JOURNAL DE LA SOCIÉTÉ FRANÇAISE DE STATISTIQUE

JEAN-MARC BARDET

La mémoire longue en économie : discussion et commentaires

Journal de la société française de statistique, tome 140, nº 2 (1999), p. 49-54

<http://www.numdam.org/item?id=JSFS_1999__140_2_49_0>

LA MÉMOIRE LONGUE EN ÉCONOMIE : DISCUSSION ET COMMENTAIRES

Jean-Marc BARDET *

Dans cet article à vocation « pédagogique », les auteurs (LM désignera par la suite indifféremment les auteurs et l'article lui-même) présentent différents modèles de processus linéaires longue-mémoire ainsi que les différentes méthodes d'estimation qui leurs correspondent, puis résument les nombreuses études concernant les utilisations économiques de ces processus. Cet article a de nombreuses vertus, notamment celles de présenter avec beaucoup de clarté et suffisamment de rigueur les objets et outils mathématiques relatifs aux processus linéaires longue-mémoire, de donner un aperu quasiment exhaustif de la littérature traitant statistiquement des processus ARFIMA, d'offrir un nombre considérable de références économiques dans lesquelles la longue dépendance est utilisée ou pourrait l'être, et enfin, d'expliquer heuristiquement la présence et l'intérêt d'une modélisation par des processus longue-mémoire (notamment pour les séries économiques et financières). Ce travail, qui mèle avec bonheur (ce qui est plutôt rare...) deux cultures (mathématique et économique) souvent étanches, est donc fort louable. Plutôt donc que de le discuter à proprement parler, nous voudrions apporter quelques compléments concernant un autre modèle de processus longue-mémoire, le bruit gaussien fractionnaire, et exposer succinctement d'autres techniques d'estimation, notamment semi-paramétriques, du paramètre de longue-mémoire. Nous pensons que ceci pourrait être utilisé avec profit dans les domaines économiques et financiers. On trouvera un exposé beaucoup plus complet sur ces questions, ainsi que les principaux résultats probabilistes et statistiques concernant les modèles longue-mémoire, dans le livre à paraître « Long-range dependence : Theory and Applications », Doukhan *et al.* (2000).

1. LE BRUIT GAUSSIEN FRACTIONNAIRE

Un modèle fréquemment étudié dans les articles sur la longue-mémoire et évoqué rapidement dans LM est le bruit gaussien fractionnaire. Pour définir (rapidement) le bruit gaussien fractionnaire, il est plus clair d'évoquer d'abord le célèbre mouvement brownien fractionnaire. Rappelons donc que $B_H = \{B_H(t), t \in \mathbb{R}_+\}$ est un mouvement brownien fractionnaire lorsque B_H est processus gaussien centré continu ($B_H(0) = 0$ presque sûrement) à

* Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse cedex 04.
email : bardet@cict.fr

accroissements stationnaires, tel que $\mathbb{E}(B_H(t) - B_H(s))^2 = \sigma^2|t-s|^{2H}$ pour tout $(s,t) \in \mathbb{R}_+^2$ et avec $H \in]0,1[$ ($\sigma^2 > 0$). On définit alors usuellement le bruit gaussien fractionnaire $X_H = \{X_H(t), t \in \mathbb{R}_+\}$ comme les accroissements de longueur 1 de B_H, soit

$$X_H(t) = B_H(t+1) - B_H(t), \text{ pour } t \in \mathbb{R}_+$$

(pour plus de détails, on pourra consulter Beran, 1994). On peut ici attirer l'attention sur deux points :

- X_H et B_H sont des processus continus. Cependant, une discrétisation à pas constant quelconque de X_H permet de définir un processus longue-mémoire discret (notons aussi que l'on pourrait définir X_H en considérant les accroissements de longueur constante quelconque autre que 1 de B_H).
- $\{X_H(0), X_H(1), ...\}$ est un processus stationnaire longue-mémoire puisque l'on montre que $\mathrm{cov}(X_H(0), X_H(k)) \sim H(2H-1)\sigma^2 k^{2-2H}$ quand $k \to \infty$ pour $H > 0.5$. Rappelons que le mouvement brownien fractionnaire B_H, lui, n'est pas un processus longue-mémoire.

Indiquons maintenant ce qui nous semble intéressant dans la modélisation de données économiques ou financières par un bruit gaussien fractionnaire (mais aussi par celle d'un mouvement brownien fractionnaire).

1. Le mouvement brownien fractionnaire est une généralisation assez naturelle (au sens de l'autosimilarité) du mouvement brownien ($H = 0.5$), que l'on rencontre pour l'instant dans l'essentiel des modélisations de données économiques ou financières (notamment sous l'hypothèse d'efficience). De plus, c'est un modèle très simple puisqu'il ne dépend que de deux paramètres. On peut penser qu'une activité de recherche importante (et déjà commencée, voir León et Ludeña, 2000), avec un large champ d'applications en finances, va se porter sur les modèles de pseudo-diffusion où la partie dB est remplacée par dB_H.

2. Il existe de nombreux théorèmes de type limite centrale dont le processus limite est un mouvement brownien fractionnaire. Plus précisément, le phénomène d'aggrégation de comportements individuels (modélisés par des processus stochastiques), peut conduire asymptotiquement à un mouvement brownien fractionnaire. Citons un exemple significatif dont on trouvera plus de détails dans Taqqu et Levy (1986) ou Taqqu *et al.* (1997). On suppose que chaque individu a le comportement suivant : on a une succession de ON et de OFF (correspondant à des connections/déconnections, achat/vente,...), les ON et les OFF pouvant être représentés par des processus de renouvellement indépendants dont la loi des durées est de type Pareto. Lorsque l'on somme un grand nombre de ces individus, à comportements indépendants, sur un temps suffisamment long, le processus agrégé, après recentrage et renormalisation, se comporte comme un mouvement brownien fractionnaire (on peut par le même procédé obtenir des processus α-stables...). Nous pensons que ce phénomène d'aggrégation de comportements individuels peut modéliser de nombreuses situations réelles, notamment économiques ou financières.

3. Nous pouvons cependant remarquer que le mouvement brownien fractionnaire relie le comportement local au comportement à long-terme ; cette relation peut s'avérer trop contraignante vis-à-vis de situations réelles (c'est la contrepartie à payer à la simplicité de ce modèle). Deux généralisations peuvent permettre une plus grande souplesse de modélisation : le mouvement brownien multifractionnaire en temps, dans lequel le paramètre H dépend du temps (voir Istas, 2000), ou le mouvement brownien multifractionnaire en fréquence, dans lequel le paramètre H dépend de la fréquence, ce qui autorise un comportement différent en haute et basse fréquence (voir Bardet, 2000b).

2. ESTIMATION DU PARAMÈTRE DE LONGUE-MÉMOIRE

Ajoutons maintenant quelques compléments à la liste des estimateurs du paramètre de longue-mémoire proposée par LM. Par la suite, on suppose connus $X_0, X_1, .., X_N$, où (X_k) est une série chronologique stationnaire.

En premier lieu, si l'on revient sur les estimateurs paramétriques, il nous semble que deux références importantes devraient être ajoutées. La première est l'article de Dahlhaus (1989) concernant l'estimateur du maximum de vraisemblance, qui montre, dans le cas gaussien, l'efficacité de cet estimateur et de l'estimateur par minimum de contraste (ou maximum de vraisemblance avec approximation de type Whittle) présenté dans Fox et Taqqu (1986) ou Avram (1988). La seconde est l'article de Giraïtis et Surgaïlis (1990) qui généralise l'emploi du minimum de contraste au cas des processus linéaires longue-mémoire, avec pour condition le fait que l'innovation soit dans $\mathbb{L}^4$ (ceci comprend donc les processus ARFIMA dans $\mathbb{L}^4$) ; la vitesse de convergence en $\sqrt{N}$ est là-encore atteinte, mais l'efficacité ne semble pas vraie hors le cas gaussien.

En second lieu, et ce sera le point principal, nous voudrions apporter quelques indications sur les principaux estimateurs semi-paramétriques existants. Un estimateur sera considéré comme semi-paramétrique en $H \in]0.5, 1[$ lorsque le modèle, supposé longue-mémoire, est précisé de la manière suivante :

- soit la covariance $r(k)$ vérifie $r(k) = L(|k|)|k|^{2H-2}$ pour tout $k \in \mathbb{Z}$,
- soit la densité spectrale $g(\lambda)$ vérifie $g(\lambda) = g_*(\lambda)|1 - e^{-\imath\lambda}|^{1-2H}$ pour $\lambda \in [-\pi, \pi[$,

avec L et g_* appartenant à des espaces fonctionnels que l'on précise. La différence avec le cadre paramétrique est donc que l'on ne peut plus spécifier ces fonctions à l'aide d'un nombre fini de paramètres. En revanche, la distribution de probabilités doit le plus souvent être précisée (gaussienne, uniforme,...). Nous pensons que trois estimateurs semi-paramétriques sont réellement intéressants. L'estimateur par log-périodogramme semble être le plus performant (voir ci-dessous). L'estimateur par ondelettes, cas particulier des variations quadratiques (voir Doukhan *et al.*, 2000) et l'estimateur du maximum de vraisemblance de type Whittle localisé (voir Robinson, 1995b), quoique moins développés, paraissent également intéressants (l'estimateur R/S, cité par LM car historiquement important, ne nous semble pas comparable aux précédents en terme de vitesse de convergence).

Nous n'évoquerons ici que les derniers travaux portant sur l'estimateur de type log-périodogramme (noté GPH plus loin) introduit par Geweke et Porter-Hudack (1983) et qui consiste en une régression par moindres carrés du logarithme du périodogramme sur le logarithme de $|1-e^{-\imath\lambda}|$ pour certaines fréquences λ à déterminer (on trouvera un remarquable état de l'art sur ces estimateurs dans Moulines et Soulier, 2000b). On peut classer les nombreuses améliorations de cet estimateur en deux ensembles.

Le premier est composé des estimateurs de type GPH localisé (on dira localisé car les fréquences prises en compte appartiennent à un intervalle proche de 0). Hurvich and Deo (1999) montre que si g_* est trois fois dérivable en 0 et si on considère les $m = \mathcal{O}(N^\alpha)$ premières fréquences, soit $\lambda_\imath = \dfrac{2i+1}{N}\pi$ pour $i = 0, 1, .., m$, alors lorsque $g_*''(0) \neq 0$, la vitesse de convergence optimale est obtenue pour $m = C.N^{4/5}$, avec C une constante dépendant des dérivées de g_* en 0; cette vitesse est en $\mathcal{O}(N^{2/5})$. Robinson (1994a), puis Lang et Azaïs (1999) ont proposé des versions modifiées de cet estimateur en ne considérant pas un certain nombre de fréquences proches de zéro et en moyennant les périodogrammes par blocs. On trouvera une version utilisant l'estimation minimax adaptative dans Giraïtis *et al.* (1997).

Il est aussi possible de considérer toutes les fréquences de Fourier; c'est le deuxième type d'estimateur GPH, dit global, introduit par Robinson (1994b) et essentiellement développé par Moulines et Soulier (1999), (2000a), et Hurvich et Brodsky (1997). La contrepartie de l'utilisation de toutes les fréquences est qu'il faut non seulement estimer H mais aussi la partie de type «courte mémoire» de la fonction $\log(g_*)$. L'idée est d'utiliser la développée de $\log(g_*)$ en série de Fourier en estimant seulement un certain nombre de coefficients de Fourier. Les résultats sont donnés aussi bien dans le cas gaussien que dans le cas des processus linéaires. Dans Iouditsky *et al.* (1999), le nombre de ces coefficients est estimé de manière adaptative, ce qui permet d'obtenir la vitesse minimax dans le cas gaussien.

Il nous semble que ces estimateurs semi-paramétriques demandent peu de spécifications quant au modèle (les bruit gaussiens fractionnaires et les processus ARFIMA vérifient notamment toujours les conditions), ce qui peut s'avérer plus probant pour montrer l'éventuelle longue-mémoire de séries économiques ou financières. On trouvera aussi une étude quantitative de ces estimateurs semi-paramétriques, associant les problèmes de biais, de vitesse de convergence, de complexité algorithmique et d'horizon à partir duquel le comportement asymptotique est légitime, dans Bardet *et al.* (2000a).

RÉFÉRENCES

[Avram (1988)] AVRAM F., On bilinear forms in Gaussian random variables and Toeplitz matrices. *Probab. Theory Related Fields*, 79 : 37–45, 1988.

[Bardet *et al.* (2000a)] BARDET J-M., LANG G., PHILIPPE A., OPPENHEIM G. et TAQQU M., Estimators of long-range dependent processes. Dans Long-range dependence : Theory and Applications, à paraître chez Birkhäuser, 2000.

[Bardet (2000b)] BARDET J-M., Les cours d'actifs financiers sont-ils autosimilaires? Preprint.

[Beran (1994)] BERAN J., Statistics for Long-Memory Processes. New York : Chapmann and Hall. 1994.

[Dahlhaus (1989)]dahl DAHLHAUS R., Efficient parameter estimation for self-similar processes. *Annals of Statistics*, 17 : 1749–1746, 1989.

[Doukhan *et al.* (2000)] DOUKHAN P., OPPENHEIM G. et TAQQU M., eds. Long-range dependence : Theory and Applications. A paraître chez Birkhäuser, 2000.

[Fox et Taqqu (1986)] FOX R. et TAQQU M. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Annals of Statistics*, 14 : 517–532, 1986.

[Giraïtis et Surgaïlis (1990)] GIRAÏTIS L. et SURGAÏLIS D., A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotical normality of Whittle's estimate. *Probab. Theory Related Fields*, 86 : 87–104, 1990.

[Giraïtis *et al.* (1997)] GIRAÏTIS L., ROBINSON P.M. et SAMAROV A., Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long-range dependence. *Journal of Time Series Analysis*, 18 : 49–60, 1997.

[Hurvich et Brodsky (1997)] HURVICH C.M. et BRODSKY J., Broadband semiparametric estimation of the memory parameter of a long memory time series using fractional exponential models. Technical report, New-York University Leonard Stern School of Business, SOR-97-2, 1997.

[Hurvich et Deo (1999)] HURVICH C.M. et Deo R., Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series. *Journal of Time Series Analysis*, 20 : 331–341, 1999.

[Iouditsky *et al.* (1999)] IOUDITSKY A., MOULINES E. et SOULIER P., Adaptive estimation of the fractional differencing coefficient. Soumis pour publication.

[Istas (2000)] ISTAS J., Identification des paramètres d'un processus gaussien fractionnaire. Preprint.

[Lang et Azaïs (1999)] LANG G. et AZAÏS J.-M., Non parametric estimation of the long-range dependence exponent for a Gaussian process. *Journal of Statistical Planning and Inferences*, 80 :59-80, 1999.

[León et Ludeña (2000)] LEÓN J.R. et LUDEÑA C., Estimating the diffusion coefficient for diffusions driven by fBm. A paraître dans *Stat. Inference and Stoch. Process.*

[Moulines et Soulier (1999)] MOULINES E. et SOULIER P., Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 27 : 1415–1439, 1999.

[Moulines et Soulier (2000a)] MOULINES E. et SOULIER P., Data driven order selection for projection estimator of the spectral density of time series with long range dependence. *Journal of Time Series Analysis*, 21 : 193–218, 2000.

[Moulines et Soulier (2000b)] MOULINES E. et SOULIER P., Semiparametric spectral estimation for fractional processes. Dans Long-range dependence : Theory and Applications, à paraître chez Birkauser, 2000.

[Robinson (1994a)] ROBINSON P., Semiparametric analysis of long-memory time series. *Annals of Statistics*, 22 : 515–539, 1994.

[Robinson (1994b)] ROBINSON P., Time series with long-range dependence. *Advances in econometrics. Proceedings of the sixth world congress*, 1994.

[Robinson (1995a)] ROBINSON P., Log-periodogram regression of time series with long range dependence. *Annals of Statistics*, 23 : 1048–1072, 1995.

[Robinson (1995b)] ROBINSON P., Gaussian semiparametric estimation of long range dependence. *Annals of Statistics*, 23 : 1630–1661, 1995.

[Taqqu et Levy (1986)] TAQQU M. et LEVY J. Using renewal processes to generate long-range dependence and high variablility. Dans Dependence in Probability and Statistics. Eberlein E., Taqqu M.S eds, Progress in Probab. and Statist., 11, Birkhäuser, 2000.

[Taqqu *et al.* (1997)] TAQQU M., WILLINGER W. et SHERMAN R., Proof of a fundamental result in self-similar traffic modelling. *Computer Communication Review*, 27 : 5–23, 1997.