Fractal Models, Financial Time Series, and Stock Volatility

May 28, 2022

Table of Contents

Table of Contents

Background

Observed features of financial time series:

1. price increments not correlated,
2. volatilities strongly correlated (i.e., power laws),
3. price increment distributions time scale dependent.

Multifractal processes describe the intermittent nature of fluid velocity field in fully developed turbulence.
Intermittency is the irregular alternation of phases of:

- apparently periodic and **chaotic** dynamics, or
- **different** forms of chaotic dynamics (crisis-induced).

Hence, their (potential) relevance in analyzing financial data.

Fractal Processes

Consider variations of a stationary stochastic process $X(t)$ at different time scales l :

$$\delta_l X(t) := X(t+l) - X(t),$$

and their absolute $q > 0$ moments:

$$M(q,l) := \mathbb{E}[|\delta_l(X(t))|^q].$$

Say X is **scale invariant** if $M(q,l)$ is a power law:

$$M(q,l) \approx C_q l^{\zeta_q}.$$

X is called **monofractal** if ζ_q is linear in q, and **multifractal**, else.
Question: How to construct processes with $M(q,l) = C_q l^{\zeta(q)}$?

Fractality from Self-Similarity

self-similarity $\leftrightarrow$ fractality $\leftrightarrow$ multiplicative cascades

If parts of a figure are small replicas of the whole, then the figure is called self-similar*. Any arbitrary part contains an exact replica of the whole figure.*

Call $X(t)$ **self-similar** if there exists $H \in \mathbb{R}$ for which

$$(\lambda^{-H} X(\lambda t)) \stackrel{law}{=} (X(t)), \ \forall \lambda > 0.$$

Then

$$M(q, l) = \mathbb{E}[|\delta_l(X(t))|^q] = C_q(\frac{l}{L})^{qH}$$

for $L = \lambda l$, $C_q = M(q, L)$. If L is fixed, then $X(t)$ is monofractal:

$$M(q, l) = C_q(\frac{l}{L})^{qH} = C'_q l^{\zeta(q)}, \ \ \zeta(q) = qH.$$

Another class of processes can generate multifractal processes:

$$P_l(\delta X) = f(l/L, P_L(\delta X)), M(q, l) = M(q, L)(\frac{l}{L})^{F(q)}.$$

Fractality from Cascades

self-similarity $\leftrightarrow$ fractality $\leftrightarrow$ multiplicative cascades

Multiplicative cascades satisfy for $(l_n)_{n\in\mathbb{N}}, \lim_{n\to\infty} l_n = 0$,

$$M(q, l_n) = C_q l_n^{\zeta(q)}.$$

In a classical such construction (Mandelbrot's), $l_n = 2^{-n}L$,

$$\delta_{l_n} X(t) = \Big(\prod_{1\leq i\leq n} W_i\Big)\delta_L X(t),$$

with $W_1, W_2, \dots, W_n$ i.i.d., random positive factors. This equation is called the **cascade paradigm**: from coarse to fine time scales, new sources of randomness emerge.

Question: Other ways to ensure $M(q, l) = C_q l^{\zeta(q)}$?

Table of Contents

Multifractal Random Measures (MRMs)

The key constructions behind **continuous cascade models** are **multifractal random measures** (MRMs). There are many such objects, and infinitely divisible laws play a crucial role in their definition (given the cascade paradigm, this can be expected). Consider the simplest one, emerging from normal distributions:

$$M_{l,T}(dt) = e^{2\omega_{l,T}(t)}dt,$$

where $T > 0$ is fixed, and $\omega_{l,T}$ is a Gaussian process with

$$\mathbb{E}[\omega_{l,T}(t)] = -\lambda^2(\log\frac{T}{l} + 1),$$

$$Cov(\omega_{l,T}(t), \omega_{l,T}(t+\tau)) = \begin{cases} \lambda^2(\log\frac{T}{l} + 1 - \frac{\tau}{l}), & \tau \in [0, l] \\ \lambda^2 \log\frac{T}{\tau}, & \tau \in [l, T] \\ 0, & \tau > T \end{cases},$$

where T is called the **integral scale**, and λ^2 the **intermittency coefficient**.

Multifractal Random Measures (continued)

Up to an extent, the parameter l replaces the sequence l_n, $\lim_{n\to\infty} l_n = 0$ in Mandelbrot's construction. In light of this, the **log-normal MRM** is defined as a weak limit:

$$M_T(dt) = \lim_{l\to 0} M_{l,T}(dt).$$

The process M_T satisfies several scale invariance properties, and the importance of

$$\zeta_M(q) = (1 + 2\lambda^2)q - 2\lambda^2 q^2$$

will soon become apparent: it is the ζ_M function associated to the multifractal process given by M_T!

Multifractal Random Walks (MRWs)

Let $B(t) \perp\!\!\!\perp M_T$ be a Brownian motion: define the **multifractal random walk**

$$X_T(t) := B(M_T[0,t]).$$

This process inherits the scale invariance properties of M_T, and is the continuous analog of cascade models for scale-invariance:

$$(i) Global : (X_{sT}(st)) \stackrel{law}{=} s(X_T(t)), \forall s > 0,$$

$$(ii) Integral : (X_T(t))_{t\in[0,sT]} \stackrel{law}{=} W_s(X_{sT}(st))_{t\in[0,sT]}, \forall s \in [0,1],$$

$$(iii) Stochastic : (X_T(st))_{t\in[0,T]} \stackrel{law}{=} W_s(X_T(t))_{t\in[0,T]}, \forall s \in [0,1],$$

for $\Omega_s \perp\!\!\!\perp M_T$ normal, $\mathbb{E}[\Omega_s] = -Var(\Omega_s) = 2\lambda^2 \log s$,
$W_s = \sqrt{s} e^{\Omega_s/2}$. Last but not least, X_T is a **multifractal process** when $\lambda^2 < 1/2$:

$$\mathbb{E}[|X_T(t)|^q] = K_q t^{\zeta_M(q)}, \forall t \le T.$$

Moments of Multifractal Random Measures (MRMs)

By definition, Brownian motion enjoys *time aggregation stability*: it satisfies a scale-invariance property, and its moments at any time can be easily computed.
What about MRMs? Key: for the **log-normal** MRM with $\lambda << 1$,

$$\log M_T(t) \approx \Omega(t),$$

where $\Omega(t)$, called the *renormalized magnitude process* of M_T, is **log-normal**.
This approximation allows for (first order in λ) estimates of the moments of M_T : e.g., its moments and covariance structure can be computed up to $O(\lambda^2)$ errors. Here $\lambda << 1$ is crucial!

Moments of log-normal MRMs

As the MRM itself, the renormalized magnitude process is given by a weak limit (using the convergence of finite dimensional distributions and tightness):

$$\Omega(t) = \lim_{l \to 0+} \Omega_l(t), \quad \Omega_l(t) = \frac{1}{\lambda} \int_0^t (\omega_{l,T}(s) - \mathbb{E}[\omega_{l,T}(s)])ds.$$

Consider the variation underlying Ω : for any interval $I = [t - \tau, t]$,

$$\Omega(I) := \delta_\tau(\Omega) = \Omega(t) - \Omega(t - \tau).$$

The covariance of the increments of Ω can be computed explicitly (similarly to $\omega_{l,T}$, it involves logarithms of ratios). When λ is close to zero, Ω can be employed to compute the moments of M :

Moments of log-normal MRMs (continued)

Theorem 2[1]**:** For any fixed intervals $I_1, I_2, \ldots, I_n$, as $\lambda \to 0$,

$$\left(\frac{1}{2\lambda}\log\frac{M(I_1)}{|I_1|}, \frac{1}{2\lambda}\log\frac{M(I_2)}{|I_2|}, \ldots, \frac{1}{2\lambda}\log\frac{M(I_n)}{|I_n|}\right) \xrightarrow{law} \left(\frac{\Omega(I_1)}{|I_1|}, \frac{\Omega(I_2)}{|I_2|}, \ldots, \frac{\Omega(I_n)}{|I_n|}\right).$$

This result yields first order approximations of the expectation and covariance of the increments of M, and subsequently of X as well:

Theorem 6[2]**:** Let $(\epsilon(n))_{n\in\mathbb{N}}$ be an i.i.d. sequence of centered normal distributions with variance σ^2. Then

$$\delta_\tau X(n\tau) \approx \sqrt{\tau}\epsilon(n) \cdot \exp\left(\lambda\frac{\delta_\tau\Omega(n\tau)}{\tau} - \lambda^2 Var\left(\frac{\delta_\tau\Omega(n\tau)}{\tau}\right)\right)$$

(their moments have the same first two terms in their Taylor series in λ).

Note: This theorem will be the key for estimation.

[1] Bacry et al., 2008.
[2] Bacry et al., 2008.

Parameter Estimation

Three parameters are relevant for the log-normal MRW:

- the integral scale T,
- the intermittency coefficient λ^2,
- the variance σ^2 (i.e., a multiplicative factor for MRW).

The asymptotic setting is given by

$$N = \frac{L}{\tau} \to \infty,$$

where the observation scale is L, sampling period τ, and number of samples N. The **low frequency** regime corresponds to $L \to \infty, \tau$ fixed, and the **high frequency** to $\tau \to 0, L$ fixed (the mixed regime $\tau \to 0, L \to \infty$ has also been considered).
The generalized method of moments (GMM) is employed for estimation.

GMM in the Low Frequency Regime

GMM requires a moment condition, which is meant to identify the parameter(s) of interest:

$$\mathbb{E}[f(X,\theta_1)] = \mathbb{E}[f(X,\theta_2)] \Longleftrightarrow \theta_1 = \theta_2.$$

Alternatively, for a given (i.e., known) θ_0,

$$\mathbb{E}[f(X,\theta)] = 0 \Longleftrightarrow \theta = \theta_0.$$

In this case, σ^2 is estimated from the variance of $\delta_l X(t)$, and T, λ^2 from the sample covariance of

$$Z_\tau(k) = \ln(\delta_\tau X(k))$$

for different time lags k.
One considerable difference among these three parameters is that λ^2 is the most reliable to estimate.

GMM in the Low Frequency Regime (continued)

Let

$$f(Z_\tau(k),\theta) = \begin{pmatrix} \exp(2Z_\tau(k)) \\ (Z_\tau(k)-\mu_\theta)(Z_\tau(k-h_1)-\mu_\theta) \\ \dots \\ (Z_\tau(k)-\mu_\theta)(Z_\tau(k-h_K)-\mu_\theta) \end{pmatrix} - \begin{pmatrix} \sigma^2\tau \\ C_\theta[h_1] \\ \dots \\ C_\theta[h_K] \end{pmatrix}$$

where $h_1, \dots h_K$ are positive lags, $\mu_\theta = \mathbb{E}_\theta[Z_\tau[k]]$, and $C_\theta[h] = \mathrm{Cov}_\theta(Z_\tau(k), Z_\tau(k-h))$. Consider the sample mean

$$g_N(\theta) = \frac{1}{N}\sum_{k=1}^{N} f(Z_\tau(k));$$

for any sequence of matrices $W_N \to W_\infty$, its GMM estimator is

$$\hat{\theta} = argmin_\theta(g_N^T W_N g_N).$$

GMM in the Low Frequency Regime (finalized)

It can be shown that under certain conditions, the estimate $\hat{\theta}$ is consistent and satisfies a CLT:

$$\sqrt{N}(\hat{\theta} - \theta) \Rightarrow N(0, \Sigma),$$

where $\Sigma = \Sigma(f, W_\infty)$. This result implies the optimal W_∞ is the inverse of the covariance of $f(Z, \theta)$, V_θ^{-1}. However, as θ_0 is unknown, constructing W_N directly is not possible.

Instead, use the following routine until the estimators $\hat{\theta}$ cluster:

1. Choose $W_N = I_N$.
2. Compute $\hat{\theta}$.
3. Set $W_N = V_{\hat{\theta}}^{-1}$. Go to step 2.

Table of Contents

Fractional Stochastic Volatility (FSV)

The **fractional Brownian Motion**[3] (fBM) with Hurst parameter $H \in (0,1)$ is a centered Gaussian process with

$$\mathbb{E}[|W^H(t+\Delta) - W^H(t)|^q] = K_q \Delta^{qH},$$

where K_q is the q^{th} moment of a standard normal. Its covariance function is $\mathbb{E}[W^H(t)W^H(s)] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t-s|^{2H})$:

- $H = 1/2$ yields Brownian motion;
- fBM is a monofractal process;
- for $H > 1/2\,(H < 1/2)$, the increments of W^H are positively (negatively) correlated.

Fractional stochastic volatility (FSV) has been proposed by Comte and Renault in 1998 as a model capturing the so-called **long memory**: FSV is an fBM with Hurst parameter $H > 1/2$.

[3] introduced by Mandelbrot and Van Ness in 1968.

Rough Fractional Stochastic Volatility (RFSV)

When $H > 1/2$, W^H exhibits **long memory** in the following sense:

$$\sum_{k \geq 1} Cov(W^H(1), W^H(k) - W^H(k-1)) = \infty,$$

since $Cov(W^H(1), W^H(k) - W^H(k-1)) = \Omega(k^{2H-2}) \geq 0$.
Long memory has changed meaning over time:

1. autocorrelation function not integrable,
2. slow decay of autocorrelation function,
3. covariance decaying like a power smaller than one.

The **RFSV** is a *rough* FSV: an fBM with $H < 1/2$.
Remark: Although $W^H, H > 1/2$ does not have long memory, statistical procedures used to detect this property suggest it does.

Rough Fractional Stochastic Volatility (RFSV, continued)

Why the RFSV model?

- implied volatility: although level and orientation of volatility surface change,[4] its overall shape, at least up to a first order approximation, typically does not; this suggests modeling volatility as a time-homogeneous process;

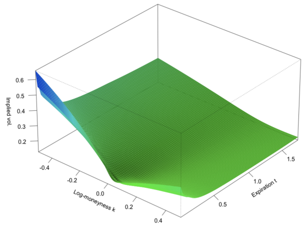


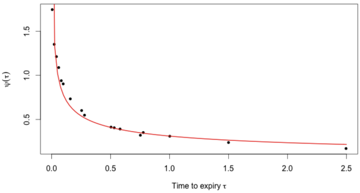
Figure 1.1: The S&P volatility surface as of June 20, 2013.

[4] for equity markets

Rough Fractional Stochastic Volatility (RFSV, continued)

Why the RFSV model (continued)?

- the volatility skew for at-the-money options[5] is of order $\tau^{H-1/2}$ when log-volatility is assumed to be an fBM; $H < 1/2$ would explain its explosion when τ[6] is small (inconsistent with the claim/belief that jumps generate it);



- better forecasts than AR, HAR, easier to estimate with

[5] the partial derivative of volatility with respect to log-moneyness.
[6] τ is the time left to expiration.

Properties RFSV

Suppose $\sigma_\Delta, \sigma_{2\Delta}, \ldots$ are volatility "observations[7]." Let

$$m(q,\Delta) := \frac{1}{N} \sum_{1 \le k \le N} |\log \sigma_{k\Delta} - \log \sigma_{(k-1)\Delta}|^q.$$

Main assumption of the RFSV: for some $s_q, b_q > 0$,

$$\lim_{\Delta \to 0} N^{q s_q} m(q,\Delta) = b_q.$$

- s_q can be viewed as a smoothness parameter (regularity in an l_q sense);
- this limit holds for $q > 0$ in probability with $s_q = H$ when $\log \sigma = W^H$;
- $m(q,\Delta)$ can be viewed as the empirical counterpart of $\mathbb{E}[|\log \sigma_\Delta - \log \sigma_0|^q]$ (assuming stationarity and a LLN).

Remark: $m(q,\Delta)$ is crucial for the estimation of H.

[7] volatility proxies: e.g., implied, realized/historical, etc.

Revised RFSV

Empirical data and a large body of literature suggest volatility is stationary: however, $\sigma_t = exp(cW^H(t))$ is not! To fix this, an Ornstein-Uhlenbeck process is employed: $\sigma_t = exp(cX_t)$,

$$dX_t = \nu dW_t^H - \alpha(X_t - m)dt$$

with $\alpha \in \mathbb{R}, \nu > 0, m > 0$.
As for its Brownian motion analog, the solution of this SDE with initial condition $X_0 = m$ can be expressed in terms of W_t :

$$X_t = \nu \int_0^t e^{-\alpha(t-s)} dW_s^H + m(1 - e^{\alpha t}).$$

Although X_t is not an fBM, it is close to a scaled one for small α :

$$\lim_{\alpha \to 0} \mathbb{E}[\sup_{t \in [0,T]} |X_t^\alpha - X_0^\alpha - \nu W^H(t)|] = 0.$$

Table of Contents

Estimation and Prediction

Given that the two models were initially employed to predict different quantities (log-returns and log-volatilies), comparing the two is not entirely appropriate/fair.
The section on prediction for MRWs shoves under the rug most details (linear regression is used, but the number of features or the sizes of the training and test sets are not specified), and due to time considerations, this will not be discussed here.
Data: SP500 index, from April 30, 2007 to May 13, 2022, 1-hour and daily.

Questions

1. Is the main assumption of RFSV reasonable?
2. Does H vary over time?
3. How does RFSV perform compared to HAR?

Estimation and Prediction: RFSV

If $W_t = W_t^H$ is a fBM with Hurst parameter $H \in (0,1)$, then

$$\mathbb{E}[W(t+\Delta)|\mathcal{F}_t] = W(t) + \frac{\cos(\pi H)}{\pi} \Delta^{H+1/2} \int_0^t \frac{W(t-s) - W(t)}{s^{H+1/2}(s+\Delta)} ds.$$

Recall $\log \sigma_t$ is modeled as CW_t. This and

$$Var(W(t+\Delta)|\mathcal{F}_t) = c\Delta^{2H}, \quad c = \frac{\Gamma(3/2 - H)}{\Gamma(2-2H)\Gamma(H+1/2)}$$

suggest the volatility estimator

$$\hat{\sigma}(t+\Delta) = exp(\hat{\log}\,\sigma(t+\Delta) + c\nu^{(2)}\Delta^{2H}),$$

where $\nu^{(2)}$ is the intercept of the linear regression $\log m(2,\Delta)$ on $\log \Delta$.

Estimation and Prediction: RFSV (continued)

$$\mathbb{E}[W(t+\Delta)|\mathcal{F}_t] = W(t) + \frac{\cos(\pi H)}{\pi}\Delta^{H+1/2}\int_0^t \frac{W(t-s)-W(t)}{s^{H+1/2}(s+\Delta)}ds,$$

$$\hat{\sigma}(t+\Delta) = exp(\hat{\log\sigma}(t+\Delta) + c\nu^{(2)}\Delta^{2H}),\ c = \frac{\Gamma(3/2-H)}{\Gamma(2-2H)\Gamma(H+1/2)}.$$

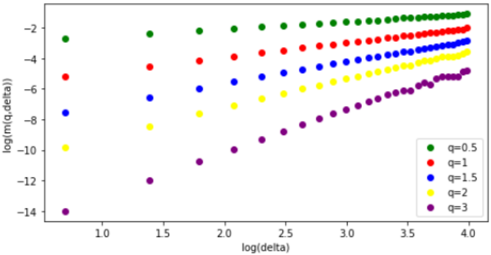
Algorithm:

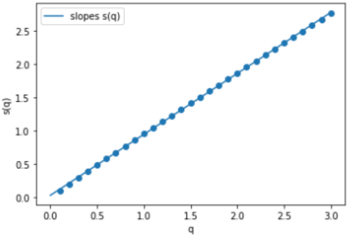
- Discretize the integral, and compute $\hat{W}(t+\Delta)$ ($\hat{\log\sigma} = \hat{W}$).
- Take q in a grid, and run a linear regression for $\log m(q,\Delta)$ on $\log\Delta$.
- Run another linear regression for the slopes $s(q)$ on q, and let $\hat{H}$ be the slope of the linear fit.
- Take $\nu^{(2)}$ as the intercept of the linear regression $\log m(2,\Delta)$ on $\log\Delta$.

1. Main assumption RFSV

For both daily and hourly data, the results match the predictions given by the model for **historical volatility**.

$$\frac{\text{Hourly SP500, June 19,2020- June 17, 2021,}}{\text{H=0.91, T=2000, window size=400}}$$

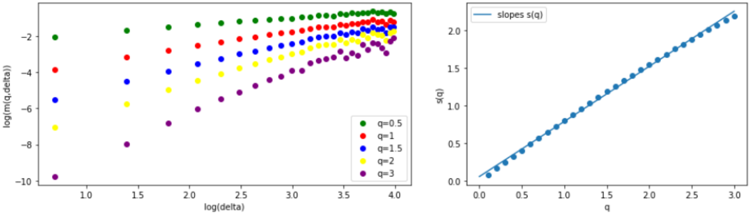




1. Main assumption RFSV (continued)

Daily SP500, September 1,2015- February 4, 2019,

H=0.73, T=1000, window size=100



2. Does H vary over time?

In the paper *Volatility is rough*, $\hat{H} \in [0.02, 0.2]$: the authors employ implied volatility and volatility estimates provided by the Oxford-Man Institute of Quantitative Finance Realized Library.

For historical volatility, RFSV performs very well, but $\hat{H}$ is generally much larger for both daily and hourly data (so technically it is oftentimes an FSV) than in the cases above:

Hourly data:

1. T$=$ 2000, window size$=$ 400, $\hat{H} \in [0.86, 0.93]$ (14 periods);
2. T$=$ 4000, window size$=$ 200, $\hat{H} \in [0.83, 0.86]$ (7 periods);
3. T$=$ 4000, window size$=$ 100, $\hat{H} \in [0.71, 0.76]$ (7 periods).

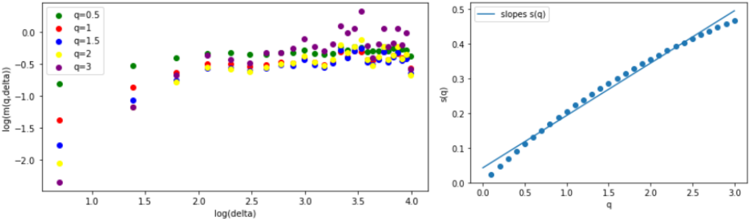
For daily data, the results are similar but $\hat{H}$ is smaller: 0.51 to 0.87.

2. Does H vary over time? (continued)

A small value of $\hat{H}$ seems to require short windows for computing historical volatility, but then the plots, which should consist of lines, become chaotic.

- T$=4000$, window size$=10$, $\hat{H} \in [0.07, 0.18]$ (7 periods);
- T$=2000$, window size$=10$, $\hat{H} \in [0.06, 0.15]$ (15 periods).

Hourly SP500, April 12, 2019- April 8, 2021,
H=0.15, T=4000, window size=10



3. RFSV versus HAR

In the paper *Volatility is rough*, the performance of RFSV is comparable to that of HAR (heterogenous autoregression) and AR. For historical volatility and its log, the former considerably outperforms the latter.

The HAR prediction formula is

$$\hat{\sigma}_{t+\Delta} = \hat{\sigma}_t + C_5 \sum_{1 \leq i \leq 5} \hat{\sigma}_{t+\Delta-i\Delta} + C_{20} \sum_{1 \leq j \leq 20} \hat{\sigma}_{t+\Delta-j\Delta},$$

where C_5, C_{20} are estimated using OLS on the training dataset (denote its size by N, and s the size of the test set).

3. RFSV versus HAR (continued)

Hourly SP500, historical volatility and log-volatility, N=500, T=2000, window size=200, s=3000

	RMSE_HAR	RMSE_RFSV	MAE_HAR	MAE_RFSV	P_val_HAR	P_val_RFSV
0	10.616272	0.091367	8.047295	0.064998	1.999662	0.000148
1	17.008795	0.120441	11.377488	0.077787	1.980537	0.000099
2	14.010327	0.106972	10.589858	0.075655	2.000433	0.000117
3	23.213487	0.156244	17.523396	0.092224	1.995808	0.000090
4	26.571312	0.158678	19.727162	0.101007	1.997394	0.000071
5	83.781255	0.391541	50.615608	0.230229	1.989150	0.000043

	RMSE_HAR	RMSE_RFSV	MAE_HAR	MAE_RFSV	P_val_HAR	P_val_RFSV
0	0.466627	0.004335	0.364927	0.003053	2.000459	0.000173
1	0.550178	0.004664	0.423940	0.003200	1.979639	0.000142
2	0.568281	0.005360	0.455053	0.003644	1.998629	0.000178
3	0.746394	0.005610	0.599340	0.003250	1.996600	0.000113
4	0.768641	0.004976	0.614509	0.003364	1.997263	0.000084
5	0.783789	0.004453	0.591353	0.003068	1.987204	0.000064

3. RFSV versus HAR (continued)

Daily SP500, historical volatility and log-volatility,
N=500, T=1000, window size=100, s=1500

	RMSE_HAR	RMSE_RFSV	MAE_HAR	MAE_RFSV	P_val_HAR	P_val_RFSV
0	20.948855	14.545714	16.857727	44.880224	1.987751	0.958322
1	89.116854	63.105333	63.484588	94.829406	1.961203	0.983411

	RMSE_HAR	RMSE_RFSV	MAE_HAR	MAE_RFSV	P_val_HAR	P_val_RFSV
0	0.447537	0.013852	0.362412	0.007897	1.990378	0.001907
1	0.781541	0.013539	0.628202	0.009465	1.955865	0.000587

Daily SP500, historical volatility and log-volatility,
N=500, T=1000, window size=200, s=1500

	RMSE_HAR	RMSE_RFSV	MAE_HAR	MAE_RFSV	P_val_HAR	P_val_RFSV
0	0.296367	0.004751	0.233250	0.002904	2.004151	0.000515
1	0.711193	0.004781	0.566757	0.003242	1.973263	0.000089

	RMSE_HAR	RMSE_RFSV	MAE_HAR	MAE_RFSV	P_val_HAR	P_val_RFSV
0	20.038942	0.285621	16.008862	0.197850	2.003558	0.000407
1	112.609457	0.684125	85.062016	0.433606	1.980984	0.000073

References

- E. Bacry et al., *Log-Normal continuous cascades: aggregation properties and estimation. Application to financial time-series*, 2008.
- J. Gatheral et al., *Volatility is rough*, 2016.
- J. F. Muzy et al., *Modelling fluctuations of financial time series: from cascade process to stochastic volatility model*, 2000.
- U. Frisch, *Turbulence*, Cambridge University Press, 1995.
- Peitgen et al., Wikipedia (self-similarity).