Estimating Fractional Brownian Motion With Noise

David Chen, Wangdong Jia, Bryce Monier, Shiyang Shen

Columbia Math Undergraduate Summer Research Program

Introduction

Suppose we have some unknown fBM with stochastic volatility, X_t, and some process with added noise that we are able to observe: $Y_t = X_t + \rho Z_t$. We would like to be able to extract the signal from the noise, and to somehow estimate both H and σ. To do so, we will consider weighted averages of Y_t at different time-points to eliminate the noise. Estimators of H and the integrated volatility have been previously shown for fBM without noise [2] and without stochastic volatility [3]; the main result of this project is to combine both generalizations.

Background and Setup

- The Riemann-Liouville fractional Brownian motion with Hurst Index H is the process given by the stochastic integral
$$B_t^H := \int_0^t (t-s)^{H-\frac{1}{2}}\, dB_s.$$
- Measurement noise will be considered as random normal variables with variance ρ^2.
- We consider the following process:
$$Y_t = X_0 + \int_0^t b_s ds + K_H^{-\frac{1}{2}} B_t^H + \rho Z_t,$$
where $\int_0^t b_s ds$ is some drift process, $K_H^{-\frac{1}{2}}$ is a normalizing constant, and $(Z_t)_{t\geq 0}$ are i.i.d. $N(0,1)$ variables.

Pre-Averaging

We attempt to remove the error from the process by considering weighted averages of the increments of Y_t to try and smooth out the random noise. The variation from these *pre-averages* are then used to construct estimators for H and $\int_0^T \sigma^2 ds$. Formally, for $g : [0,1] \to \mathbb{R}$, some constant θ, and some stochastic process W_t, put

$$k_n = \frac{n^\kappa}{\theta},\ g_j^n = g\left(\frac{j}{k_n}\right)$$

$$\overline{W}(g)_i^n = \sum_{j=1}^{k_n-1} g_j^n \left(W_{\frac{i+j-1}{n}} - W_{\frac{i+j-2}{n}}\right) = \sum_{j=1}^{k_n-1} g_j^n \Delta_{i+j-1}^n W$$

$$\widehat{W}(g)_i^n = \sum_{j=1}^{k_n} \left(g_j^n - g_{j-1}^n\right)^2 \left(\Delta_{i+j-1}^n W\right)^2 = \sum_{j=1}^{k_n} \left(\Delta g_j^n\right)^2 \left(\Delta_{i+j-1}^n W\right)^2.$$

Main Theorem

Given some fBM with measurement error,

$$Y_t = X_t + \rho Z_t$$

if we have the following conditions,

1. $f : \mathbb{R}^L \to \mathbb{R}$ is C^2 with all partial derivatives up to order 2 of at most polynomial growth.
2. $g : [0,1] \to \mathbb{R}$ is C^2.
3. b, σ are of size 1 (that is, $||b_s||_{L_p}, ||\sigma_s||_{L_p}$ are bounded) and adapted.
4. σ is L^2-continuous.
5. $\kappa \in \left(\frac{2H}{2H+1}, 1\right)$.

then we get the following convergence:

$$V(g)_T^{n,f}(Y) = \frac{1}{n}\sum_{i=1}^{\lfloor nT \rfloor} f\left(\frac{\overline{Y}(g)_i^n}{(k_n/n)^H}, \frac{\widehat{Y}(g)_i^n}{(k_n/n)^{2H}}\right) \xrightarrow{\mathbb{P}} \int_0^T \mu_f\left(\sigma_s^2 \eta^H(g), \Theta^2\rho^2 \int_0^1 g'(r)^2 dr\right) ds$$

where

$$\mu_f(v_1, v_2) = \mathbb{E}\left[f\left(v_1 Z_2 + v_2 Z_2, 2v_2\right)\right],\ Z_1, Z_2 \sim N(0,1) \text{ i.i.d.}$$

$$\Theta = \begin{cases} 0 & \kappa \neq \frac{2H}{2H+1} \\ \theta & \kappa = \frac{2H}{2H+1} \end{cases}$$

$$\eta^H(g) = 2H \int_0^1 g(x)\left(g(1)(1-x)^{2H-1} + \int_x^1 (y-x)^{2H-1} g'(y) dy\right) dx$$

Sketch of the Proof

We make a series of adjustments to $V(g)_T^{n,f}(Y)$ to get closer to our final result. We first remove the drift: if $U_t = K_H^{-\frac{1}{2}} B_t^H + \rho Z_t$,

$$\lim_{n\to\infty} \mathbb{E}\left[\left|V(g)_T^{n,f}(Y) - V(g)_T^{n,f}(U)\right|\right] = 0.$$

We show also that one can truncate the the domain of integration to within ϵ of where we first begin averaging, i.e. $\frac{i}{n} - \epsilon$ while making only an error bounded by ϵ. We now *discretize* the volatility by showing that

$$\lim_{\epsilon\to 0}\lim_{n\to\infty} \mathbb{E}\left[\left|\frac{1}{n}\sum_{i=\lfloor n\epsilon\rfloor+1}^{\lfloor nT\rfloor}\left[V(g)_T^{n,f}(Y) - f\left(\frac{\sigma_{\frac{i}{n}-\epsilon}\overline{B^H}(g)_i^{n,\epsilon} + \rho\overline{Z}(g)_i^n}{(k_n/n)^H}, \frac{\rho^2\widehat{Z}(g)_i^n}{(k_n/n)^{2H}}\right)\right]\right|\right] = 0$$

We now center to the conditional expectation:

$$\frac{1}{n}\sum_{i=\lfloor n\epsilon\rfloor+1}^{\lfloor nT\rfloor}\left[f\left(\frac{\sigma_{\frac{i}{n}-\epsilon}\overline{B^H}(g)_i^{n,\epsilon} + \rho\overline{Z}(g)_i^n}{(k_n/n)^{2H}}, \frac{\rho^2\widehat{Z}(g)_i^n}{(k_n/n)^H}\right) - \mathbb{E}\left[f\left(\frac{\sigma_{\frac{i}{n}-\epsilon}\overline{B^H}(g)_i^{n,\epsilon} + \rho\overline{Z}(g)_i^n}{(k_n/n)^H}, \frac{\rho^2\widehat{Z}(g)_i^n}{(k_n/n)^{2H}}\right) \Big| \mathcal{F}_{\frac{i-1}{n}}\right]\right] \xrightarrow{\mathbb{P}} 0$$

Once we compute the expectation and take limits ($n \to \infty$, $\epsilon \to 0$), we arrive at what we wanted.

Estimating H

Changing the freqency of the averaging allows us to create an estimator for H; consider taking $f(x,y) = x^2 - \frac{1}{2}y$ and a suitable κ such that $\Theta = 0$, such that

$$V(g)_T^{n,f}(Y),\ V(g)_T^{\frac{n}{2},f}(Y) = \frac{2}{n}\sum_{i=1}^{\lfloor nT/2\rfloor}\left(\frac{\overline{Y}(g)_i^{n/2}}{(2k_n/n)^H}\right)^2$$

both converge in probability to $\int_0^T \sigma^2 \eta^H(g) ds$, so

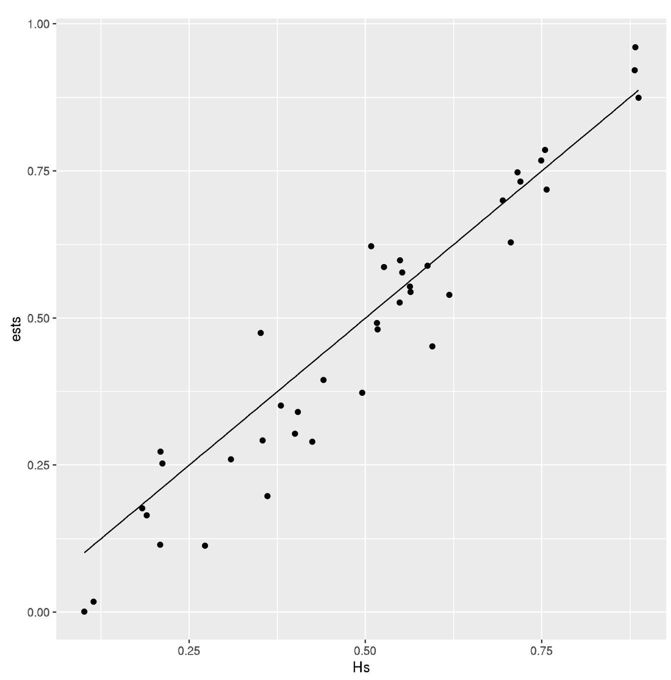
$$\frac{1}{2(1-\kappa)}\log_2\left(\frac{\widetilde{V}(g)_T^{\frac{n}{2},f}(Y)}{\widetilde{V}(g)_T^{n,f}(Y)}\right) \xrightarrow{\mathbb{P}} H.$$

where $\widetilde{V}(g)_T^{n,f}(Y)$ is simply $V(g)_T^{n,f}(Y)$ without the normalizing factor. Furthermore, we can estimate the integrated volatility:

$$\frac{V(g)_T^{n,f}(Y)}{\eta^{\hat{H}_n}(g)} \to \int_0^T \sigma_s^2 ds.$$

In Practice

Estimating simulated datasets with $n = 10^5$ and random Hurst indices, the following is $\hat{H}_n$ plotted against H. Plotted line is just the identity, which we expect our estimators to converge to.



Consistent estimations of the volatility, for which convergence is slow, is an ongoing effort; related code may be found at [4].

Acknowledgements and References

First, we would like to thank Dr. Carsten Chong for his guidance and support in this project before, during, and after the research, as well as the Columbia Mathematics Undergraduate Summer Research Program.

1. Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., Vetter, M.: Microstructure noise in the continuous case: the pre-averaging approach. Stochastic Process. Appl. 119 (2009), no. 7, 2249-2276.
2. Corcuera, J. M., Hedevang, E., Pakkanen, M. S., Podolskij, M.: Asymptotic theory for Brownian semi-stationary processes with application to turbulence. Stochastic Process. Appl. 123 (2013), no. 7, 2552-2574.
3. Gloter, A, and Hoffmann, M.: Estimation of the Hurst parameter from discrete noisy data. Ann. Statist. 35 (2007), no. 5, 1947-1974.
4. https://github.com/DaisiesForAMouse/REU-Writeup

Contact: {dc3451,bjm2190,wj2315,ss6101}@columbia.edu