

Un test Bootstrap dans un modèle $AR(1)$

Patrice BERTAIL*

RÉSUMÉ. – Nous présentons une procédure bootstrap de test de racine unité dans un modèle $AR(1)$, basée sur le rééchantillonnage des résidus. Le Bootstrap est mis en œuvre sous l'hypothèse nulle et sous des alternatives locales afin de déterminer la puissance du test : ceci permet d'éviter les problèmes de non stabilité du Bootstrap observé dans les cas non stationnaires. Nous montrons la validité asymptotique de la méthode et analysons par simulations les performances à distance finie.

A Bootstrap Test in an $AR(1)$ Model

ABSTRACT. – We study a Bootstrap test of the unit-root hypothesis, based on residual resampling, in a simple $AR(1)$ model. The Bootstrap is performed under the null hypothesis and under local alternatives which allows to construct a power function estimator: this avoids the non-stable properties of the Bootstrap observed in the non stationary cases. We study the asymptotic validity of the method theoretically and give some empirical evidences by simulations.

* P. BERTAIL : INRA : Institut National de la Recherche Agronomique.

Cet article reprend et prolonge une partie d'un article « Test Bootstrap dans les modèles linéaires et non linéaires » présenté au XXI^e Journées Statistiques de l'ASU [1988] ainsi que le document de travail Insee n° 9203 [1991]. Je remercie M. Gourieroux C., Tassi P., Trognon A. et Raoult J. P. pour leurs nombreuses suggestions.

1 Introduction

L'objet de cet article est de présenter et démontrer la validité asymptotique d'une procédure de type Bootstrap permettant de tester la valeur du paramètre β de régression dans un modèle autorégressif d'ordre 1 ($AR(1)$).

Considérons le problème de test

$$(1) \quad H_0 : \beta = \beta_0 \text{ contre } H_a : \beta \neq \beta_0$$

dans le modèle $AR(1)$:

$$(2) \quad y_t = \beta y_{t-1} + \varepsilon_t, \quad t = 1, \dots, T,$$

où ε_t est un bruit blanc indépendant de loi F de variance $V_F \varepsilon_t = \sigma^2$. On suppose dans la suite que le processus y_t est généré à partir d'une observation y_0 constante. On pose $c = \sigma^{-2} y_0^2$. On pourrait, sans grands changements aussi dans le cas $|\beta| < 1$, supposer que y_0 est une variable aléatoire d'espérance nulle et de variance $(1 - \beta^2)^{-1} \sigma^2$ afin de rendre le processus (y_t) stationnaire.

Soient $\hat{\beta}_T$ et $\hat{\sigma}_T^2$ les estimateurs des moindres carrés ordinaires :

$$\hat{\beta}_T = \left(\sum_{t=1}^T y_{t-1}^2 \right)^{-1} \sum_{t=1}^T y_t y_{t-1}$$

et

$$\hat{\sigma}_T^2 = T^{-1} \sum_{t=1}^T (y_t - \hat{\beta}_T y_{t-1})^2.$$

Les résidus estimés centrés dans le modèle (2)

$$\tilde{\varepsilon}_t = \hat{\varepsilon}_t - T^{-1} \sum_{t=1}^T \hat{\varepsilon}_t,$$

permettent de définir une fonction de répartition empirique \tilde{F}_T dans lequel on effectue le rééchantillonnage, suivant l'idée de FREEDMAN [1981] [1984]. On notera par ailleurs \hat{F}_T la fonction de répartition empirique des résidus $\hat{\varepsilon}_t$.

Le processus fictif bootstrap y_t^* intégrant l'hypothèse nulle $H_0 : \beta = \beta_0$ s'obtient de manière récursive :

$$(3) \quad \begin{cases} y_0^* = y_0, \\ y_t^* = \beta_0 y_{t-1}^* + \varepsilon_t^*, \quad t = 1, \dots, T, \quad \varepsilon_t^* \text{ i.i.d. de loi } \tilde{F}_T, \end{cases}$$

soit encore par la relation :

$$y_t^* = \sum_{i=1}^{t-1} \beta_0^i \varepsilon_{t-i}^* + \beta_0^t y_0.$$

On calcule alors l'estimateur Bootstrap des moindres carrés ordinaires de β sous H_0 dans le modèle (3), $\hat{\beta}_T^*$, par régression de y_t^* sur y_{t-1}^* . L'idée du Bootstrap est de considérer que la distribution de $\hat{\beta}_T^* - \beta_0$ donne une bonne approximation de la distribution de $\hat{\beta}_T - \beta_0$ sous H_0 .

Nous montrerons que cette procédure de test Bootstrap est asymptotiquement valide quelle que soit la valeur de c et du paramètre testée. Ceci permet entre autre de résoudre les problèmes de non stabilité du Bootstrap dans le modèle $AR(1)$ non stationnaire observés par BASAWA, MALLIK, McCORMICK, REEVES et TAYLOR [1991], lorsque le processus bootstrap est simulé en utilisant $\hat{\beta}_T$ au lieu de β_0 . Nous en déduisons une méthode de construction du diagramme de puissances locales du test de racine unité.

Nous comparons par simulation les résultats obtenus avec ce test Bootstrap aux résultats théoriques de DICKEY et FULLER [1979] [1981] et EVANS et SAVIN [1981 a, b] [1984], en terme de distributions et de puissance, dans le cas où le bruit-blanc de l' $AR(1)$ est gaussien. Ce test Bootstrap donne des résultats à distance finie très satisfaisants. Ces quelques résultats empiriques confirment que la distribution Bootstrap est beaucoup plus proche de la distribution exacte que ne l'est l'approximation asymptotique.

2 Bootstrap et modèle autorégressif

2.1. Distributions asymptotique, Bootstrap et à distance finie

Le modèle (2) a été abondamment étudié dans la littérature : voir par exemple WHITE [1958] [1959], ANDERSON [1959], RAO [1978], EVANS-SAVIN [1981a] [1981b] [1984], DICKEY et FULLER [1981], PHILLIPS [1987a] [1987b]) et DIEBOLD et NERLOVE [1988] pour une synthèse et une bibliographie complète. Il est possible dans un cadre très large d'hypothèses d'obtenir la distribution asymptotique de l'estimateur des m.c.o.

Lorsqu'on prend la partie principale en T dans l'expression dérivant du calcul de l'information de Fischer des estimateurs (WHITE [1958]) :

$$(4) \quad E_F \left(\sum_{t=1}^T y_{t-1}^2 / \sigma^2 \right) = \frac{T}{1 - \beta^2} - \frac{1 - \beta^{2T}}{(1 - \beta^2)^2} + \frac{y_0^2}{\sigma^2} \left(\frac{1 - \beta^{2T}}{(1 - \beta^2)} \right)$$

si $|\beta| \neq 1$,

$$= \frac{T(T+1)}{2} + T \frac{y_0^2}{\sigma^2} \quad \text{si } |\beta| = 1.$$

suivant les valeurs prises par β , on obtient la vitesse de convergence de l'estimateur des m.c.o. de sorte que l'on peut montrer (voir WHITE [1958])

et EVANS et SAVIN [1981a]) :

$$(5) \quad \text{si } |\beta| < 1 \quad \left(\frac{T}{1 - \beta^2} \right)^{1/2} (\hat{\beta}_T - \beta) \xrightarrow[T \rightarrow \infty]{L} N(0, 1)$$

$$(6) \quad \text{si } |\beta| = 1 \quad \frac{T}{\sqrt{2}} (\hat{\beta}_T - \beta) \xrightarrow[T \rightarrow \infty]{L} L_1$$

$$(7) \quad \text{si } |\beta| > 1 \quad \frac{\beta^T}{(\beta^2 - 1)} (\hat{\beta}_T - \beta) \xrightarrow[T \rightarrow \infty]{L} L_{\beta,c}.$$

Dans le cas stationnaire (5), la distribution limite est gaussienne et ne dépend pas de la condition initiale (voir ANDERSON [1959]). Pour $|\beta| = 1$, WHITE [1959] a montré que la distribution asymptotique de l'estimateur des m.c.o. centré dilaté, L_1 , de forme très dissymétrique, ne dépendait pas de la condition initiale y_0 et a donné une caractérisation de la loi limite en terme de processus de Wiener (voir PHILLIPS [1987a] pour une généralisation dans le cas d'erreurs non i.i.d., de type α -mélangeantes). Dans le cas explosif (7), $L_{\beta,c}$ est une loi non-dégénérée, dépendant notamment de la loi des résidus et de la condition initiale retenue par l'intermédiaire de c . Ainsi WHITE [1958] montre que dans le modèle gaussien, $L_{\beta,c}$ a pour densité :

$$(8) \quad f_{\beta,c} = \frac{\exp(-q)}{\pi(1+x^2)} \left(1 + \sum_{i=1}^{\infty} \frac{(4q)^i (i!)}{(1+x^2)^i (2i!)} \right)$$

avec $q = c(\beta^2 - 1)/2$. La distribution limite pour $|\beta| > 1$ peut donc avoir des queues épaisses ; dans le cas gaussien, avec $c = 0$, la loi limite est donc une loi de Cauchy. Il est important de noter que les paramètres c et β (si $c \neq 0$) interviennent dans la distribution limite des estimateurs dans le cas explosif.

Le comportement à distance finie est nettement différent ; il n'existe pas de sauts brusques entre les distributions. On obtient une déformation progressive d'une loi normale vers une loi dissymétrique de plus en plus aplatie (à cause des différences de vitesse de convergence). Pour des valeurs proches de 1, la distribution exacte est souvent très éloignée de la distribution asymptotique *i.e.* d'une loi de Gauss même pour des tailles d'échantillons assez élevées : la forme de la distribution exacte est plus proche de celle de L_1 accusant une forte dissymétrie, ceci d'autant plus que c est grand (EVANS et SAVIN [1981a]).

La méthode du Bootstrap introduite par EFRON [1979] [1982] étant connue pour donner à distance finie des approximations plus fiables de la vraie loi que l'approximation asymptotique (voir par exemple FREEDMAN [1981], HALL [1992] pour le modèle linéaire), il paraît particulièrement intéressant de l'étudier dans ce contexte.

2.2. Validité asymptotique du Bootstrap

Nous établissons dans ce paragraphe la validité asymptotique de la procédure de test dans le cas $AR(1)$ pur, quelle que soit la valeur de β_0 .

Notons que la validité de la méthode dans le cas stationnaire peut se déduire assez directement, à quelques adaptations près des preuves, des résultats de FREEDMAN [1984]. Considérant un modèle autorégressif stationnaire $AR(p)$, FREEDMAN [1984] montre en effet, en construisant la version Bootstrap de l' $AR(p)$ à partir de l'estimateur des moindres carrés ordinaires et des résidus estimés centrés du modèle, la validité asymptotique de la méthode, permettant ainsi de nombreuses applications (FREEDMAN et PETERS [1984a] [1984b]). BOSE [1988] démontre la validité au second ordre. Ces résultats permettent d'obtenir des intervalles non-paramétriques de prédiction du processus correct au second ordre (voir STINE [1985], FINDLEY [1986], SCHUCANY et THOMBS [1990]).

BASAWA, MALLIK, McCORMICK, TAYLOR [1989] ont montré, dans une approche directe, la validité asymptotique (en terme de convergence vague *i.e.* de convergence de la fonction caractéristique sur des compacts) du Bootstrap de l'e.m.c.o. dans un $AR(1)$ explosif avec premier terme Y_0 égal à zéro : leur résultat assure le fonctionnement du Bootstrap au sens où ils montrent que la loi de $(\hat{\beta}_T^2 - 1)^{-1} \hat{\beta}_T^T (\hat{\beta}_T^* - \hat{\beta}_T)$ conditionnelle à $(y_t)_{t=0, \dots, T}$ et la loi de $(\beta^2 - 1)^{-1} \beta^T (\hat{\beta}_T - \beta)$, avec $\beta > 1$, sont asymptotiquement les mêmes. Leur preuve suit ANDERSON [1959] et repose sur le développement de la fonction caractéristique de la distribution Bootstrap. Cependant il n'est pas sûr que leur résultat soit encore valable si $c \neq 0$, car la condition $c = 0$ joue un rôle important dans les considérations asymptotiques. Notons par exemple que si $c = 0$ la densité $f_{\beta,c}$ donnée en (8) ne dépend plus de β et est donc en un certain sens stable.

BASAWA, MALLIK, McCORMICK, REEVES et TAYLOR [1991] ont montré l'échec du Bootstrap dans le cas $\beta = 1$ lorsque la version Bootstrap est construite avec l'e.m.c.o. : ceci est en particulier dû à la forte instabilité du modèle, en présence de racine unité. Par ailleurs, les simulations que nous avons menées dans le cas de l' $AR(1)$ stationnaire, avec résidus gaussiens, montrent que le Bootstrap donne des résultats très décevants lorsque le paramètre de l'autorégressif s'approche de 1, même pour des tailles d'échantillon assez grandes ($T = 100$, $T = 500$). Ceci semble en particulier dû au fait que le processus Bootstrap est simulé à distance finie à partir de l'estimateur $\hat{\beta}_T$ biaisé.

L'optique test (qui en pratique est la plus utilisée, notamment dans la détection de racines unités) permet de contourner ces problèmes et de valider l'utilisation du Bootstrap dans les cas où le processus est explosif (avec premier terme constant quelconque) ou présente une racine unité. Le fait de travailler sous H_0 permet d'englober les cas stationnaire, instable et explosif et de prendre en compte l'effet de la valeur initiale.

L'outil fondamental utilisé dans la preuve est la distance de Mallows DM_p (voir MALLOWES [1972], BICKEL et FREEDMAN [1981]). Nous en rappelons quelques propriétés utiles en annexe 1, auxquelles nous nous référons dans la démonstration. La notion de convergence utilisée ici, $\xrightarrow[T \rightarrow \infty]{L^*}$ convergence en loi conditionnelle, presque sûre dans la suite (voir MALLOWES [1972] et la propriété A.2) est différente et plus forte que la convergence vague conditionnelle, presque sûre, utilisée par BASAWA, MALLIK, McCORMICK, TAYLOR [1989].

THÉOREME 1 : Soit un autorégressif pur de bruit blanc fort associé ε_t admettant des moments d'ordre 2. Pour toute valeur $\beta_\alpha \in \mathbb{R}$, si $\hat{\beta}_T^*$ est l'estimateur des moindres-carrés ordinaires Bootstrap, défini conditionnellement au processus initial, construit sous l'hypothèse H_α : $\beta = \beta_\alpha$, on a

$$\text{Si } |\beta_\alpha| < 1 \quad \left(\frac{T}{1 - \beta_\alpha^2} \right)^{1/2} (\hat{\beta}_T^* - \beta_\alpha) \xrightarrow[T \rightarrow \infty]{L^*} N(0, 1),$$

$$\text{Si } |\beta_\alpha| = 1 \quad \frac{T}{\sqrt{2}} (\hat{\beta}_T^* - \beta_\alpha) \xrightarrow[T \rightarrow \infty]{L^*} L_1,$$

$$\text{Si } |\beta_\alpha| > 1 \quad \frac{\beta_\alpha^T}{\beta_\alpha^2 - 1} (\hat{\beta}_T^* - \beta_\alpha) \xrightarrow[T \rightarrow \infty]{L^*} L_{\beta_\alpha, c}$$

où L_1 et $L_{\beta_\alpha, c}$, sont définis en (6) et (7).

Démonstration : On pose $g(T)$ égal au taux de convergence de l'estimateur des m.c.o. dans les différents cas de figure étudiés

$$\text{Si } |\beta_\alpha| < 1 \quad g(T) = \left(\frac{T}{1 - \beta_\alpha^2} \right)^{1/2}$$

$$\text{Si } |\beta_\alpha| = 1 \quad g(T) = \frac{T}{\sqrt{2}}$$

$$\text{Si } |\beta_\alpha| > 1 \quad g(T) = \left(\frac{\beta_\alpha^{2T}}{(1 - \beta_\alpha^2)^2} \right)^{1/2}$$

On montre que, sous l'hypothèse H_α : $\beta = \beta_\alpha$, $g(T) (\hat{\beta}_T^* - \beta_\alpha)$ conditionnellement aux observations et $g(T) (\hat{\beta}_T - \beta_\alpha)$ ont asymptotiquement la même loi (voir propriété A.1). On a :

$$g(T) (\hat{\beta}_T^* - \beta_\alpha) = \frac{g(T)^{-1} \sum_{t=1}^T y_{t-1}^* \varepsilon_t^*}{g(T)^{-2} \sum_{t=1}^T y_t^{*2}}$$

$$g(T) (\hat{\beta}_T - \beta_\alpha) = \frac{g(T)^{-1} \sum_{t=1}^T y_{t-1} \varepsilon_t}{g(T)^{-2} \sum_{t=1}^T y_t^2}$$

Sachant que $g(T)^{-1} \sum_{t=1}^T y_{t-1} \varepsilon_t$ et $g(T)^{-2} \sum_{t=1}^T y_t^2$ ont des lois limites non dégénérées dans tous les cas de figure (voir WHITE [1958] pour le cas $|\beta_\alpha| \leq 1$, ANDERSON [1959] pour le cas $|\beta_\alpha| > 1$), il suffit de montrer que

$g(T)^{-2} \sum_{t=1}^T y_{t-1}^* \varepsilon_t^*$ et $g(T)^{-1} \sum_{t=1}^T y_{t-1} \varepsilon_t$ (respectivement $g(T)^{-2} \sum_{t=1}^T y_t^{*2}$ et $g(T)^{-2} \sum_{t=1}^T y_t^2$) convergent (le premier en loi conditionnelle) vers une même loi presque sûrement dans la suite. Ces résultats sont vérifiés si on montre que sous H_α :

$$DM_2 \left(\mathcal{L}_{\tilde{F}_T} \left(g(T)^{-1} \sum_{t=1}^T y_{t-1}^* \varepsilon_t^* \right), \mathcal{L}_F \left(g(T)^{-1} \sum_{t=1}^T y_{t-1} \varepsilon_t \right) \right) \xrightarrow[T \rightarrow \infty]{\text{p.s.}} 0.$$

et

$$DM_1 \left(\mathcal{L}_{\tilde{F}_T} \left(g(T)^{-2} \sum_{t=1}^T y_t^{*2} \right), \mathcal{L}_F \left(g(T)^{-2} \sum_{t=1}^T y_t^2 \right) \right) \xrightarrow[T \rightarrow \infty]{\text{p.s.}} 0.$$

où $\mathcal{L}_{\tilde{F}_T} \left(g(T)^{-1} \sum_{t=1}^T y_{t-1}^* \varepsilon_t^* \right)$ désigne la distribution de $g(T)^{-1} \sum_{t=1}^T y_{t-1}^* \varepsilon_t^*$ sous la loi \tilde{F}_T des résidus, conditionnellement à y_1, \dots, y_T et $\mathcal{L}_F \left(g(T)^{-1} \sum_{t=1}^T y_{t-1} \varepsilon_t \right)$ la distribution de $g(T)^{-1} \sum_{t=1}^T y_{t-1} \varepsilon_t$, lorsque ε_t a pour loi F etc...

LEMME 2 : Il existe une constante K' telle que :

$$\begin{aligned} & DM_2 \left(\mathcal{L}_{\tilde{F}_T} \left(g(T)^{-1} \sum_{t=1}^T y_{t-1}^* \varepsilon_t^* \right), \mathcal{L}_F \left(g(T)^{-1} \sum_{t=1}^T y_{t-1} \varepsilon_t \right) \right) \\ & \leq K' DM_2(\tilde{F}_T, F) \quad \text{p.s.} \end{aligned}$$

Démonstration : On a par définition de la distance de Mallows :

$$\begin{aligned} DM_2 &= DM_2 \left(\mathcal{L}_{\tilde{F}_T} \left(g(T)^{-1} \sum_{t=1}^T y_{t-1}^* \varepsilon_t^* \right), \mathcal{L}_F \left(g(T)^{-1} \sum_{t=1}^T y_{t-1} \varepsilon_t \right) \right)^2 \\ &\leq \inf_{\substack{\Lambda(u_t, \infty) = F \\ \Lambda(\infty, u_t^*) = \tilde{F}_T \\ \Lambda(u_t, u_t^*) = \Lambda \\ (u_t, u_t^*) \text{ i.i.d.}}} E_\Lambda \left(\left(g(T)^{-1} \sum_{t=1}^T y_{t-1}^*(u^*) u_t^* - g(T)^{-1} \sum_{t=1}^T y_{t-1}(u) u_t \right)^2 \right). \end{aligned}$$

avec les constructions

$$\begin{aligned} y_t^*(u^*) &= \beta_\alpha y_{t-1}^*(u^*) + u_t^*, & y_\alpha^*(u^*) &= y_0, & t &= 1, \dots, T \\ y_t(u) &= \beta_\alpha y_{t-1}(u) + u_t, & y_0(u) &= y_0, & t &= 1, \dots, T \end{aligned}$$

Il suffit donc de montrer que le second terme de l'inégalité tend vers 0.

Mais en utilisant l'inégalité $\left(\sum_{i=1}^2 x_i\right)^2 \leq 2 \sum_{i=1}^2 x_i^2$ on a

$$\begin{aligned} E_{\Lambda} & \left(\left(g(T)^{-1} \sum_{t=1}^T y_{t-1}^*(u^*) u_t^* - g(T)^{-1} \sum_{t=1}^T y_{t-1}(u) u_t \right)^2 \right) \\ & = E_{\Lambda} \left(\left(g(T)^{-1} \sum_{t=1}^T (y_{t-1}^*(u^*) - y_{t-1}(u)) u_t^* \right. \right. \\ & \quad \left. \left. - g(T)^{-1} \sum_{t=1}^T y_{t-1}(u) (u_t - u_t^*) \right)^2 \right) \\ & \leq 2 E_{\Lambda} \left(g(T)^{-1} \sum_{t=1}^T (y_{t-1}^*(u^*) - y_{t-1}(u)) u_t^* \right)^2 \\ & \quad + 2 E_{\Lambda} \left(g(T)^{-1} \sum_{t=1}^T y_{t-1}(u) (u_t - u_t^*) \right)^2 \end{aligned}$$

Les couples (u_t^*, u_t) étant i.i.d. de lois marginales respectives F et \tilde{F}_T , de loi jointe Λ , u_t est indépendant de $(y_{t-1}^*(u^*) - y_{t-1}(u))$ de même que $y_{t-1}(u)$ est indépendant de $(u_t - u_t^*)$. Par ailleurs si $t' > t$ alors $u_{t'}^*$ est indépendant de u_t^* , de u_t , de $y_{t-1}^*(u^*)$ et de $y_{t-1}(u)$ d'où, en remarquant que $E_{\Lambda} u_t^{*2} = \hat{\sigma}_T^2$,

$$\begin{aligned} (9) \quad DM2 & \leq 2 g(T)^{-2} \sum_{t=1}^T E_{\Lambda} ((y_{t-1}^*(u^*) - y_{t-1}(u))^2) \hat{\sigma}_T^2 \\ & \quad + 2 g(T)^{-2} E_{\Lambda} ((u_t - u_t^*)^2) \sum_{t=1}^T E_{\Lambda} (y_{t-1}^2(u)). \end{aligned}$$

Sous forme moyenne mobile $y_t(u)$ et $y_t^*(u^*)$ s'écrivent :

$$y_t(u) = \sum_{i=0}^{t-1} \beta_{\alpha}^i u_{t-i} + \beta_{\alpha}^t y_0 \quad \text{et} \quad y_t^*(u^*) = \sum_{i=0}^{t-1} \beta_{\alpha}^i u_{t-i}^* + \beta_{\alpha}^t y_0.$$

Le calcul de $\sum_{t=1}^T E_{\Lambda} (y_{t-1}^2(u))$ donne :

$$\text{si } |\beta_{\alpha}| \neq 1$$

$$\sigma^2 (1 - \beta_{\alpha}^2)^{-1} (T - (1 - \beta_{\alpha}^{2T}) (1 - \beta_{\alpha}^2)^{-1}) + y_0^2 (1 - \beta_{\alpha}^{2T}) (1 - \beta_{\alpha}^2)^{-1},$$

$$\text{si } |\beta_{\alpha}| = 1$$

$$\sigma^2 T(T+1)/2 + T y_0^2$$

Par construction de $g(T)^2$ qui est la partie principale de ce terme à σ^2 près (voir (4)),

$$(10) \quad 2g(T)^{-2} \sum_{t=1}^T E_{\Lambda}(y_{t-1}^2)(u) \xrightarrow[T \rightarrow \infty]{\text{p.s.}} 2\sigma^2$$

De plus on a :

$$y_t^*(u^*) - y_t(u) = \sum_{i=0}^{t-1} \beta_{\alpha}^i (u_{t-i}^* - u_{t-i}),$$

d'où :

$$\begin{aligned} E_{\Lambda}(y_t^*(u^*) - y_t(u))^2 &\leq 2 \sum_{i=0}^{t-1} \beta_{\alpha}^{2i} E_{\Lambda}(u_{t-i}^* - u_{t-i})^2 \\ &\leq 2 \frac{1 - \beta_{\alpha}^{2t}}{1 - \beta_{\alpha}^2} E_{\Lambda}(u_t^* - u_t)^2 \quad \text{si } |\beta_{\alpha}| \neq 1 \\ &\leq 2t E_{\Lambda}(u_t^* - u_t)^2 \quad \text{si } |\beta_{\alpha}| = 1 \end{aligned}$$

Comme on a

$$\sum_{t=0}^{T-1} \frac{1 - \beta_{\alpha}^{2t}}{1 - \beta_{\alpha}^2} = T/(1 - \beta_{\alpha}^2) - (1 - \beta_{\alpha}^2)^{-2} \beta_{\alpha}^2 (1 - \beta_{\alpha}^{2T})$$

et

$$\sum_{t=0}^T t = T(T-1)/2,$$

on en déduit que :

$$(11) \quad g(T)^{-2} \sum_{t=1}^T E_{\Lambda}(y_{t-1}^*(u^*) - y_{t-1}(u))^2 \leq K E_{\Lambda}(u_t^* - u_t)^2,$$

où K est une constante. En rassemblant les résultats (9)-(11) et passant à l'inf sur l'ensemble des couples (u_t^*, u_t) iid on obtient :

$$DM_2 \leq K' DM_2(\tilde{F}_T, F).$$

où K' est une constante.

Si l'on montre que $DM_2(\tilde{F}_T, F) \xrightarrow[T \rightarrow \infty]{\text{p.s.}} 0$, on en déduit la

convergence de $g(T)^{-1} \sum_{t=1}^T y_{t-1}^* \varepsilon_t^*$ vers la même loi asymptotique que

$g(T)^{-1} \sum_{t=1}^T y_{t-1} \varepsilon_t$. Ceci est fait dans le lemme suivant.

LEMME 3 : Sous les hypothèses du théorème 1 on a :

$$DM_2(\tilde{F}_T, F) \xrightarrow[T \rightarrow \infty]{\text{p.s.}} 0.$$

Démonstration : On notera que chez BASAWA, McCORMICK, TAYLOR [1989], la condition technique permettant de contrôler les restes dans le développement des fonctions caractéristiques, s'apparente à $DM_2(\tilde{F}_T, F) \xrightarrow[T \rightarrow \infty]{p.s.} 0$. Comme \tilde{F}_T s'obtient par simple centrage de la fonction de répartition des résidus estimés et que les estimateurs des m.c.o. sont convergents on a

$$DM_2(\tilde{F}_T, \hat{F}_T) \xrightarrow[T \rightarrow \infty]{p.s.} 0$$

car d'après la propriété A.3

$$DM_2(\tilde{F}_T, \hat{F}_T) = \int_0^1 (\tilde{F}_T^{-1}(t) - (\hat{F}_T^{-1}(t))^2) dt \leq \left| \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t \right|$$

tend p.s. vers 0.

Soit F_T la fonction de répartition empirique des ε_t . En utilisant l'inégalité triangulaire et le fait que $DM_2(F, F_T) \xrightarrow[T \rightarrow \infty]{p.s.} 0$ (propriété A.2), il suffit de montrer

$$DM_2(\hat{F}_T, F_T) \xrightarrow[T \rightarrow \infty]{p.s.} 0,$$

Considérant les v.a. (X, Y) de loi de probabilité jointe $\Lambda = n^{-1} \sum_{t=1}^T \delta_{(\hat{\varepsilon}_t, \varepsilon_t)}$

où δ_y désigne la masse de dirac en y , alors leurs lois marginales sont \hat{F}_T et F_T et on a

$$DM_2(\hat{F}_T, F_T)^2 = E_{\Lambda} (X - Y)^2 \leq \frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_t - \varepsilon_t)^2$$

On peut alors remarquer :

$$\frac{1}{T} \sum_{t=1}^T (\hat{\varepsilon}_t - \varepsilon_t)^2 = \frac{1}{T} \|P_{y_{-1}} \varepsilon\|^2$$

où $P_{y_{-1}}$ est le projecteur sur le vecteur des composantes passées :

$$y_{-1} = \begin{pmatrix} y_0 \\ | \\ y_{T-1} \end{pmatrix}.$$

Poursuivant le calcul on obtient :

$$\begin{aligned} \|P_{y_{-1}} \varepsilon\|^2 &= \|y_{-1} (y'_{-1} y_{-1})^{-1} y'_{-1} \varepsilon\|^2 = \|y_{-1} y'_{-1} \varepsilon / (y'_{-1} y_{-1})\|^2 \\ &= (y'_{-1} \varepsilon / (y'_{-1} y_{-1}))^2 \|y_{-1}\|^2 = (g(T) (\hat{\beta}_T - \beta_{\alpha}))^2 \frac{1}{g(T)^2} \|y_{-1}\|^2 \end{aligned}$$

Or $g(T) (\hat{\beta}_T - \beta_{\alpha})$ et $\frac{1}{g(T)^2} \|y_{-1}\|^2$ sont stochastiquement bornés p.s. (voir WHITE [1958], ANDERSON [1959]). \square

LEMME 4 : Sous les conditions du théorème 1, on a :

$$DM_1 \left(\mathcal{L}_{\tilde{F}_T} \left(g(T)^{-2} \sum_{t=1}^T y_t^{*2} \right), \mathcal{L}_F \left(g(T)^{-2} \sum_{t=1}^T y_t^2 \right) \right) \xrightarrow[T \rightarrow \infty]{\text{p.s.}} 0$$

Démonstration : Reprenant les notations de la preuve du lemme 2, on obtient

$$DM_1 \leq E_\Lambda \left(\left| g(T)^{-2} \sum_{t=1}^T y_t^{*2}(u^*) - g(T)^{-2} \sum_{t=1}^T y_t^2(u) \right| \right)$$

avec $\Lambda(u_t, \infty) = F$, $\Lambda(\infty, u_t^*) = \tilde{F}_T$, $\Lambda(u_t, u_t^*) = \Lambda(u_t, u_t^*)$ i.i.d.
Cauchy-Schwarz donne

$$DM_1 \leq \left(E_\Lambda g(T)^{-2} \sum_{t=1}^T (y_t^*(u^*) - y_t(u))^2 \right)^{1/2} \\ \times E_\Lambda \left(g(T)^{-2} \sum_{t=1}^T (y_t^*(u^*) + y_t(u))^2 \right)^{1/2}$$

Combinant le lemme 2 et le lemme 3 on obtient

$$g(T)^{-2} \sum_{t=1}^T E_\Lambda (y_{t-1}^*(u^*) - y_{t-1}(u))^2 \leq K E_\Lambda (u_t^* - u_t)^2 \xrightarrow[T \rightarrow \infty]{\text{p.s.}} 0$$

Par ailleurs, on a la majoration :

$$E_\Lambda \left(g(T)^{-2} \sum_{t=1}^T (y_t^*(u^*) + y_t(u))^2 \right) \\ \leq 2 g(T)^{-2} \left(E_\Lambda \left(\sum_{t=1}^T y_t^{*2}(u^*) \right) + E_\Lambda \left(\sum_{t=1}^T y_t^2(u) \right) \right)$$

Or on sait que $g(T)^{-2} E_\Lambda \sum_{t=1}^T y_t^2(u)$ est stochastiquement borné par construction de $g(T)$. D'autre part en reprenant les calculs du lemme 2 avec u^* au lieu de u , on obtient

$$E_{\tilde{F}_T} y_{t-1}^*(u^*)^2 = \sum_{i=1}^{t-1} \beta_\alpha^{2i} E_{\tilde{F}_T} u_{t-i}^{*2} \\ \leq 2 \beta_\alpha^2 \frac{1 - \beta_\alpha^{2t}}{1 - \beta_\alpha^2} E_{\tilde{F}_T} u_t^{*2} \quad \text{si } |\beta_\alpha| \neq 1 \\ \leq 2t E_{\tilde{F}_T} u_t^{*2} \quad \text{si } |\beta_\alpha| = 1.$$

Par définition de $g(T)$ et comme $E_{\tilde{F}_T} u_t^{*2}$ est la variance empirique des résidus estimés centrés (qui converge, en utilisant le lemme 3, p.s. vers σ^2)

on en déduit que $g(T)^{-2} \sum_{t=1}^T E_{\tilde{F}_T} y_t^* (u^*)^2$ est p.s. borné d'où le résultat du lemme. \square

Les lemmes 2 et 3 permettent donc de conclure que :

$$DM_2 \left(\mathcal{L}_{\tilde{F}_T} \left(g(T)^{-1} \sum_{t=1}^T y_{t-1}^* \varepsilon_t^* \right), \mathcal{L}_F \left(g(T)^{-1} \sum_{t=1}^T y_{t-1} \varepsilon_t \right) \right) \xrightarrow[T \rightarrow \infty]{\text{p.s.}} 0.$$

Le lemme 4 montre que :

$$DM_1 \left(g(T)^{-2} \sum_{t=1}^T y_t^{*2}, g(T)^{-2} \sum_{t=1}^T y_t^2 \right) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0$$

ce qui achève la démonstration du théorème. \square

Remarques :

- Le choix de $y_0^* = y_0$ est important dans la démonstration du résultat uniquement dans le cas explosif $|\beta| \geq 1$. En effet si, dans (3), y_0^* est choisi constant différent de y_0 , alors le résultat

$$g(T)^{-2} \sum_{t=1}^T E_{\Lambda} (y_{t-1}^* - y_{t-1})^2 \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0,$$

qui joue un rôle fondamental dans la démonstration du théorème, n'est pas vérifiée. Ceci n'est pas étonnant dans la mesure où l'on a vu dans le paragraphe 1 que la condition initiale n'intervenait pas dans les distributions limites des estimateurs des m.c.o. si $|\beta| \leq 1$, mais avait une influence décisive sur la loi limite dans le cas $|\beta| > 1$. Cependant, vu les résultats de EVANS et SAVIN [1981b] [1984] sur l'importance de la constante dans la détermination des distributions à distance finie, il semble préférable dans tous les cas de garder $y_0^* = y_0$.

- La validité au second ordre de la méthode dans le cas stationnaire peut s'obtenir directement en adaptant les résultats de BOSE [1988] au processus Bootstrap défini sous H_0 . Le choix de $y_0^* = y_0$ est alors fondamental. On notera que dans l'optique test, il n'est pas nécessaire de standardiser par un écart-type estimé la distribution de l'e.m.c.o. pour pouvoir récupérer les propriétés au second ordre car, sous H_0 , la distribution asymptotique ne dépend pas de paramètres de nuisance. La validité au second ordre pour des processus non-stationnaire ou explosif semble plus difficile à obtenir car le développement d'Edgeworth usuel sur la gaussienne n'est plus valable.

Mise en œuvre pratique du test :

Le théorème précédent justifie l'utilisation des quantiles de la distribution Bootstrap pour construire une région critique pour le test (1). Ne pouvant pas déterminer pratiquement les quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$ de la loi $G_{T,0}^*$ de $\hat{\beta}_T^*$ e.m.c.o. construit sous H_0 (parce que cela nécessiterait trop de calcul, en fait T^T) on est amené à donner une approximation de ces quantiles par une technique de type Monte-carlo. La mise en œuvre du test se décompose donc ainsi :

- ayant calculé l'estimateur des m.c.o. dans le modèle (2) on construit la fonction de répartition empirique des résidus estimés centrés : \tilde{F}_T .

- La méthode de rééchantillonnage des résidus dans \tilde{F}_T , i.e. de façon pratique, le tirage avec remise dans l'ensemble des résidus centrés estimés, conduit à construire un processus du type (3) qui permet de calculer l'estimateur bootstrap des m.c.o. La répétition B fois de cette phase permet d'obtenir un ensemble d'estimations sous l'hypothèse nulle $\beta_{T,0}^{*(b)}$, pour $b = 1, \dots, B$, dont la fonction de répartition empirique $G_{T,0_B}^*(u)$ approxime la vraie distribution $G_{T,0}^*$ de $\hat{\beta}_T^*$ sous H_0 . On utilise alors une méthode de type percentile (EFRON [1979]) pour construire une région d'acceptation approximative au niveau α du test $H_0 : \beta = \beta_0$:

- Soient $g_{T,0_B}^*(\alpha/2)$ et $g_{T,0_B}^*(1 - \alpha/2)$ respectivement les quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$ de $G_{T,0_B}^*$:

$$g_{T,0_B}^*(\beta) = \inf \{t, G_{T,0_B}^*(t) \geq \beta\}$$

L'intervalle $[g_{T,0_B}^*(1 - \alpha/2), g_{T,0_B}^*(\alpha/2)]$ est une région d'acceptation approchée du test bilatéral. La règle de décision s'en déduit naturellement : on accepte l'hypothèse H_0 si l'estimateur des moindres carrés ordinaires $\hat{\beta}_T$ appartient à $[g_{T,0_B}^*(\alpha/2), g_{T,0_B}^*(1 - \alpha/2)]$, on rejette sinon.

- Le résultat de validité de la méthode quelque soit la valeur du paramètre testée permet aisément de construire un diagramme de puissance estimé du test (1) dans l'optique de BERAN [1986]. En effet, si l'on considère une alternative locale β_a , on peut construire un processus Bootstrap sous H_a et obtenir sa distribution Bootstrap. Un même tirage des résidus permet de construire le processus Bootstrap sous H_0 et sous des alternatives locales $\beta_a \neq \beta_0$. La puissance du test contre une alternative β_a s'obtient simplement en comptant le nombre de valeurs de l'e.m.c.o. bootstrap sous H_a (ramené aux nombres B de simulations effectuées) qui sont dans la région critique construite.

Même si l'on dispose de résultats au second ordre, les résultats théoriques obtenus n'en restent pas moins asymptotiques. Il convient donc d'étudier la validité à distance finie de la méthode. Ceci est fait dans le paragraphe suivant qui montrent, par simulation, que dans certains cas le test Bootstrap peut être mieux adapté que le test asymptotique.

3 Une étude par simulation

3.1. Préliminaire

Les simulations ont été mises en œuvre sur le modèle $AR(1)$ simple avec résidus i.i.d. normaux, afin de comparer les résultats des simulations avec la théorie exposée brièvement en 2.1. On s'intéresse au test de l'hypothèse $H_0 : \beta = 1$.

Étant donnée la dissymétrie des distributions limites et des distributions à distance finie, il ne semble pas judicieux, pour un seuil α donné, de prendre des régions d'acceptation du test basées sur les quantiles d'ordre

$\alpha/2$ et $1 - \alpha/2$, mais plus intéressant de choisir des quantiles α_1 et $1 - \alpha_2$ avec $\alpha_1 + \alpha_2 = \alpha$ de manière à obtenir une région d'acceptation la plus petite possible ou un test ayant une puissance maximale pour une alternative donnée (les deux critères n'étant pas forcément compatibles).

3.2. Simulations

Pour $\alpha = 5\%$ nous avons choisi de construire 9 intervalles de confiance correspondant à $\alpha_1 \in \{0.5\%, 1.0\%, 1.5\%, 2.0\%, 2.5\%, 3.0\%, 3.5\%, 4.0\%, 4.5\%\}$. Le tableau ci-dessous obtenu sur une simulation donne les différentes régions d'acceptation du test bootstrap, ainsi que leur longueur $l_{T,0_B}^*(\alpha_1)$. Le nombre de rééchantillonnage utilisé pour construire les distributions Bootstrap empiriques est $B = 3000$. On a choisi ici $\beta = 1$, $T = 25$, $\varepsilon_t \sim N(0, 1)$, $y_0 = 0$. Le paramètre β n'est pas bien estimé par les m.c.o. dans la simulation retenue : $\hat{\beta}_T = 0.747$, $\hat{\sigma}_{\hat{\beta}} = 0.136$. Le test de DICKEY-FULLER conduit néanmoins à ne pas rejeter l'hypothèse de racine-unité. On pourra consulter d'autres résultats de simulation dans BERTAIL [1992].

TABLEAU 1

Régions d'acceptation du test de racine unité, basée sur la distribution Bootstrap de l'e.m.c.o.

α_1 %	$1 - \alpha_2$ %	$g_{T,0_B}^*(\alpha_1)$	$g_{T,0_B}^*(1 - \alpha_2)$	$l_{T,0_B}^*(\alpha_1)$
0.5	95.5	0.468	1.057	0.589
1.0	96.0	0.520	1.061	0.540
1.5	96.5	0.572	1.065	0.493
2.0	97.0	0.612	1.017	0.459
2.5	97.5	0.650	1.076	0.426
3.0	98.0	0.668	1.078	0.410
3.5	98.5	0.678	1.085	0.407
4.0	99.0	0.689	1.091	0.402
4.5	99.5	0.702	1.098	0.396

On notera que dans tous les cas (même pour $\alpha_1 = 4.5\%$) on accepte l'hypothèse nulle. La longueur de l'intervalle décroît lorsque α_1 augmente, ce qui invite à prendre des régions d'acceptation plus centrées autour de la valeur 1. Il convient néanmoins dans les applications de ne pas choisir $\alpha_1 = 5\%$, à cause de la nature discrète de la distribution empirique dans laquelle s'effectue le rééchantillonnage ; on risquerait d'obtenir une valeur aberrante pour la borne supérieure de l'intervalle de confiance dans un problème de test bilatéral.

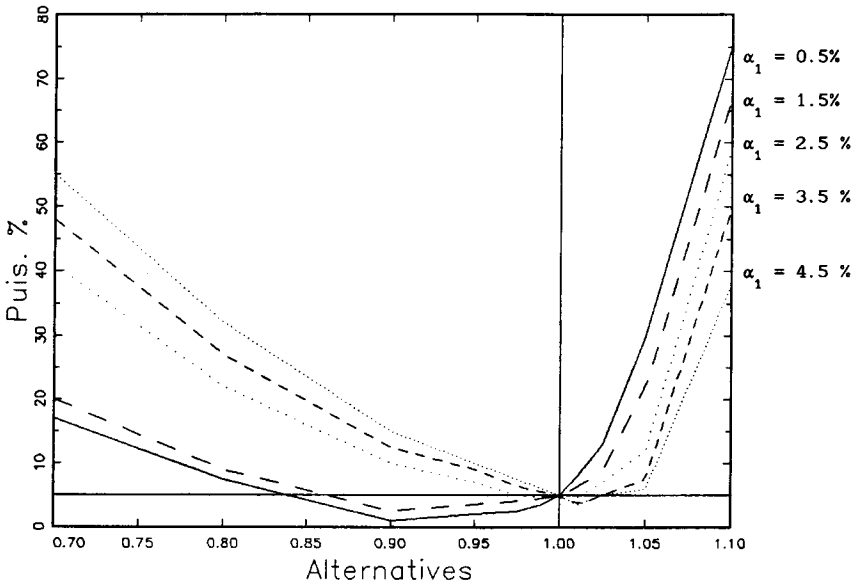
Les problèmes liés à la multiplicité des granicules disparaissent si, au lieu d'utiliser la fonction de répartition empirique, on utilise un estimateur à noyau de F :

$$\tilde{F}_T(x) = \frac{1}{T} \sum_{t=1}^T K\left(\frac{x - \tilde{\varepsilon}_t}{h}\right)$$

où K est le noyau intégré (par exemple $K(y) = (2\pi)^{-1/2} \int_{-\infty}^y \exp(-u^2/2)$) et h est le paramètre de lissage. Il est clair que la procédure de test et la démonstration du théorème fonctionne toujours dans un tel cadre pourvu que la constante de lissage soit bien choisie. Cependant, quelques simulations effectuées dans le cas du modèle $AR(1)$ ont permis de mettre en évidence l'importance du paramètre de lissage h dans l'obtention de la distribution Bootstrap de l'estimateur des m.c.o. : les résultats peuvent être considérablement différents pour de petites variations de h .

3.3. Puissance simulée

Les diagrammes des puissances des tests associés aux différentes régions de confiance construites en 3.2., estimées par la méthode du Bootstrap sont fortement dissymétriques (voir graphique 1).



GRAPHIQUE 1
Puissance estimée du test : cas $AR(1)$.

Il montre bien que le vrai problème est de tester $\beta = 1$ contre $\beta < 1$. Ce sont pour des alternatives $\beta_a < 1$ proches de 1 que l'on obtient les puissances les plus faibles. Le test est d'autant plus fortement biaisé (pour des alternatives $\beta_a < 1$ proches de 1) que l'on choisit la valeur de α_1 petite. Cependant il peut même, pour certaines valeurs de α_1 , être biaisé de part et d'autre de la valeur 1. Il semble ressortir que l'intervalle de confiance construit avec la valeur $\alpha_1 = 4.5\%$ est le plus recommandable (aussi bien en terme de puissance qu'en terme de longueur d'intervalle).

EVANS, SAVIN [1981a] donnent quelques résultats à distance finie sur la puissance du test de racine unité basé sur l'e.m.c.o. dans un $AR(1)$ avec

erreurs gaussiennes, ce qui permet d'avoir quelques éléments de comparaison (pour des régions d'acceptation symétriques basées sur les quantiles d'ordre $\alpha/2$ et $1 - \alpha/2$, $\alpha = 5\%$). On notera que les puissances obtenues par EVANS et SAVIN [1981] mettent aussi en évidence un biais du test.

Pour $T = 25$, le tableau suivant compare les puissances calculées pour des alternatives proches de 1 par EVANS et SAVIN [1981a] (P.E.S.) lorsque l'intervalle est construit avec les quantiles d'ordre 97.5 et 2.5, avec les puissances obtenues par la méthode du Bootstrap (P.B.).

TABLEAU 2

Comparaison des puissances exactes (P.E.S.) et des puissances estimées par la méthode du Bootstrap (P.B.) en %.

Alt	0.90	0.95	0.99	1.00	1.01	1.025	1.05
P.E.S. . .	7.79	5.06	4.72	5.00	5.70	7.59	16.45
P.B. . . .	7.52	5.02	4.80	5.00	5.45	6.87	14.25

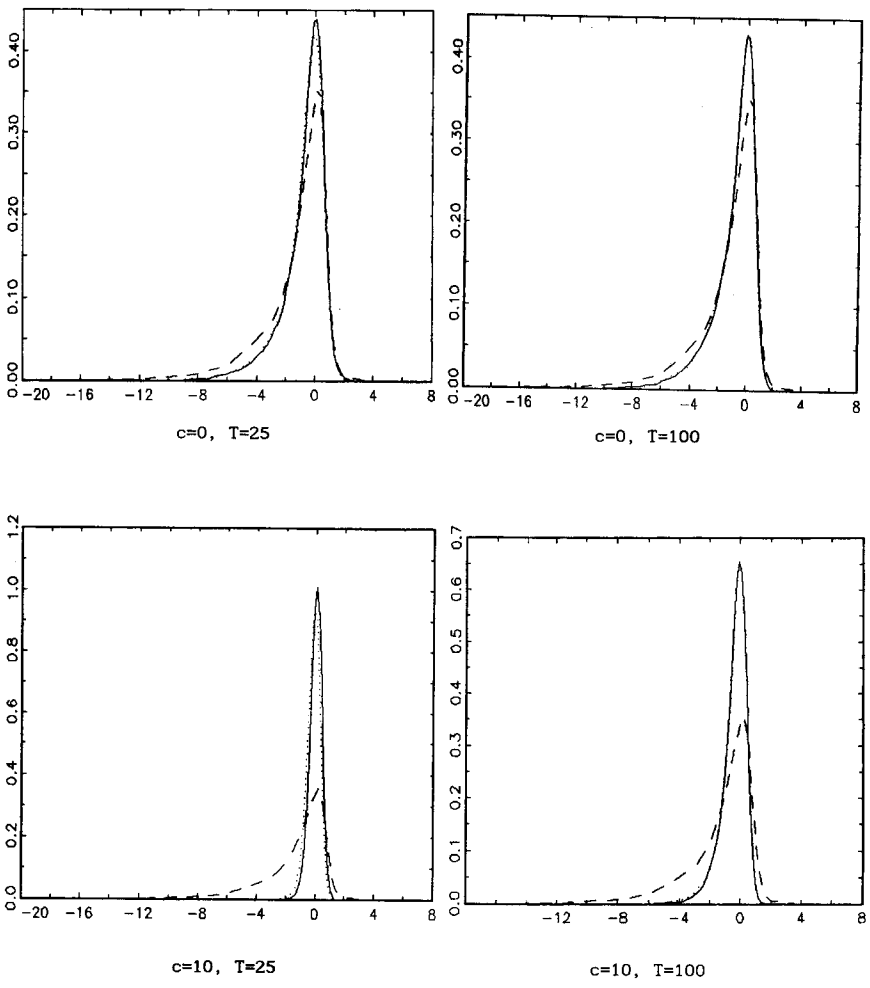
Ces résultats (confirmés par toutes les autres simulations) sont assez satisfaisants : on notera néanmoins que dans ce cas précis la qualité de l'estimation de la puissance pour des alternatives supérieures à 1 est moins grande.

3.4. Étude des distributions Bootstrap

Les graphiques 2, 3, 4 et 5 donnent la comparaison des estimations de la densité Bootstrap, centrée dilatée de l'estimateur des m.c.o., après lissage (le paramètre de lissage choisi est celui qui minimise l'écart-quadrique moyen) sous l'hypothèse nulle $H_0 : \beta = 1$ pour $T = 100$ (et $T = 25$ pour $\beta_0 = 1$ et $\beta_0 = 0.95$), de la distribution asymptotique et de la distribution à distance finie (estimée à partir de 10000 tirages), pour deux valeurs différentes de c . Pour tous les graphiques, la légende est la suivante :

- Distribution Bootstrap
- - - Distribution Asymptotique L_1
- Distribution Exacte

Dans tous les cas, la distribution Bootstrap est très proche de la distribution à distance finie et nettement meilleure que l'approximation asymptotique. Pour $\beta = 1$ et $c = 0$ (graphique 2), la distribution Bootstrap et la distribution limite, lorsque T augmente, sont relativement proches. Pour $c = 10$, même avec une taille d'échantillon grande $T = 100$ (compte tenu de la convergence en T^1), l'asymptotique rend mal compte du comportement à distance finie contrairement à la distribution Bootstrap qui donne une très bonne approximation de la vraie distribution. Dans le cas $\beta_0 = 0.95$, l'approximation gaussienne est d'autant plus mauvaise que T est petit et c proche de 0. La distribution Bootstrap permet bien de prendre en compte l'impact de la condition initiale sur la forme de la distribution. Pour des valeurs $|\beta| > 1$ très proche de 1 ($\beta = 1.01$) on peut formuler des remarques



GRAPHIQUE 2
Comparaison des densités exactes, Bootstrap et asymptotiques, $\beta_0 = 1$.

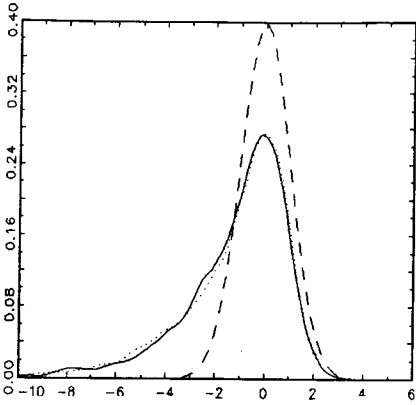
semblables même pour une taille d'échantillonnage assez élevée $T = 100$. Lorsque β s'éloigne de 1 (graphique 4), les distributions exactes, Bootstrap et asymptotiques (construites pour $c = 10$ à partir de la relation (8) la somme infinie étant tronquée à $i = 10$) sont pratiquement confondues : rappelons que dans ce cas la vitesse de convergence est en $(\beta^2 - 1)/\beta_0^{-T}$ (ce qui ici donne une erreur de l'asymptotique sur la vraie distribution de l'ordre 10^{-5}).

Dans tous les cas, les tests usuels (test de student et test du signe) rejettent l'hypothèse de nullité de la moyenne de ces distributions ce qui signifie que l'estimateur choisi est biaisé. Pour $T = c = 25$, si le nombre de rééchantillonnage est important ($B > 3000$) la méthode du Bootstrap donne des ordres de grandeurs semblables à ceux obtenus par calculs avec l'approximation de WHITE [1961] comme le montre le tableau ci-dessous.

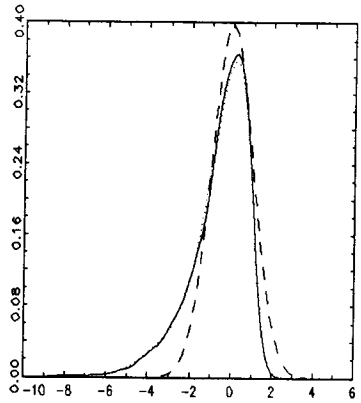
TABLEAU 3

*Comparaison des estimations du biais de $\hat{\beta}_T$ par rapport à β_0 .
Approximation de White et estimation Bootstrap.*

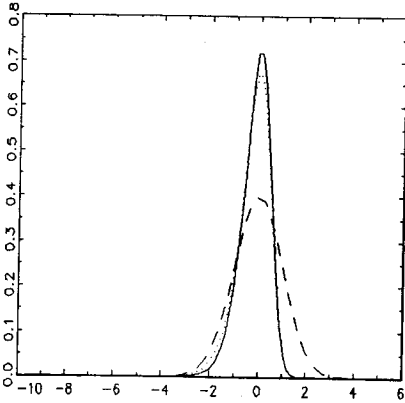
β_0	0.9	0.95	0.975	0.99	1.0
Approximation du biais (WHITE [1961])	-0.066	-0.070	-0.072	-0.073	-0.074
Méthode du Bootstrap $B=3\ 000$	-0.059	-0.068	-0.067	-0.069	-0.071



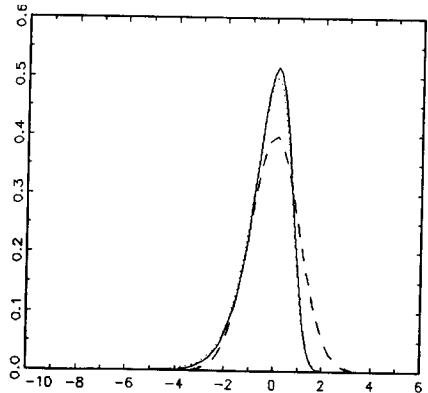
$c=0, T=25$



$c=0, T=100$



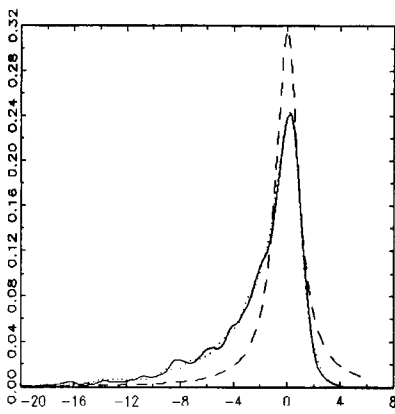
$c=10, T=25$



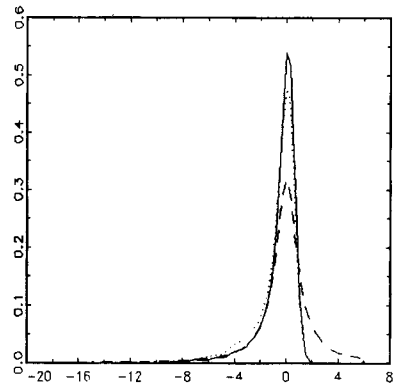
$c=10, T=100$

GRAPHIQUE 3

Comparaison des densités exactes, Bootstrap et asymptotiques, $\beta_0=0.95$.



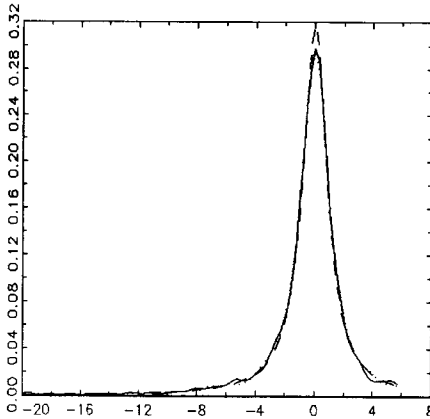
c=0, T=100



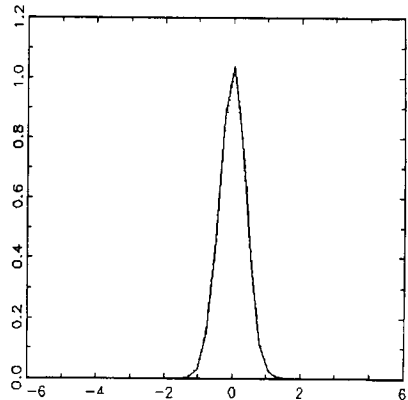
c=10, T=100

GRAPHIQUE 4

Comparaison des densités exactes, Bootstrap et asymptotiques, $\beta_0 = 1.01$.



c=0, T=100



c=10, T=100

GRAPHIQUE 5

Comparaison des densités exactes, Bootstrap et asymptotiques, $\beta_0 = 1.1$.

4 Conclusion

Les simulations effectuées confirment le bien-fondé théorique de la méthode et mettent en évidence l'apport du Bootstrap dans le problème de test de racine unité. Les distributions Bootstrap rendent beaucoup mieux compte du comportement à distance finie des estimateurs. Ceci est particulièrement sensible pour des alternatives $\beta_a < 1$. Nous

renvoyons à A. BOSE [1988] pour des résultats montrant que, dans un autorégressif stationnaire ($\beta < 1$), la distribution Bootstrap coïncide avec un développement d'Edgeworth à l'ordre 1 en $o(T^{1/2})$. Ceci est particulièrement important en terme de puissance du test de l'hypothèse $H_0 : \beta = 1$ mais aussi si l'on veut tester $H_0 : \beta = \beta_0$ avec $\beta_0 < 1$ proche de 1.

L'étude de la validité au second ordre du Bootstrap de l'estimateur des m.c.o. dans le cas non-stationnaire semble plus difficile (les limites n'étant pas gaussiennes), car on ne dispose que de très peu de résultats sur les développements d'Edgeworth dans ce cadre. Il serait néanmoins intéressant d'obtenir un développement (dont le premier terme serait la fonction de répartition de la loi asymptotique) et d'étudier l'analogie Bootstrap.

L'application du Bootstrap à un problème de test dans un modèle $AR(1)$ présente de nombreux intérêts dont celui de permettre l'obtention d'estimation des distributions plus proches des vraies lois que les distributions asymptotiques généralement utilisées. Cette étude peut donner lieu à divers prolongements et applications, par exemple le test bootstrap de racine unité dans un modèle autorégressif avec présence de variables exogènes stationnaires peut facilement être mis en œuvre dans ce cadre. On peut par ailleurs s'interroger sur une généralisation naturelle de la méthode de test Bootstrap sous l'hypothèse nulle à des modèles et des problèmes plus complexes (par exemple pour tester la présence d'une ou plusieurs racines unités dans un modèle $AR(p)$ $p \geq 2$) : désirant tester une hypothèse H_0 complexe dans un modèle dynamique, les résultats obtenus ici suggèrent de construire la version bootstrap du modèle en utilisant l'estimateur contraint au lieu de l'estimateur non contraint. Elle soulève par ailleurs d'autres points théoriques : comment mettre en œuvre la procédure Bootstrap si les erreurs ne sont pas i.i.d. ? Comment montrer la validité au second ordre du Bootstrap dans le cas de limite non gaussienne ?

Quelques propriétés de la distance de Mallows

On note $F(\mathcal{R})$ l'espace des fonctions de répartition (les résultats restent valables si $F(\mathcal{R})$ est un espace de probabilités) définies sur un Banach \mathcal{R} de cotype 2 (typiquement un espace de Hilbert) muni d'une norme $\| \cdot \|$, $F_p(\mathcal{R}) = \left\{ F \in F(\mathcal{R}), \int \|x\|^p dF < \infty \right\}$.

Soient $(F, G) \in \mathcal{F}_p(\mathcal{R}) \times \mathcal{F}_p(\mathcal{R})$, la distance de Mallows d'ordre p entre F et G est définie par :

$$DM_p(F, G) = \text{Inf}_{(X, Y)} (E_\Lambda (\|X - Y\|^p))^{1/p}$$

avec $\Lambda(X, \infty) = F, \Lambda(\infty, Y) = G$

où $\Lambda = \Lambda(X, Y)$ est la loi jointe du couple (X, Y) , $\Lambda(X, \infty)$ la loi marginale de X , $\Lambda(\infty, Y)$ la loi marginale de Y . On a alors les résultats suivants :

PROPRIÉTÉ A.1 : Soient $F_n \in F_p(\mathcal{R})$ et $F \in F_p(\mathcal{R})$ alors on a l'équivalence suivante :

$$DM_p(F_n, F) \xrightarrow[n \rightarrow \infty]{} 0 \Leftrightarrow F_n \xRightarrow[n \rightarrow \infty]{} F$$

et $\int \|x\|^p dF_n \xrightarrow[n \rightarrow \infty]{} \int \|x\|^p dF$

et $F_p(\mathcal{R})$ muni de DM_p est complet.

Démonstration : voir BICKEL et FREEDMAN [1981]. \square

$\xRightarrow[n \rightarrow \infty]{} F$ désigne la convergence faible au sens de BILLINGSLEY [1968]. La caractérisation précédente montre que de la convergence au sens de DM_p est plus forte que la convergence faible et implique l'équintégrabilité de la fonction $\|x\|^p$ (voir aussi SHORACK et WELLNER [1986]).

PROPRIÉTÉ A.2 : Soit (X_1, X_2, \dots, X_n) un échantillon de n variables aléatoires i.i.d. de loi $F \in F_p(\mathcal{R})$, de fonction de répartition empirique F_n alors :

$$DM_p(F_n, F) \xrightarrow[n \rightarrow \infty]{\text{p.s.}} 0.$$

Démonstration : voir BICKEL et FREEDMAN [1981]. \square

PROPRIÉTÉ A.3 : Soient (X_1, X_2, \dots, X_n) n v.a.r. i.i.d. de loi F de fonction de répartition empirique F_n et de fonction de répartition empirique centrée \tilde{F}_n alors

$$DM_p(F_n, \tilde{F}_n) \leq \left| n^{-1} \sum_{i=1}^n X_i \right|$$

Démonstration : On désigne par $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ la statistique d'ordre associée à (X_1, X_2, \dots, X_n) . D'après VALLANDER [1973], on a dans le cas réel :

$$DM_p(F, G) = \left(\int |F^{-1}(t) - G^{-1}(t)|^p dt \right)^{1/p}$$

avec $F^{-1}(t) = \inf(x/F(x) \geq t)$. On en déduit

$$\begin{aligned} DM_p(F_n, \tilde{F}_n) &= \left(\int |F_n^{-1}(t) - \tilde{F}_n^{-1}(t)|^p dt \right)^{1/p} \\ &= \left(\sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{i}{n}} |F_n^{-1}(t) - \tilde{F}_n^{-1}(t)|^p dt \right)^{1/p} = \left(\frac{1}{n} \sum_{i=1}^n |X_{(i)} - Y_{(i)}|^p \right)^{1/p} \end{aligned}$$

avec $Y_i = X_i - n^{-1} \sum_{i=1}^n X_i$ et $Y_{(i)} = X_{(i)} - n^{-1} \sum_{i=1}^n X_i$ d'où le résultat.

● Références bibliographiques

- ANDERSON, T. W. (1959). - "On the Asymptotic Distribution of Estimates of Parameters of Stochastic Difference Equations", *Ann. Math. Statist.*, 30, pp. 676-687.
- BASAWA, I. V., MALLIK, A. K., McCORMICK, W. P., TAYLOR, R. L. (1989). - "Bootstrapping in Explosive Autoregressive Processes", *Ann. Statist.*, 17, pp. 1479-1486.
- BASAWA, I. V., MALLIK, A. K., McCORMICK, W. P., REEVES, J. H., TAYLOR, R. L. (1991). - "Bootstrapping Unstable Autoregressive Processes", *Ann. Statist.*, 19, pp. 1098-1101.
- BERAN, R. (1986). - "Simulated Power Functions", *Ann. Statist.*, 14, pp. 151-173.
- BERTAIL, P. (1992). - "La méthode du Bootstrap : quelques applications et résultats théoriques", *Thèse*, Université Paris IX.
- BICKEL, P. J., FREEDMAN, D. A. (1981). - "Some Asymptotic Theory for The Bootstrap", *Ann. Statist.*, 9, pp. 1196-1217.
- BILLINGSLEY, P. (1968). - "Convergence of Probability Measures", Wiley, New York.
- BOSE, A. (1988). - "Edgeworth Correction by Bootstrap In Autoregressions", *Ann. Statist.*, 16, pp. 1709-1722.
- DICKEY, D. A., FULLER, W. F. (1979). - "Distribution of The Estimators for Autoregressive Time Series with Unit Roots", *J.A.S.A.*, 74, pp. 427-431.
- DICKEY, D. A., FULLER, W. F. (1981). - "Likelihood Ratio for Autoregressive Time Series with Unit Roots", *Econometrica*, 49, pp. 1057-1072.
- DIEBOLD, F. X., NERLOVE, M. (1988). - "Unit Roots in Economic Time-Series: a Selective Survey, *Econometrics: Co-Integration, Spurious Regressions and Unit Roots*", JAI Press.
- EFRON, B. (1982). - "The Jackknife, the Bootstrap, and Other Resampling Plans", *CBMS-NF SIAM*, 38.
- EVANS, G. B. A., SAVIN, N. E. (1981a). - "The Calculation of The Limiting Distribution of the Least Squares Estimators of the Parameter In a Random Walk", *Ann. Statist.*, 9, pp. 1114-1148.
- EVANS, G. B. A., SAVIN, N. E. (1981b). - "Testing For Unit Roots: I", *Econometrica*, 49, pp. 753-779.

- EVANS, G. B. A., SAVIN, N. E. (1984). – “Testing For Unit Roots: II”, *Econometrica*, 52, pp. 1241-1269.
- FINDLEY, D. F. (1986). – “On Bootstrap Estimate of Forecast Mean Square Errors For Autoregressive Processes”, in 17th Symposium On The Interface, ed. M. Allen, Amsterdam, pp. 11-17.
- FREEDMAN, D. (1981). – “Bootstrapping Regression Models”, *Ann. Statist.*, 9, pp. 1218-1228.
- FREEDMAN, D. (1984). – “On Bootstrapping 2-Stage-Least-Squares Estimates In Stationary Models”, *Ann. Statist.*, 12, pp. 827-842.
- FREEDMAN, D., PETERS, S. C. (1984a). – “Bootstrapping a Regression Equation: Some Empirical Results”, *JASA*, 79.
- FREEDMAN, D., PETERS, S. C. (1984b). – “Bootstrapping An Econometric Model: Some Empirical Results”, *J. of Bus. Econ. Statist.*, 2, pp. 150-158.
- HALL, P. (1992). – “The Bootstrap and Edgeworth Expansion”, Wiley, New York.
- IMHOF, P. I. (1961). – “Computing the Distribution of Quadratic Forms in Normal Variables”, *Biometrika*, 48, pp. 419-426.
- MALLOWS, C. L. (1972). – “A Note on Asymptotic Joint Normality”, *Ann. Statist.*, 43, pp. 508-515.
- PHILLIPS, P. C. B. (1987a). – “Towards a Unified Asymptotic Theory of Autoregression”, *Biometrika*, 74, pp. 535-547.
- PHILLIPS, P. C. B. (1987b). – “Times Series with Unit Roots”, *Econometrica*, 55, pp. 277-301.
- RAO, C. R. (1978). – “Asymptotic Distribution of an Estimator of The Boundary Parameter of an Unstable Process”, *Ann. Statist.*, 6, pp. 185-190.
- SCHUCANY, W. R., THOMBS, L. A. (1990). – “Bootstrap Prediction Intervals for Autoregression”, *JASA*, 85, pp. 486-492.
- SHORACK, G. R., WELLNER, J. A. (1986). – “Empirical Processes with Applications to Statistics”, Wiley, New York.
- STINE, R. A. (1985). – “Bootstrap Prediction Intervals for Regression”, *JASA*, 80, pp. 1026-1031.
- VALLANDER, S. S. (1973). – “Calculation of the Wasserstein Metric”, *Theory Probab. Appl.*, 18, pp. 784-786.
- WHITE, J. S. (1958). – “The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case”, *Ann. Math. Statist.*, 29, pp. 1188-1197.
- WHITE, J. S. (1959). – “The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case II”, *Ann. Math. Statist.*, 30, pp. 831-834.
- WHITE, J. S. (1961). – “Asymptotic Expansions for the Mean and Variance of the Serial Correlation Coefficient”, *Biometrika*, 48, pp. 85-94.