STATISTICAL METHODS FOR FINANCE - COURSEWORK

MSC. MATHEMATICS AND FINANCE

Volatility is rough

Authors:

Paul Wattellier

Oussama Saadi

Ines Vanheuverswyn

Ahmed Reda Seghrouchni

Thomas Aujoux

Supervisor:

C. SALVI

November 2023 - December 2023

Contents

Introduction

In the domain of financial derivatives, the modeling of asset prices and their volatility, is a central challenge. The log-prices are often represented as continuous semi-martingales, formalized as:

$$dY_t = \mu_t dt + \sigma_t dW_t, \tag{1}$$

where Y_t is the log-price of an asset at time t, μ_t is the drift component, σ_t represents the volatility, and dW_t is an increment of a Brownian motion.

One traditional method of modeling volatility involves using a constant or a deterministic function of t and Y_t. However, as anticipated, such a straightforward model has proven to be inefficient in practical applications. A more realistic approach, better aligning with the complexities of reality, is to model volatility as a stochastic process.

The concept of fractional volatility and of fractional Brownian motion processes (fBM)) has been introduced to model log-volatility. A fractional Brownian motion $B_H(t)$ is a continuous Gaussian process with mean zero, self-similarity, and stationary increments. It is characterized by the Hurst index H, where $0 < H < 1$, influencing the smoothness of its paths. The covariance function exhibits long-range dependence, making it a generalization of the classical Brownian motion ($H = 1/2$).

The parameter H is often referred to as the Hurst index and determines the level of roughness or smoothness of the paths. For H>1/2, the paths are more likely to exhibit persistent trends, while for H<1/2, the paths are more likely to be rough and fluctuate rapidly. The choice of $H > 1/2$ ensures long memory properties, making it a potentially better fit for modeling the persistence observed in market volatility.

The implied volatility surface, crucial for option pricing, is often at odds with predictions from conventional models. This surface, denoted as $\sigma_{BS}(k, \tau)$ for an option with log-moneyness k and time to expiration τ, reveals discrepancies when modeled with constant or deterministic volatility functions. In practice, this surface exhibits complex behaviors and patterns that standard models, such as Black-Scholes or even local volatility models, struggle to replicate accurately.

Motivated by these challenges, this paper introduces the Rough Fractional Stochastic Volatility (RFSV) model as an alternative model to capture the intricacies of financial market volatility. The coursework is structured to first present empirical findings on volatility smoothness, followed by the detailed specification of the RFSV model, an investigation into the notion of spurious long memory in volatility, and an evaluation of the model's forecasting capabilities. The final sections delve into the microstructural underpinnings of volatility irregularities and conclude with key insights and implications of the research.

1 Estimating the Smoothness of the Volatility Process

In this section the paper focuses on estimating the smoothness of the volatility process using the historical data of four assets.

1.1 Theory behind

To measure the smoothness of the volatility process based on the observed discrete data the paper introduce $m(q,\Delta)$ for $q \geq 0$:

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

With $\sigma_0, \sigma_\Delta, \ldots, \sigma_{k\Delta}, \ldots$, the volatility process defined on the interval $[0,T]$ where $k \in \{0, \lfloor bT/\Delta \rfloor\}$. Setting $N = \lfloor bT/\Delta \rfloor$.
The paper make the assumption that for some $s_q > 0$ and $b_q > 0$, as Δ tends to zero:

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

This convergence implies the volatility process belongs to the Besov smoothness space $B^{q,\infty}_{s_q}$, indicating the regularity of the process in the L^q norm. Functions in $B^{q,\infty}_{s_q}$ exhibit the Hölder property with parameter h for any $h < s$.
For a fractional Brownian motion (fBM) with Hurst parameter H, the relationship $N^{qs_q} m(q,\Delta) \to b_q$ holds, and the sample paths of the fBM belong to $B^{q,\infty}_H$ almost surely.
Under the hypothesis of stationary increments (the statistical properties of the increments of the process remain constant over time), and the applicability of the law of large numbers $m(q,\Delta)$ serves as the empirical counterpart of $E[|\log(\sigma_\Delta) - \log(\sigma_0)|^q]$.
Spot volatility values are estimated through proxies due to the impracticality of directly observing the volatility process. The minimal Δ is set to one day, and two daily spot volatility proxies are considered.

1.2 Methodology Summary

To provide valuable insights into the smoothness characteristics of the underlying volatility process, the paper considers the regressions of log$m(q,\Delta)$ against $\log\Delta$, considering a range of Δ values for multiple q value. To compute $m(q,\Delta)$, the volatility discrete values are needed, 2 different approach are given:

- For DAX and Bund future contracts an integrated variance estimator derived from a model incorporating uncertainty zones is employed. The estimator focuses on the 10 am to 11 am London time interval, producing re-normalized estimates of integrated variance as proxies for unobservable spot variance.

- For the S&P and NASDAQ indice, daily realized variance estimates from the Oxford-Man Institute of Quantitative Finance Realized Library are utilized. These estimates, covering the entire trading day, introduce an expected upward bias in estimating the smoothness of the volatility process, attributed to the regularization effect of integration.

1.3 Results

The plots of $\log m(q, \Delta)$ against $\log \Delta$ for different q values (q = 0.5, 1, 1.5, 2, 3) exhibits straight lines, indicating a scaling property in expectation for log-volatility increments.
The slope ζ_q is found to scale with q, with H values of 0.125 for DAX and 0.082 for Bund.
Similar results are reported for the S&P and NASDAQ indices using precomputed 5-minute realized variance estimates. The scaling property observed for DAX and Bund persists, with slightly higher estimated smoothness values ($H = 0.142$ for S&P and $H = 0.139$ for NASDAQ). Expectedly, smoothness estimates may be biased upward due to the use of whole-day realized variance.
The analysis extends to other indices in the Oxford-Man dataset, revealing a universal scaling behavior for $m(q, \Delta)$.

All assets exhibit a Gaussian distribution for increments of log-volatility, consistent with prior literature. Re-scaling by Δ^H aligns empirical distributions with normal fits.

1.4 Comments and Critique

1.4.1 Comments on the results

The consistent observation of scaling properties in the log-log plots indicates a power-law relationship. The patterns in volatility are repeated at different time scales supporting the claim that the volatility process exhibits a certain level of self-similarity or fractal behavior.
Also the universality of the scaling behavior supported by the finding across the different indexes of the Oxford-Man data-set show that this property of underlying structure of volatility processes is common among various financial instruments.

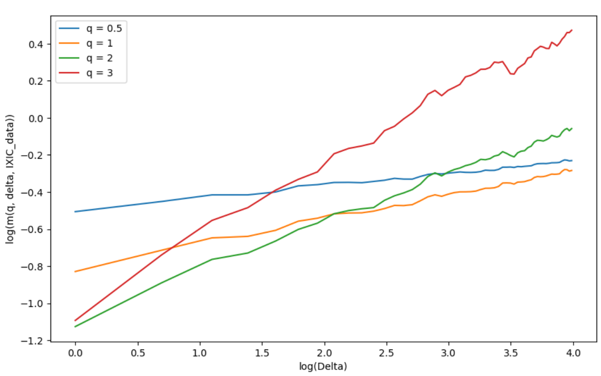
1.4.2 Comments on the methodology

To approximate the volatility for DAX and Bund future contracts an integrated variance estimation has been used. This method is commonly employed in financial practice. However, it may exhibit sensitivity to extreme events, and different estimators could introduce varying biases and efficiencies. Also, the authors used realized variance as a direct measure of volatility, this can introduce bias due to non-synchronous trading. The proxy spot volatility assumes that a fixed time of day captures representative volatility and may introduce bias if significant intra-day volatility patterns exist. The Monte Carlo Simulation used in the paper allows for testing the robustness of the methodology.

1.4.3 Our results

While reproducing the same procedure as in the paper, we observed a difference in the estimated values of the Hurst parameter (H) for different q. In Figure 1 a comparison of the regression for the NASDAQ index can be observed, the different H for the NASDAQ index can be found in the annex. The change in H with q indicates that the process exhibits varying levels of persistence or smoothness at different orders of moments.

In our analysis, we observed a distribution of log-volatility increments for the S&P index that closely aligns with the theoretical expectations of a normal (Gaussian) distribution. This finding is consistent with the results reported the paper. The histograms depicting increments for various lag intervals (Δ) exhibited shapes similar to normal distributions.



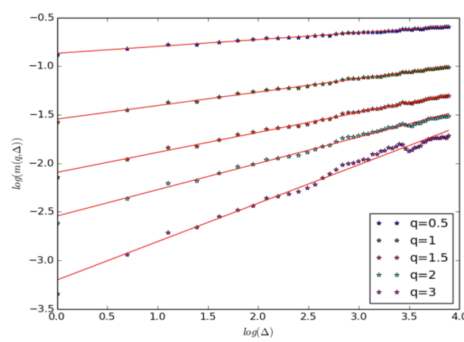
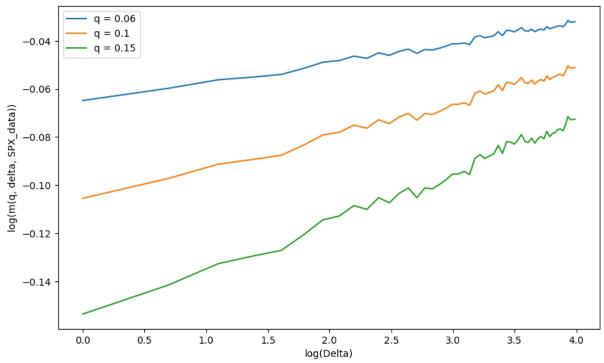


Figure 2.5: $\log m(q, \Delta)$ as a function of $\log(\Delta)$, NASDAQ.

Figure 1: Comparison between our findings (left) and the paper results (right) for the regressions with the NASDAQ index

Our results for the SPX index, when utilizing small q, exhibit a similarity with those reported in the referenced paper. This alignment is particularly encouraging, given that the analysis in the rest of the paper employs $q = 0.14$.



q	H_{estimate}	p-value
0.06	0.008383	1.019847×10^{-49}
0.10	0.013940	1.434118×10^{-50}
0.15	0.020856	1.504066×10^{-51}

Table 1: Results for SPX Data

2 Specification of the RFSV Model

Informed by the scaling property and near-Gaussian distribution of log-volatility increments of various financial assets, we introduce a model that encapsulates these empirical characteristics.

2.1 Modeling Log-Volatility with Fractional Brownian Motion

The increments of log-volatility can be aptly modeled by fractional Brownian motion (fBM), as indicated by the relation:

$$\log \sigma_{t+\Delta} - \log \sigma_t = \nu(W^H_{t+\Delta} - W^H_t), \tag{2}$$

where ν is a positive constant, and W^H denotes a fractional Brownian motion with a Hurst parameter H that matches the empirically observed smoothness of volatility. This model facilitates the expression of volatility σ_t as:

$$\sigma_t = \sigma \exp(\nu W^H_t), \tag{3}$$

with σ being another positive constant.

2.2 Fractional Ornstein-Uhlenbeck Process for Stationarity

To ensure stationarity, a desired property for both mathematical tractability and the practical application of the model over extended time periods, we employ the fractional Ornstein-Uhlenbeck (fOU) process. The fOU process is defined as the solution to the stochastic differential equation:

$$dX_t = \nu dW^H_t - \alpha(X_t - m)dt, \tag{4}$$

where ν and α are positive parameters, m represents the long-term mean level, and the process X_t is given by the integral expression:

$$X_t = \nu \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_s^H + m. \tag{5}$$

Here, the integral is understood in the sense of a pathwise Riemann-Stieltjes integral.

Our final model for volatility over a given time interval $[0, T]$ is thus:

$$\sigma_t = \exp(X_t), \quad t \in [0, T], \tag{6}$$

with the fOU process X_t satisfying the equation above for specified values of $\nu, \alpha, m,$ and $H < \frac{1}{2}$. This model maintains stationarity and, when α is small relative to the time scale $1/T$, the log-volatility locally resembles a fractional Brownian motion.

2.3 Refinement of the RFSV Model

The RFSV model refines volatility representation by employing a fractional Ornstein-Uhlenbeck process, ensuring that volatility exhibits both empirical consistency and mathematical stationarity. Volatility is thus modeled as:

$$\sigma_t = \exp(X_t), \quad X_t = \nu \int_{-\infty}^{t} e^{-\alpha(t-s)} dW_s^H + m, \tag{7}$$

where σ_t is the spot volatility, X_t the log-volatility, W_t^H a fractional Brownian motion, and m the long-term mean.

2.4 The Role of the Hurst Parameter and Long Memory

Crucial to the RFSV model is the Hurst parameter H, constrained to values less than $\frac{1}{2}$. The choice of $H < \frac{1}{2}$ is fundamental, as it aligns with the empirical observation of volatility being a process with long memory—where past values significantly influence the future values over an extended period. This relationship is a stark contrast to classical models, which typically assume short memory characterized by a rapid decay of autocorrelation. However, The document examines the constancy of the variable H through the segmentation of the dataset into equally sized portions, establishing a fixed H for subsequent lotests. A potential critique involves demonstrating the temporal stability of H and deriving an estimate by employing broader segmentation and diverse estimators, thus providing a more robust evaluation of its constancy.

2.5 Autocovariance Functions of the RFSV Model

The RFSV model predicts that the autocovariance function of log-volatility displays a specific form. When the process $\{X_t\}$ follows the fOU process, the autocovariance of the increments of $\{X_t\}$ is expected to be linear with respect to Δ^{2H}. This is a direct consequence of the roughness parameter $H < \frac{1}{2}$, which also induces the empirical scaling law observed in the financial data.

2.6 Simulation-Based Analysis and Robustness

A simulation-based analysis of the RFSV model reveals its robustness and the consistent application of the roughness parameter H. The model's predictions about the behavior of log-volatility and its increments have been validated through comprehensive simulation studies, which replicate the key statistical properties of the empirical market data.

3 Spurious long memory of volatility?

This section of the paper addresses the common belief in financial econometrics that volatility exhibits long memory, meaning its autocovariance function decays very slowly. The authors challenge this notion using the Rough Fractional Stochastic Volatility (RFSV) model. They demonstrate that even though the RFSV model does not inherently possess long memory properties, standard statistical procedures used to detect volatility persistence often misidentify the model's output as exhibiting long memory. This finding suggests that such statistical tests may be unreliable and could lead to incorrect conclusions about the nature of financial volatility.

In fact, the section delves into the behavior of the autocovariance function of log-volatility, $\text{Cov}[\log(\sigma_t), \log(\sigma_{t+\Delta})]$ in the RFSV model, showing that they do not decay as a power law, contrary to the typical characteristics of long memory processes.

In previous sections, it was shown that both in the data and in the model:

$$\text{Cov}[\log(\sigma_t), \log(\sigma_{t+\Delta})] \approx A - B\Delta^{2H}$$

$$\text{Cov}[\sigma_t, \sigma_{t+\Delta}] \approx Ce^{-B\Delta^{2H}} - D$$

for some constants A, B, C and D. Thus, neither in the model nor in the data does the autocovariance function decay as a power law. And neither the data nor the model exhibits long memory.

This misinterpretation arises due to the specific decay pattern of the autocovariance function in the RFSV model, which does not align with the typical power-law decay characteristic of long memory processes:

$$\text{Cov}[\log(\sigma_t), \log(\sigma_{t+\Delta})] \neq \Delta^{-\gamma} \text{ for } \gamma < 1$$

Another point that is discussed is that the RFSV model's representation of volatility suggests persistence but not in the classical power-law sense of long memory.
Therefore, the findings indicate a need for reevaluation of traditional econometric methods for analyzing financial volatility, as they may overestimate the presence of long memory.

As a conclusion, the RFSV model, while displaying some form of persistence, does not exhibit long memory in the classical power law sense. This raises questions about the use of standard long memory detection techniques in financial econometrics.

4 Forecasting using the RFSV model

In this part of the report, we will see a concrete application of the Rough Fractional Stochastic Volatility (RFSV) model, this approach will alow us to predict both log-volatility and variance. This model has a great particularity, this is the use of the fractional Brownian motion using particular parameters. In this part of the report, we will see a forecasting methodology using this model, and at the end we will compare it with other classical models like the Autoregressive (AR) and Heterogeneous Autoregressive (HAR) models.

4.1 Forecasting using the RFSV model

We will now see how to forecast log-volatility using RFSV model. We will first use the formula that involves fractional Brownian motion (fBM) with :

$$E[W^H_{t+\Delta}|\mathcal{F}_t] = \frac{\cos(H\pi)\Delta^{H+\frac{1}{2}}}{\pi}\int_{-\infty}^{t}\frac{W^H_s\,ds}{(t-s+\Delta)(t-s)^{H+\frac{1}{2}}} \tag{8}$$

After rewriting the expression using different approximation, we find the formula that we will use in the next part of this section for the prediction:

$$E[\log\sigma^2_{t+\Delta}|\mathcal{F}_t] = \frac{\cos(H\pi)\Delta^{H+\frac{1}{2}}}{\pi}\int_{-\infty}^{t}\frac{\log\sigma^2_s\,ds}{(t-s+\Delta)(t-s)^{H+\frac{1}{2}}} \tag{9}$$

We will now compare the prediction that we will find with the predictions of the AR and HAR forecasts. We can recall in this report the formula for the prediction of the two other models:

$$\log(\sigma^2_{t+\Delta}) = K^\Delta_0 + \sum_{i=0}^{p} C^\Delta_i \log(\sigma^2_{t-i}) \quad \text{(AR(p))} \tag{10}$$

$$\log(\sigma^2_{t+\Delta}) = K^\Delta_0 + C^\Delta_0\log(\sigma^2_t) + C^\Delta_5\sum_{i=1}^{5}\log(\sigma^2_{t-i}) + C^\Delta_{20}\sum_{i=1}^{20}\log(\sigma^2_{t-i}) \quad \text{(HAR)} \tag{11}$$

To conclude this part of the report we can clearly see that the model RFSV has better performances than the two other models performance, especially for longer forecast horizons, in order to compare the different models we used the Ratio P with this formula:

$$P = \frac{1}{N-\Delta}\sum_{k=500}^{N-\Delta}\left(\log(\sigma_k^{2+\Delta}) - \log(\sigma_k^{2+\Delta})'\right)^2$$

4.2 Variance prediction

In the part we will see the prediction of variance using the RFSV model. We will begin be saying the approximation for log-variance based on fractional Brownian motion. This will give us the variance prediction formula and the conditional variance formula:

Variance prediction formula:

$$\sigma^2_{t+\Delta} = \exp(\log \sigma^2_{t+\Delta} + 2c\nu^2\Delta^{2H}) \tag{12}$$

Conditional variance formula:

$$\text{Var}[W^H_{t+\Delta}|\mathcal{F}_t] = c\Delta^{2H} \tag{13}$$

The formula show the conditional Gaussian nature with a specific conditional variance. We will now compare the performances of the different models with the AR and HAR like before. To conclude this section, we can see the robustness of the RFSV model in capturing variance dynamics.

Conclusion

The comprehensive exploration of the paper offers pivotal insights into financial modeling. Central to this study is the Rough Fractional Stochastic Volatility (RFSV) model, challenging traditional views of market volatility. This research emphasizes that volatility may or may not exhibit long-term memory, contradicting previous beliefs and impacting financial econometrics. The RFSV model demonstrates superior forecasting abilities compared to standard models like AR and HAR, particularly in predicting volatility trends. Additionally, the influence of market microstructure on volatility's irregular behavior is highlighted, suggesting complex interactions in financial markets.

5 Annexes

Table 2: Regression Results

q	H_{estimate}
0.5	0.066818
1.0	0.131452
1.5	0.193748
2.0	0.253554
3.0	0.365360