

## Alternate estimators for $H$

### The $m(q, \Delta)$ estimator for fBM

Let  $X_t = \sigma B_t^H$ , and set  $SS_n^{(q)} := \frac{1}{n} \sum_{i=1}^n |X_{i\Delta} - X_{(i-1)\Delta}|^q$ . Then

$$\mathbb{E}(SS_n^{(q)}) = \frac{1}{n} \mathbb{E} \left( \sum_{i=1}^n |X_{i\Delta} - X_{(i-1)\Delta}|^q \right) = \Delta^{qH} \frac{1}{n} \sum_{i=1}^n \mathbb{E}(|X_i - X_{i-1}|^q) = \sigma^q \mathbb{E}(|Z|^q) \Delta^{qH} = \sigma^q K_q \Delta^{qH} \quad (1)$$

where  $\Delta = \frac{1}{n}$ ,  $K_q = \mathbb{E}(|Z|^q) = \frac{2^{q/2}}{\sqrt{\pi}} \Gamma(\frac{q+1}{2})$ , (for  $q > -1$ ) and  $Z \sim N(0, 1)$ . This naturally leads to the estimates  $(\hat{H}_n, \hat{\sigma}_n)$  for  $(H, \sigma)$  defined by

$$SS_n^{(q)} = \hat{\sigma}_n^q K_q \Delta^{q\hat{H}_n}$$

if we have computed  $SS_n^{(q)}$  for at least two  $\Delta$ -values. Taking logs we see that

$$\log SS_n^{(q)} = q \log \hat{\sigma}_n + \log K_q + q\hat{H}_n \log \Delta$$

so we can perform **linear regression** on  $\log SS_n^{(q)}$  vs  $\log \Delta$  for a range of  $\Delta$ -values (i.e. using a log-log plot, see plot overleaf). Then for the line of best fit, the **slope** will equal  $q\hat{H}_n$  ( $q$  is chosen by you, e.g.  $q = 1, 2, 2.5, 3$  etc), and the **intercept** at  $\log \Delta = 0$  is  $q \log \hat{\sigma}_n + \log K_q$ , from which we can compute  $\hat{\sigma}_n$  since  $K_q$  has an explicit formula above. This is the  $m(q, \Delta)$  estimator discussed in [GJR18]. One can then also compute the  **$R^2$ -statistic** for the regression (which measures how close the data is to the line of best fit), and try to estimate the **sample variance** of  $\hat{H}_n$  and  $\hat{\sigma}_n$ .

To approximate the effect of using **realized variance** with  $m$  subwindows to estimate  $V_t$  (as in Part 2), we can use the Central Limit Theorem approximation from FM02:

$$V_{i\Delta} = V_0 e^{\sigma B_{i\Delta}^H} (1 + \sqrt{\frac{2}{m}} \varepsilon_i)$$

where the  $\varepsilon_i$ 's are i.i.d.  $N(0, 1)$  (and independent of  $B^H$ ), so (using that  $\log(1+x) = x + O(x^2)$ ), we see that  $\log V_{i\Delta} = \log V_0 + \sigma B_{i\Delta}^H + \sqrt{\frac{2}{m}} \varepsilon_i + O(\frac{1}{m}) = \log V_0 + X_{i\Delta} + \sqrt{\frac{2}{m}} \varepsilon_i + O(\frac{1}{m})$ .

For convenience we now define  $\tilde{X}_{i\Delta} = X_{i\Delta} + \sqrt{\frac{2}{m}} \varepsilon_i$ . Then  $\tilde{X}_{i\Delta} - \tilde{X}_{(i-1)\Delta} = X_{i\Delta} - X_{(i-1)\Delta} + \sqrt{\frac{2}{m}} (\varepsilon_i - \varepsilon_{(i-1)}) \sim N(0, \sigma^2 \Delta^{2H} + \frac{4}{m})$ , and setting  $\widetilde{SS}_n^{(2)} := \frac{1}{n} \sum_{i=1}^n |\tilde{X}_{i\Delta} - \tilde{X}_{(i-1)\Delta}|^2$ , adding the effect of  $\varepsilon$  into the computation above we find that

$$\mathbb{E}(\widetilde{SS}_n^{(2)}) = \sigma^2 \Delta^{2H} + \frac{4}{m}$$

(since  $K_q = 1$  for  $q = 2$ ) so we now regress  $\log(\widetilde{SS}_n^{(2)} - \frac{4}{m})$  vs  $\log \Delta$ , which provides a smart adjustment to  $\hat{H}_n$  (this adjustment is made in

<https://colab.research.google.com/drive/1jJGf4bVWETJqWRIMZ6STjsJ9jINgEaPd#scrollTo=e9fn6ABHNf52>

### The $m(q, \Delta)$ estimator for the RL process

If  $X_t = \sigma Z_t^H$  where  $Z_t^H = \sqrt{2H} \int_0^t (t-s)^{H-\frac{1}{2}} dW_s$  is an RL process, then for  $q = 2$  and  $H \in (0, \frac{1}{2})$  one can show that

$$\mathbb{E}(SS_n^{(2)}) = \frac{1}{n} \mathbb{E} \left( \sum_{i=1}^n |X_{i\Delta} - X_{(i-1)\Delta}|^2 \right) = \sigma^q \Delta^{2H} \frac{1}{n} \sum_{i=1}^n \mathbb{E}(|Z_i^H - Z_{i-1}^H|^2) \sim \sigma^2 c_H \Delta^{2H}$$

as  $n \rightarrow \infty$  where  $c_H = -\frac{4H\Gamma(\frac{1}{2}+H)\Gamma(-2H)}{\Gamma(\frac{1}{2}-H)} > 0$ . Hence for the regression we now consider

$$\log(SS_n^{(2)}) = 2 \log \hat{\sigma}_n + \log c_{\hat{H}} + 2\hat{H}_n \log \Delta$$

so now the intercept is  $2 \log \hat{\sigma}_n + \log c_{\hat{H}}$ , which leads to an adjusted estimate for  $\hat{\sigma}_n$

### The Han-Schied [HS21] estimator

Let  $X_t = \sigma B_t^H$  and let

$$\theta_{m,k} = 2^{\frac{m}{2}} (2X_{\frac{2k+1}{2^{m+1}}} - X_{\frac{k}{2^m}} - X_{\frac{k+1}{2^m}}) = -2^{\frac{m}{2}} (X_{\frac{2(k+1)}{2^{m+1}}} - 2X_{\frac{2k+1}{2^{m+1}}} - X_{\frac{2k}{2^{m+1}}})$$

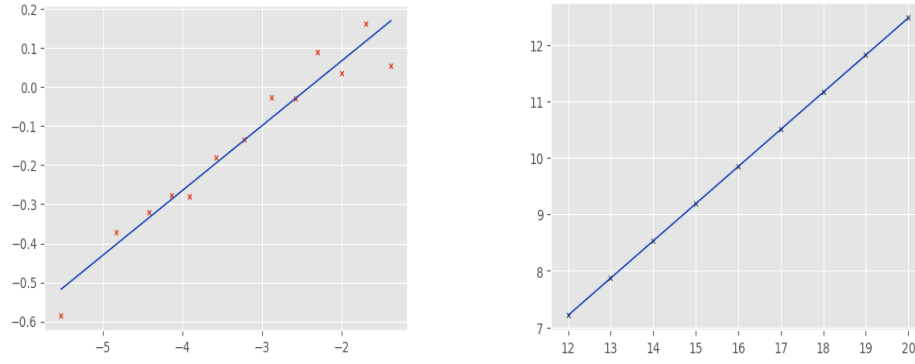


Figure 1: On the left, we see estimates of  $H$  for the SPX using the  $m(q, \Delta)$  method for the SPX from 3rdJan22-15thJul24 for  $q = 2$  for which  $\hat{H} = 0.0830$  and  $\hat{\sigma} = 1.221$  (see similar plots in [GJR18]). On the right we see the linear regression in (4) for the Han-Schied method ( $\log s_n$  vs  $n$ ) for a true fBM path with  $2^{20}$  time points, for which  $\hat{H} = 0.0508$ , and  $\hat{\sigma} = 1.010$ .

(note the similarity of the second expression to a 2nd order finite difference estimate). Then (with some tedious algebra) using the formula for  $R(s, t) = \mathbb{E}(B_s^H B_t^H)$ , one can check that

$$\mathbb{E}(\theta_{m,k}^2) = \sigma^2 2^{m-2H(1+m)} (4 - 4^H). \quad (2)$$

Then setting  $s_n^2 = \sum_{m=0}^{n-1} \sum_{k=0}^{2^m-1} \theta_{m,k}^2$ , we see that  $\mathbb{E}(s_n^2) = \sum_{m=0}^{n-1} 2^m \mathbb{E}(\theta_{m,k}^2)$  (since (2) does not depend on  $k$ ) which simplifies to

$$\mathbb{E}(s_n^2) = \sigma^2 (4^{n(1-H)} - 1) \sim \sigma^2 4^{n(1-H)} = \sigma^2 2^{2n(1-H)} \quad (3)$$

as  $n \rightarrow \infty$ , which suggests an estimator  $\hat{H}_n$  defined by  $s_n = \hat{\sigma}_n 2^{n(1-\hat{H}_n)}$  which (assuming  $\hat{\sigma}_n = O(1)$  as  $n \rightarrow \infty$ ) we can re-arrange as

$$\hat{H}_n = 1 - \frac{1}{n} \log_2 \left( \frac{s_n}{\hat{\sigma}_n} \right) = 1 - \frac{1}{n} \log_2 s_n + O\left(\frac{1}{n}\right)$$

where  $\log_2$  denotes the base-2 logarithm, so (ignoring the  $O(\frac{1}{n})$  remainder term), we recover the Han-Schied[HS21] estimator  $\hat{H}_n = 1 - \frac{1}{n} \log_2 s_n$ . Then

$$\begin{aligned} \mathbb{E}(\hat{H}_n) &= 1 - \frac{1}{n} \mathbb{E}(\log_2(s_n)) = 1 - \frac{1}{2n} \mathbb{E}(\log_2(s_n^2)) \geq 1 - \frac{1}{2n} \log_2 \mathbb{E}(s_n^2) = 1 - \frac{1}{2n} \log_2(\sigma^2 (4^{n(1-H)} - 1)) \\ &\geq 1 - \frac{1}{2n} \log_2(\sigma^2 (4^{n(1-H)})) \\ &= 1 - \frac{1}{2n} \log_2(\sigma^2) - \frac{1}{2n} \log_2(4^{n(1-H)}) \\ &= H - \frac{1}{2n} \log_2(\sigma^2) \end{aligned}$$

and the final line is  $> H$  if  $\sigma < 1$ , so  $\mathbb{E}(\hat{H}_n) > H$  if  $\sigma < 1$ . See also discussion on optimal scaling factors in [HS21].

Or we can jointly estimate  $H$  and  $\sigma$  by performing linear regression since

$$\log s_n = \log \hat{\sigma}_n + n(1 - \hat{H}_n) \log 2 \quad (4)$$

but we now have to compute  $\log s_n$  for a range of different  $n$ -values to get a line of best fit, for which the slope is  $(1 - \hat{H}_n) \log 2$  and the intercept is  $\log \hat{\sigma}_n$ .

## The rough Heston model

The driftless rough Heston model satisfies

$$V_t = V_0 + \nu \int_0^t (t-u)^{H-\frac{1}{2}} \sqrt{V_u} dW_u.$$

Then  $\mathbb{E}(V_t) = V_0$ , and  $V$  has covariance function:

$$\begin{aligned} \mathbb{E}((V_s - V_0)(V_t - V_0)) &= \nu^2 \mathbb{E}\left(\int_0^s (s-u)^{H-\frac{1}{2}} \sqrt{V_u} dW_u \cdot \int_0^t (t-r)^{H-\frac{1}{2}} \sqrt{V_r} dW_r\right) \\ &= \nu^2 \int_0^s (s-u)^{H-\frac{1}{2}} (t-u)^{H-\frac{1}{2}} \mathbb{E}(V_u) du \\ &= V_0 \nu^2 \int_0^s (s-u)^{H-\frac{1}{2}} (t-u)^{H-\frac{1}{2}} du = V_0 \nu^2 \bar{R}(s, t) \end{aligned}$$

for  $0 \leq s \leq t$ , where  $\bar{R}(s, t)$  is the covariance function for the Riemann-Liouville (RL) process  $Z_t = \int_0^t (t-u)^{H-\frac{1}{2}} dW_u$  used for the rough Bergomi model (note  $Z$  is a Gaussian process but  $V$  is not), but the explicit formula for  $\bar{R}(s, t)$  is more complicated than the  $R(s, t)$  formula for fBM.

We also have the **Mandelbrot-van Ness** representation for fBM:

$$W_t^H = c_H \left( \int_{-\infty}^0 ((t-s)^{H-\frac{1}{2}} - (-s)^{H-\frac{1}{2}}) dW_s + \int_0^t (t-s)^{H-\frac{1}{2}} dW_s \right) = c_H (A_t + Z_t)$$

for  $t \geq 0$ , in terms of an RL process  $Z$  (and note that  $A_t$  is known at time zero for all  $t \geq 0$ ), and  $c_H = \left( \frac{2H \Gamma(\frac{3}{2}-H)}{\Gamma(H+\frac{1}{2}) \Gamma(2-2H)} \right)^{\frac{1}{2}}$ . Note also that  $A_t$  and  $Z_t$  are independent.

## References

- [BFG16] Bayer, C., P.K.Friz, and J.Gatheral, “Pricing under rough volatility”, *Quantitative Finance*, 16(6), pp. 887–904, 2016.
- [Cha14] Chang, Y-C., “Efficiently Implementing the Maximum Likelihood Estimator for Hurst Exponent”, *Mathematical Problems in Engineering*, vol. 2014
- [CD22] Cont, R. and P.Das, “Rough Volatility: Fact or Artefact?”, preprint, 2022
- [F23] Forde, M., “Statistical issues and calibration problems under rough and Markov volatility”, <https://martinforde.github.io/Talk.pdf>
- [GJR18] Gatheral, J., T.Jaisson and M.Rosenbaum, “Volatility is rough”, *Quantitative Finance*, 18(6), 2018
- [HS21] Han, X. and A.Schied, “The Hurst roughness exponent and its model-free estimation”, preprint, 2021
- [HS24] Han, X. and A.Schied, “Estimating the roughness exponent of stochastic volatility from discrete observations of the integrated variance”, preprint, 2024