Alternate estimators for H

The $m(q,\Delta)$ estimator from [GJR18]

For the first task, let $SS_n^{(q)} := \frac{1}{n}\sum_{i=1}^n |B^H_{i\Delta} - B^H_{(i-1)\Delta}|^q$. Then

$$\mathbb{E}(SS_n^{(q)}) \;=\; \frac{1}{n}\mathbb{E}(\sum_{i=1}^n |B^H_{i\Delta} - B^H_{(i-1)\Delta}|^q) \;=\; \Delta^{qH}\frac{1}{n}\sum_{i=1}^n \mathbb{E}(|B^H_i - B^H_{i-1}|^q) \;=\; \mathbb{E}(|Z|^q)\Delta^{qH} \;=\; K_q\Delta^{qH} \tag{1}$$

where $K_q = \mathbb{E}(|Z|^q) = \frac{2^{q/2}}{\sqrt{\pi}}\Gamma(\frac{q+1}{2})$, with $q > -1$ and $Z \sim N(0,1)$. From this we can then derive the estimator for the first task. Alternatively, if let $X = \sigma B^H_t$, and now assume H and σ are unknown, then (1) changes to

$$\mathbb{E}(SS_n^{(q)}) \;=\; \sigma^q K_q \Delta^{qH}$$

which leads to the estimate

$$SS_n^{(q)} \;=\; \hat{\sigma}^q K_q \Delta^{q\hat{H}}\,.$$

Taking logs we see that

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

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

The Han-Schied [HS21] estimator

Let $X_t = \sigma B^H_t$. Then

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

Then from the formula for $R(s,t) = \mathbb{E}(B^H_s B^H_t)$, one can check that

$$\mathbb{E}(\theta^2_{m,k}) \;=\; \sigma^2 2^{m-2H(1+m)}(4 - 4^H) \tag{2}$$

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

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

as $n \to \infty$, which suggests the estimator given by

$$s_n \;=\; \hat{\sigma}\,2^{n(1-\hat{H})}$$

which we can re-arrange as

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

where $\log_2$ denotes the base-2 logarithm, so (ignoring the $O(\frac{1}{n})$ remainder term), we recover the Han-Schied[HS21] estimator $\hat{H} = 1 - \frac{1}{n}\log_2 s_n$ (which in principle also works for general processes which aren't fBM if n is sufficiently large). Or we can jointly estimate H and σ by performing linear regression again

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

but we now have to compute $\log s_n$ for a range of different n-values to get a line-of-best-fit.

You can then draw histograms of $\hat{H}$ if you simulate M fBM paths and compute the sample variance for $\hat{H}$ (or a confidence interval), or compute $\hat{H}$ for real data, e.g. using the SPX data file or data from yahoo finance.

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