Quantitative Finance
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rquf20

A new microstructure noise index

Mathieu Rosenbaum [a]
[a] CREST and CMAP-École Polytechnique Paris , Route de Saclay, 91128 Palaiseau Cedex, France
Published online: 14 Jul 2010.

To cite this article: Mathieu Rosenbaum (2011) A new microstructure noise index, Quantitative Finance, 11:6, 883-899, DOI: 10.1080/14697680903514352

To link to this article: http://dx.doi.org/10.1080/14697680903514352

Quantitative Finance, Vol. 11, No. 6, June 2011, 883–899

A new microstructure noise index

MATHIEU ROSENBAUM*

CREST and CMAP-École Polytechnique Paris, Route de Saclay,
91128 Palaiseau Cedex, France

(*Received 21 April 2008; in final form 18 November 2009*)

We introduce a new microstructure noise index for financial data. This index, the computation of which is based on the *p*-variations of the considered asset or rate at different time scales, can be interpreted in terms of Besov smoothness spaces. We study the behavior of our new index using empirical data. It gives rise to phenomena that a classical signature plot is unable to detect. In particular, with our data set, it enables us to separate the sampling frequencies into three zones: no microstructure noise for low frequencies, increasing microstructure noise from low to high frequencies, and some kind of additional regularity on the finest scales. We then investigate the index from a theoretical point of view, under various contexts of microstructure noise, trying to reproduce the facts observed on the data. We show that this can be partially done using models involving additive correlated errors or rounding error. Accurate reproduction seems to require either both kinds of error together or some unusual form of rounding error.

Keywords: Continuous time finance; Market microstructure; Financial econometrics; Inference for stochastic processes; Stochastic volatility; Fractional Brownian motion

1. Introduction

1.1. Microstructure noise and signature plot

Microstructure noise is usually defined as what makes the observed price differ from the 'theoretical price' at fine scales. This 'theoretical price' is often modeled by a continuous time semi-martingale, as suggested by classical financial theory. In the literature, the reasons why this noise occurs are given as the bid–ask spread and the discreteness of the prices. Imagine we observe a financial asset Y_t on a time interval $[0, T]$. We want to know if there is any microstructure noise in our data. To answer this question, we define microstructure noise as a form of irregularity in the high-frequency data that disappears in the lower frequencies. Consider the sample

$$\{Y_{k\Delta}, k = 0, \ldots, \lfloor T/\Delta \rfloor\},$$

where Δ denotes the minimum time to wait to be able to record new data (in our data, Δ will be equal to one second). For simplicity, we take $\Delta = 2^{-N}$ with N a positive integer and T and T/Δ are also positive integers. The usual way to study the presence of some kind of irregularity in high frequencies is to compute the signature plot, that is the function

$$q \to Z_q,$$

with $q \in [0, N]$ and

$$Z_q = \sum_{k=0}^{T2^{N-q}-1} \left| Y_{(k+1)2^{q-N}} - Y_{k2^{q-N}} \right|^2$$

(absolute signature plot) (1)

or†

$$Z_q = \sum_{k=0}^{T2^{N-q}-1} \left| \log(Y_{(k+1)2^{q-N}}) - \log(Y_{k2^{q-N}}) \right|^2$$

(relative signature plot). (2)

*Email: mathieu.rosenbaum@polytechnique.edu

†The question of which is the 'best' of the two measures is probably not so relevant for high-frequency data for a short period. Indeed,

$$\log(Y_{(i+1)/n}) - \log(Y_{i/n}) \sim \frac{Y_{(i+1)/n} - Y_{i/n}}{Y_{i/n}}$$

and therefore it is almost proportional to $Y_{(i+1)/n} - Y_{i/n}$ as soon as the variation of the price in the period is not too large.

Quantitative Finance
ISSN 1469–7688 print/ISSN 1469–7696 online © 2011 Taylor & Francis
http://www.informaworld.com
DOI: 10.1080/14697680903514352

The signature plot was made popular by Andersen *et al.* (2000). As soon as Y (or $\log Y$) is a continuous Ito semi-martingale, for fixed T and given q, this quantity converges in probability, as N tends to infinity, to the quadratic variation of Y (or $\log Y$) on $[0, T]$ (see Jacod 2008 for a general study of this type of quantity). Hence, in the low frequencies, where there is no microstructure noise, the signature plot of an asset is quite flat as soon as the number of data is sufficient. On the contrary, it is not flat in the presence of noise. A drawback of this measure is that it is not absolute, one just observes where the signature plot is flat and where it is not.

1.2. A new microstructure noise index

In this paper, we propose an alternative to the signature plot. For a given real $p>0$, we now consider the functions

$$q \to S_q^p, \quad q \to S_q'^p,$$

with

$$S_q^p = \frac{1}{p}\left\{1+\log_2\left(\frac{V_{q+1}^p}{V_q^p}\right)\right\} \quad \text{(absolute microstructure index)} \tag{3}$$

and

$$S_q'^p = \frac{1}{p}\left\{1+\log_2\left(\frac{V_{q+1}'^p}{V_q'^p}\right)\right\} \quad \text{(relative microstructure index)}, \tag{4}$$

where

$$V_q^p = \sum_{k=0}^{T2^{N-q}-1} |Y_{(k+1)2^{q-N}} - Y_{k2^{q-N}}|^p$$

and

$$V_q'^p = \sum_{k=0}^{T2^{N-q}-1} |\log(Y_{(k+1)2^{q-N}}) - \log(Y_{k2^{q-N}})|^p.$$

Using Besov smoothness spaces, we will see in section 2.1 that the microstructure indexes S_q^p and $S_q'^p$ may be considered as regularity measures associated with the subsampling frequency 2^{N-q}. Hence, if the function $q \to S_q^p$ (or $q \to S_q'^p$) is constant, we say that, in our sense, there is no microstructure noise in the data. Practically speaking, we may hope to observe a constant regularity for sufficiently low frequencies and more specific behavior for high frequencies, reflecting the presence of microstructure noise. Note that close quantities have recently been used by Aït-Sahalia and Jacod (2009) in the context of jumps.

In section 2 we provide an interpretation of our new indexes in terms of Besov spaces and explain their virtues. In section 3, we study them using financial data: the Bund future contract from the Eurex market and the Euro/US dollar exchange rate from the Reuters database. What we learn from the data is that the regularity remains constant for large sampling scales (longer than 15–20 minutes), is decreasing when going to finer scales (from 15 minutes to 10 seconds) and is increasing when going to the finest scales (from 10 to 1 second). We then show in sections 4 and 5 that models involving additive correlated noise or rounding error enable us to reproduce most of the facts observed concerning the data. Nevertheless, we will see in section 6 that the use of complementary microstructure functions seems to indicate that an accurate reproduction requires either both kinds of error together or some unusual form of rounding error.

2. Properties of the index

2.1. Interpretation of the index

In this section, we show why the quantity S_q^p can be viewed as a regularity measure. For that purpose, we consider here that T is equal to one and we use Besov smoothness spaces $\mathcal{B}_{p,\infty}^s([0,1])$ as defined in the appendix. Following the assumptions of Ciesielski *et al.* (1993) and Rosenbaum (2009a), if $t \to Y_t$ belongs to $\mathcal{B}_{p,\infty}^s([0,1])$ and does not belong to $\mathcal{B}_{p,\infty}^{s+\varepsilon}([0,1])$, for any positive ε, we can consider the following asymptotic approximation:

$$V_q^p \simeq c2^{(q-N)(ps-1)},$$

with c a positive constant value (see the appendix for details). Consequently, we easily see that, in this case,

$$S_q^p \simeq s.$$

For example, such results hold for a fractional Brownian with Hurst index H when taking s equal to H. Thus, we can interpret the preceding index S_q^p in the following way: based on the subsampled data at period $m=2^q$, in terms of Besov spaces $\mathcal{B}_{p,\infty}^s([0,1])$, the regularity s of the underlying continuous time process suggested by the data is S_q^p.

2.2. Virtues of the index

- The index is model free. We do not have to assume that the underlying process follows a semi-martingale-type process or a 'noisy' semi-martingale process. On the contrary, the signature plot really makes sense only in the context of semi-martingales. For example, for fixed T, as N goes to infinity, for any $1/2<H<1$, the signature plot of a fractional Brownian motion with Hurst parameter H converges almost surely to zero, whereas, for any $p>0$, the microstructure index converges almost surely to H.
- The value of the index depends on the subsampling period and consequently we can discriminate between sampling periods using the index. For example, if we fix a reference sampling period, e.g. 1 hour, and we compute the associated index S_{ref}, we could say that the microstructure noise occurs in the sampling

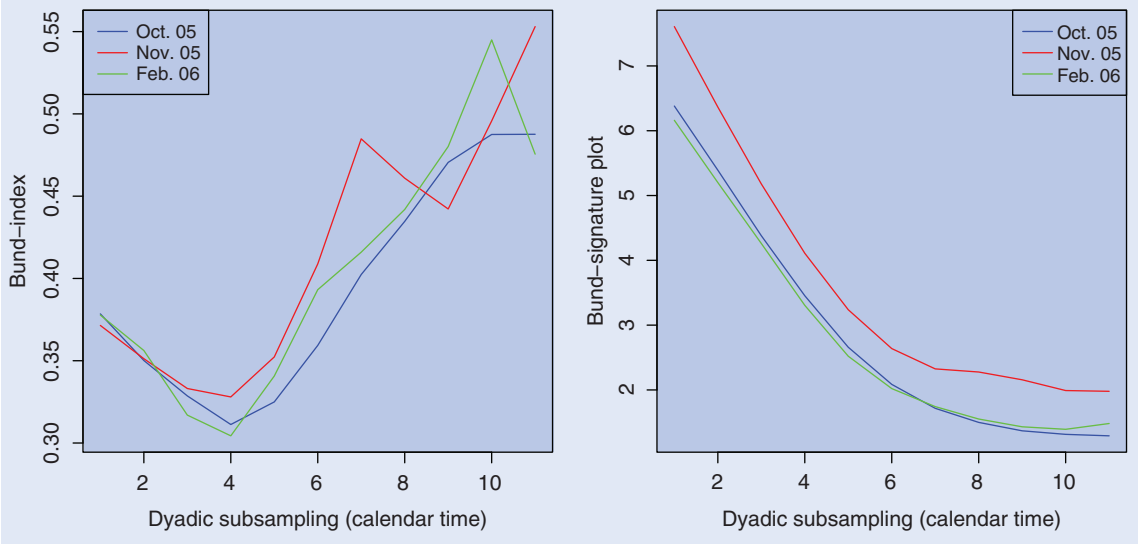

Figure 1. Microstructure noise index (left) and signature plot (right) for the Bund, Oct. 05, Nov. 05, Feb. 06.

period $2^q < 1$ hour if S_q^p is significantly different from S_{ref}.†

- In the case of a continuous semi-martingale, for any value of p, the value of the index S_q^p is asymptotically equal to 1/2 (convergence in probability, see Jacod 2008 for details). Hence, if one now considers that the continuous semi-martingale is the reference case, the index is in some sense absolute: the farther from 1/2 the index is, the more the noise is important.
- The calculation of the index is based on well-known financial quantities, the p-variations, often considered in practice as empirical moments. The most classical case is obtained for $p = 2$. Indeed, the microstructure noise index can then be seen as a function of two realized volatilities considered at different time scales. Therefore, for simplicity, we will take from now on $p = 2$ until section 6.
- From a practical point of view, the index could probably be made more robust by using averages over all possible starting points for the subsampled statistics, in the spirit of Zhang *et al.* (2005). However, since it complicates the analysis we do not consider such versions of the index.

3. Empirical study

We compute the indexes on a financial asset and an exchange rate.

The Bund future contract from the Eurex market, every second, for October 2005, November 2005, February 2006, October 2006, November 2006, and February 2007 (we skip the months of December and January to avoid particular effects due to the roll of the contract and to the beginning of the year). We fix the start of a day at 9 a.m. CET and the end of a day at 7 p.m. CET‡ (once again, we want to avoid particular effects due to the beginning or the end of the day). We consider here that the value at time t is the last traded value. Results for the bid and midquote prices are given by Rosenbaum (2007). Each day, we compute

$$V_q^{2,d} = \sum_{k=0}^{2^{N-q}-1} \left| Y^d_{(k+1)2^{q-N}} - Y^d_{k2^{q-N}} \right|^2$$

and

$$V_q'^{2,d} = \sum_{k=0}^{2^{N-q}-1} \left| \log(Y^d_{(k+1)2^{q-N}}) - \log(Y^d_{k2^{q-N}}) \right|^2,$$

where Y^d denotes the price on day d. Then, in the empirical study, V_q^2 (respectively $V_q'^2$) is defined as the sum over each day of $V_q^{2,d}$ (respectively $V_q'^{2,d}$).

The Euro/US dollar exchange rate, every second, bid prices from the Reuters database, from 15/08/2005 to 01/10/2005. All our computations are carried out with the software *R*. The results are given in figures 1–4 for the absolute microstructure noise index, with dyadic subsamplings $m = 2^k$ seconds, $k = 0, \ldots, 12$ (the value 1 on the x axis corresponds to $k = 0$). The results for the relative microstructure index are similar and can be found in Rosenbaum (2007).

Thus, we observe an under diffusive behavior ($S_q^2 < 1/2$) in the very high frequencies and a regularity that is close to 1/2 in the lower frequencies. It is also surprising that the graphs are decreasing for q between 0 and 3. This particular behavior does not appear at all in

†In the following, we will give results for specific models aiming at quantifying the distance between S_q^p and S_{ref} from a statistical point of view.
‡Precisely, starting from 9 a.m. we take $2^{15} + 1$ data per day.

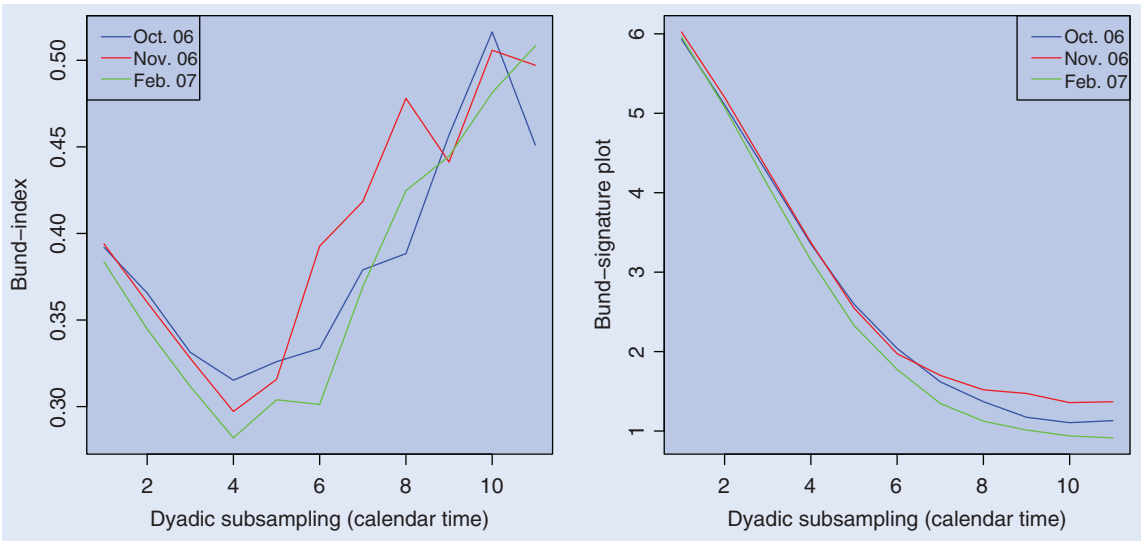

Figure 2. Microstructure noise index (left) and signature plot (right) for the Bund, Oct. 06, Nov. 06, Feb. 07.

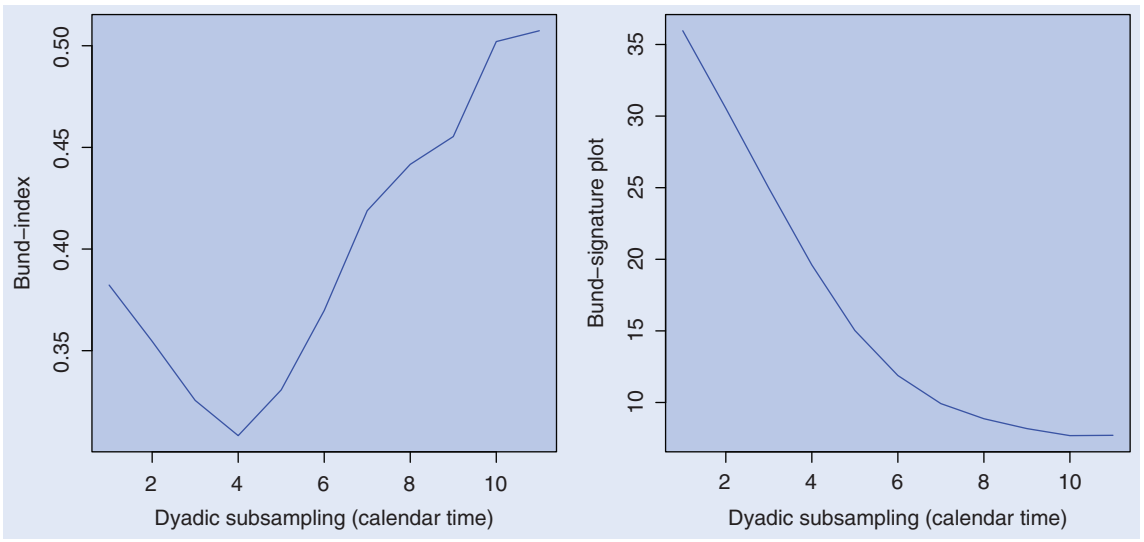

Figure 3. Microstructure noise index (left) and signature plot (right) for the Bund, aggregated data: Oct. 05, Nov. 05, Feb. 06, Oct. 06, Nov. 06, Feb. 07.

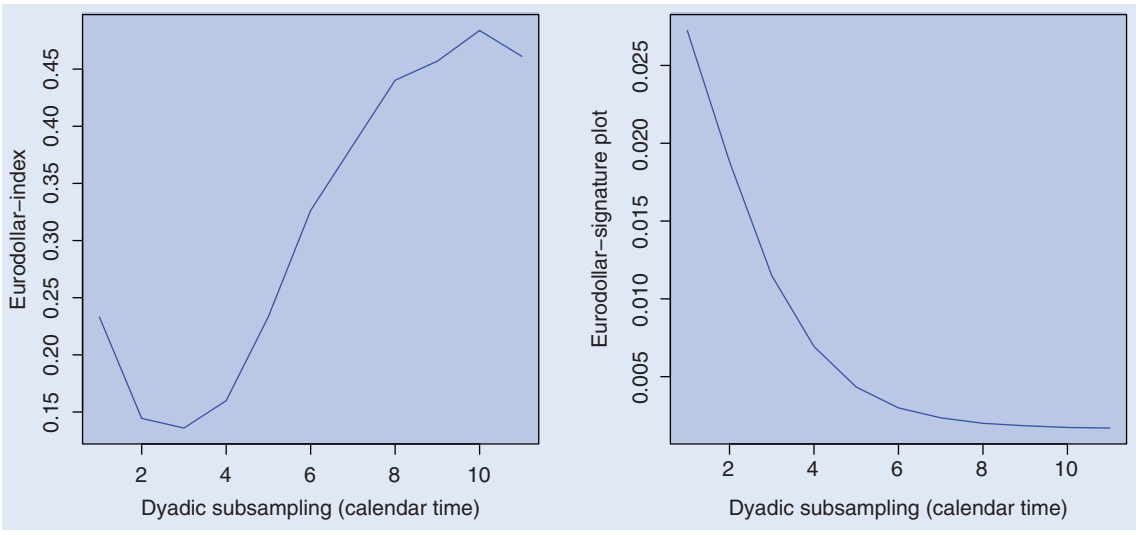

Figure 4. Microstructure noise index (left) and signature plot (right) for the bid Eurodollar 15/08/2005–01/10/2005.

the signature plots and can be linked to ultra-high-frequency persistence in the prices. Indeed, when the market is trading on a price, there is a duration to wait before this price changes and the average duration is significantly positive. The fact that the price stays constant for a small time period might be an explanation for this additional ultra-high-frequency regularity.

The curves show that the regularity is less in the high frequencies, which can be considered a microstructure noise effect. Thus, in our approach, we can conclude from these graphs that there is an essentially decreasing microstructure noise in the Bund and the Euro/US dollar exchange rate if the subsampling period is less than 2^{10} seconds.

The question now is to explain this behavior of the graphs. The idea is to consider models that behave as classical financial models (e.g. diffusions) at low frequencies, but differently at high frequencies. In the following section we study our index in the classical framework of additive microstructure noise. We then study models with rounding error. Note that, in this paper, our goal is not to study complex continuous time dynamics for the theoretical price. We focus on the impact of noise on the observed price.

4. Models with additive noise

4.1. Introduction

Models with additive microstructure noise have been widely studied (see, in particular, Gloter and Jacod 1997, Aït-Sahalia *et al.* 2005, Zhang *et al.* 2005, 2006, Hansen and Lunde 2006, Bandi and Russel 2008, and Andersen *et al.* 2009). We consider here

$$Y_{k\Delta} = c\exp(X_{k\Delta} + \varepsilon_k^\Delta), \quad k = 0, \ldots, T/\Delta, \tag{5}$$

where $X_{k\Delta}$ is the 'theoretical log-price', c is an initial condition (we take $c = 115$ to be consistent with the Bund data) and ε_k^Δ is an additive centered noise with variance V^2, independent of X. For simplicity, we assume

$$X_{k\Delta} = \sigma W_{k\Delta}, \tag{6}$$

where W is a Brownian motion and σ is equal to 10^{-2} (this value is chosen in order to be consistent with the Bund data). A stochastic volatility model for X is treated by Rosenbaum (2007). We work under the following specifications for the noise.

- (M1) ε_k^Δ are i.i.d. Gaussian variables.
- (M2) ε_k^Δ follow an $AR(1)$ model $\varepsilon_k^\Delta = r\varepsilon_{k-1}^\Delta + \xi_k^\Delta$, $r<1$, where ξ_k^Δ are i.i.d. centered Gaussian variables.
- (M3) ε_k^Δ follow a fractional Gaussian noise† with Hurst parameter $H = 0.1$.
- (M4) ε_k^Δ follow a fractional Gaussian noise with Hurst parameter $H = 0.9$.

Models M2 and M4 introduce positive correlations between the noise terms. This is a possible way to add an 'ultra-high-frequency regularity' and to try to reproduce the fact that the microstructure index is decreasing for q between 0 and 3.

4.2. Graphical overview

To obtain an initial idea of the behavior of the absolute microstructure index and of the absolute signature plot in such models, we give graphs for the four preceding models (figures 5–8). For each model, we compute 50 simulations with $T = 1$ and $\Delta = 2^{-19}$ and give the average absolute microstructure index and signature plot for V equal to 0, 4×10^{-5}, 8×10^{-5}, 12×10^{-5}, 16×10^{-5} and 2×10^{-4}. Moreover, we take $r = 0.8$ for model M2. As expected, model M1 does not enable us to reproduce the shape of the graphs observed for real data. We can see from the graphs for models M2 and M4 that the additional ultra-high-frequency regularity can be obtained by means of positive correlations between the noise terms. The use of an AR-type noise is more common than fractional noise, and we focus in the next section on model M2. Model M3 is a control experiment that shows that if we consider a negatively correlated noise, we can obtain an additional high-frequency irregularity.

4.3. Dependent noise: Theoretical considerations

In this section we investigate model M2 from a theoretical point of view. For simplicity, we work here with the log prices. Therefore, we consider the model where the observed log prices $Y_{k\Delta}$ are given by

$$Y_{k\Delta} = X_{k\Delta} + \varepsilon_k^\Delta, \quad k = 0, \ldots, T/\Delta, \tag{7}$$

where $X_t = \sigma W_t$ and ε_k^Δ are independent of X and are the Gaussian stationary solution of an $AR(1)$ model:

$$\varepsilon_k^\Delta = r\varepsilon_{k-1}^\Delta + \gamma\zeta_k^\Delta, \quad k \geq 1, \tag{8}$$

with $0 \leq r < 1$. The ζ_k^Δ are i.i.d. centered Gaussian variables with variance 1, such that ζ_k^Δ is independent of the past of ε_j^Δ, $j<k$, and ε_0^Δ is chosen appropriately. This kind of framework has been investigated by Aït-Sahalia *et al.* (2009) in the purpose of volatility estimation, in the spirit of Zhang *et al.* (2005) and Zhang (2006). Using computations of the realized volatility, our objective is to give an approximate equation for the microstructure noise index in this context.

4.3.1. Asymptotic for T and Δ. We study the signature plot and the microstructure index at low and high

†Recall that a fractional Gaussian noise ($\varepsilon_k^\Delta,\ k \geq 0$) is defined by $\varepsilon_k^\Delta = W_{k+1}^H - W_k^H$, where ($W_t^H,\ t \geq 0$) is a fractional Brownian motion with Hurst parameter H.

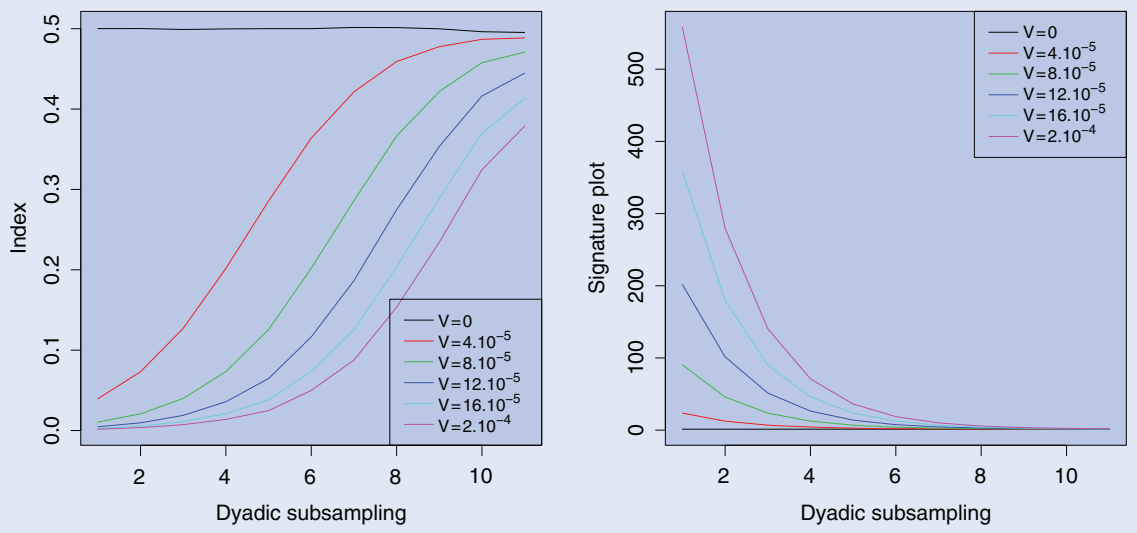

Figure 5. Model M1, microstructure noise index (left) and signature plot (right).

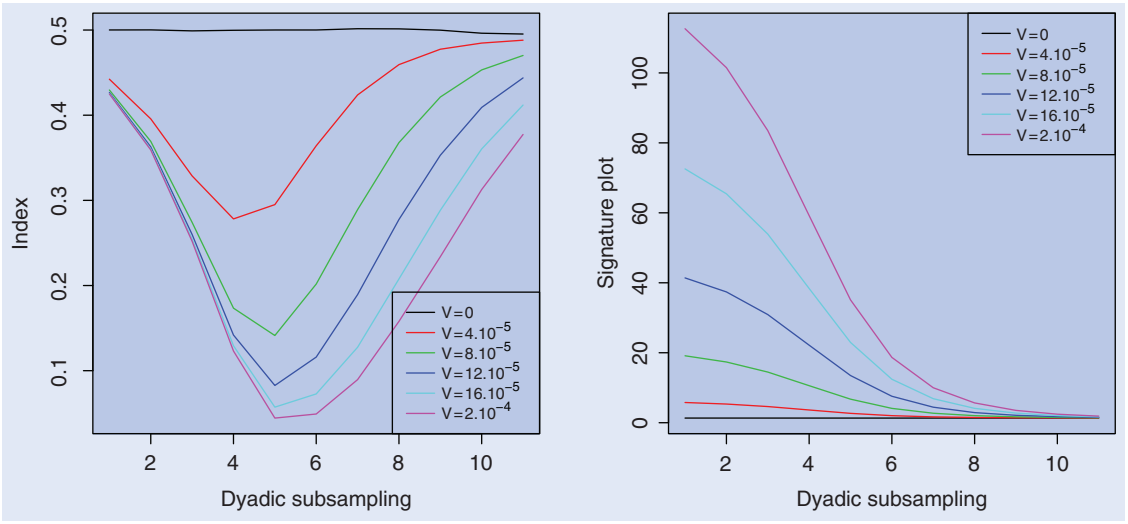

Figure 6. Model M2, microstructure noise index (left) and signature plot (right).

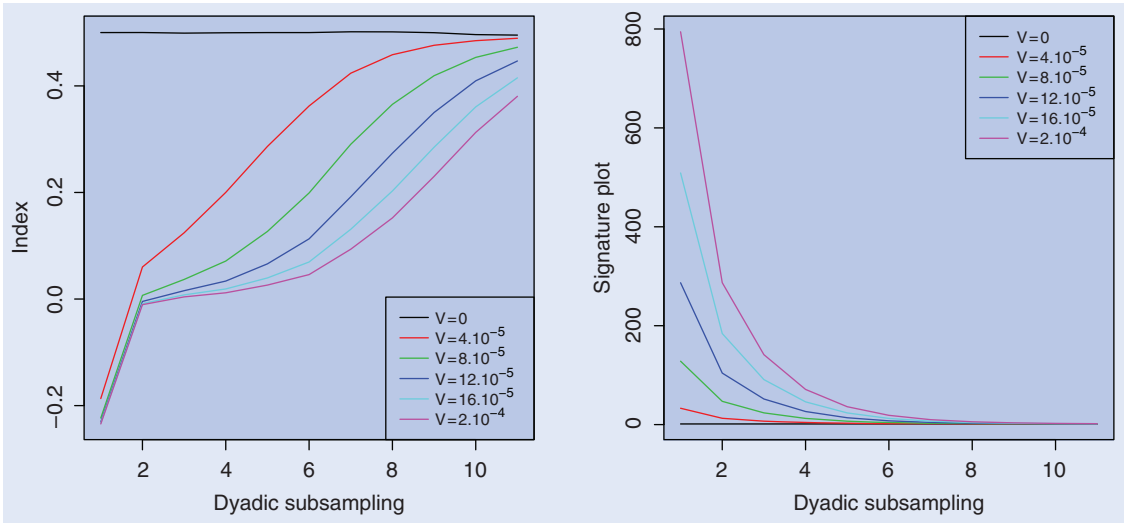

Figure 7. Model M3, microstructure noise index (left) and signature plot (right).

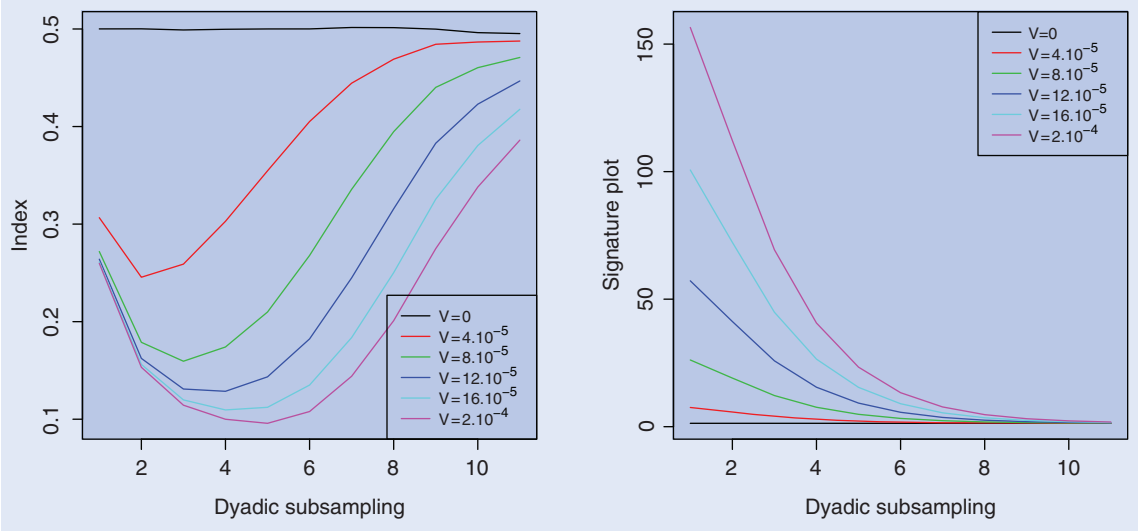

Figure 8. Model M4, microstructure noise index (left) and signature plot (right).

frequencies. For the low frequencies, we need to have enough time between two data so that the microstructure noise disappears. Hence, it is natural to assume for the low frequencies that $\Delta = \Delta_n$ tends to infinity and therefore $T = T_n$ tends to infinity. On the contrary, we need close data at high frequencies. Hence, in that case, we will assume that $\Delta = \Delta_n$ tends to zero.

4.3.2. First asymptotic for the noise: Constant variance of the noise. Our noise sequence is defined on the finest scale Δ. We first study asymptotics where the variance of the noise does not depend on Δ.

Realized volatility. Let m be a positive integer and

$$RV(T,\Delta,m) = \sum_{i=0}^{T(m\Delta)^{-1}} (Y_{im\Delta} - Y_{(i-1)m\Delta})^2. \tag{9}$$

We also define positive sequences T_n and Δ_n. We have the following result, the proof of which is given in appendix B.

Theorem 4.1: *In model M2 there exist N_1 and N_2 such that $\mathbb{E}[N_1^2 + N_2^2] < +\infty$ and*

$$RV(T,\Delta,m) = \sigma^2 \sum_{i=1}^{T(m\Delta)^{-1}} \left(W_{im\Delta} - W_{(i-1)m\Delta}\right)^2 + 2\gamma^2 \frac{1-r^m}{1-r^2} T(m\Delta)^{-1} + T^{1/2}(m\Delta)^{-1/2} N_1 + N_2 T^{1/2}.$$

If T is fixed, as Δ_n tends to zero,

$$RV(T,\Delta_n,m) = 2\gamma^2 \frac{1-r^m}{1-r^2} T(m\Delta_n)^{-1} + \mathcal{O}_p(m\Delta_n)^{-1/2}.$$

If T_n, Δ_n and $T_n\Delta_n^{-1}$ tend to infinity and $T_n^{1/2}\Delta_n^{-3/2}$ tends to $\beta \geq 0$,

$$T_n^{1/2}(2\sigma^2 m\Delta_n)^{-1/2}\left(T_n^{-1} RV(T_n,\Delta_n,m) - \sigma^2\right) \to_{\mathcal{L}} \sqrt{2}(\sigma^2 m)^{-1/2}\gamma^2\beta\frac{1-r^m}{1-r^2} + \mathcal{N}(0,1).$$

If T_n, Δ_n, $T_n\Delta_n^{-1}$ and $T_n^{1/2}\Delta_n^{-3/2}$ tend to infinity,

$$T_n^{-1} RV(T_n,\Delta_n,m) = \sigma^2 + 2\gamma^2\frac{1-r^m}{1-r^2}(m\Delta_n)^{-1} + \mathcal{O}_p\left((m\Delta_n/T_n)^{1/2}\right).$$

This result enables us to derive the behavior of the microstructure noise index.

Microstructure noise index. For the microstructure noise index, we have the following corollary of the preceding theorem.

Corollary 4.2: *In model M2, if T is fixed, as Δ_n tends to zero,*

$$S_q'^2 = \frac{1}{2}\log_2\left(1 + r^{2^q}\right) + \mathcal{O}_p\left((2^q\Delta_n)^{1/2}\right).$$

If T_n, Δ_n and $T_n\Delta_n^{-1}$ tend to infinity and $T_n^{1/2}\Delta_n^{-3/2}$ tend to $\beta \geq 0$,

$$S_q'^2 = \frac{1}{2} + \mathcal{O}_p\left((2^q\Delta_n/T_n)^{1/2}\right).$$

If T_n, Δ_n, $T_n\Delta_n^{-1}$ and $T_n^{1/2}\Delta_n^{-3/2}$ tend to infinity,

$$S_q'^2 = \frac{1}{2} - \frac{\gamma^2(1-r^{2^q})^2\Delta_n^{-1}}{2^{q+1}(1-r^2)} + \mathcal{O}_p\left((2^q\Delta_n/T_n)^{1/2} + (2^q\Delta_n)^{-2}\right).$$

It is important to remark that, in the case where T is fixed and Δ_n tends to zero, we in fact have

$$2S_q'^2 = 1 \underbrace{-1}_{\text{due to the noise}} + \underbrace{\log_2(1+r^{2^q})}_{\text{due to the correlation}} + \mathcal{O}_p((2^q\Delta_n)^{1/2}).$$

Hence we reproduce the decreasing behavior at the beginning of the graphs. In the low-frequency asymptotic,

the index goes to 1/2, which agrees with the empirical results.

Parameters estimation. Under the preceding specification, we have the following obvious corollary of theorem 4.1.

Corollary 4.3: *Let*

$$\widehat{r} = 2^{2S_0'^2} - 1 \tag{10}$$

and

$$\widehat{\gamma^2} = T^{-1}\Delta V_1'^2. \tag{11}$$

In model M2, if T is fixed, as Δ_n tends to zero, the sequences $\Delta_n^{-1/2}(\widehat{r} - r)$ and $\Delta_n^{-1/2}(\widehat{\gamma^2} - \gamma^2)$ are tight.

4.3.3. Second asymptotic for the noise: Variance of the noise depending on the finest scale. It is also natural to assume that ε_i^Δ are of order $\Delta^{1/2}$. Indeed, our noise process is built using the finest period and, under this specification, $\varepsilon_{(i+1)}^\Delta - \varepsilon_i^\Delta$ is of order $\Delta^{1/2}$, which is the order of magnitude of the increment of the Brownian motion on $[i\Delta, (i+1)\Delta]$. This asymptotic is justified as soon as we assume that the quadratic variation of the observed price is bounded, which is very reasonable in practice. In that case, since the quadratic variation of the theoretical price is bounded, that of the noise also has to be bounded. In other words, we consider that the signature plot is not going to infinity at high frequencies.

Realized volatility. Similar computations as in the preceding theorem lead to the following result.

Theorem 4.4: *In model M2, if $\gamma^2 = \gamma^2(\Delta) = \eta^2\Delta$, there exist N_1 and N_2 such that $\mathbb{E}[N_1^2 + N_2^2] < +\infty$ and*

$$\begin{aligned} RV(T,\Delta,m) = \sigma^2 &\sum_{i=1}^{T(m\Delta)^{-1}} (W_{im\Delta} - W_{(i-1)m\Delta})^2 \\ &+ 2\eta^2 \frac{1-r^m}{1-r^2} Tm^{-1} + m^{-1/2}(\Delta T)^{1/2} N_1 \\ &+ (\Delta T)^{1/2} N_2. \end{aligned}$$

Hence, under this specification for the noise, the signature plot does not explode around $q = 0$. An important remark is that, in this framework, if T is fixed, as Δ_n tends to zero, from V_0', V_1' and V_2' one can estimate σ^2, r and η^2 with accuracy $\Delta_n^{1/2}$.

Microstructure noise index. For the microstructure noise index, we have the following corollary of the preceding theorem.

Corollary 4.5: *In model M2, if $\gamma^2 = \gamma^2(\Delta_n) = \eta^2\Delta_n$ and if T is fixed, as Δ_n tends to zero, we have*

$$S_q'^2 = \tilde{S}_q'^2 + \mathcal{O}_p((\Delta_n T)^{1/2}),$$

with

$$\tilde{S}_q'^2 = \frac{1}{2} + \frac{1}{2}\log_2\left\{\frac{2^q + (1-r^{2^{q+1}})\eta^2(\sigma^2)^{-1}(1-r^2)^{-1}}{2^q + 2(1-r^{2^q})\eta^2(\sigma^2)^{-1}(1-r^2)^{-1}}\right\}.$$

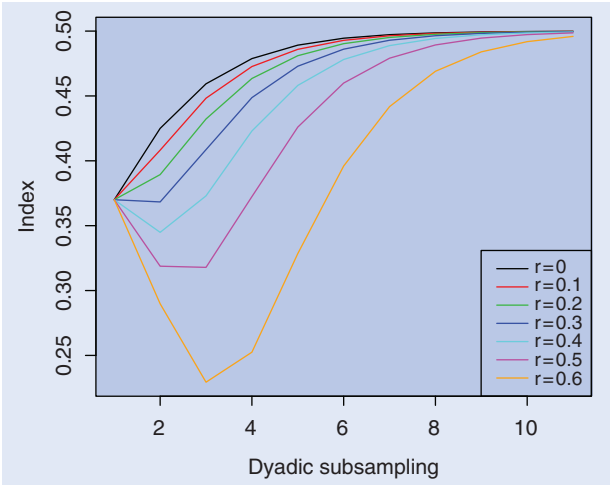

Figure 9. Approximate noise index $\tilde{S}_q'^2$.

Hence, the approximate noise index $\tilde{S}_q'^2$ only depends on q, r and the noise signal ratio $\eta^2(\sigma^2)^{-1}(1-r^2)^{-1}$. Moreover, $\tilde{S}_q'^2 \leq 1/2$. Note that if $\eta^2 = 0$ and $\tilde{S}_q'^2 = 1/2$ and if $q = 0$ and $r = 0$,

$$\tilde{S}_0'^2 = \frac{1}{2} + \frac{1}{2}\log_2\left\{\frac{1+\eta^2(\sigma^2)^{-1}}{1+2\eta^2(\sigma^2)^{-1}}\right\}.$$

We also have the following relation:

$$\frac{\eta^2(\sigma^2)^{-1}}{1-r^2} = \frac{1 - 2^{2\tilde{S}_0'^2 - 1}}{(2^{2\tilde{S}_0'^2} - 1 - r)(1-r)}. \tag{12}$$

To make sense, if $\tilde{S}_0'^2 < 1/2$, the preceding equation implies

$$r \leq 2^{2\tilde{S}_0'^2} - 1.$$

Figure 9 presents the graphs of the function $\tilde{S}_q'^2$ for $\tilde{S}_0'^2 = 0.37$, $r = (0, 0.1, 0.3, 0.4, 0.5, 0.6)$ and $\eta^2(\sigma^2)^{-1}(1-r^2)$ associated by (12).

Approximation of the microstructure noise index. In this section, we give approximations of the noise index from a practical point of view in the case where $\gamma^2 = \gamma^2(\Delta) = \eta^2\Delta$. For small q, the signature plot shows that

$$\sigma^2 \ll 2\eta^2 \frac{1-r^{2^q}}{1-r^2} 2^{-q}$$

and for large q,

$$\sigma^2 \gg 2\eta^2 \frac{1-r^{2^q}}{1-r^2} 2^{-q}.$$

We have the following result.

Corollary 4.6: *If $\sigma^2 \to 0$ while the other parameters are fixed,*

$$\tilde{S}_q'^2 = \frac{2^{q-1}\sigma^2}{\eta^2}\left(\frac{1-r^2}{1+r^{2^q}}\right) + \frac{1}{2}\log_2(1+r^{2^q}) + o(\sigma^2).$$

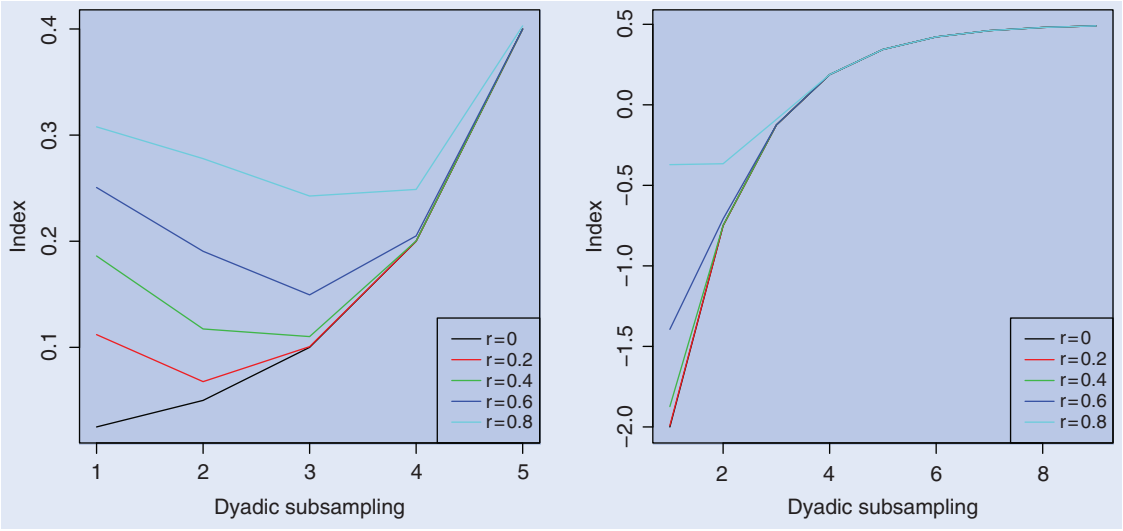

Figure 10. High-frequency approximation (left) and low-frequency approximation (right).

If $q \to +\infty$ while the other parameters are fixed,

$$\tilde{S}_q^{\prime 2} = \frac{1}{2} - \frac{\eta^2}{(1-r^2)2^{q+1}\sigma^2}\{1 - r^{2^q}(2 - r^{2^q})\} + o(2^{-q}).$$

Note that we have in the first case

$$\tilde{S}_q^{\prime 2} = \underbrace{\frac{1}{2} - \frac{1}{2} + \frac{2^{q-1}\sigma^2}{\eta^2}}_{\text{due to the noise}} + \underbrace{\frac{1}{2}\log_2(1 + r^{2^q}) - \frac{2^{q-1}\sigma^2}{\eta^2}\left(\frac{r^{2^q} + r^2}{1 + r^{2^q}}\right)}_{\text{due to the correlation}} + o(\sigma^2),$$

and in the second case

$$\tilde{S}_q^{\prime 2} = \frac{1}{2} \underbrace{- \frac{\eta^2}{2^{q+1}\sigma^2}}_{\text{due to the noise}} + \underbrace{\frac{\eta^2}{(1-r^2)2^{q+1}\sigma^2}\{r^{2^q}(2 - r^{2^q}) - r^2\}}_{\text{due to the correlation}} + o(2^{-q}).$$

For illustration, figure 10 shows the graphs of the two preceding approximations of $\tilde{S}_q^{\prime 2}$ for $(\sigma^2/\eta^2)(1-r^2) = 0.05$ and $r = (0, 0.2, 0.4, 0.6, 0.8)$.

5. Models with rounding error

5.1. Introduction

Models with a rounding error have been studied by Delattre and Jacod (1997), Delattre (1997), Li and Mykland (2007) and Rosenbaum (2009b). A model with rounding error is a simple and natural way to obtain discrete prices and diffusive behavior on a large sampling scale. Moreover, it is striking to see how rounded diffusions appear visually like tick by tick financial data (figures 11 and 12).

We define the notation $x^{(0)} = x$ and for $\beta \neq 0$, $x^{(\beta)} = \beta\lfloor x/\beta \rfloor$. We now observe

$$Y_{k\Delta} = \left(c \exp\{X_{k\Delta} + V\varepsilon_k^\Delta\}\right)^{(\alpha)}, \quad i = 0, \ldots, T/\Delta, \tag{13}$$

where $X_{k\Delta}$ is the 'theoretical log price', $V\varepsilon_k^\Delta$ is an additive centered noise with variance V^2, independent of X, and α is the rounding level, equal to 10^{-2} for the Bund. As previously, we take $c = 115$. We assume

$$X_k^\Delta = \sigma W_k^\Delta, \tag{14}$$

where W is a Brownian motion and σ is equal to 10^{-2}. A stochastic volatility model for X is treated by Rosenbaum (2007). We work under the following specifications for the noise.

- (M1′) ε_k^Δ are i.i.d. Gaussian variables.
- (M2′) ε_k^Δ follow an $AR(1)$ model $\varepsilon_k^\Delta = r\varepsilon_{k-1}^\Delta + \xi_k^\Delta$, where ξ_k^Δ are i.i.d. centered Gaussian variables.
- (M3′) ε_k^Δ follow a fractional Gaussian noise with Hurst parameter $H = 0.1$.
- (M4′) ε_k^Δ follow a fractional Gaussian noise with Hurst parameter $H = 0.9$.
- (M5′) ε_k^Δ are i.i.d. with law $\sqrt{12}\mathcal{U}[0, 1]$, where $\mathcal{U}[0, 1]$ denotes the uniform law on $[0, 1]$.

Note that, from a theoretical point of view, model M5′ was investigated by Jacod *et al.* (2009).

5.2. Graphical overview

Figures 13–17 show graphs corresponding to those in section 4.2. The graphs are similar to those obtained for the additive microstructure noise. We see that for model M2′, we need a stronger additive noise component to obtain the decreasing behavior of the beginning of the empirical graphs. Indeed, it has to be sufficiently important to 'compensate' for the rounding effect.

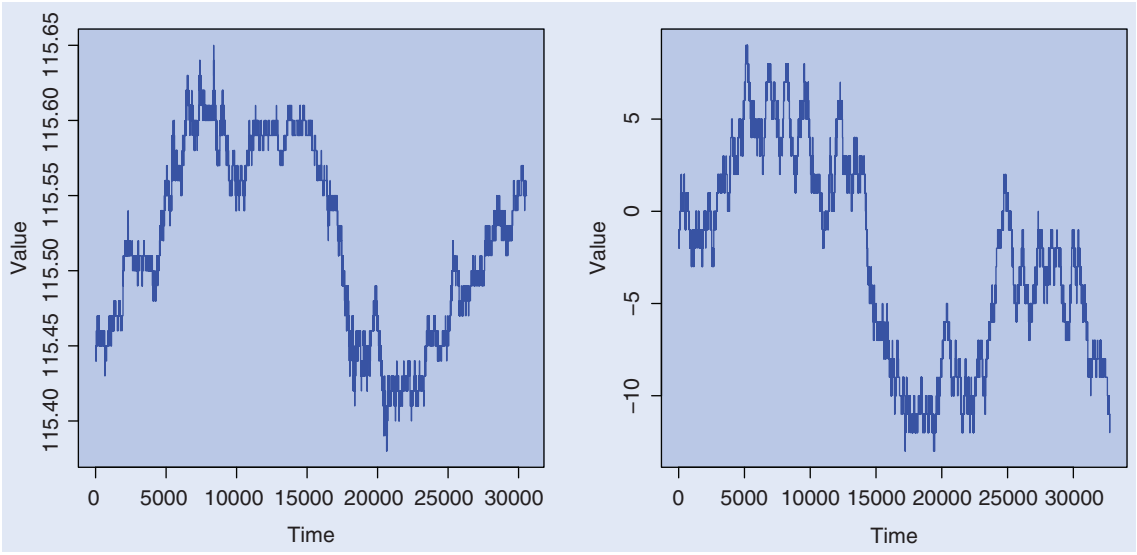

Figure 11. Bund contract, 2007-05-06, one data every second (left), $\lfloor 20W_t \rfloor$, frequency $= 2^{15}$ on [0,1] (right).

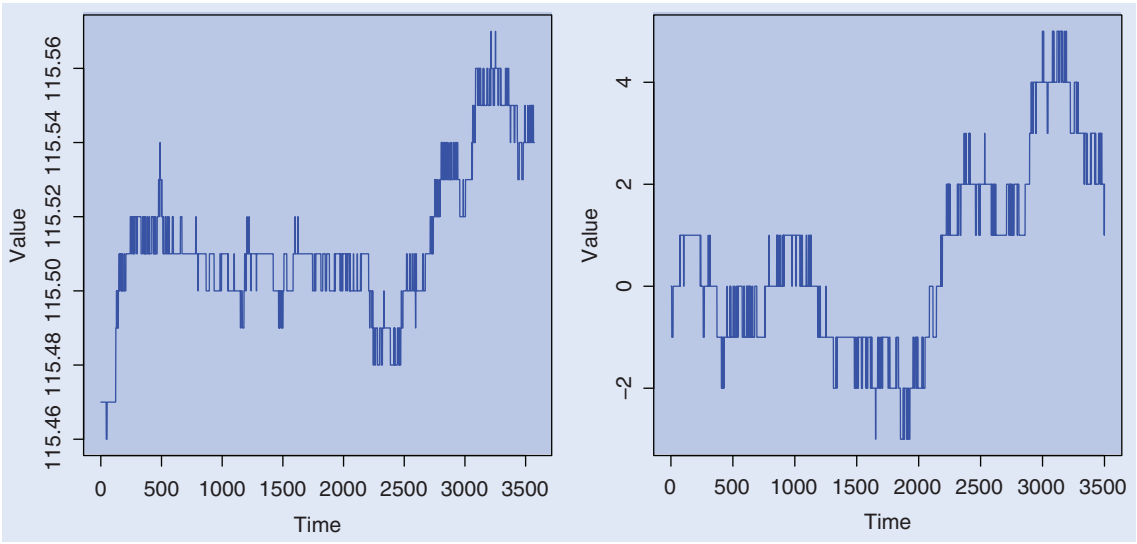

Figure 12. Bund contract, 2007-05-06, from 10 to 11 a.m., one data every second (left), $\lfloor 20W_t \rfloor$, frequency $= 2^{15}$ on [0,0.1] (right).

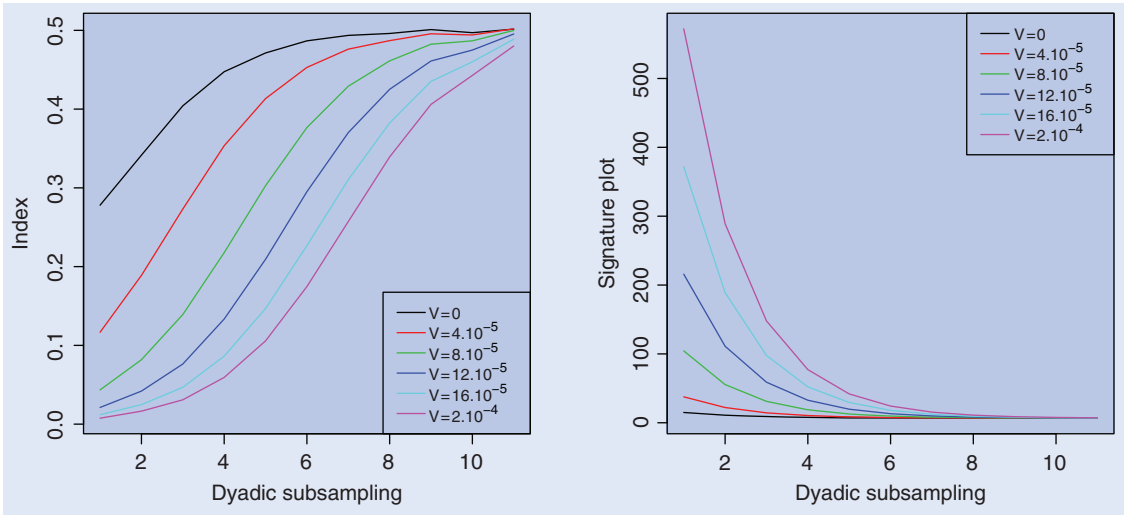

Figure 13. Model M1′, microstructure noise index (left) and signature plot (right).

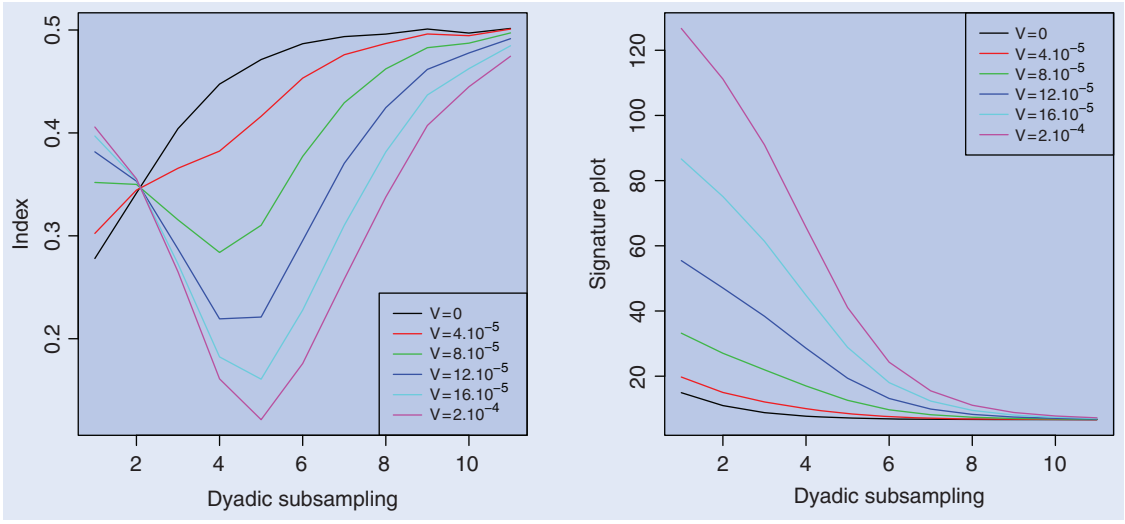

Figure 14. Model M2′, microstructure noise index (left) and signature plot (right).

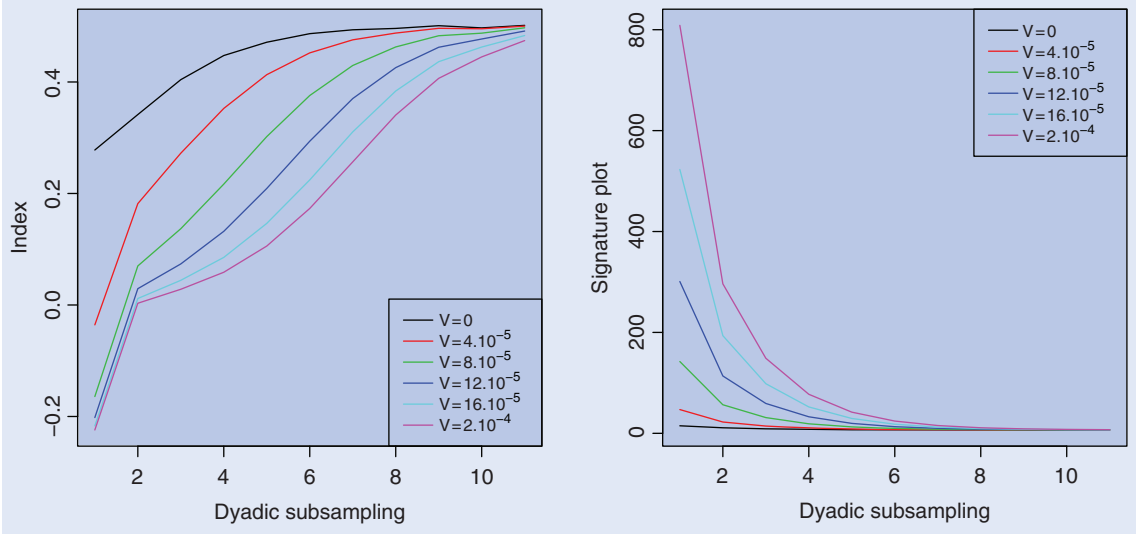

Figure 15. Model M3′, microstructure noise index (left) and signature plot (right).

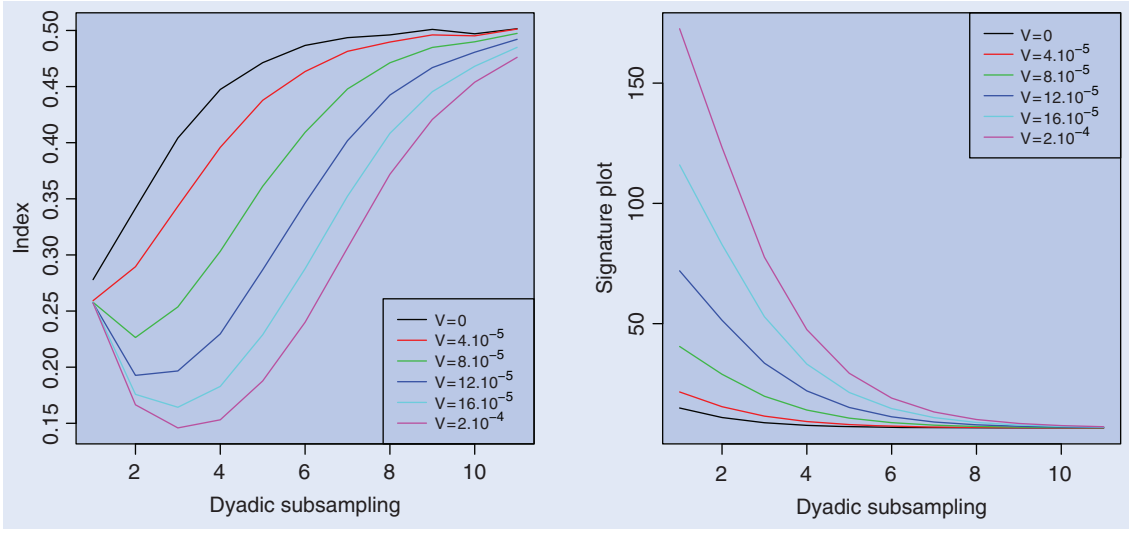

Figure 16. Model M4′, microstructure noise index (left) and signature plot (right).

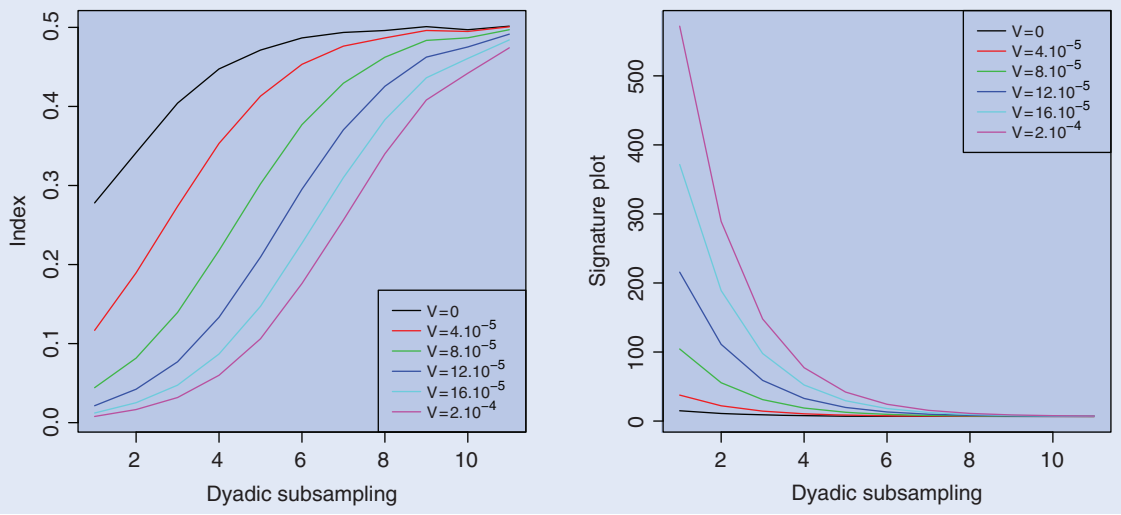

Figure 17. Model M5′, microstructure noise index (left) and signature plot (right).

5.3. *Pure rounding: Theoretical considerations*

Consider the model where the observed (absolute) price $Y_{k\Delta}$ is equal to $X^{(\alpha)}_{k\Delta}$, where X is a sufficiently regular underlying diffusion process. Let m be a subsampling period. At high frequencies (m small), the rounded theoretical price increment,

$$X^{(\alpha)}_{m(k+1)\Delta} - X^{(\alpha)}_{mk\Delta},$$

largely differs from the quantity $X_{m(k+1)\Delta} - X_{m\Delta}$, whereas at low frequencies (m large) the two quantities are very close. This difference appears in our index. We now describe theoretically the behavior of the absolute signature plot and of the absolute microstructure index in the model of a diffusion observed with rounding error. We take here $T=1$ and $\Delta = 1/n$. Let α_n be a positive sequence tending to zero as n goes to infinity. On a filtered probability space $(\Omega, (\mathcal{F}_t)_{t\in[0,1]}, \mathbb{P})$, we consider $(X_t)_{t\in[0,1]}$ of the form

$$X_t = x_0 + \int_0^t \sigma(X_s, s)\,\mathrm{d}W_s + \int_0^t a_s\,\mathrm{d}s, \tag{15}$$

where $(W_t)_{t\in[0,1]}$ is a $(\mathcal{F}_t)$-standard Brownian motion, $(a_t)_{t\in[0,1]}$ a progressively measurable process with respect to $(\mathcal{F}_t)_{t\in[0,1]}$, $(x,y)\to\sigma(x,t)$ a real deterministic function and x_0 a real constant. We observe the sample

$$(X^{(\alpha_n)}_{i/n}, i = 0, \ldots, n), \tag{16}$$

where

$$X^{(\alpha_n)}_{i/n} = \alpha_n \lfloor X_{i/n}/\alpha_n \rfloor,$$

and α_n tends to zero. We denote by $\mathcal{C}^k(I)$ the set of k times continuously differentiable functions on $I\subseteq\mathbb{R}$. We assume that $\sigma(x,t) = g_1(x)g_2(t)$, with g_1 and g_2 two positive functions such that $g_1\in\mathcal{C}^2(\mathbb{R})$ and $g_2\in\mathcal{C}^1([0,1])$. Let $\beta_n = \alpha_n\sqrt{n}$,

$$\gamma(\sigma,\beta) = \int_0^1 \mathrm{d}u \int_{\mathbb{R}} \mathrm{d}y\, h(y)((\beta u + \sigma y)^{(\beta)})^2,$$

where h denotes the density of a standard Gaussian variable and

$$\tilde{V}^p_n = \sum_{k=0}^{n-1} \left| X^{(\alpha_n)}_{(k+1)/n} - X^{(\alpha_n)}_{k/n} \right|^p.$$

From Delattre (1997) and Rosenbaum (2007, 2009b) we have the following result.

Theorem 5.1: *In model* (15) *and* (16),

$$\tilde{V}^2_n = \int_0^1 \gamma(\sigma(X_t,t),\beta_n)\,\mathrm{d}t + \mathcal{O}_p(\alpha_n \vee n^{-1/2}).$$

Moreover, as β_n *tends to zero,*

$$\gamma(\sigma(X_t,t),\beta_n) \sim \sigma(X_t,t)^2 + \beta_n^2/6,$$

and as β_n *tends to infinity,*

$$\gamma(\sigma(X_t,t),\beta_n) \sim \sqrt{\frac{2}{\pi}}\sigma(X_t,t)\beta_n.$$

From the preceding theorem, we immediately obtain the following corollary.

Corollary 5.2: *Assume that our data are generated by model* (15) *and* (16) *and suppose that* $\sigma(X_t,t)=\sigma$. *If* $\beta_n\to+\infty$,

$$S^2_q \to_p 1/4.$$

If $\beta_n\to 0$,

$$S^2_q = 1/2 - \frac{\alpha_n^2 n 2^{-q}}{12\sigma^2} + \mathcal{O}_p(n^{-1/2}2^{-q/2}) + o(\alpha_n^2 n 2^{-q}).$$

Let α be the rounding error on our data. The case where β_n tends to infinity corresponds to the situation where $\alpha(n2^{-q})^{1/2}$ is 'large', that is high frequencies, and the case where β_n tends to zero corresponds to the situation where $\alpha(n2^{-q})^{1/2}$ is 'small', that is low frequencies. Once again, the asymptotic value of the index is 1/2 at low frequencies, agreeing with the empirical results. For high frequencies, the asymptotic value is 1/4 in the pure

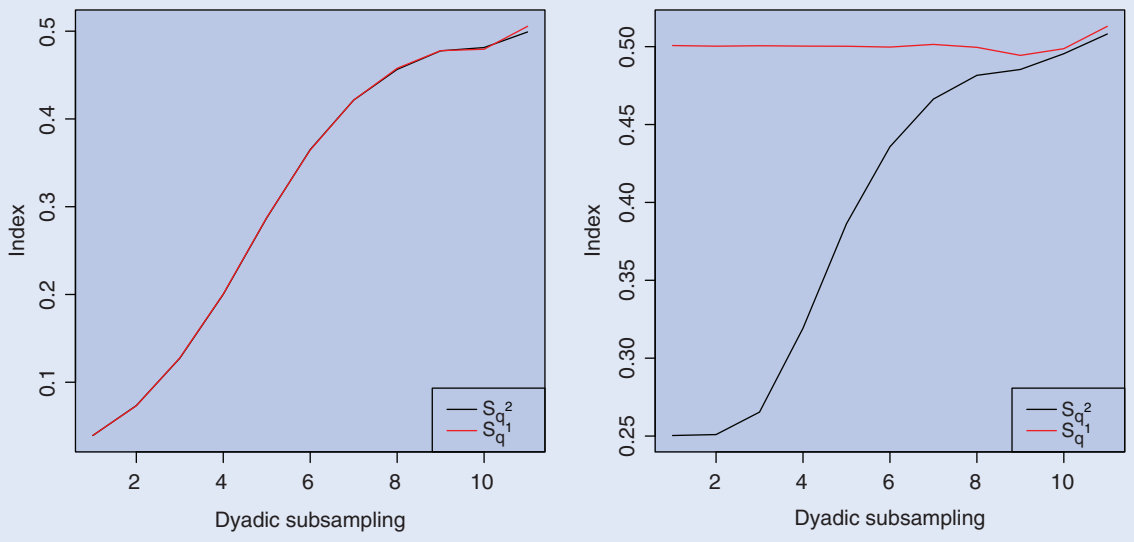

Figure 18. Model M1, microstructure noise indexes S_q^1 and S_q^2 (left) and model M1′, microstructure noise indexes S_q^1 and S_q^2 (right).

rounding case. This is quite close to the values obtained for the empirical graphs of the Bund for the frequency where the curves start to increase.

6. Complementary microstructure functions

6.1. Theoretical considerations

We present in this section results based on the behavior of the 1-variation V_q^1. Theoretical results on the 1-variation for discretely observed Ito semi-martingales are given by Jacod (2008). The 1-variation is less commonly used than the quadratic variation but, in our context, it has the advantage of being less impacted by the rounding effect than the quadratic variation. Indeed, from Delattre (1997) and Rosenbaum (2007, 2009b), we have the following theorem, which is a very specific property of the rounding error.

Theorem 6.1: *In model* (15) *and* (16),

$$\frac{1}{\sqrt{n}}\tilde{V}_n^1 = \sqrt{\frac{2}{\pi}}\int_0^1 \sigma(X_t,t)\,\mathrm{d}t + \mathcal{O}_p(\alpha_n \vee n^{-1/2}).$$

This result has a very interesting consequence for the microstructure noise index.

Corollary 6.2: *Assume that our data are generated by model* (15) *and* (16), *then, for all q, as* $n \to +\infty$,

$$S_q^1 = 1/2 + \mathcal{O}_p(\alpha_n \vee n^{-1/2}).$$

This point is very particular. Indeed, in the case of an additive noise, roughly speaking, S_q^1 and S_q^2 exhibit the same behavior. This is no longer the case for the rounding error. We illustrate the point using the graphs in figure 18, which are the averages of 50 computations of S_q^1 and S_q^2 using model M1 with $V = 4 \times 10^{-5}$ and model M1′ with $V = 0$.

6.2. Data analysis

Figures 19–21 show graphs of $q \to S_q^1$ and $q \to (\pi/2)^{1/2}(n2^{-q})^{-1/2}V_q^1$ for the last traded price for the Bund. The graphs for the bid price can be found in Rosenbaum (2007). Note that, in the context of rounding, one may believe that the appropriate quantity to consider is the bid price. Indeed, if one assumes that the theoretical price lies between the bid price and the ask price, and that the bid–ask spread is constant and equal to one tick,† then the bid price is almost surely the right measure of the rounded theoretical price.

6.3. Comments, perspectives and conjectures

6.3.1. Rounding and additive error. In this section, we form conjectures concerning the behavior of the 1-variation under rounding and additive error. The theoretical part remains to be done. The behavior of the 1-variation-based signature plot is quite surprising. Indeed, a shape similar to those for the realized volatility-based signature plot would have been expected. Note that the increasing part of the graph is even more pronounced for the bid prices. To understand this fact, we consider that there are two sources of error. First an additive error and then a rounding error. Let α_n be a sequence tending to zero. We consider the following toy model for the observed price:

$$Y_{i/n} = (\sigma W_{i/n} + V\varepsilon_i^n)^{(\alpha_n)}, \quad i = 0, \ldots, n, \qquad (17)$$

†This is (almost) the case for the Bund.

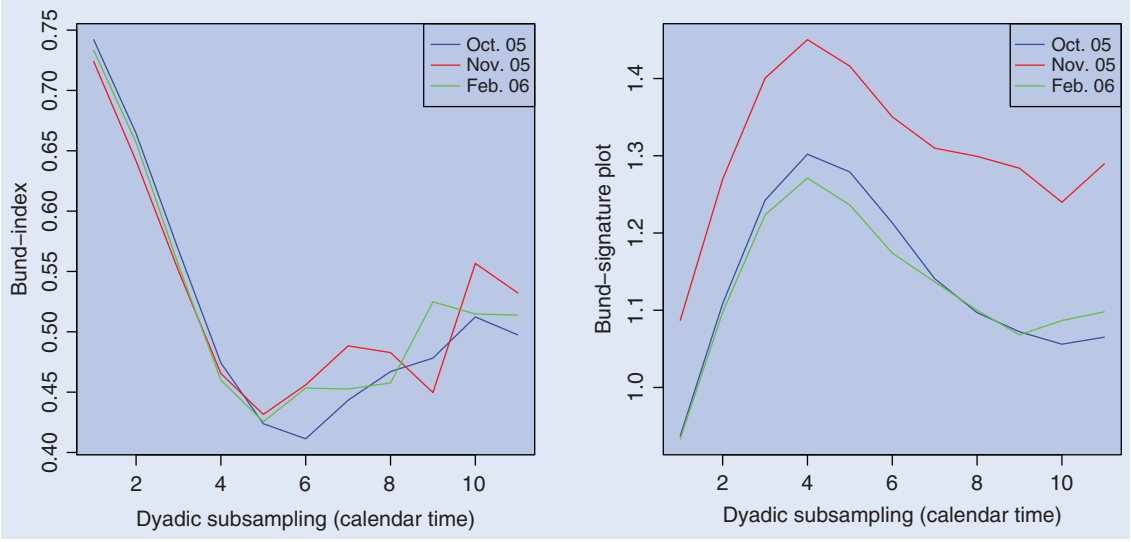

Figure 19. Microstructure noise index S_q^1 (left) and signature plot for $p=1$ (right), for the Bund, last traded price, Oct. 05, Nov. 05, Feb. 06.

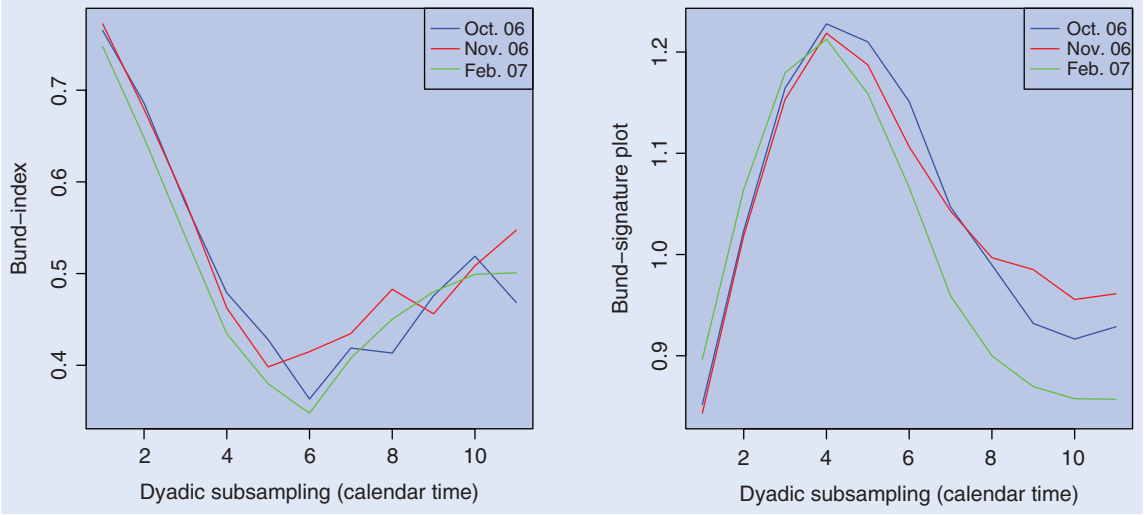

Figure 20. Microstructure noise index S_q^1 (left) and signature plot for $p=1$ (right), for the Bund, last traded price, Oct. 06, Nov. 06, Feb. 07.

where ε_i^n are centered Gaussian variables, independent of W and such that $\gamma_\varepsilon^{n,m} = \mathbb{E}[(\varepsilon_{mi}^n - \varepsilon_{m(i-1)}^n)^2]$ does not depend on i. Let

$$\mathcal{V}(m,n) = (\pi/2)^{1/2}\sqrt{\frac{m}{n}}\sum_{i=0}^{n/m-1}|Y_{(i+1)m/n} - Y_{im/n}|. \qquad (18)$$

Assume that as $n \to +\infty$, $\mathcal{V}(m,n)$ converges in probability to its expectation if there is no rounding error, that is

$$\mathcal{V}(m,n)^2 \sim \sigma^2 + V^2\frac{n}{m}\gamma_\varepsilon^{n,m}.$$

We want to reproduce the behavior of the 1-variation-based signature plot. We focus on the increasing part of the graph. First note that we cannot obtain it if $\gamma_{\varepsilon,W}^{n,m} = 0$ and if $\gamma_\varepsilon^{n,m} = c$. We need that the following inequality holds:

$$\gamma_\varepsilon^{n,m} < \frac{1}{2}\gamma_\varepsilon^{n,2m}.$$

This inequality cannot be obtained if ε_i^n follow an $AR(1)$ model. However, if we assume that there exists a continuous noise process $\tilde{\varepsilon}_t$ such that for all n, m and i, $\varepsilon_{mi}^n = \tilde{\varepsilon}_{mi/n}$, if $\tilde{\varepsilon}_t = cB_t^H$, with B_t^H a fractional Brownian motion,† then

$$\varepsilon_{mi}^n - \varepsilon_{m(i-1)}^n = c\left(B_{mi/n}^H - B_{m(i-1)/n}^H\right)$$

†Once again, a stationary process such as a fractional Ornstein–Uhlenbeck process is probably more suitable, but with the theoretical part of this section remaining, we only focus on very simple examples.

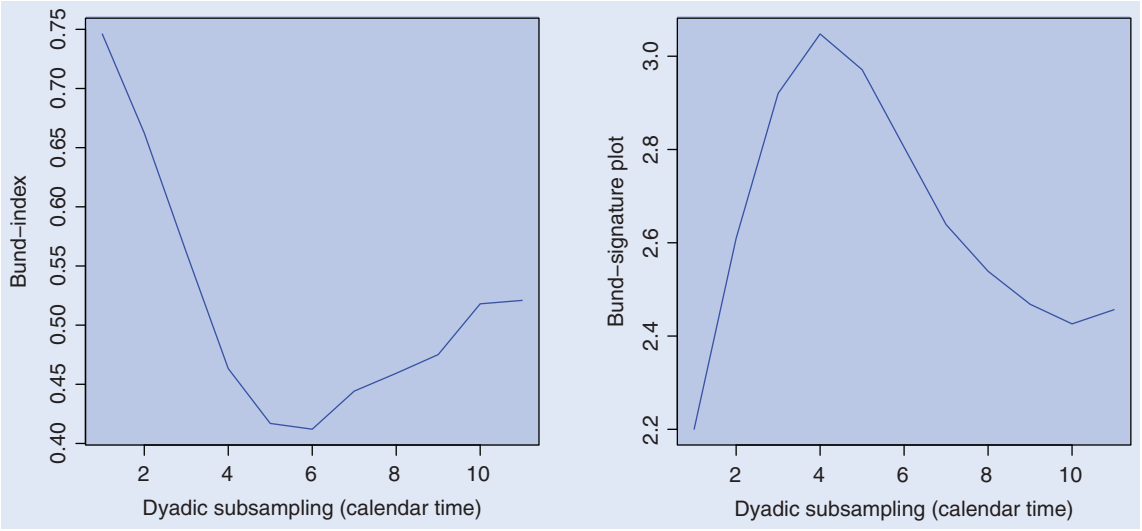

Figure 21. Microstructure noise index S_q^1 (left) and signature plot for $p=1$ (right), for the Bund, last traded price, aggregated data: Oct. 05, Nov. 05, Feb. 06, Oct. 06, Nov. 06, Feb. 07.

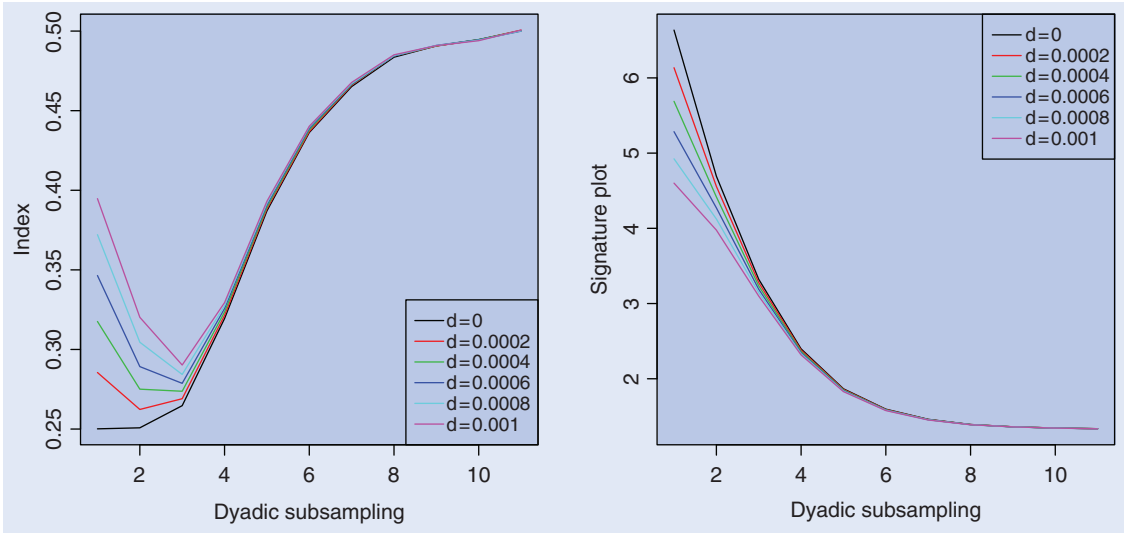

Figure 22. Microstructure noise index S_q^2 (left) and signature plot for $p=2$ (right).

and

$$\varepsilon^n_{2mi} - \varepsilon^n_{2m(i-1)} = c\left(B^H_{2mi/n} - B^H_{2m(i-1)/n}\right).$$

Hence the inequality is true as soon as $H>1/2$. It is important to remark also that the increasing behavior of the beginning of the 1-variation-based signature plot is not incompatible with a decreasing realized volatility-based signature plot.

6.3.2. Model with uncertainty zones. A drawback of the previous model is that, for very high frequencies, because of the local time of the semi-martingale together with the rounding effect, the observed price 'jumps a lot'. One can probably deal with this using, for example, non-centered noise. We present here a (perhaps) more natural model to treat the local time problem, allowing for ultra-high-frequency persistence. The observed price is Y and the theoretical price is X. We define our model as follows. For a fixed parameter d,

- if $X^{(\alpha)}_{i/n} \neq Y_{(i-1)/n}$, then, if $|X_{i/n} - X^{(\alpha)}_{i/n}| < d$, $Y_{i/n} = Y_{(i-1)/n}$ else $Y_{i/n} = X^{(\alpha)}_{i/n}$,
- if $X^{(\alpha)}_{i/n} = Y_{(i-1)/n}$, then $Y_{i/n} = X^{(\alpha)}_{i/n}$.

The observed price can then change only if the theoretical price is out of the *uncertainty zone* of width d above its rounded value. We obtain the graphs shown in figures 22 and 23 for the microstructure noise indexes and signature plots for the model where the theoretical price is

$$X_t = c\exp\sigma W_t, \tag{19}$$

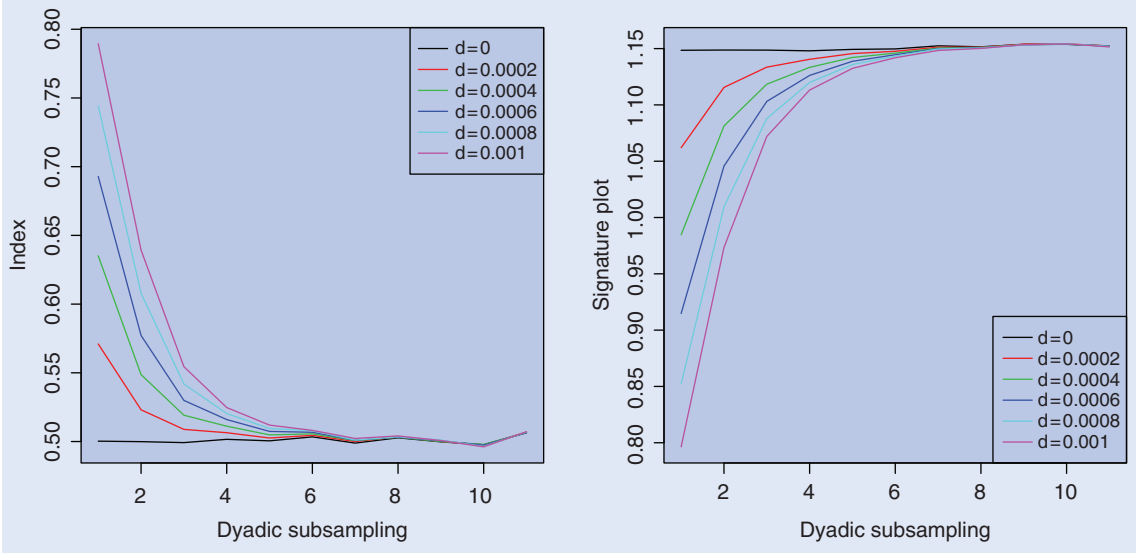

Figure 23. Microstructure noise index S_q^1 (left) and signature plot for $p=1$ (right).

with c and σ defined as previously (average over 50 simulations).

The decreasing part of the 1-variation-based signature plot can probably be obtained by introducing correlation between the price increments and the size of the uncertainty zones. Note that this model with uncertainty zones has been extended and made more sophisticated by Robert and Rosenbaum (2009a, b).

Acknowledgements

I am grateful to Nour Meddahi and Marc Hoffmann for helpful discussions and comments. I also thank Renaud Drappier, Jean-Marc Duprat, Patrick Guével, Nicolas Michon, Ian Solliec and Sebouh Takvorian from BNP-Paribas for providing and discussing the data. This work was partially supported by BNP-Paribas, FIRST-ETG (Fixed Income Research Strategic Team-Electronic Trading Group).

References

Aït-Sahalia, Y. and Jacod, J., Testing for jumps in a discretely observed process. *Ann. Statist.*, 2009, **37**, 184–222.

Aït-Sahalia, Y., Mykland, P.A. and Zhang, L., How often to sample a continuous time process in the presence of market microstructure noise. *Rev. Financial Stud.*, 2005, **18**, 351–416.

Aït-Sahalia, Y., Mykland, P.A. and Zhang, L., Ultra high frequency estimation with dependent microstructure noise. *J. Econometr.*, 2009, in press.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P., Great realizations. *Risk*, 2000, **13**, 105–108.

Andersen, T., Bollerslev, T. and Meddahi, N., Realized volatility forecasting and market microstructure noise. *J. Econometr.*, 2009, in press.

Bandi, F.M. and Russel, J.R., Microstructure noise, realized variance and optimal sampling. *Rev. Econ. Stud.*, 2008, **75**, 339–369.

Barndorff-Nielsen, O.E. and Shephard, N., Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *JRSS B*, 2002, **64**, 253–280.

Ciesielski, Z., Kerkyacharian, G. and Roynette, B., Quelques espaces fonctionnels associés à des processus gaussiens. *Stud. Math.*, 1993, **107**, 172–204.

Delattre, S., Estimation du coefficient de diffusion d'un processus de diffusion avec erreurs d'arrondi. PhD thesis, University of Paris, 1997.

Delattre, S. and Jacod, J., A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli*, 1997, **3**, 1–28.

Gloter, A. and Jacod, J., Diffusions with measurement errors, I. Local asymptotic normality. II. Optimal estimators. *ESAIM PS*, 1997, **5**, 225–260.

Gonçalves, S. and Meddahi, N., Bootstrapping realized volatility. *Econometrica*, 2009, **77**, 283–306.

Hansen, P.R. and Lunde, A., Realized variance and market microstructure noise. *J. Business Econ. Statist.*, 2006, **24**, 127–161.

Jacod, J., Asymptotic properties of realized power variations and related functionals of semimartingales. *Stoch. Proc. Appl.*, 2008, **118**, 517–559.

Jacod, J. and Protter, P., Asymptotic error distributions for the Euler method for stochastic differential equations. *Ann. Probab.*, 1998, **26**, 267–307.

Jacod, J., Li, Y., Mykland, P.A., Podolskij, M. and Vetter, M., Microstructure noise in the continuous case: The pre-averaging approach. *Stoch. Proc. Appl.*, 2009, **119**, 2249–2276.

Li, Y. and Mykland, P., Are volatility estimators robust with respect to modeling assumptions? *Bernoulli*, 2007, **13**, 601–622.

Meddahi, N., A theoretical comparison between integrated and realized volatility. *J. Appl. Econometr.*, 2002, **17**, 475–508.

Robert, C.Y. and Rosenbaum, M., A new approach for the dynamics of ultra high frequency data: the model with uncertainty zones. Working Paper, 2009a (to appear in *J. Financ. Econom.*).

Robert, C.Y. and Rosenbaum, M., Volatility and covariation estimation when microstructure noise and trading times are endogenous. Working Paper, 2009b (to appear in *Math. Finan.*).

Rosenbaum, M., Étude de quelques problèmes d'estimation statistique en finance. PhD thesis, 2007.

Rosenbaum, M., First order p-variations and Besov spaces. *Stat. Probab. Lett.*, 2009a, **79**, 55–62.

Rosenbaum, M., Integrated volatility and round off error. *Bernoulli*, 2009b, **15**, 687–720.
Zhang, L., Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach. *Bernoulli*, 2006, **12**, 1019–1043.
Zhang, L., Mykland, P.A. and Aït-Sahalia, Y., A tale of two time scales: determining integrated volatility with noisy high-frequency data. *J. Am. Statist. Assoc.*, 2005, **100**, 1394–1411.

Appendix A: Besov spaces

Let Δ_h^n be the operator defined by $\Delta_h^1 f(x) = f(x+h) - f(x)$ and $\Delta_h^n f(x) = \Delta_h^1(\Delta_h^{n-1}) f(x)$. The nth order L^p modulus of smoothness of f on $[0,1]$ is

$$\omega_n(f,t)_p = \sup_{|h|\le t} \|\Delta_h^n f\|_{\mathrm{L}^p(\Omega_{h,n})},$$

where $\Omega_{h,n} = \{x \in [0,1];\ x + kh \in [0,1],\ k = 0,\dots,n\}$. For $1 \le p \le +\infty$, $s > 0$, the Besov space $\mathcal{B}^s_{p,\infty}([0,1])$ consists of those functions $f \in \mathrm{L}^p[0,1]$ such that

$$\sup_{j\ge 0}\left\{2^{sj}\omega_n(f,2^{-j})_p\right\}_{j\ge 0} < \infty,$$

where $n \in \mathbb{N}$ and $s < n$. It is a Banach space when equipped with the norm

$$\|f\|_{\mathcal{B}^s_{p,\infty}([0,1])} = \|f\|_{\mathrm{L}^p} + \sup_{j\ge 0}\left\{2^{sj}\omega_n(f,2^{-j})_p\right\}_{j\ge 0}.$$

For a real function f on $[0,1]$ and $0 < p < +\infty$, $J_j^p(f)$ is defined by

$$J_j^p(f) = \sum_{k=1}^{2^j} |f(k2^{-j}) - f(\{k-1\}2^{-j})|^p.$$

For $p = +\infty$ we consider the following usual modification: for a sequence $(a_j)_{j\ge 0}$ real, $(\sum_j |a_j|^p)^{1/p} = \sup_j |a_j|$ for $p = +\infty$, we have the following results (Ciesielski *et al.* 1993, Rosenbaum 2009a).

Theorem A.1: *Let $0 < s < 1$, $s > 1/p$, $1 \le p \le \infty$. The usual norm on $\mathcal{B}^s_{p,q}([0,1])$ is equivalent to the norm defined by*

$$\|f\| = \max\left\{|f(0)|,\ \sup_{j\ge 0} 2^{j(s-1/p)}\{J_j^p(f)\}^{1/p}\right\}.$$

Appendix B: Proof of theorem 4.1

Proof: With a slight abuse of notation, we write ε_k for ε_k^Δ:

$$\begin{aligned} Y_{(i+1)m\Delta} - Y_{im\Delta} &= X_{(i+1)m\Delta} - X_{im\Delta} + \varepsilon_{(i+1)m} - \varepsilon_{im} \\ &= X_{(i+1)m\Delta} - X_{im\Delta} + (r^m - 1)\varepsilon_{im} + \gamma\sum_{j=0}^{m-1} r^j \zeta_{m(i+1)-j} \\ &= X_{(i+1)m\Delta} - X_{im\Delta} + (r^m - 1)\varepsilon_{im} + \gamma\sqrt{\frac{1-r^{2m}}{1-r^2}}v_{i,m}, \end{aligned}$$

where $v_{i,m}$, $i \ge 0$ are i.i.d. centered Gaussian variables with variance 1, such that $v_{i,m}$ is independent of the past of ε_j^n, $j \le im$. Let $\mu_{i,m} = (\mathbb{E}[\varepsilon_i^2])^{-1/2}\varepsilon_{im}$. Since $E[\mu_{i,m}\mu_{j,m}] = r^{m|i-j|}$, by Mehler's formula, we have

$$\mathbb{E}\Big[(\mu_{i,m}^2 - 1)(\mu_{j,m}^2 - 1)\Big] = 2r^{2m|i-j|}.$$

Hence,

$$\mathbb{E}\left[\left(\sum_{i=1}^{T(m\Delta)^{-1}} \mu_{i-1,m}^2 - 1\right)^2\right] \le cT(m\Delta)^{-1}.$$

Consequently,

$$\sum_{i=1}^{T(m\Delta)^{-1}} \varepsilon_{(i-1)m}^2 = \mathbb{E}[\varepsilon_i^2]T(m\Delta)^{-1} + T^{1/2}(m\Delta)^{-1/2}R_1,$$

with $E[R_1^2] \le c$. We also readily obtain that

$$\sum_{i=1}^{T(m\Delta)^{-1}} (X_{im\Delta} - X_{(i-1)m\Delta})\varepsilon_{(i-1)m} = cR_2T^{1/2},$$

with $E[R_2^2] \le c$ and

$$\sum_{i=1}^{T(m\Delta)^{-1}} (X_{im\Delta} - X_{(i-1)m\Delta})v_{i,m} = cR_3T^{1/2},$$

with $E[R_3^2] \le c$. Using conditional expectations, we also obtain

$$\sum_{i=1}^{T(m\Delta)^{-1}} \varepsilon_{(i-1)m}v_{i-1,m} = T^{1/2}(m\Delta)^{-1/2}R_4,$$

with $E[R_4^2] \le c$. Finally, using that $\mathbb{E}[\varepsilon_{im}^2] = \gamma^2(1-r^2)^{-1}$, we obtain

$$\begin{aligned} RV(T,\Delta,m) = \sigma^2 &\sum_{i=1}^{T(m\Delta)^{-1}} (W_{im\Delta} - W_{(i-1)m\Delta})^2 \\ &+ 2\gamma^2\frac{1-r^m}{1-r^2}T(m\Delta)^{-1} \\ &+ T^{1/2}(m\Delta)^{-1/2}R_5 + R_6T^{1/2}, \end{aligned}$$

with $E[R_5^2 + R_6^2] \le c$. The result follows in the case where T is fixed and Δ tends to zero. For the second kind of asymptotics, as Δ/T tends to zero, the Lindeberg condition holds in order to prove that

$$\begin{aligned} &2^{-1/2}T_n^{-1/2}(m\Delta_n)^{1/2} \\ &\times \sum_{i=1}^{T_n(m\Delta_n)^{-1}} \left\{(m\Delta_n)^{-1}\left(W_{im\Delta_n} - W_{(i-1)m\Delta_n}\right)^2 - 1\right\} \to_{\mathcal{L}} \mathcal{N}(0,1). \end{aligned}$$

□