

Volatility is rough

Jim Gatheral

Baruch College, City University of New York
jim.gatheral@baruch.cuny.edu

Thibault Jaisson*

CMAP, École Polytechnique Paris
thibault.jaisson@polytechnique.edu

Mathieu Rosenbaum

LPMA, Université Pierre et Marie Curie (Paris 6)
mathieu.rosenbaum@upmc.fr

October 13, 2014

Abstract

Estimating volatility from recent high frequency data, we revisit the question of the smoothness of the volatility process. Our main result is that log-volatility behaves essentially as a fractional Brownian motion with Hurst exponent H of order 0.1, at any reasonable time scale. This leads us to adopt the fractional stochastic volatility (FSV) model of Comte and Renault [16]. We call our model Rough FSV (RFSV) to underline that, in contrast to FSV, $H < 1/2$. We demonstrate that our RFSV model is remarkably consistent with financial time series data; one application is that it enables us to obtain improved forecasts of realized volatility. Furthermore, we find that although volatility is not long memory in the RFSV model, classical statistical procedures aiming at detecting volatility persistence tend to conclude the presence of long memory in data generated from it. This sheds light on why long memory of volatility has been widely accepted as a stylized fact. Finally, we provide a quantitative market microstructure-based foundation for our findings, relating the roughness of volatility to high frequency trading and order splitting.

Keywords: High frequency data, volatility smoothness, fractional Brownian motion, fractional Ornstein-Uhlenbeck, long memory, volatility persistence, volatility forecasting, option pricing, volatility surface, Hawkes processes, high frequency trading, order splitting.

*Thibault Jaisson gratefully acknowledges financial support from the chair “Risques Financiers” of the Risk Foundation and the chair “Marchés en Mutation” of the French Banking Federation.

1 Introduction

1.1 Volatility modeling

In the derivatives world, log-prices are often modeled as continuous semi-martingales. For a given asset with log-price Y_t , such a process takes the form

$$dY_t = \mu_t dt + \sigma_t dW_t,$$

where μ_t is a drift term and W_t is a one-dimensional Brownian motion. The term σ_t denotes the volatility process and is the most important ingredient of the model. In the Black-Scholes framework, the volatility function is either constant or a deterministic function of time. In Dupire's local volatility model, see [22], the local volatility $\sigma(Y_t, t)$ is a deterministic function of the underlying price and time, chosen to match observed European option prices exactly. Such a model is by definition time-inhomogeneous; its dynamics are highly unrealistic, typically generating future volatility surfaces (see Section 1.3 below) completely unlike those we observe. A corollary of this is that prices of exotic options under local volatility can be substantially off-market. On the other hand, in so-called stochastic volatility models, the volatility σ_t is modeled as a continuous Brownian semi-martingale. Notable amongst such stochastic volatility models are the Hull and White model [32], the Heston model [31], and the SABR model [29]. Whilst stochastic volatility dynamics are more realistic than local volatility dynamics, generated option prices are not consistent with observed European option prices. We refer to [26] and [39] for more detailed reviews of the different approaches to volatility modeling. More recent market practice is to use local-stochastic-volatility (LSV) models which both fit the market exactly and generate reasonable dynamics.

1.2 Fractional volatility

In terms of the smoothness of the volatility process, the preceding models offer two possibilities: very regular sample paths in the case of Black-Scholes, and volatility trajectories with regularity close to that of Brownian motion for the local and stochastic volatility models. Starting from the stylized fact that volatility is a long memory process, various authors have proposed models that allow for a wider range of regularity for the volatility. In a pioneering paper, Comte and Renault [16] proposed to model log-volatility using fractional Brownian motion (fBM for short), ensuring long memory by choosing the Hurst parameter $H > 1/2$. A large literature has subsequently developed around such fractional volatility models, for example [12, 15, 44].

The fBM $(W_t^H)_{t \in \mathbb{R}}$ with Hurst parameter $H \in (0, 1)$, introduced in [36], is a centered self-similar Gaussian process with stationary increments satisfying for any $t \in \mathbb{R}$, $\Delta \geq 0$, $q > 0$:

$$\mathbb{E}[|W_{t+\Delta}^H - W_t^H|^q] = K_q \Delta^{qH}, \quad (1.1)$$

with K_q the moment of order q of the absolute value of a standard Gaussian variable. For $H = 1/2$, we retrieve the classical Brownian motion. The sample paths of W^H are Hölder-continuous with exponent r , for any $r < H$ ¹. Finally, when $H > 1/2$, the increments of the fBM are positively correlated and exhibit long memory in the sense that

$$\sum_{k=0}^{+\infty} \text{Cov}[W_1^H, W_k^H - W_{k-1}^H] = +\infty.$$

Indeed, $\text{Cov}[W_1^H, W_k^H - W_{k-1}^H]$ is of order k^{2H-2} as $k \rightarrow \infty$. Note that in the case of the fBM, there is a one to one correspondence between regularity and long memory through the Hurst parameter H .

As mentioned earlier, the long memory property of the volatility process has been widely accepted as a stylized fact since the seminal analyses of Ding, Granger and Engle [20], Andersen and Bollerslev [1] and Andersen et al. [3]. Initially, it appears that the term *long memory* referred to the slow decay of the autocorrelation function (of absolute returns for example), anything slower than exponential. Over time however, it seems that this term has acquired the more precise meaning that the autocorrelation function is not integrable, see [8], and even more precisely that it decays as a power-law with exponent less than 1. Much of the more recent literature, for example [7, 11, 13], assumes long memory in volatility in this more technical sense. Indeed, meaningful results can probably only be obtained under such a specification, since it is not possible to estimate the asymptotic behavior of the covariance function without assuming a specific form. Nevertheless, analyses such as that of Andersen et al. [3] use data that predate the advent of high-frequency electronic trading, and the evidence for long memory has never been sufficient to satisfy remaining doubters such as Mikosch and Stărică in [38]. To quote Rama Cont in [17]:

... the econometric debate on the short range or long range nature of dependence in volatility still goes on (and may probably never be resolved)...

One of our contributions in this paper is (we believe) to finally resolve this question, showing that the autocorrelation function of volatility does not behave as a power law, at least at usual time scales of observation. This implies

¹Actually H corresponds to the regularity of the process in a more accurate way: in terms of Besov smoothness spaces, see Section 2.1.

that when stated in term of the asymptotic behavior of the autocorrelation function, the long memory question can simply not be answered. Nevertheless, we are able to provide explicit expressions enabling us to analyze thoroughly the dependence structure of the volatility process.

1.3 The shape of the implied volatility surface

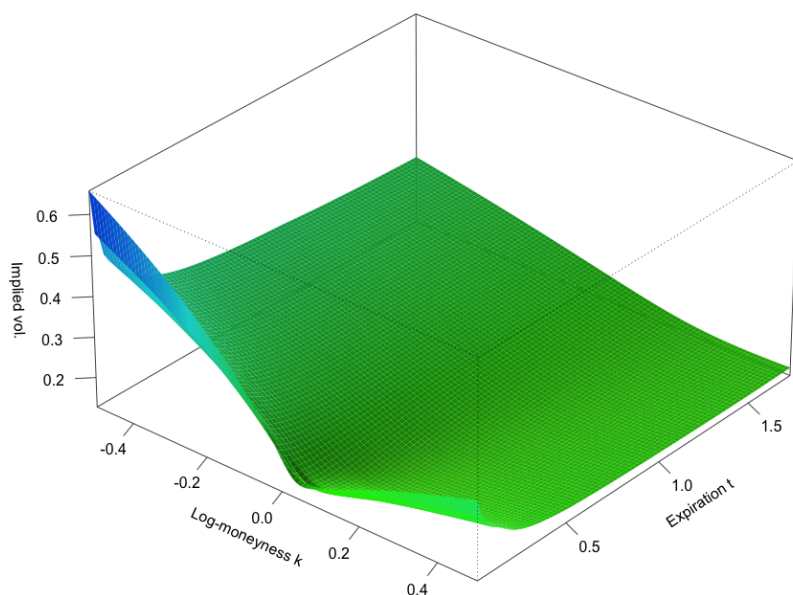


Figure 1.1: The S&P volatility surface as of June 20, 2013.

As is well-known, the implied volatility $\sigma_{BS}(k, \tau)$ of an option (with log-moneyness k and time to expiration τ) is the value of the volatility parameter in the Black-Scholes formula required to match the market price of that option. Plotting implied volatility as a function of strike price and time to expiry generates the *volatility surface*, explored in detail in, for example, [26]. A typical such volatility surface generated from a “stochastic volatility inspired” (SVI) [27] fit to closing SPX option prices as of June 20, 2013² is shown in Figure 1.1. It is a stylized fact that, at least in equity markets, although the level and orientation of the volatility surface do change over time, the general overall shape of the volatility surface does not change, at least to a first approximation. This suggests that it is desirable to model

²Closing prices of SPX options for all available strikes and expirations as of June 20, 2013 were sourced from OptionMetrics (www.optionmetrics.com) via Wharton Research Data Services (WRDS).

volatility as a time-homogenous process, *i.e.* a process whose parameters are independent of price and time.

However, conventional time-homogenous models of volatility such as the Hull and White, Heston, and SABR models do not fit the volatility surface. In particular, as shown in Figure 1.2, the observed term structure of at-the-money ($k = 0$) volatility skew

$$\psi(\tau) := \left| \frac{\partial}{\partial k} \sigma_{\text{BS}}(k, \tau) \right|_{k=0}$$

is well-approximated by a power-law function of time to expiry τ . In contrast, conventional stochastic volatility models generate a term structure of at-the-money (ATM) skew that is *constant* for small τ and behaves as a sum of decaying exponentials for larger τ .

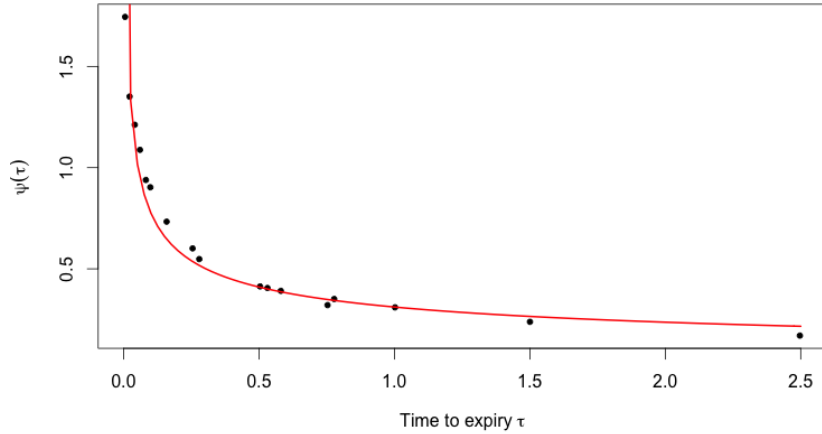


Figure 1.2: The black dots are non-parametric estimates of the S&P ATM volatility skews as of June 20, 2013; the red curve is the power-law fit $\psi(\tau) = A\tau^{-0.4}$.

In Section 3.3 of [25], as an example of the application of his martingale expansion, Fukasawa shows that a stochastic volatility model where the volatility is driven by fractional Brownian motion with Hurst exponent H generates an ATM volatility skew of the form $\psi(\tau) \sim \tau^{H-1/2}$, at least for small τ . This is interesting in and of itself in that it provides a counterexample to the widespread belief that the explosion of the volatility smile as $\tau \rightarrow 0$ (as clearly seen in Figures 1.1 and 1.2) implies the presence of jumps [10]. The main point here is that for a model of the sort analyzed by Fukasawa to generate a volatility surface with a reasonable shape, we would need

to have a value of H close to zero. As we will see in Section 2, our empirical estimates of H from time series data are in fact very small.

The volatility model that we will specify in Section 3.1, driven by fBM with $H < 1/2$, therefore has the potential to be not only consistent with the empirically observed properties of the volatility time series but also consistent with the shape of the volatility surface. In this paper, we focus on the modeling of the volatility time series. A more detailed analysis of the consistency of our model with option prices is left for a future article.

1.4 Main results and organization of the paper

In Section 2, we report our estimates of the smoothness of the log-volatility for selected assets. This smoothness parameter lies systematically between 0.08 and 0.2 (in the sense of Hölder regularity for example). Furthermore, we find that increments of the log-volatility are approximately normally distributed and that their moments enjoy a remarkable monofractal scaling property. This leads us to model the log of volatility using a fBM with Hurst parameter $H < 1/2$ in Section 3. Specifically we adopt the fractional stochastic volatility (FSV) model of Comte and Renault [16]. We call our model Rough FSV (RFSV) to underline that, in contrast to FSV, we take $H < 1/2$. We also show in the same section that the RFSV model is remarkably consistent with volatility time series data. The issue of volatility persistence is considered through the lens of the RFSV model in Section 4. Our main finding is that although the RFSV model does not have any long memory property, classical statistical procedures aiming at detecting volatility persistence tend to conclude the presence of long memory in data generated from it. This sheds new light on the supposed long memory in the volatility of financial data. In Section 5, we apply our model to forecasting volatility. In particular, we show that RFSV volatility forecasts outperform conventional AR and HAR volatility forecasts. Finally, in Section 6, we present a market microstructure explanation for the regularities we observe in the volatility process at the macroscopic scale. We show that the empirical behavior of volatility may be explained in terms of order splitting and the high degree of endogeneity of the market ascribed to algorithmic trading. Some proofs are relegated to the appendix.

2 Smoothness of the volatility: empirical results

In this section we report estimates of the smoothness of the volatility process for four assets: The DAX and Bund futures contracts, for which we estimate integrated variance directly from high frequency data using an estimator based on the model with uncertainty zones, [42, 43], and the S&P and

NASDAQ indices, for which we use precomputed realized variance estimates from the Oxford-Man Institute of Quantitative Finance Realized Library³.

2.1 Estimating the smoothness of the volatility process

Let us first pretend that we have access to discrete observations of the volatility process, on a time grid with mesh Δ on $[0, T]$: $\sigma_0, \sigma_\Delta, \dots, \sigma_{k\Delta}, \dots$, $k \in \{0, \lfloor T/\Delta \rfloor\}$. Set $N = \lfloor T/\Delta \rfloor$, then for $q \geq 0$, we define

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

In the spirit of [46], our main assumption is that for some $s_q > 0$ and $b_q > 0$, as Δ tends to zero,

$$N^{qs_q} m(q, \Delta) \rightarrow b_q. \quad (2.1)$$

Under additional technical conditions, Equation (2.1) essentially says that the volatility process belongs to the Besov smoothness space $\mathcal{B}_{q,\infty}^{s_q}$ and does not belong to $\mathcal{B}_{q,\infty}^{s'_q}$, for $s'_q > s_q$, see [45]. Hence s_q can really be viewed as the regularity of the volatility when measured in l_q norm. In particular, functions in $\mathcal{B}_{q,\infty}^s$ for every $q > 0$ enjoy the Hölder property with parameter h for any $h < s$. For example, if $\log(\sigma_t)$ is a fBM with Hurst parameter H , then for any $q \geq 0$, Equation (2.1) holds in probability with $s_q = H$ and it can be shown that the sample paths of the process indeed belong to $\mathcal{B}_{q,\infty}^H$ almost surely. Assuming the increments of the log-volatility process are stationary and that a law of large number can be applied, $m(q, \Delta)$ can also be seen as the empirical counterpart of

$$\mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q].$$

Of course, the volatility process is not directly observable, and an exact computation of $m(q, \Delta)$ is not possible in practice. We must therefore proxy spot volatility values by appropriate estimated values. Since the minimal Δ will be equal to one day in the sequel, we proxy the (true) spot volatility daily at a fixed given time of the day (11 am for example). Two daily spot volatility proxies will be considered:

- For our ultra high frequency intraday data (DAX future contracts and Bund future contracts⁴, 1248 days from 13/05/2010 to 01/08/2014⁵),

³<http://realized.oxford-man.ox.ac.uk/data/download>. The Oxford-Man Institute's Realized Library contains a selection of daily non-parametric estimates of volatility of financial assets, including realized variance (rv) and realized kernel (rk) estimates. A selection of such estimators is described and their performances compared in, for example, [28].

⁴For every day, we only consider the future contract corresponding to the most liquid maturity.

⁵Data kindly provided by QuantHouse EUROPE/ASIA, <http://www.quanthouse.com>.

we use the estimator of the integrated variance from 10 am to 11 am London time obtained from the model with uncertainty zones, see [42, 43]. After renormalization, the resulting estimates of integrated variance over very short time intervals can be considered as good proxies for the unobservable spot variance. In particular, the one hour long window on which they are computed is small compared to the extra day time scales that will be of interest here.

- For the S&P and NASDAQ indices⁶, we proxy daily spot variances by daily realized variance estimates from the Oxford-Man Institute of Quantitative Finance Realized Library (3,540 trading days from January 3, 2000 to March 31, 2014). Since these estimates of integrated variance are for the whole trading day, we expect estimates of the smoothness of the volatility process to be biased upwards, integration being a regularizing operation. We compute the extent of this bias by simulation in Section 3.4.

In the following, we retain the notation $m(q, \Delta)$ with the understanding that we are only proxying the (true) spot volatility as explained above. We now proceed to estimate the smoothness parameter s_q for each q by computing the $m(q, \Delta)$ for different values of Δ and regressing $\log m(q, \Delta)$ against $\log \Delta$. Note that for a given Δ , several $m(q, \Delta)$ can be computed depending on the starting point. Our final measure of $m(q, \Delta)$ is the average of these values.

2.2 DAX and Bund futures contracts

DAX and Bund futures are amongst the most liquid assets in the world and moreover, the model with uncertainty zones used to estimate volatility is known to apply well to them, see [19]. So we can be confident in the reliability of our volatility proxy. Nevertheless, as an extra check, we will confirm the quality of our volatility proxy by Monte Carlo simulation in Section 3.4.

Plots of $\log m(q, \Delta)$ vs $\log \Delta$ for different values of q , are displayed for the DAX in Figure 2.1, and for the Bund in Figure 2.2.

⁶And also the CAC40, Nikkei and FTSE indices in some specific parts of the paper.

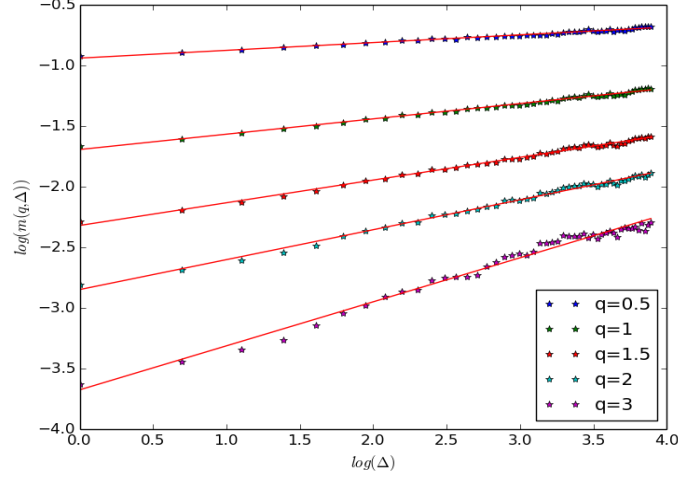


Figure 2.1: $\log m(q, \Delta)$ as a function of $\log \Delta$, DAX.

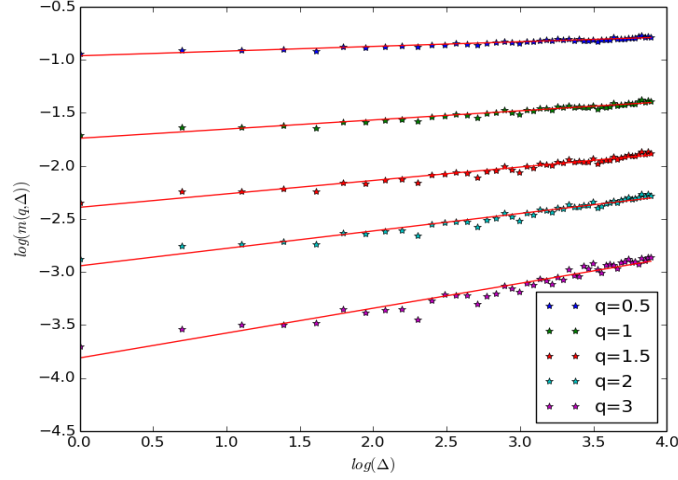


Figure 2.2: $\log m(q, \Delta)$ as a function of $\log \Delta$, Bund.

For both DAX and Bund, for a given q , the points essentially lie on a straight line. Under stationarity assumptions, this implies that the log-volatility increments enjoy the following scaling property in expectation:

$$\mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q] = K_q \Delta^{\zeta_q},$$

where $\zeta_q > 0$ is the slope of the line associated to q . Moreover, the smoothness parameter s_q does not seem to depend on q . Indeed, plotting ζ_q against

q , we obtain that $\zeta_q \sim H q$ with H equal to 0.125 for the DAX and to 0.082 for the Bund, see Figure 2.3.

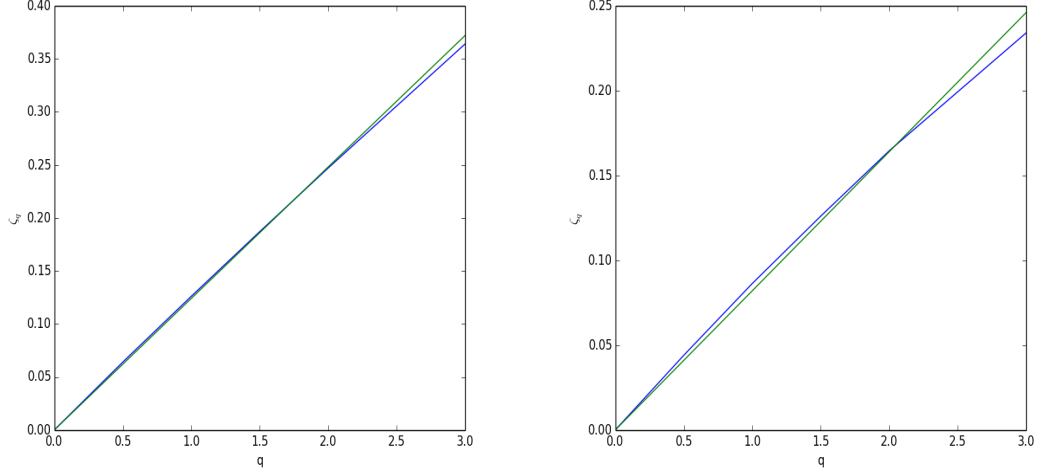


Figure 2.3: ζ_q (blue) and $0.125 \times q$ (green), DAX (left); ζ_q (blue) and $0.082 \times q$ (green), Bund (right).

We remark that the graphs for ζ_q are actually very slightly concave. However, we observe the same small concavity effect when we replace the log-volatility by simulations of a fBM with the same number of points. We conclude that this effect relates to finite sample size and is thus not significant.

2.3 S&P and NASDAQ indices

We report in Figure 2.4 and Figure 2.5 similar results for the S&P and NASDAQ indices. The variance proxies used here are the precomputed 5-minute realized variance estimates for the whole trading day made publicly available by the Oxford-Man Institute of Quantitative Finance.

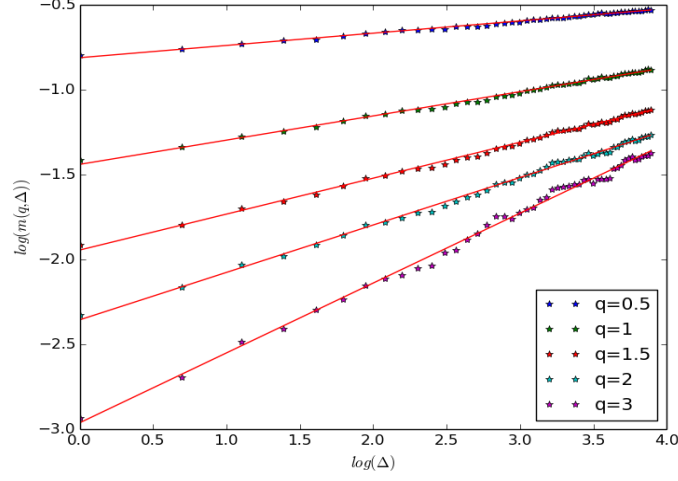


Figure 2.4: $\log m(q, \Delta)$ as a function of $\log \Delta$, S&P.

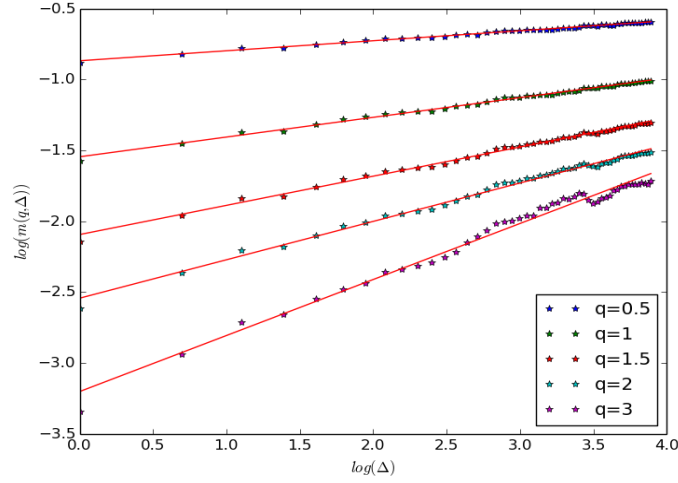


Figure 2.5: $\log m(q, \Delta)$ as a function of $\log(\Delta)$, NASDAQ.

We observe the same scaling property for the S&P and NASDAQ indices as we observed for DAX and Bund futures and again, the s_q do not depend on q . However, the estimated smoothnesses are slightly higher here: $H = 0.142$ for the S&P and $H = 0.139$ for the NASDAQ, see Figure 2.6.

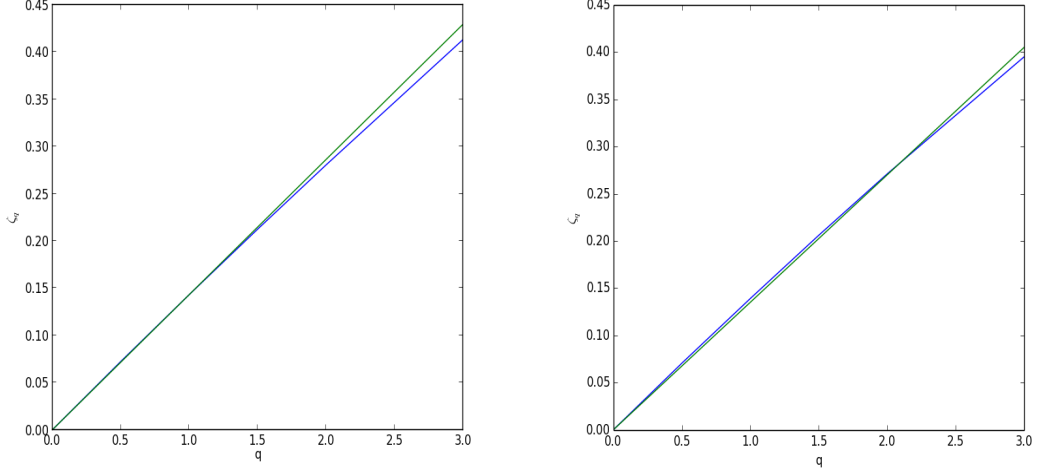


Figure 2.6: ζ_q (blue) and $0.142 \times q$ (green), S&P (left); ζ_q (blue) and $0.139 \times q$ (green), NASDAQ (right).

Once again, we do expect these smoothness estimates to be biased high because we are using whole-day realized variance estimates, as explained earlier in Section 2. Finally, we remark that as for DAX and Bund futures, the graphs for ζ_q are slightly concave.

2.4 Other indices

Repeating the analysis of Section 2.3 for each index in the Oxford-Man dataset, we find the $m(q, \Delta)$ present a universal scaling behavior. For each index and for $q = 0.5, 1, 1.5, 2, 3$, by doing a linear regression of $\log(m(q, \Delta))$ on $\log(\Delta)$ for $\Delta = 1, \dots, 30$, we obtain estimates of ζ_q that we summarize in Table B.1 in the appendix.

2.5 Distribution of the increments of the log-volatility

Having established that all our underlying assets exhibit essentially the same scaling behavior⁷, we focus in the rest of the paper only on the S&P index, unless specified otherwise. That the distribution of increments of log-volatility is close to Gaussian is a well-established stylized fact reported for example in the papers [2] and [3] of Andersen et al. Looking now at the histograms of the increments of the log-volatility in Figure 2.7 with the fitted normal density superimposed in red, we see that, for any Δ , the empirical distributions of log-volatility increments are verified as being close to

⁷We have also verified that this scaling relationship holds for Crude Oil and Gold futures with similar smoothness estimates ζ_q .

Gaussian. More impressive still is that rescaling the 1-day fit of the normal density by Δ^H generates (blue dashed) curves that are very close to the red fits of the normal density, consistent with the observed scaling.

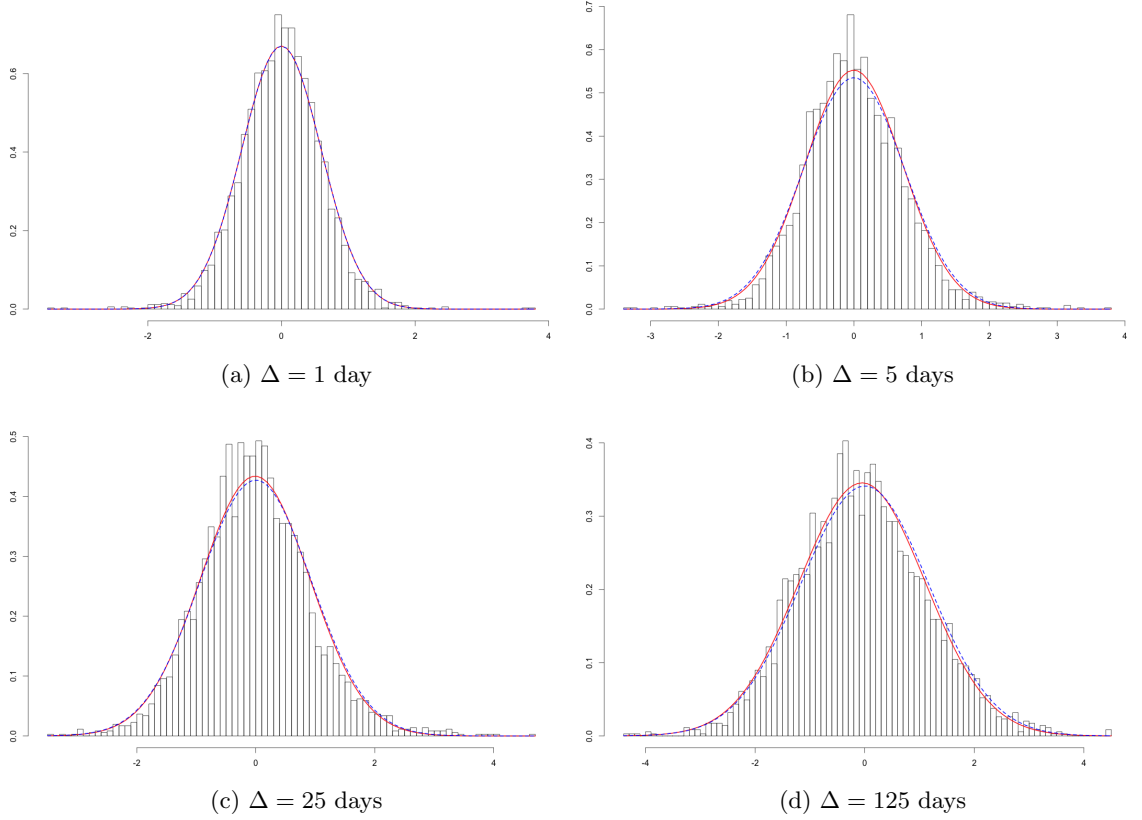


Figure 2.7: Histograms for various lags Δ of the (overlapping) increments $\log \sigma_{t+\Delta} - \log \sigma_t$ of the S&P log-volatility; normal fits in red; normal fit for $\Delta = 1$ day rescaled by Δ^H in blue.

The slight deviations from the Normal distribution observed in Figure 2.7 are again consistent with the computation of the empirical distribution of the increments of a fractional Brownian motion on a similar number of points.

2.6 Does H vary over time?

In order to check whether our estimations of H depends on the time interval, we split the Oxford-Man realized variance dataset into two halves and reestimate H for each half separately. The results are presented in Table B.2 in the appendix. We note that although the estimated H all lie between 0.06 and 0.20, they seem to be higher in the second period which includes

the financial crisis.

3 A simple model compatible with the empirical smoothness of the volatility

In this section, we specify the Rough FSV model and demonstrate that it reproduces the empirical facts presented in Section 2.

3.1 Specification of the RFSV model

In the previous section, we showed that, empirically, the increments of the log-volatility of various assets enjoy a scaling property with constant smoothness parameter and that their distribution is close to Gaussian. This naturally suggests the simple model:

$$\log \sigma_{t+\Delta} - \log \sigma_t = \nu (W_{t+\Delta}^H - W_t^H), \quad (3.1)$$

where W^H is a fractional Brownian motion with Hurst parameter equal to the measured smoothness of the volatility and ν is a positive constant. We may of course write (3.1) under the form

$$\sigma_t = \sigma \exp \{ \nu W_t^H \}, \quad (3.2)$$

where σ is another positive constant.

However this model is not stationary, stationarity being desirable both for mathematical tractability and also to ensure reasonableness of the model at very large times. This leads us to impose stationarity by modeling the log-volatility as a fractional Ornstein-Uhlenbeck process (fOU process for short) with a very long reversion time scale.

A stationary fOU process (X_t) is defined as the stationary solution of the stochastic differential equation

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

where $m \in \mathbb{R}$ and ν and α are positive parameters, see [12]. As for usual Ornstein-Uhlenbeck processes, there is an explicit form for the solution which is given by

$$X_t = \nu \int_{-\infty}^t e^{-\alpha(t-s)} dW_s^H + m. \quad (3.3)$$

Here the stochastic integral with respect to fBM is simply a pathwise Riemann-Stieltjes integral, see again [12].

We thus arrive at the final specification of our Rough Fractional Stochastic Volatility (RFSV) model for the volatility on the time interval of interest $[0, T]$:

$$\sigma_t = \exp \{X_t\}, \quad t \in [0, T], \quad (3.4)$$

where (X_t) satisfies Equation (3.3) for some $\nu > 0$, $\alpha > 0$, $m \in \mathbb{R}$ and $H < 1/2$ the measured smoothness of the volatility. Such a model is indeed stationary. However, if $\alpha \ll 1/T$, the log-volatility behaves locally (at time scales smaller than T) as a fBM. This observation is formalized in Proposition 3.1 below.

Proposition 3.1. *Let W^H be a fBM and X^α defined by (3.3) for a given $\alpha > 0$. As α tends to zero,*

$$\mathbb{E} \left[\sup_{t \in [0, T]} |X_t^\alpha - X_0^\alpha - \nu W_t^H| \right] \rightarrow 0.$$

The proof is given in Appendix A.1.

Proposition 3.1 implies that in the RFSV model, if $\alpha \ll 1/T$, and we confine ourselves to the interval $[0, T]$ of interest, we can proceed as if the the log-volatility process were a fBM. Indeed, simply setting $\alpha = 0$ in (3.3) gives (at least formally) $X_t - X_s = \nu(W_t^H - W_s^H)$ and we immediately recover our simple non-stationary fBM model (3.1).

The following corollary implies that the (exact) scaling property of the fBM is approximately reproduced by the fOU process when α is small.

Corollary 3.1. *Let $q > 0$, $t > 0$, $\Delta > 0$. As α tends to zero, we have*

$$\mathbb{E}[|X_{t+\Delta}^\alpha - X_t^\alpha|^q] \rightarrow \nu^q K_q \Delta^{qH}.$$

The proof is given in Appendix A.2.

RFSV versus FSV

We recognize our RFSV model (3.4) as a particular case of the classical FSV model of Comte and Renault [16]. The key difference is that here we take $H < 1/2$ and $\alpha \ll 1/T$, whereas to accommodate the assumption of long memory, Comte and Renault have to choose $H > 1/2$. The analysis of Fukasawa referred to earlier in Section 1.3 implies in particular that if $H > 1/2$, the volatility skew function $\psi(\tau)$ is *increasing* in time to expiration τ (at least for small τ), which is obviously completely inconsistent with the approximately $1/\sqrt{\tau}$ skew term structure that is observed. To generate a decreasing term structure of volatility skew for longer expirations, Comte and Renault are then forced to choose $\alpha \gg 1/T$. Consequently, for very short expirations ($\tau \ll 1/\alpha$), models of the Comte and Renault type with

$H > 1/2$ still generate a term structure of volatility skew that is inconsistent with the observed one, as explained for example in Section 4 of [15].

In contrast, the choice $H < 1/2$ enables us to reproduce both the observed smoothness of the volatility process and generate a term structure of volatility skew in agreement with the observed one. The choice $H < 1/2$ is also consistent with what is improperly called mean reversion by practitioners, which is the fact that if volatility is unusually high, it tends to decline and if it is unusually low, it tends to increase. Finally, taking α very small implies that the dynamics of our process is close to that of a fBM, see Proposition 3.1. This last point is particularly important. Indeed, recall that at the time scales we are interested in, the important feature we have in mind is really this fBM like-behavior of the log-volatility.

We could no doubt have considered other stationary models satisfying Proposition 3.1 and Corollary 3.1, where log-volatility behaves as a fBM at reasonable time scales; the choice of the fOU process is probably the simplest way to accommodate this local behavior together with the stationarity property.

3.2 RFSV model autocovariance functions

From Proposition 3.1 and Corollary 3.1, we easily deduce the following corollary, where $o(1)$ tends to zero as α tends to zero.

Corollary 3.2. *Let $q > 0$, $t > 0$, $\Delta > 0$. As α tends to zero,*

$$\text{Cov}[X_t^\alpha, X_{t+\Delta}^\alpha] = \text{Var}[X_t^\alpha] - \frac{1}{2} \nu^2 \Delta^{2H} + o(1).$$

Consequently, in the RFSV model, for fixed t , the covariance between X_t and $X_{t+\Delta}$ is linear with respect to Δ^{2H} . This result is very well satisfied empirically. For example, in Figure 3.1, we see that for the S&P, the empirical autocovariance function of the log-volatility is indeed linear with respect to Δ^{2H} . Note in passing that at the time scales we consider, the term $\text{Var}[X_t^\alpha]$ is higher than $\frac{1}{2} \nu^2 \Delta^{2H}$ in the expression for $\text{Cov}[X_t^\alpha, X_{t+\Delta}^\alpha]$.

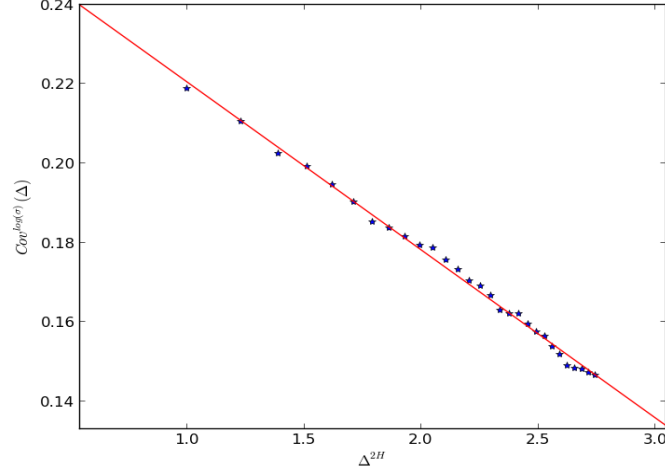


Figure 3.1: Autocovariance of the log-volatility as a function of Δ^{2H} for $H = 0.14$, S&P.

Thanks to [12], we even have an exact formula for the autocovariance function of the log-volatility in the RFSV model:

$$\begin{aligned} \text{Cov}[\log \sigma_t, \log \sigma_{t+\Delta}] &= \frac{H(2H-1)\nu^2}{2\alpha^{2H}} \left\{ e^{-\alpha\Delta} \Gamma(2H-1) \right. \\ &\quad \left. + e^{-\alpha\Delta} \int_0^{\alpha\Delta} \frac{e^u}{u^{2-2H}} du + e^{\alpha\Delta} \int_{\alpha\Delta}^{\infty} \frac{e^{-u}}{u^{2-2H}} du \right\} \end{aligned} \quad (3.5)$$

and

$$\text{Var}[\log \sigma_t] = \frac{H(2H-1)\nu^2}{\alpha^{2H}} \Gamma(2H-1),$$

where Γ denotes the Gamma function.

Having computed the autocovariance function of the log-volatility, we now turn our attention to the volatility itself. We have

$$\mathbb{E}[\sigma_{t+\Delta}\sigma_t] = \mathbb{E}[e^{X_t^\alpha + X_{t+\Delta}^\alpha}],$$

with X^α defined by Equation (3.3). Since X^α is a Gaussian process, we deduce that

$$\mathbb{E}[\sigma_{t+\Delta}\sigma_t] = e^{\mathbb{E}[X_t^\alpha] + \mathbb{E}[X_{t+\Delta}^\alpha] + \text{Var}[X_t^\alpha]/2 + \text{Var}[X_{t+\Delta}^\alpha]/2 + \text{Cov}[X_t^\alpha, X_{t+\Delta}^\alpha]}.$$

Applying Corollary 3.2, we obtain that when α is small, $\mathbb{E}[\sigma_{t+\Delta}\sigma_t]$ is approximately equal to

$$e^{2\mathbb{E}[X_t^\alpha] + 2\text{Var}[X_t^\alpha]} e^{-\nu^2 \frac{\Delta^{2H}}{2}}. \quad (3.6)$$

It follows that in the RFSV model, $\log(\mathbb{E}[\sigma_{t+\Delta}\sigma_t])$ is also linear in Δ^{2H} . This property is again very well satisfied on data, as shown by Figure 3.2, where we plot the logarithm of the empirical counterpart of $\mathbb{E}[\sigma_{t+\Delta}\sigma_t]$ against Δ^{2H} , for the S&P with $H = 0.14$.

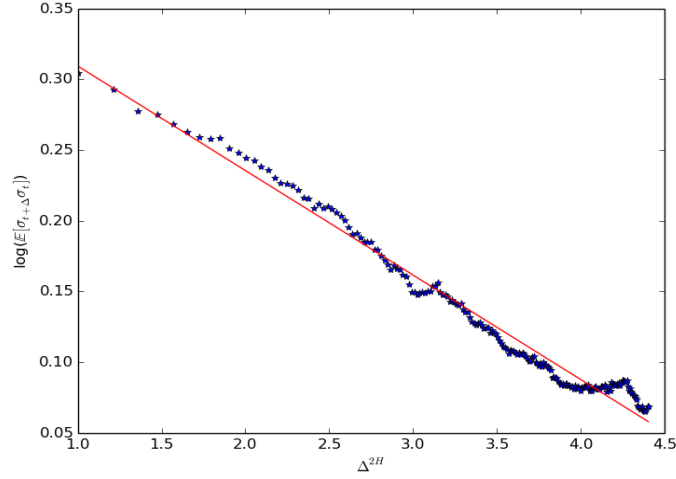


Figure 3.2: Empirical counterpart of $\log(\mathbb{E}[\sigma_{t+\Delta}\sigma_t])$ as a function of Δ^{2H} , S&P.

We note that putting Δ^{2H} on the x-axis of Figure 3.2 is really crucial in order to retrieve linearity. In particular, a corollary of (3.6) is that the autocovariance function of the volatility does not decay as a power law as widely believed; see Figure 3.3 where we show that a log-log plot of the autocovariance function does not yield a straight line.

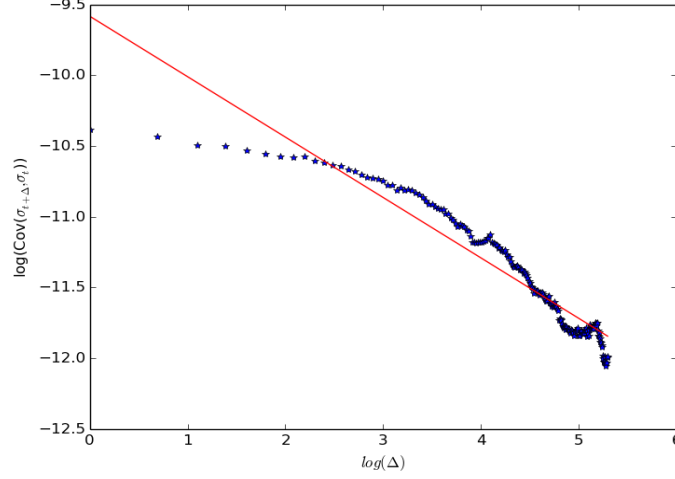


Figure 3.3: Empirical counterpart of $\log(\text{Cov}[\sigma_{t+\Delta}, \sigma_t])$ as a function of $\log(\Delta)$, S&P.

3.3 RFSV versus FSV again

To further demonstrate the incompatibility of the classical long memory FSV model with volatility data, consider the quantity $m(2, \Delta)$. Recall that in the data (see Section 2) we observe the linear relationship $\log m(2, \Delta) \approx \zeta_2 \log \Delta + k$ for some constant k . Also, in both FSV and RFSV, we can consider

$$\begin{aligned} m(2, \Delta) &= \mathbb{E} [(\log \sigma_{t+\Delta} - \log \sigma_t)^2] \\ &= 2 (\text{Var}[\log \sigma_t] - \text{Cov}[\log \sigma_t, \log \sigma_{t+\Delta}]). \end{aligned}$$

Therefore, using Equation (3.5), we have a closed form formula for $m(2, \Delta)$.

In Figure 3.4, we plot $m(2, \Delta)$ with the parameters $H = 0.53$, corresponding to the FSV model parameter estimate of Chronopoulou and Viens in [14], and $\alpha = 0.5$ to ensure some visible decay of the volatility skew. The slope of $m(2, \Delta)$ in the FSV model for small lags is driven by the value of H ; the lag at which $m(2, \Delta)$ begins to flatten and stationarity kicks in corresponds to a time scale of order $1/\alpha$. It is clear from the picture that to fit the data, we must have $\alpha \ll 1/T$ and the value of H must be set by the initial slope of the regression line, which as reported earlier in Section 2 is $\zeta_2 = 2 \times 0.14$.

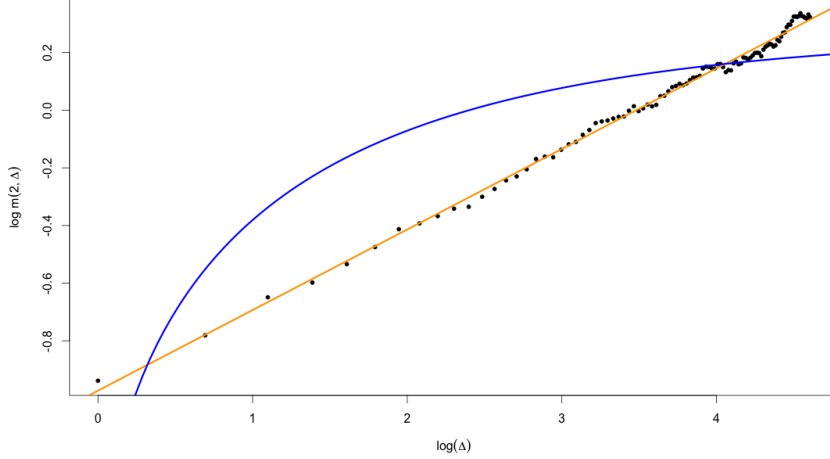


Figure 3.4: Long memory models such as the FSV model of Comte and Renault are not compatible with S&P volatility data. Black points are empirical estimates of $m(2, \Delta)$; the blue line is the FSV model with $\alpha = 0.5$ and $H = 0.53$; the orange line is the RFSV model with $\alpha = 0$ and $H = 0.14$.

3.4 Simulation-based analysis of the RFSV model

Our goal in this section is to show that in terms of smoothness measures, one obtains on simulated data from the RFSV model the same behaviors as those observed on empirical data. In particular, we would like to be able to quantify the positive bias associated with estimating H from whole-day realized variance data as in Section 2.3 relative to using data from a one-hour window as in Section 2.2.

We simulate the RFSV model for 2,000 days (chosen to be between the lengths of our two datasets). In order to account for the overnight effect, we simulate the volatility σ_t ⁸ and efficient price P_t ⁹ over the whole day. The parameters: $H = 0.14$, $\nu = 0.3$, $m = X_0 = -5$ and $\alpha = 5 \times 10^{-4}$, are chosen to be consistent with our empirical estimates from Section 2. To model microstructure effects such as the discreteness of the price grid, we consider that the observed price process is generated from P_t using the uncertainty zones model of [42] with tick value 5×10^{-4} and parameter $\eta = 0.25$.

Exactly as in Section 2, for each of the 2,000 days, we consider two volatility

⁸To simulate the fBM, we use a spectral method with 40,000,000 points (20,000 points per day). We then simulate X taking $X_{(n+1)\delta} - X_{n\delta} = \nu(W_{(n+1)\delta}^H - W_{n\delta}^H) + \alpha\delta(m - X_{n\delta})$ (with $\delta = 1/20000$).

⁹ $P_{(n+1)\delta} - P_{n\delta} = P_{n\delta}\sigma_{n\delta}\sqrt{\delta}U_n$ where the U_n are iid standard Gaussian variables.

proxies obtained from the observed price and based on:

- The integrated variance estimator using the model with uncertainty zones over one hour windows, from 10 am to 11 am.
- The 5 minutes realized variance estimator, over eight hours windows (the trading day).

We now repeat our analysis of Section 2, generating graphs analogous to Figures 2.1, 2.2, 2.4 and 2.5 obtained on empirical data. Figure 3.5 compares smoothness measures obtained using the uncertainty zones estimator on one-hour windows with those obtained using the realized variance estimator on 8-hour windows.

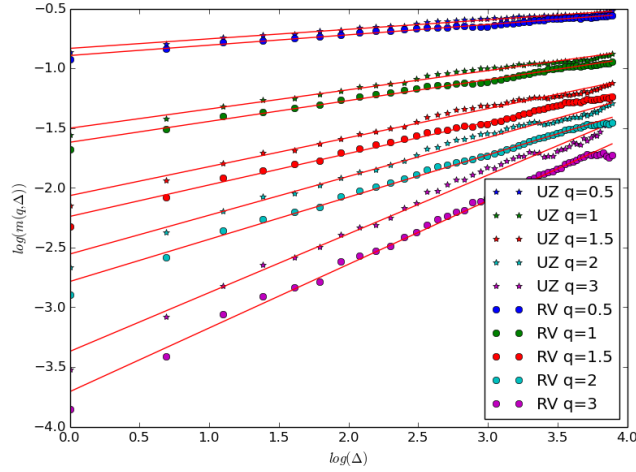


Figure 3.5: $\log(m(q, \Delta))$ as a function of $\log(\Delta)$, simulated data, with realized variance and uncertainty zones estimators.

When the uncertainty zones estimator is applied on a one-hour window ($1/24$ of a simulated day) as in Section 2.2, we estimate $H = 0.16$, which is close to the true value $H = 0.14$ used in the simulation. The results obtained with the realized variance estimator over daily eight-hour windows ($1/3$ of a simulated day) do exhibit the same scaling properties that we see in the empirical data with a smoothness parameter that does not depend on q . However, the estimated H is biased slightly higher at around 0.18. As discussed in Section 2.1, this extra positive bias is no surprise and is due to the regularizing effect of the integral operator over the longer window. We note also that the estimated values of ν (“volatility of volatility” in some sense) obtained from the intercepts of the regressions, are lower with the longer time windows, again as expected. A detailed computation of the bias in the

estimated H associated with the choice of window length in an analogous but more tractable model is presented in Appendix C.

We end this section by presenting in Figure 3.6 a sample path of the model-generated volatility (spot volatility direct from the simulation rather than estimated from the simulated price series) together with a graph of S&P volatility over 3,500 days.

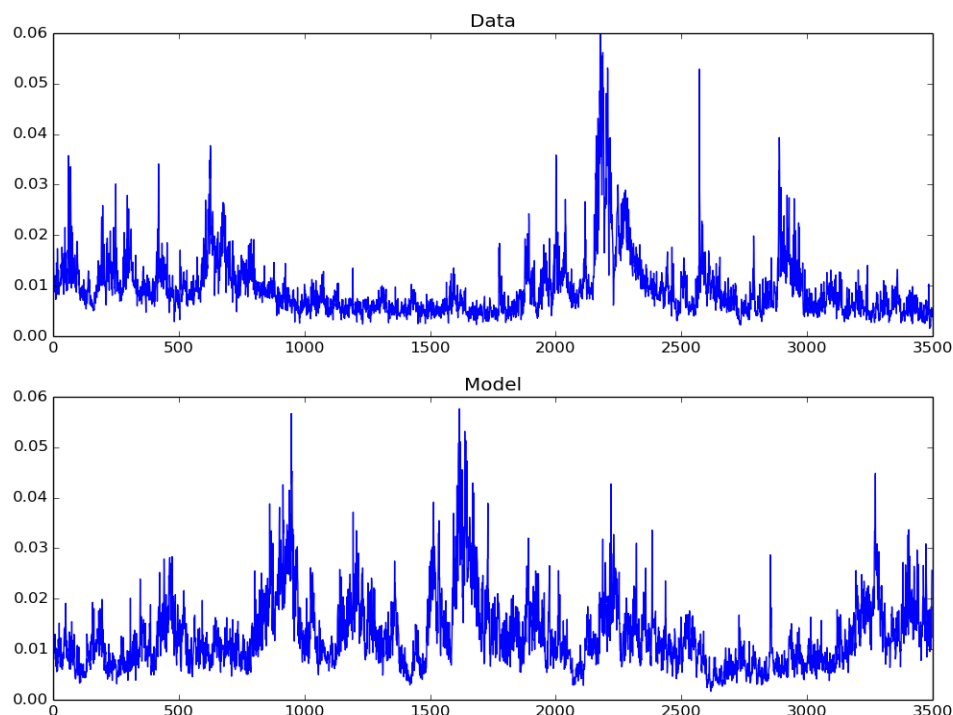


Figure 3.6: Volatility of the S&P (above) and of the model (below).

A first reaction to Figure 3.6 is that the simulated and actual graphs look very alike. In particular, in both of them, persistent periods of high volatility alternate with low volatility periods. On closer inspection of the empirical volatility series, we observe that the sample path of the volatility on a restricted time window seems to exhibit the same kind of qualitative properties as those of the global sample path (for example periods of high and low activity). This fractal-type behavior of the volatility has been investigated both empirically and theoretically in, for example, [5, 9, 37].

At the visual level, we observe that this fractal-type behavior is also reproduced in our model, as we now explain. Denote by $L^{x,H}$ the law of the geometric fractional Brownian motion with Hurst exponent H and volatility x on $[0, 1]$, that is $(e^{xW_t^H})_{t \in [0,1]}$. Then, when α is very small, the rescaled volatility process on $[0, \Delta]$: $(\sigma_{t\Delta}/\sigma_0)_{t \in [0,1]}$, has approximately the law $L^{\nu\Delta^H, H}$. Now remark that for H small, the function u^H increases very slowly. Thus, over a large range of observation scales Δ , the rescaled volatility processes on $[0, \Delta]$ have approximately the same law. For example, between an observation scale of one day and five years (1250 open days), the coefficient x characterizing the law of the volatility process is “only” multiplied by $1250^{0.14} = 2.7$. It follows that in the RFSV model, the volatility process over one day resembles the volatility process over a decade.

4 Spurious long memory of volatility?

We revisit in this section the issue of long memory of volatility through the lens of our model. As mentioned earlier in the introduction, the long memory of volatility is widely accepted as a stylized fact. Specifically, this means that the autocovariance function $\text{Cov}[\log(\sigma_t), \log(\sigma_{t+\Delta})]$ (or sometimes $\text{Cov}[\sigma_t, \sigma_{t+\Delta}]$) goes slowly to zero as $\Delta \rightarrow \infty$ and often even more precisely, that it behaves as $\Delta^{-\gamma}$, with $\gamma < 1$ as $\Delta \rightarrow \infty$.

In previous sections, we showed that both in the data and in our model,

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

and

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

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 memory¹⁰, see again Figure 3.3.

We now revisit some standard statistical procedures aimed at identifying long memory that have been used in the financial econometrics literature. In the sequel, we apply these both to the data and to sample paths of the RFSV model. Such procedures are of course designed to identify long memory under rather strict modeling assumptions; spurious results may obviously then be obtained if the model underlying the estimation procedure

¹⁰In fact the notion of empirical long memory does not make much sense outside the power law case. Indeed the empirical values of covariances at very large time scales are never measurable and thus one cannot conclude if the series of covariances converges in general. All that we say here is that the autocovariance of the (log-)volatility does not behave as a power law.

is misspecified .

With the same model parameters as in Section 3.4, we simulate our model over 3,500 days, which corresponds to the size of our dataset. Consider first the procedure in [3], where the authors test for long memory in the volatility by studying the scaling behavior of the quantity

$$V(t) = \text{Var} \left[\int_0^t \sigma_s^2 ds \right]$$

with respect to t . In the model they consider, if $V(t)$ behaves asymptotically as $t^{2-\gamma}$ with $\gamma < 1$, then the autocorrelation function of the log-volatility should behave as $t^{-\gamma}$. Figure 4.1 presents the graph of the logarithm of the empirical counterpart of $V(t)$ against the logarithm of t , on the S&P data and within our simulation framework.

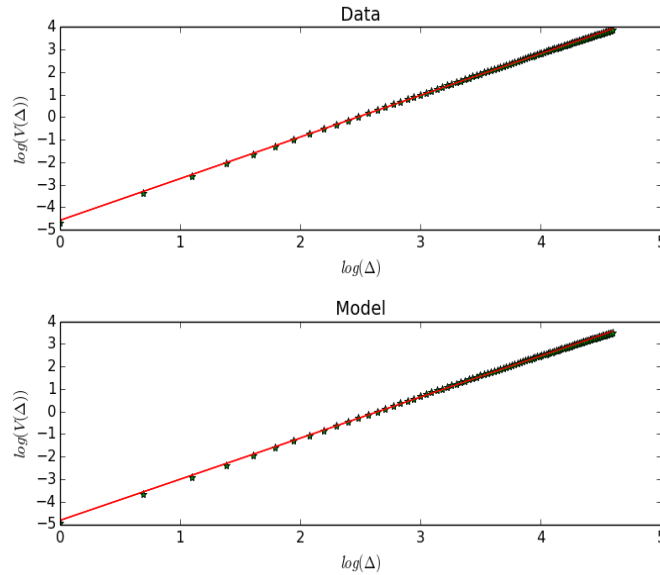


Figure 4.1: Empirical counterpart of $\log(V(t))$ as a function of $\log(t)$ on S&P (above) and simulation (below).

We note from Figure 4.1 that both our simulated model and market data lead to very similar graphs, close to straight lines with slope 1.86. Accordingly, in the setting of [3], we would deduce power law behavior of the autocorrelation function with exponent 0.14 and therefore long memory. Thus, if the data are generated by a model like the RFSV model, one can easily be wrongly convinced that the volatility time series exhibits long memory.

In [4], the authors deduce long memory in the volatility by showing that the process ε_t obtained by fractional differentiation of the log-volatility $\varepsilon_t = (1 - L)^d \log(\sigma_t)$, with $d = 0.4$ (which is considered as a reasonable value) and L the lag operator, behaves as a white noise. To check for this, they simply compute the autocorrelation function of ε_t . We give in Figure 4.2 the autocorrelation functions of the logarithm of σ_t and ε_t , again both on the data and on the simulated path.

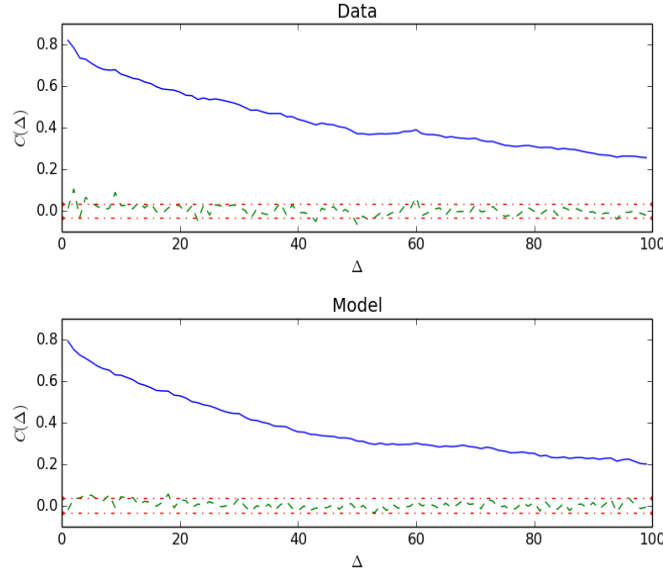


Figure 4.2: Autocorrelation functions of $\log(\sigma_t)$ (in blue) and ε_t (in green) and the Bartlett standard error bands (in red), for S&P data (above) and for simulated data (below).

Once again, the data and the simulation generate very similar plots. We conclude that this procedure for estimating long memory is just as fragile as the first, and it is easy to wrongly deduce volatility long memory when applying it.

In conclusion, it seems that classical estimation procedures identify spurious long memory of volatility in the RFSV model. Moreover, these procedures estimate the same long memory parameter from data generated from a suitably calibrated RFSV model as they estimate from empirical data. Once again, our conclusion is that although the (log-)volatility may exhibit some form of persistence, it does not present any long memory in the classical power law sense.

5 Forecasting using the RFSV model

In this section, we present an application of our model: forecasting the log-volatility and the variance.

5.1 Forecasting log-volatility

The key formula on which our prediction method is based is the following one:

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

where W^H is a fBM with $H < 1/2$ and \mathcal{F}_t the filtration it generates, see Theorem 4.2 of [41]. By construction, over any reasonable time scale of interest, as formalized in Corollary 3.1, we may approximate the fOU volatility process in the RFSV model as $\log \sigma_t^2 \approx 2\nu W_t^H + C$ for some constants ν and C . Our prediction formula for log-variance then follows:¹¹

$$\mathbb{E}[\log \sigma_{t+\Delta}^2 | \mathcal{F}_t] = \frac{\cos(H\pi)}{\pi} \Delta^{H+1/2} \int_{-\infty}^t \frac{\log \sigma_s^2}{(t-s+\Delta)(t-s)^{H+1/2}} ds. \quad (5.1)$$

This formula, or rather its approximation through a Riemann sum (we assume in this section that the volatilities are perfectly observed, although they are in fact estimated), is used to forecast the log-volatility 1, 5 and 20 days ahead ($\Delta = 1, 5, 20$).

We now compare the predictive power of formula (5.1) with that of AR and HAR forecasts, in the spirit of [18]¹². Recall that for a given integer $p > 0$, the AR(p) and HAR predictors take the following form (where the index i runs over the series of daily volatility estimates):

- AR(p):

$$\widehat{\log(\sigma_{t+\Delta}^2)} = K_0^\Delta + \sum_{i=0}^p C_i^\Delta \log(\sigma_{t-i}^2).$$

- HAR :

$$\widehat{\log(\sigma_{t+\Delta}^2)} = K_0^\Delta + C_0^\Delta \log(\sigma_t^2) + C_5^\Delta \frac{1}{5} \sum_{i=0}^5 \log(\sigma_{t-i}^2) + C_{20}^\Delta \frac{1}{20} \sum_{i=0}^{20} \log(\sigma_{t-i}^2).$$

¹¹The constants 2ν and C cancel when deriving the expression.

¹² Note that we do not consider GARCH models here since we have access to high frequency volatility estimates and not only to daily returns. Indeed, it is shown in [4] that forecasts based on the time series of realized variance outperform GARCH forecasts based on daily returns.

We estimate AR coefficients using the R `stats` library¹³ on a rolling time window of 500 days. In the HAR case, we use standard linear regression to estimate the coefficients as explained in [18]. In the sequel, we consider $p = 5$ and $p = 10$ in the AR formula. Indeed, these parameters essentially give the best results for the horizons at which we wish to forecast the volatility (1, 5 and 20 days). For each day, we forecast volatility for five different indices¹⁴.

We then assess the quality of the various forecasts by computing the ratio P between the mean squared error of our predictor and the (approximate) variance of the log-variance:

$$P = \frac{\sum_{k=500}^{N-\Delta} \left(\log(\sigma_{k+\Delta}^2) - \widehat{\log(\sigma_{k+\Delta}^2)} \right)^2}{\sum_{k=500}^{N-\Delta} \left(\log(\sigma_{k+\Delta}^2) - \mathbb{E}[\log(\sigma_{t+\Delta}^2)] \right)^2},$$

where $\mathbb{E}[\log(\sigma_{t+\Delta}^2)]$ denotes the empirical mean of the log-variance over the whole time period.

	AR(5)	AR(10)	HAR(3)	RFSV
SPX2.rv $\Delta = 1$	0.317	0.318	0.314	0.313
SPX2.rv $\Delta = 5$	0.459	0.449	0.437	0.426
SPX2.rv $\Delta = 20$	0.764	0.694	0.656	0.606
FTSE2.rv $\Delta = 1$	0.230	0.229	0.225	0.223
FTSE2.rv $\Delta = 5$	0.357	0.344	0.337	0.320
FTSE2.rv $\Delta = 20$	0.651	0.571	0.541	0.472
N2252.rv $\Delta = 1$	0.357	0.358	0.351	0.345
N2252.rv $\Delta = 5$	0.553	0.533	0.513	0.504
N2252.rv $\Delta = 20$	0.875	0.795	0.746	0.714
GDAXI2.rv $\Delta = 1$	0.237	0.238	0.234	0.231
GDAXI2.rv $\Delta = 5$	0.372	0.362	0.350	0.339
GDAXI2.rv $\Delta = 20$	0.661	0.590	0.550	0.498
FCHI2.rv $\Delta = 1$	0.244	0.244	0.241	0.238
FCHI2.rv $\Delta = 5$	0.378	0.373	0.366	0.350
FCHI2.rv $\Delta = 20$	0.669	0.613	0.598	0.522

Table 5.1: Ratio P for the AR, HAR and RFSV predictors.

We note from Table 5.1 that the RFSV forecast consistently outperforms the AR and HAR forecasts, especially at longer horizons. Moreover, our forecasting method is more parsimonious since it only requires the parameter

¹³More precisely, we use the default Yule-Walker method.

¹⁴In addition to S&P and NASDAQ, we also investigate CAC40, FTSE and Nikkei, over the same time period as S&P and NASDAQ. For simplicity, the parameter H used in our predictor is computed only once for each asset, using the whole time period. This yields similar results to using a moving time window adapted in time.

H to forecast the log-variance. Compare this with the AR and HAR methods, for which coefficients depend on the forecast time horizon and must be recomputed if this horizon changes.

Remark that our predictor can be linked to that of [21], where the issue of the prediction of the log-volatility in the multifractal random walk model of [5] is tackled. In this model,

$$\mathbb{E}[\log(\sigma_{t+\Delta}^2)|\mathcal{F}_t] = \frac{1}{\pi} \sqrt{\Delta} \int_{-\infty}^t \frac{\log(\sigma_s^2)}{(t-s+\Delta)\sqrt{t-s}} ds,$$

which is the limit of our predictor when H tends to zero.

Note also that our prediction formula may be rewritten as

$$\mathbb{E}[\log(\sigma_{t+\Delta}^2)|\mathcal{F}_t] = \frac{\cos(H\pi)}{\pi} \int_0^{+\infty} \frac{\log(\sigma_{t-\Delta u}^2)}{(u+1)u^{H+1/2}} du.$$

For a given small $\varepsilon > 0$, let r be the smallest real number such that

$$\int_r^{+\infty} \frac{1}{(u+1)u^{H+1/2}} du \leq \varepsilon.$$

Then we have, with an error of order ε ,

$$\mathbb{E}[\log(\sigma_{t+\Delta}^2)|\mathcal{F}_t] \approx \frac{\cos(H\pi)}{\pi} \int_0^r \frac{\log(\sigma_{t-\Delta u}^2)}{(u+1)u^{H+1/2}} du.$$

Consequently, the volatility process needs to be considered (roughly) down to time $t - \Delta r$ if one wants to forecast up to time Δ in the future. The relevant regression window is thus linear in the forecasting horizon. For example, for $r = 1$, $\varepsilon = 0.35$ which is not so unreasonable. In this case, as is well-known to practitioners, to predict volatility one week ahead, one should essentially look at the volatility over the last week. If trying to predict the volatility one month ahead, one should look at the volatility over the last month.

5.2 Variance prediction

Recall that $\log \sigma_t^2 \approx 2\nu W_t^H + C$ for some constant C . In [41], it is shown that $W_{t+\Delta}^H$ is conditionally Gaussian with conditional variance

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

with

$$c = \frac{\Gamma(3/2 - H)}{\Gamma(H + 1/2) \Gamma(2 - 2H)}.$$

Thus, we obtain the following natural form for the RFSV predictor of the variance:

$$\widehat{\sigma_{t+\Delta}^2} = \exp \left\{ \log \widehat{\sigma_{t+\Delta}^2} + 2 c \nu^2 \Delta^{2H} \right\}$$

where $\log(\widehat{\sigma_{t+\Delta}^2})$ is the estimator from Section 5.1 and ν^2 is estimated as the exponential of the intercept in the linear regression of $\log(m(2, \Delta))$ on $\log(\Delta)$.

As in the previous paragraph, we compare in Table 5.2 the performance of the RFSV forecast with those of AR and HAR forecasts (constructed on variance rather than log-variance this time).

	AR(5)	AR(10)	HAR(3)	RFSV
SPX2.rv $\Delta = 1$	0.520	0.566	0.489	0.475
SPX2.rv $\Delta = 5$	0.750	0.745	0.723	0.672
SPX2.rv $\Delta = 20$	1.070	1.010	1.036	0.903
FTSE2.rv $\Delta = 1$	0.612	0.621	0.582	0.567
FTSE2.rv $\Delta = 5$	0.797	0.770	0.756	0.707
FTSE2.rv $\Delta = 20$	1.046	0.984	0.935	0.874
N2252.rv $\Delta = 1$	0.554	0.579	0.504	0.505
N2252.rv $\Delta = 5$	0.857	0.807	0.761	0.729
N2252.rv $\Delta = 20$	1.097	1.046	1.011	0.964
GDAXI2.rv $\Delta = 1$	0.439	0.448	0.399	0.386
GDAXI2.rv $\Delta = 5$	0.675	0.650	0.616	0.566
GDAXI2.rv $\Delta = 20$	0.931	0.850	0.816	0.746
FCHI2.rv $\Delta = 1$	0.533	0.542	0.470	0.465
FCHI2.rv $\Delta = 5$	0.705	0.707	0.691	0.631
FCHI2.rv $\Delta = 20$	0.982	0.952	0.912	0.828

Table 5.2: Ratio P for the AR, HAR and RFSV predictors.

We find again that the RFSV forecast typically outperforms HAR and AR, although it is worth noting that the HAR forecast is already visibly superior to the AR forecast.

6 The microstructural foundations of the irregularity of the volatility

We gather in this section some ideas which may help to understand why the observed volatility appears so irregular. The starting point is the analysis of the order flow through Hawkes processes. These processes are extensions of Poisson processes where the intensity at a given time depends on the

location of the past jumps. More precisely, let us consider a time period starting at 0 and denote by N_t the number of transactions between 0 and t . Assuming the point process N_t follows a Hawkes process means its intensity at time t , λ_t , takes the form:

$$\lambda_t = \mu + \sum_{0 < J_i < t} \phi(t - J_i),$$

where the J_i are the past jump times, μ is a positive constant and ϕ is a non negative deterministic function called kernel.

When trying to calibrate such models on high frequency data, two main phenomena almost systematically occur:

- The L^1 norm of ϕ is close to one, see [23, 24, 30, 35].
- The function ϕ has a power law tail, see [6, 30].

The first of these two facts means the degree of endogeneity of the market is very high, that is one given order endogenously generates many other orders, see [23, 24, 30]. This recent feature of financial markets is obviously related to electronic high frequency trading, where market participants automatically react to other participants orders through their algorithms. The second observation tells us that generally, a given order influences other orders over a long time period. This is likely due to the splitting of large orders. Indeed, many orders are actually part of a metaorder whose full execution can take a large amount of time.

We believe these two phenomena together lead to a superposition effect inducing this irregular volatility. Indeed, it is explained in [33, 34] that the macroscopic scaling limit of Hawkes processes with power law tail and kernel with L^1 norm close to one can be seen as an integrated fractional process, with Hurst parameter H smaller than $1/2$. This signifies that at large sampling scales, the dynamics of the cumulated order flow is well approximated by an integrated fractional process, with $H < 1/2$. Then, it is clearly established that there is a linear relation between cumulated order flow and integrated variance. Thus we retrieve here that because of this superposition effect, the volatility should behave as a fractional process with $H < 1/2$.

7 Conclusion

Using daily realized variance estimates as proxies for daily spot (squared) volatilities, we uncovered two startlingly simple regularities in the resulting

time series. First we found that the distributions of increments of log-volatility are approximately Gaussian, consistent with many prior studies. Secondly, we established the monofractal scaling relationship

$$\mathbb{E}[|\log(\sigma_\Delta) - \log(\sigma_0)|^q] = K_q \nu^q \Delta^{qH}, \quad (7.1)$$

where H can be seen as a measure of smoothness characteristic of the underlying volatility process; typically, $0.06 < H < 0.2$. The simple scaling relationship (7.1) naturally suggests that log-volatility may be modeled using fractional Brownian motion.

The resulting Rough Fractional Stochastic Volatility (RFSV) model turns out to be formally almost identical to the FSV model of Comte and Renault [16], with one major difference: In the FSV model, $H > 1/2$ to ensure long memory whereas in the RFSV model $H < 1/2$, typically, $H \approx 0.1$. Moreover, in the FSV model, the mean reversion coefficient α has to be large compared to $1/T$ to ensure a decaying volatility skew; in the RFSV model, the volatility skew decays naturally just like the observed volatility skew, $\alpha \ll 1/T$ and indeed for time scales of practical interest, we may proceed as if α were exactly zero.

We further showed that applying standard statistical estimators to volatility time series simulated with the RFSV model would lead us to erroneously deduce the presence of long memory, with parameters similar to those found in prior studies. Despite that volatility in the RFSV model (or in the data) is not long memory, we can therefore explain why long memory of volatility is widely accepted as a stylized fact.

As an application of the RFSV model, we showed how to forecast volatility at various times cales, at least as well as Fulvio Corsi's impressive HAR estimator, but with only one parameter – H !

Finally, we explained how the RFSV model could emerge as the scaling limit of a Hawkes process description of order flow.

In future work, we will explore the implications of the RFSV model (written under the physical measure \mathbb{P}), for option pricing (under the pricing measure \mathbb{Q}). In particular, following Mandelbrot and Van Ness, the fBM that appears in the definition (3.4) of the RFSV model may be represented as a fractional integral of a standard Brownian motion as follows [36]:

$$W_t^H = \int_0^t \frac{dW_s}{(t-s)^\gamma} + \int_{-\infty}^0 \left[\frac{1}{(t-s)^\gamma} - \frac{1}{(-s)^\gamma} \right] dW_s, \quad (7.2)$$

with $\gamma = \frac{1}{2} - H$. The observed anticorrelation between price moves and volatility moves may then be modeled naturally by anticorrelating the Brow-

nian motion W that drives the volatility process with the Brownian motion driving the price process. As already shown by Fukasawa [25], such a model with a small H reproduces the observed decay of at-the-money volatility skew with respect to time to expiry, asymptotically for short times. We will show that an appropriate extension of Fukasawa's model, consistent with the RFSV model, fits the entire implied volatility surface remarkably well, not just for short expirations. Moreover, despite that it would seem from (7.2) that knowledge of the entire path $\{W_s : s < t\}$ of the Brownian motion would be required, it turns out that the statistics of this path necessary for option pricing are traded and thus easily observed.

A Proofs

A.1 Proof of Proposition 3.1

Starting from Equation (3.3) and applying integration by parts, we get

$$X_t^\alpha = \nu W_t^H - \int_{-\infty}^t \nu \alpha e^{-\alpha(t-s)} W_s^H ds + m.$$

Therefore,

$$(X_t^\alpha - X_0^\alpha) - \nu W_t^H = - \int_0^t \nu \alpha e^{-\alpha(t-s)} W_s^H ds - \int_{-\infty}^0 \nu \alpha (e^{-\alpha(t-s)} - e^{\alpha s}) W_s^H ds.$$

Consequently,

$$\sup_{t \in [0, T]} |(X_t^\alpha - X_0^\alpha) - \nu W_t^H| \leq \nu \alpha T \hat{W}_T^H + \int_{-\infty}^0 \nu \alpha (e^{\alpha s} - e^{-\alpha(T-s)}) \hat{W}_s^H ds,$$

where $\hat{W}_t^H = \sup_{s \in [0, t]} |W_s^H|$. Using the maximum inequality of [40], we get

$$\mathbb{E} \left[\sup_{t \in [0, T]} |(X_t^\alpha - X_0^\alpha) - \nu W_t^H| \right] \leq c(\nu \alpha T T^H + \int_{-\infty}^0 \nu \alpha (T \alpha e^{\alpha s}) |s|^H ds),$$

with c some constant. The term on the right hand side is easily seen to go to zero as α tends to zero.

A.2 Proof of Corollary 3.1

We first recall Equation (2.2) in [12] which writes:

$$\text{Cov}[X_{t+\Delta}^\alpha, X_t^\alpha] = K \int_{\mathbb{R}} e^{i\Delta x} \frac{|x|^{1-2H}}{\alpha^2 + x^2} dx,$$

with $K = \nu^2 \Gamma(2H + 1) \sin(\pi H) / (2\pi)$ ¹⁵. Now remark that

$$\mathbb{E}[(X_{t+\Delta}^\alpha - X_t^\alpha)^2] = 2\text{Var}[X_t^\alpha] - 2\text{Cov}[X_{t+\Delta}^\alpha, X_t^\alpha].$$

Therefore,

$$\mathbb{E}[(X_{t+\Delta}^\alpha - X_t^\alpha)^2] = 2K \int_{\mathbb{R}} (1 - e^{i\Delta x}) \frac{|x|^{1-2H}}{\alpha^2 + x^2} dx.$$

This implies that for fixed Δ , $\mathbb{E}[|X_{t+\Delta}^\alpha - X_t^\alpha|^2]$ is uniformly bounded by

$$2K \int_{\mathbb{R}} (1 - e^{i\Delta x}) \frac{|x|^{1-2H}}{x^2} dx.$$

Moreover, $X_{t+\Delta}^\alpha - X_t^\alpha$ is a Gaussian random variable and thus for every q , its $(q + 1)^{th}$ moment is uniformly bounded (in α) so that the family $|X_{t+\Delta}^\alpha - X_t^\alpha|^q$ is uniformly integrable. Therefore, since by Proposition 3.1,

$$|X_{t+\Delta}^\alpha - X_t^\alpha|^q \rightarrow \nu^q |W_{t+\Delta}^H - W_t^H|^q, \text{ in law,}$$

we get the convergence of the sequence of expectations.

¹⁵This covariance is real because it is the Fourier transform of an even function.

B Estimations of H

B.1 On different indices

Index	$\zeta_{0.5/0.5}$	ζ_1	$\zeta_{1.5/1.5}$	$\zeta_2/2$	$\zeta_3/3$
SPX2.rv	0.128	0.126	0.125	0.124	0.124
FTSE2.rv	0.132	0.132	0.132	0.131	0.127
N2252.rv	0.131	0.131	0.132	0.132	0.133
GDAXI2.rv	0.141	0.139	0.138	0.136	0.132
RUT2.rv	0.117	0.115	0.113	0.111	0.108
AORD2.rv	0.072	0.073	0.074	0.075	0.077
DJI2.rv	0.117	0.116	0.115	0.114	0.113
IXIC2.rv	0.131	0.133	0.134	0.135	0.137
FCHI2.rv	0.143	0.143	0.142	0.141	0.138
HSI2.rv	0.079	0.079	0.079	0.080	0.082
KS11.rv	0.133	0.133	0.134	0.134	0.132
AEX.rv	0.145	0.147	0.149	0.149	0.149
SSML.rv	0.149	0.153	0.156	0.158	0.158
IBEX2.rv	0.138	0.138	0.137	0.136	0.133
NSEI.rv	0.119	0.117	0.114	0.111	0.102
MXX.rv	0.077	0.077	0.076	0.075	0.071
BVSP.rv	0.118	0.118	0.119	0.120	0.120
GSPTSE.rv	0.106	0.104	0.103	0.102	0.101
STOXX50E.rv	0.139	0.135	0.130	0.123	0.101
FTSTI.rv	0.111	0.112	0.113	0.113	0.112
FTSEMIB.rv	0.130	0.132	0.133	0.134	0.134

Table B.1: Estimates of ζ_q for all indices in the Oxford-Man dataset.

B.2 On different time intervals¹⁶

Index	H (first half)	H (second half)
SPX2.rk	0.115	0.158
FTSE2.rk	0.140	0.156
N2252.rk	0.083	0.134
GDAXI2.rk	0.154	0.168
RUT2.rk	0.098	0.149
AORD2.rk	0.059	0.114
DJI2.rk	0.123	0.151
IXIC2.rk	0.094	0.156
FCHI2.rk	0.140	0.146
HSI2.rk	0.072	0.129
KS11.rk	0.109	0.147
AEX.rk	0.168	0.151
SSMI.rk	0.206	0.183
IBEX2.rk	0.122	0.149
NSEI.rk	0.112	0.124
MXX.rk	0.068	0.118
BVSP.rk	0.074	0.134
GSPTSE.rk	0.075	0.147
STOXX50E.rk	0.138	0.132
FTSTL.rk	0.080	0.171
FTSEMIB.rk	0.133	0.140

Table B.2: Estimates of H over two different time intervals for all indices in the Oxford-Man dataset

C The effect of smoothing

Although we are really interested in the model

$$\log \sigma_{t+\Delta} - \log \sigma_t = \nu \left(W_{t+\Delta}^H - W_t^H \right),$$

consider the more tractable (fractional Stein and Stein or fSS) model:

$$v_{t+\Delta} - v_t = \alpha \left(W_{t+\Delta}^H - W_t^H \right),$$

where $v_t = \sigma^2$. We cannot observe v_t but suppose we can proxy it by the average

$$\hat{v}_t^\delta = \frac{1}{\delta} \int_0^\delta v_u du.$$

¹⁶Note that we used realized kernel rather than realized variance estimates to generate Table B.2. Results obtained using different variance estimators are almost indistinguishable.

We would, for example, like to estimate $m(2, \Delta) = \mathbb{E} [(v_{t+\Delta} - v_t)^2]$. However, we need to proxy spot variance with integrated variance so instead we have the estimate

$$\begin{aligned}
m^\delta(2, \Delta) &= \mathbb{E} [(\hat{v}_{t+\Delta}^\delta - \hat{v}_t^\delta)^2] \\
&= \frac{1}{\delta^2} \mathbb{E} \left[\left(\int_0^\delta (v_{u+\Delta} - v_u) du \right)^2 \right] \\
&= \frac{\alpha^2}{\delta^2} \int_0^\delta \int_0^\delta \mathbb{E} [(W_{u+\Delta}^H - W_u^H) (W_{s+\Delta}^H - W_s^H)] du ds \\
&= \int_0^\delta \int_0^\delta \{|u - s + \Delta|^{2H} - |u - s|^{2H}\} du ds, \tag{C.1}
\end{aligned}$$

where the last step uses that:

$$\mathbb{E} [W_u^H W_s^H] = \frac{1}{2} \{u^{2H} + s^{2H} - |u - s|^{2H}\},$$

and the symmetry of the integral.

We assume that the length δ of the smoothing window is less than one day so $\Delta > \delta$. Then easy computations give

$$\begin{aligned}
&\int_0^\delta \int_0^\delta |u - s + \Delta|^{2H} du ds \\
&= \frac{1}{2H+1} \frac{1}{2H+2} \{(\Delta + \delta)^{2H+2} - 2\Delta^{2H+2} + (\Delta - \delta)^{2H+2}\}
\end{aligned}$$

and

$$\int_0^\delta \int_0^\delta |u - s|^{2H} du ds = \frac{2}{2H+1} \frac{1}{2H+2} \delta^{2H+2}.$$

Substituting back into (C.1) gives

$$\begin{aligned}
m^\delta(2, \Delta) &= \alpha^2 \Delta^{2H} \frac{1}{2H+1} \frac{1}{2H+2} \frac{1}{\theta^2} \{(1 + \theta)^{2H+2} - 2 - 2\theta^{2H+2} + (1 - \theta)^{2H+2}\} \\
&=: \alpha^2 \Delta^{2H} f(\theta).
\end{aligned}$$

where $\theta = \delta/\Delta$.

Figure C.1 shows the effect of smoothing on the estimated variance in the fSS model. Keeping δ fixed, as Δ increases, $f(\theta) = f(\delta/\Delta)$ increases towards one. Thus, in a linear regression of $\log m^\delta(2, \Delta)$ against $\log \Delta$, we will obtain a higher effective H (from the higher slope) and a lower effective (“volatility of volatility”) α , exactly as we observed in the RSFV model simulations in Section 3.4.

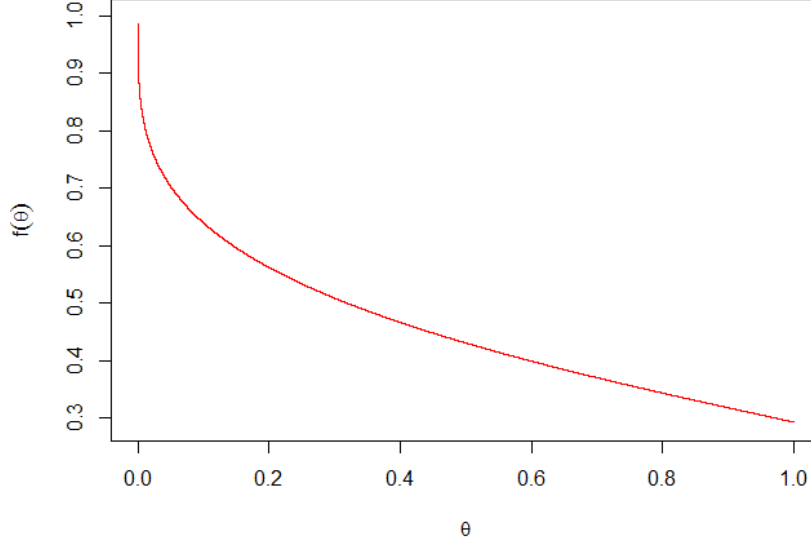


Figure C.1: $f(\theta)$ vs $\theta = \delta/\Delta$ with $H = 0.14$.

Numerical example

In the simulation of the RSFV model in Section 3.4, we have $H = 0.14$, $\delta_1 = 1/24$ for the UZ estimate and $\delta_2 = 1/3$ for the RV estimate. We now reproduce a fSS analogue of the RFSV simulation plots of $m(2, \Delta)$ in Figure 3.5. Specifically, for each $\Delta \in \{1, 2, \dots, 100\}$, with $\alpha = 0.3$ and $\delta = \delta_1$ or $\delta = \delta_2$, we compute the $m^\delta(2, \Delta)$ and regress $\log m^\delta(2, \Delta)$ against $\log \Delta$. The regressions are shown in Figure C.2 and results tabulated in Table C.1.

In Figure C.2 and Table C.1, we observe similar qualitative and quantitative biases from our fSS model simulation as we observe in our simulation of the RSFV model with equivalent parameters in Section 3.4.

Estimate	Est. α	Est. H
Exact ($\delta = 0$)	0.300	0.140
UZ ($\delta = 1/24$)	0.263	0.161
RV ($\delta = 1/3$)	0.230	0.184

Table C.1: Estimated model parameters from the regressions shown in Figure C.2.

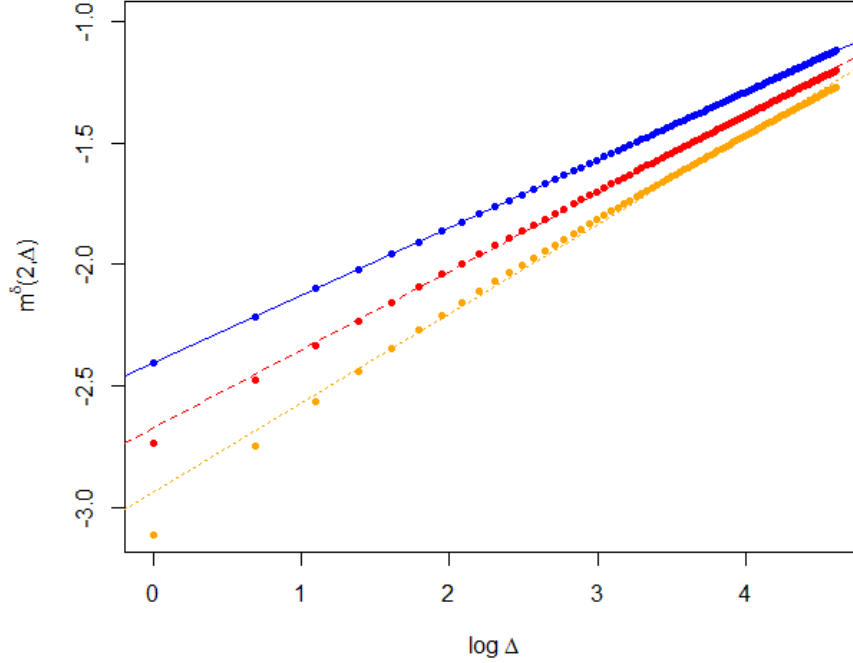


Figure C.2: Analogue of Figure 3.5 in the fSS model: The blue solid line is the true $m(2, \Delta)$; the red long-dashed line is the UZ estimate $m^{\delta_1}(2, \Delta)$; the orange short-dashed line is the RV estimate $m^{\delta_2}(2, \Delta)$.

References

- [1] T. G. Andersen and T. Bollerslev. Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance*, 4(2):115–158, 1997.
- [2] T. G. Andersen, T. Bollerslev, F. X. Diebold, and H. Ebens. The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76, 2001.
- [3] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, 96(453):42–55, 2001.
- [4] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.
- [5] E. Bacry and J. F. Muzy. Log-infinitely divisible multifractal processes. *Communications in Mathematical Physics*, 236(3):449–475, 2003.

- [6] E. Bacry and J.-F. Muzy. Hawkes model for price and trades high-frequency dynamics. *Quantitative Finance*, 14(7):1147–1166, 2014.
- [7] S. R. Bentes and M. M. Cruz. Is stock market volatility persistent? A fractionally integrated approach. 2011.
- [8] J. Beran. *Statistics for long-memory processes*, volume 61. CRC Press, 1994.
- [9] J.-P. Bouchaud and M. Potters. *Theory of financial risk and derivative pricing: From statistical physics to risk management*. Cambridge University Press, 2003.
- [10] P. Carr and L. Wu. What type of process underlies options? A simple robust test. *Journal of Finance*, 58(6):2581–2610, 2003.
- [11] Z. Chen, R. T. Daigler, and A. M. Parhizgari. Persistence of volatility in futures markets. *Journal of Futures Markets*, 26(6):571–594, 2006.
- [12] P. Cheridito, H. Kawaguchi, and M. Maejima. Fractional Ornstein-Uhlenbeck processes. *Electron. J. Probab*, 8(3):14, 2003.
- [13] A. Chronopoulou. Parameter estimation and calibration for long-memory stochastic volatility models. In F. G. Viens, M. C. Mariani, and I. Florescu, editors, *Handbook of Modeling High-Frequency Data in Finance*, pages 219–231. John Wiley & Sons, 2011.
- [14] A. Chronopoulou and F. G. Viens. Estimation and pricing under long-memory stochastic volatility. *Annals of Finance*, 8(2-3):379–403, 2012.
- [15] F. Comte, L. Coutin, and E. Renault. Affine fractional stochastic volatility models. *Annals of Finance*, 8(2-3):337–378, 2012.
- [16] F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Mathematical Finance*, 8(4):291–323, 1998.
- [17] R. Cont. Volatility clustering in financial markets: Empirical facts and agent-based models. In G. Teyssière and A. P. Kirman, editors, *Long Memory in Economics*, pages 289–309. Springer Berlin Heidelberg, 2007.
- [18] F. Corsi. A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196, 2009.
- [19] K. Dayri and M. Rosenbaum. Large tick assets: Implicit spread and optimal tick size. *Working paper*, 2013.
- [20] Z. Ding, C. W. Granger, and R. F. Engle. A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, 1(1):83–106, 1993.

- [21] J. Duchon, R. Robert, and V. Vargas. Forecasting volatility with the multifractal random walk model. *Mathematical Finance*, 22(1):83–108, 2012.
- [22] B. Dupire. Pricing with a smile. *Risk Magazine*, 7(1):18–20, 1994.
- [23] V. Filimonov and D. Sornette. Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E*, 85(5):056108, 2012.
- [24] V. Filimonov and D. Sornette. Apparent criticality and calibration issues in the Hawkes self-excited point process model: Application to high-frequency financial data. *arXiv preprint arXiv:1308.6756*, 2013.
- [25] M. Fukasawa. Asymptotic analysis for stochastic volatility: Martingale expansion. *Finance and Stochastics*, 15(4):635–654, 2011.
- [26] J. Gatheral. *The volatility surface: A practitioner’s guide*, volume 357. John Wiley & Sons, 2006.
- [27] J. Gatheral and A. Jacquier. Arbitrage-free SVI volatility surfaces. *Quantitative Finance*, 14(1):59–71, 2014.
- [28] J. Gatheral and R. C. Oomen. Zero-intelligence realized variance estimation. *Finance and Stochastics*, 14(2):249–283, 2010.
- [29] P. S. Hagan, D. Kumar, A. S. Lesniewski, and D. E. Woodward. Managing smile risk. *Wilmott Magazine*, pages 84–108, 2002.
- [30] S. J. Hardiman, N. Bercot, and J.-P. Bouchaud. Critical reflexivity in financial markets: A Hawkes process analysis. *arXiv preprint arXiv:1302.1405*, 2013.
- [31] S. L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, 1993.
- [32] J. Hull and A. White. One-factor interest-rate models and the valuation of interest-rate derivative securities. *Journal of Financial and Quantitative Analysis*, 28(02):235–254, 1993.
- [33] T. Jaisson and M. Rosenbaum. Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability*, to appear, 2013.
- [34] T. Jaisson and M. Rosenbaum. Fractional diffusions as scaling limits of nearly unstable heavy-tailed Hawkes processes. *Working paper*, 2014.
- [35] M. Lallouache and D. Challet. Statistically significant fits of Hawkes processes to financial data. *Available at SSRN 2450101*, 2014.

- [36] B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM review*, 10(4):422–437, 1968.
- [37] R. N. Mantegna and H. E. Stanley. *Introduction to econophysics: Correlations and complexity in finance*. Cambridge University Press, 2000.
- [38] T. Mikosch and C. Stărică. Is it really long memory we see in financial returns. In P. Embrechts, editor, *Extremes and integrated risk management*, pages 149–168. Risk Books, 2000.
- [39] M. Musiela and M. Rutkowski. *Martingale methods in financial modelling*, volume 36. Springer, 2006.
- [40] A. Novikov and E. Valkeila. On some maximal inequalities for fractional Brownian motions. *Statistics & Probability Letters*, 44(1):47–54, 1999.
- [41] C. J. Nuzman and V. H. Poor. Linear estimation of self-similar processes via Lamperti’s transformation. *Journal of Applied Probability*, 37(2):429–452, 2000.
- [42] C. Y. Robert and M. Rosenbaum. A new approach for the dynamics of ultra-high-frequency data: The model with uncertainty zones. *Journal of Financial Econometrics*, 9(2):344–366, 2011.
- [43] C. Y. Robert and M. Rosenbaum. Volatility and covariation estimation when microstructure noise and trading times are endogenous. *Mathematical Finance*, 22(1):133–164, 2012.
- [44] M. Rosenbaum. Estimation of the volatility persistence in a discretely observed diffusion model. *Stochastic Processes and their Applications*, 118(8):1434–1462, 2008.
- [45] M. Rosenbaum. First order p-variations and Besov spaces. *Statistics & Probability Letters*, 79(1):55–62, 2009.
- [46] M. Rosenbaum. A new microstructure noise index. *Quantitative Finance*, 11(6):883–899, 2011.