

SUR-APPRENTISSAGE ET CHOIX DE MODÈLE POUR L'ESTIMATION DE L'INDICE DE HURST D'UN MOUVEMENT BROWNIEN MULTIFRACTIONNAIRE

Pierre R BERTRAND¹ & Marie-Eliette DURY²

¹ UMR CNRS 6620, Laboratoire de Mathématiques, Université Clermont Auvergne.

E-mail: Pierre.Bertrand@math.univ-bpclermont.fr

² École d'Économie, Université Clermont Auvergne. E-mail: m-eliette.dury@udamail.fr

Résumé. Le mouvement brownien multifractionnaire (mBm) est une généralisation du mouvement brownien fractionnaire (fBm) pour laquelle l'indice de Hurst H est remplacé par une fonction du temps $H(t)$. On observe un indice de Hurst dépendant du temps dans différentes applications. En finance quantitative, on constate que l'indice de Hurst varie entre 0,45 et 0,65. Le mouvement brownien avec $H = 1/2$ correspond à l'hypothèse d'efficience des marchés, tandis que les périodes avec un indice de Hurst différent de $1/2$ s'expliquent par des arguments d'économie comportementale.

Pour un indice de Hurst $H(t)$ dépendant du temps, les méthodes d'estimation classiques consistent à localiser l'estimation de cet indice sur un petit voisinage autour de l'instant t pour obtenir un estimateur $\hat{H}(t)$.

On ignore alors si les fluctuations de $\hat{H}(t)$ reflètent la réalité ou sont un artefact statistique correspondant à un sur-apprentissage. Nous construisons un test qui rejette asymptotiquement cet estimateur naïf $\hat{H}(t)$.

Mots-clés. Mouvement Brownien fractionnaire, Mouvement brownien multifractionnaire, indice de Hurst, sur-apprentissage, sélection de modèles

Abstract. The multifractional Brownian motion (mBm) can be viewed as a generalisation of the fractional Brownian motion (fBm) where the Hurst index H is replaced by a time-varying function $H(t)$. A time-varying Hurst index is encountered in different kinds of applications: In quantitative finance for instance, it has been shown that the Hurst index estimated on sliding windows is varying with time between 0.45 and 0.65. Theoretical explanations are developed by economists. To sum up, periods with a Hurst index that significantly differs from $H = 1/2$ can be explained by behavioural finance, whereas the case $H = 1/2$ corresponds to Brownian motion and the hypothesis of efficiency of financial markets.

For such a time-varying Hurst index, the methods of estimation developed up to now localize the estimation of Hurst index on a small vicinity giving a naive estimator $\hat{H}(t)$. Actually, we can not know whether fluctuations of $\hat{H}(t)$ reflect reality or are just artifact of the statistics. We build a test that asymptotically rejects the naive estimator $\hat{H}(t)$.

Keywords. Fractional Brownian motion, Multifractional Brownian motion, Hurst index, Overfitting, Model selection

1 Introduction

Le mouvement brownien multifractionnaire (mBm) est une généralisation du mouvement brownien fractionnaire (fBm) pour laquelle l'indice de Hurst H est remplacé par une fonction du temps $H(t)$, cf. [5, 14]. On rencontre un indice de Hurst variable dans différents types d'applications :

- En turbulence, [13] modélise la vitesse du vent par un mBm avec un indice de Hurst dépendant du temps.
- Dans une étude statistique en dynamique magnétosphérique [15], un changement brusque de l'indice de Hurst peut être observé quelques heures avant un orage magnétique dans le vent solaire.
- En biologie des systèmes, [11, 12] modélisent la diffusion uni-dimensionnelle par un mBm avec un indice de Hurst constant par morceaux ou variable.
- En finance quantitative, [8] a montré que l'indice de Hurst estimé sur des fenêtres glissantes varie en fonction du temps entre 0,45 et 0,65.

Concernant le dernier point, des justifications théoriques ont été développées par des économistes dans [7, 10]. En résumé, les possibilités d'arbitrage existent pour le fBm lorsque l'indice de Hurst H est constant et connu à l'avance, mais n'existent plus lorsque celui-ci varie en fonction du temps, voire est aléatoire. De plus, les périodes pour lesquelles l'indice de Hurst est significativement différent de $H = 1/2$ qui correspond au mouvement brownien et à l'hypothèse d'efficacité du marché, peuvent être analysées via la finance comportementale. Lorsque $H(t) < 1/2$, le marché réagit de manière excessive, témoignant un manque de confiance des investisseurs, alors que lorsque $H(t) > 1/2$, le marché réagit trop faiblement, reflétant un excès de confiance.

2 Estimation de l'indice de Hurst et sur-apprentissage

La méthode classique pour estimer un indice de Hurst dépendant du temps $H(t)$ consiste à localiser l'estimation de l'indice de Hurst \hat{H} sur un voisinage $\mathcal{V}(t, \varepsilon)$, asymptotiquement petit, du point t , cf. [4, 9]. Ceci donne l'impression que l'estimateur $\hat{H}(t)$ est lui-même un processus stochastique, ce qui a suggéré l'introduction de modèles de mouvement brownien fractionnaire avec un indice de Hurst stochastique, cf. [1, 2, 3]. Il est malheureusement impossible d'affirmer qu'il s'agit vraiment d'un indice de Hurst irrégulier et non d'un phénomène de sur-apprentissage, comme le montre la figure 1. Cette fluctuation autour de la valeur théorique de l'indice de Hurst (*i.e.* $H(t) = 0,7$ pour tout t) s'explique par un théorème central limite fonctionnel. Prenons par exemple la variation quadratique généralisée et l'estimateur par régression linéaire sur plusieurs échelles défini dans [9]. On a alors le théorème central limite fonctionnel suivant :

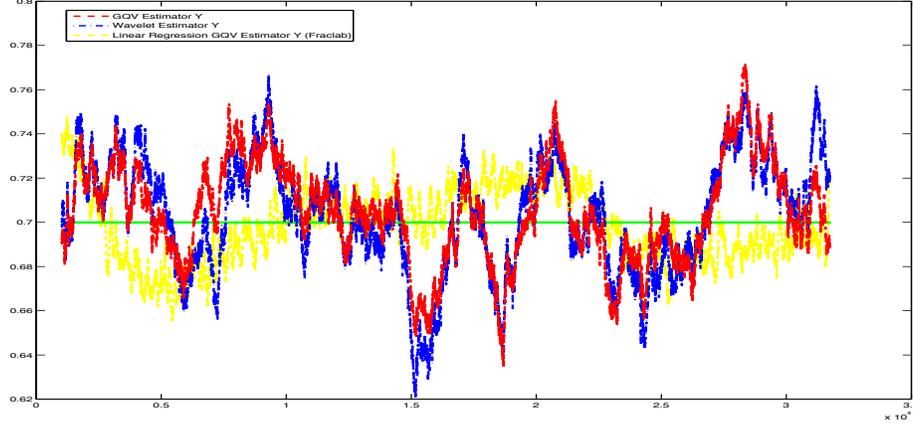


Figure 1: Estimation d'un indice de Hurst dépendant du temps $H(t)$ pour un fBm avec un indice de Hurst constant $H = 0,7$.

Théorème 2.1 (Coeurjolly, 2005–2006) Soit $(t_k = k/n)_{k=1,\dots,n}$ une famille de temps d'observation, $X = B_{H(t)}$ un mouvement brownien multifractionnaire d'indice de Hurst $0 < H(t) < 1$ avec $H \in C^1([0, 1], (0, 1))$, $a = (1; -2; 1)$ le filtre d'ordre 2 dont les variations quadratiques sont

$$V_n(t, a) := \frac{1}{v_n} \sum_{t_k \in \mathcal{V}(t, \varepsilon_n)} |X(t_k) - 2X(t_{k-1}) + X(t_{k-2})|^2.$$

Alors, en posant

$$\hat{H}_n(t) = \frac{A^t}{2AA^t} \left(\ln(V_n(t, a)) \right)_{j=1,\dots,M}$$

on obtient $\hat{H}_n(t) \xrightarrow[n \rightarrow \infty]{p.s.} H$ et

$$\sqrt{2n\varepsilon_n} \times \left(\hat{H}_n(t) - H(t) \right) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathbb{G}'(t) \quad (1)$$

où $\mathbb{G}'(t)$ est processus gaussien centré de structure de variance/covariance donnée par

$$\begin{aligned} \text{Var}(\mathbb{G}'(t)) &= \gamma(H(t)) \quad \text{pour tout } t \in (0, 1), \\ \text{cov}(\mathbb{G}'(t_1), \mathbb{G}'(t_2)) &= 0 \quad \text{pour tout } (t_1, t_2) \in (0, 1)^2 \quad \text{avec } t_1 \neq t_2, \end{aligned}$$

avec

$$\begin{aligned}
\gamma(H) &= \left(\frac{1}{\pi_H^a(0)^2} \sum_{k \in \mathbb{Z}} \pi_H^a(k)^2 \right) \times \frac{A^t(UU^t)A}{4\|A\|^4} \quad (2) \\
&= \frac{-1}{2(4 - 2^{2H})} \sum_{k \in \mathbb{Z}} \left(|k - 2|^{2H} - 4|k - 1|^{2H} + 6|k|^{2H} - 4|k + 1|^{2H} + |k + 2|^{2H} \right) \\
&\quad \times \frac{A^t(UU^t)A}{4\|A\|^4} \\
\pi_H^a(k) &:= -\frac{1}{2} \sum_{q=0}^2 \sum_{q'=0}^2 a_q a_{q'} |q - q' + k|^{2H},
\end{aligned}$$

le vecteur ligne A étant défini par $A_j = \ln(j) - \frac{1}{M} \sum_{\nu=1}^M \ln(\nu)$ pour $j = 1, \dots, M$ et $U = (1, \dots, 1)$.

Donc, même si $H(t)$ est constant, on trouve un estimateur qui est un processus $\widehat{H}_n(t)$ qui converge vers un processus stochastique gaussien, ce qui explique le phénomène de sur-apprentissage de la figure 1.

3 Test d'admissibilité et sélection d'un modèle d'indice de Hurst dépendant du temps

On déduit également du théorème 2.1 que la norme \mathbb{L}^2 de la différence converge après renormalisation, *i.e.*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{n} \sum_{k=1}^n |\widehat{H}_n(t_k) - H(t_k)|^2 \right] = \int_0^1 \gamma(H(t)) dt. \quad (3)$$

Preuve. cf. [6]. \square

De plus on a le théorème central limite suivant

Théorème 3.1 *Avec les hypothèses et notations du théorème 2.1, on a*

$$\frac{(2n\varepsilon_n) \times \left[\frac{1}{n} \sum_{k=1}^n |\widehat{H}_n(t_k) - H(t_k)|^2 \right] - \int_0^1 \gamma(H(t)) dt}{\left[\left(\frac{2}{n} \right) \times \int_0^1 \gamma(H(t))^2 dt \right]^{1/2}} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1), \quad (4)$$

où $\gamma(H)$ est donné par la formule (2).

Preuve. La démonstration est donnée dans [6]. \square

On en déduit un test d'adéquation pour un indice de Hurst dépendant du temps. Plus précisément, on veut tester si $\tilde{H}(\cdot)$ est un modèle admissible, c'est-à-dire tester

$$(H_0) : \tilde{H}(\cdot) = H(\cdot) \quad \textit{versus} \quad (H_1) : \tilde{H}(\cdot) \neq H(\cdot).$$

On utilise la statistique de test $T_n(\tilde{H})$ définie par

$$T_n(\tilde{H}) = \frac{(2n\varepsilon_n) \times \left[\frac{1}{n} \sum_{k=1}^n |\hat{H}_n(t_k) - \tilde{H}(t_k)|^2 \right] - \int_0^1 \gamma(\tilde{H}(t)) dt}{\left(\binom{2}{n} \times \int_0^1 \gamma(\tilde{H}(t))^2 dt \right)^{1/2}}. \quad (5)$$

Le théorème 3.1 signifie que sous l'hypothèse nulle, on a

$$T_n(\tilde{H}) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, 1).$$

Par contre, il est plus délicat de calculer la puissance du test, car $H(\cdot) \in \mathcal{C}([0, 1])$ est un espace vectoriel topologique de dimension infini.

En conclusion, on a mis en évidence un phénomène de sur-apprentissage pour l'indice de Hurst dépendant du temps par l'estimateur obtenu par localisation de l'estimateur de la variation quadratique généralisée. On a également proposé un test d'adéquation d'un modèle d'indice de Hurst dépendant du temps. Il sera ensuite possible d'utiliser ce test d'adéquation pour développer des méthodes de sélection d'un modèle d'indice de Hurst variable avec le temps.

References

- [1] Ayache, A, Bertrand, PR, Lévy-Véhel J (2007), A Central Limit Theorem for the Generalized Quadratic Variation of the Step Fractional Brownian Motion. *Statistical Inference for Stochastic Processes* **10**, 1–27.
- [2] Ayache, A, Taqqu, MS (2005), Multifractional processes with random exponent. *Publ. Mat.* **49**, 459–486.
- [3] Ayache, A, Jaffard, S, Taqqu, MS (2007) Wavelet construction of Generalized Multifractional processes. *Revista Matemática Iberoamericana*, **23**, No 1, 327–370.

- [4] Bardet, JM, Surgailis, D (2013). Nonparametric estimation of the local Hurst function of multifractional processes. *Stochastic Processes and Applications*, 123, 1004–1045.
- [5] Benassi A, Jaffard S, Roux D (1997), Elliptic Gaussian random processes. *Revista Matemática Iberoamericana*, 13(1):19-90.
- [6] Bertrand, PR, Dury, ME, Haouas, N (2015) Selection of sparse multifractional model. *Hal-01194347*.
- [7] Bianchi, S, Pantanella, A, Pianese, A (2015) Efficient Markets and Behavioral Finance: a comprehensive multifractal model. *Advances in Complex Systems* 18.
- [8] Bianchi, S (2005) Pathwise Identification of the memory function of multifractional Brownian motion with application to finance. *International Journal of Theoretical and Applied Finance* Vol. 8, No. 2, 255–281.
- [9] Coeurjolly, JF (2005), Identification of multifractional Brownian motion, *Bernoulli* **11** (6), 987–1008.
- [10] Frezza, M (2012) Modeling the time-changing dependence in stock markets. *Chaos Solitons & Fractals* 45(12):1510–1520
- [11] Lim SC, Teo LP (2009) Modeling Single-File Diffusion by Step Fractional Brownian Motion and Generalized Fractional Langevin Equation. *Journal of Statistical Mechanics Theory and Experiment*
- [12] Marquez-Lago TT, Leier A, Burrage K. Anomalous diffusion and multifractional Brownian motion: simulating molecular crowding and physical obstacles in systems biology *IET Systems Biology* 6(4):134–142.
- [13] Papanicolaou, G, Sølna, K (2002) Wavelet based estimation of local Kolmogorov turbulence. In: P Doukhan, G Oppenheim and MS Taqqu (eds), *Long-range Dependence: Theory and Applications*, Birkhäuser, Basel, (2002), 473-506.
- [14] Peltier, RF, Lévy-Véhel, J (1995), Multifractional Brownian motion: definition and preliminary results, Research Report RR-2645, INRIA, Rocquencourt.
- [15] Wanliss JA, Dobias, P (2007) Space storm as a phase transition *Journal of Atmospheric and Solar-Terrestrial Physics* 69 (2007) 675-684.