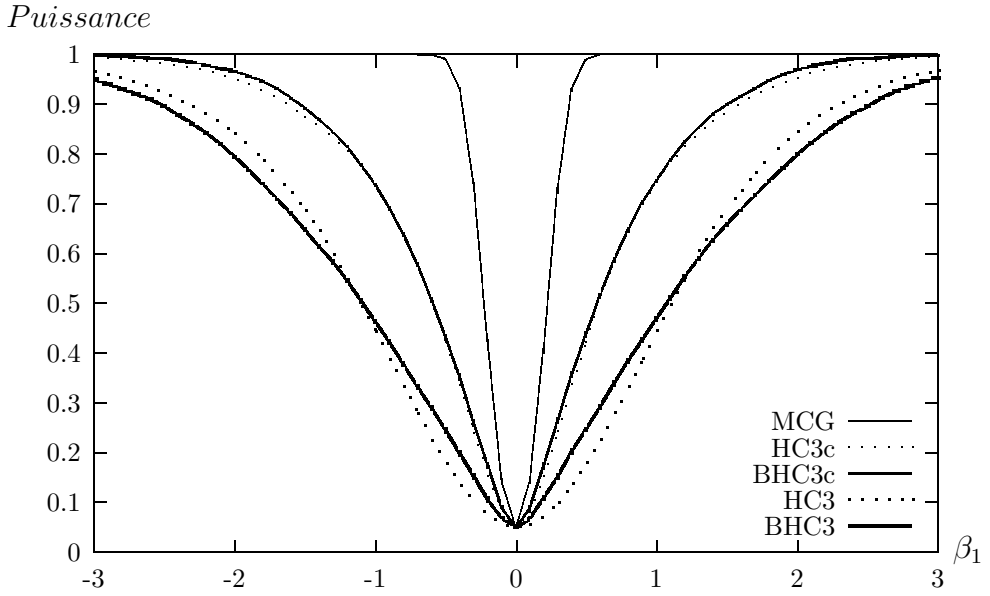

FIG. 5 – Traçé de la puissance corrigée du niveau, résidus non-contraints et contraints.

La figure 5 représente la puissance corrigée du niveau des tests asymptotique et bootstrap, basés sur la version HC_3 du HCCME, calculée avec les résidus contraints ou non-contraints. Si on compare les tests bootstrap basés sur les résidus non-contraints et contraints, c'est à dire `BHC3` contre `BHC3c`, on constate un large gain de puissance dans l'utilisation des résidus contraints. Ce graphique montre qu'un test basé sur les résidus contraints est largement plus puissant qu'un test basé sur les résidus non-contraints, néanmoins on peut voir que la puissance reste beaucoup plus faible que celle d'un test asymptotique efficace `MCG`.

Un dernier aspect intéressant concerne la différence de puissance obtenue par l'utilisation d'un test bootstrap plutôt qu'un test asymptotique. Davidson et MacKinnon (2002) montrent que si la puissance est corrigée du niveau, la différence entre la puissance d'un test bootstrap et celle d'un test asymptotique devrait être du même ordre que l'ERP du test bootstrap. En d'autres termes, si l'ERP d'un test bootstrap est négligeable, la différence de puissance entre le test asymptotique et le test bootstrap devrait être également négligeable. Dans la figure 5,

si on compare les tests bootstrap et asymptotique, respectivement `HC3` contre `BHC3` puis `HC3c` contre `BHC3c`, on constate que les différences de puissance sont relativement faibles.

Des expériences supplémentaires montrent que si on considère les différentes versions du HCCME, la différence de puissance est très faible entre les tests bootstrap basés sur les résidus contraints. On obtient un graphique similaire à celui de la figure 3, avec des écarts de puissance très faibles, mais ou la version HC_3 du HCCME montre une puissance très légèrement supérieure à celle des autres.

6 Conclusion

Dans la pratique, les tests asymptotique robustes à l'hétéroscédasticité basés sur la matrice de covariance de Eicker (1963) et White (1980), sont habituellement construits sur la base des résidus non-contraints. Nos résultats expérimentaux montrent que ces tests ne sont pas fiables en échantillon fini. Nous montrons que si ces tests sont construits sur la base des résidus contraints et si la méthode du wild bootstrap est utilisée, ils sont fiables et performants, même en présence d'effet à fort effet de levier. L'utilisation de tests bootstrap basés sur les résidus contraints est préférable pour deux raisons essentielles. Premièrement, la distorsion du niveau de ces tests est très faible, et la puissance est similaire à celle des tests asymptotique correspondants. Deuxièmement, on constate qu'une augmentation de puissance est significative si les résidus contraints sont utilisés à la place des résidus non-contraints.

Références

Beran, R. (1986). Discussion of "Jackknife bootstrap and other resampling methods in regression analysis" by C.F.J. Wu. *Annals of Statistics*, **14**, 1295–1298.

Beran, R. (1988). "Prepivoting test statistics : a bootstrap view of asymptotic refinements". *Journal of the American Statistical Association* , **83**, 687–697.

Chesher, A. et I. Jewitt (1987). "The bias of a heteroskedasticity consistent covariance matrix estimator". *Econometrica*, **55**, 1217–1222.

Davidson, R. et E. Flachaire (2001). "The wild bootstrap, tamed at last". working paper STICERD, London School of Economics, Darp58.

Davidson, R. et J. G. MacKinnon (1985). "Heteroskedasticity-robust tests in regression directions". *Annales de l'INSEE*, **59/60**, 183–218.

Davidson, R. et J. G. MacKinnon (1993). *Estimation and Inference in Econometrics* . New York : Oxford University Press.

Davidson, R. et J. G. MacKinnon (1999). "The size distortion of bootstrap tests". *Econometric Theory* , **15**, 361–376.

Davidson, R. et J. G. MacKinnon (2002). "The power of bootstrap and asymptotic tests". unpublished paper.

Davidson, R. et J. G. MacKinnon (2003). *Econometric Theory and Methods*. New York : Oxford University Press.

Eicker, B. (1963). "Limit theorems for regression with unequal and dependant errors". *Annals of Mathematical statistics*, **34**, 447–456.

Flachaire, E. (2000). "Les méthodes du bootstrap dans les modèles de régression". *Économie et Prévision*, **142**, 183–194.

Godfrey, L. G. et C. D. Orme (2001). "Significance levels of heteroskedasticity-robust tests for specification and misspecification : some results on the use of wild bootstraps". paper presented at ESEM'2001, Lausanne.

Horowitz, J. L. (1997). "Bootstrap methods in econometrics : theory and numerical performance". In *Advances in Economics and Econometrics : Theory and Application*, Volume 3, pp. 188–222. David M. Kreps and Kenneth F. Wallis (eds), Cambridge, Cambridge University Press.

Horowitz, J. L. (2000). "The bootstrap". In *Handbook of Econometrics*, Volume 5. J. J. Heckman and E. E. Leamer (eds), Elsevier Science.

Li, H. et G. S. Maddala (1993, dec). "Bootstrapping cointegrating regressions". Paper presented at the Fourth Meeting of the European Conference Series in Quantitative Economics and Econometrics.

Long, J. S. et L. H. Ervin (2000). "Heteroscedasticity consistent standard errors in the linear regression model". *The American Statistician*, **54**, 217–224.

MacKinnon, J. G. (2002). "Bootstrap inference in econometrics". *Canadian Journal of Economics*, **35**, 615–645.

MacKinnon, J. G. et H. L. White (1985). "Some heteroskedasticity consistent covariance matrix estimators with improved finite sample properties". *Journal of Econometrics*, **21**, 53–70.

Nankervis, J. C. et N. E. Savin (1994). "The level and power of the bootstrap- t test in the AR(1) model with trend". Manuscript, Department of Economics, University of Surrey and University of Iowa.

van Giersbergen, N. P. A. et J. F. Kiviet (2002). "How to implement bootstrap hypothesis testing in static and dynamic regression model : test statistic versus confidence interval approach". *Journal of Econometrics*, **108**, 133–56.

White, H. (1980). "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity". *Econometrica*, **48**, 817–838.

Wu, C. F. J. (1986). "Jackknife bootstrap and other resampling methods in regression analysis". *Annals of Statistics*, **14**, 1261–1295.