Available online at www.sciencedirect.com

SciVerse ScienceDirect

Stochastic Processes and their Applications 123 (2013) 2728–2751

stochastic processes and their applications

www.elsevier.com/locate/spa

Measuring the relevance of the microstructure noise in financial data

Cecilia Mancini

Dipartimento di Scienze per l'Economia e l'Impresa, Università di Firenze, Italy

Available online 16 April 2013

Abstract

We show that the Truncated Realized Variance (TRV) of a SemiMartingale (SM) converges to zero when observations are contaminated by noise. Under the additive i.i.d. noise assumption, a central limit theorem is also proved. In consequence it is possible to construct a feasible test allowing us to measure, for a given path of a given data generating process at a given observation frequency, the relevance of the noise in the data when we want to estimate the *efficient* process integrated variance IV. We thus can optimally select the observation frequency at which we can "safely" use TRV. The performance of our test is verified on simulated data. We are especially interested in the application of the test to financial data, and a comparison conducted with Bandi and Russel (2008) and Ait-Sahalia, Mykland and Zhang (2005) mean square error criteria shows that, in order to estimate IV, in many cases we can rely on TRV for lower observation frequencies than previously indicated when using Realized Variance (RV). The advantages of our method are at least two: on the one hand the underlying model for the efficient data generating process is less restrictive in that jumps are allowed (in the form of an Itô SM). On the other hand our criterion is pathwise, rather than based on an average estimation error, allowing for a more precise estimation of IV because the choice of the optimal frequency is based on the observed path. Further analysis on both simulated and empirical financial data is conducted in Lorenzini (2012) [15] and is also still in progress.

Keywords: Semimartingales with jumps; Noisy data; Integrated variance; Threshold estimation; Test to select optimal sampling frequency; Financial asset prices

E-mail address: cecilia.mancini@unifi.it.

0304-4149/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.spa.2013.04.003

1. Introduction

We can observe noisy data $Y_{t_i} = X_{t_i} + \varepsilon_{t_i}$ of an *efficient* data generating process (DGP) X, which is assumed to be an Itô semimartingale (SM) with continuous martingale part $\int \sigma_s dW_s$. If we want to estimate the *integrated variance* $IV \doteq \int_0^T \sigma_s^2 ds$ of X, we have to decide whether to use an estimator explicitly accounting for the contribution of the noise process ε (e.g. by pre-averaging the observations as in [19]) or to directly apply an estimator which is consistent in the absence of noise. This depends on whether the noise is relevant or not in our data, which is determined by the magnitude of $\mathrm{Var}(\varepsilon_{t_i})$ but also by the frequency at which we pick the observations. We are especially interested in the application of our results to financial data, where X_{t_i} represents the logarithm of the *efficient* price of an asset at time t_i and $\varepsilon_{t_i}, i = 1..n$ are called *microstructure noises*. Given a time series generated by a Brownian semimartingale (BSM), i.e. a SM without jumps, it is well known that the realized variance (RV_h) converges to IV, as the observation frequency h tends to 0. If the BSM observations are noisy, we can look at the signature plot (SP) of the realized variance as a function of h to decide whether at a predetermined frequency the noise contamination is relevant or not [8]: when the noise is judged to be negligible, we rely on RV_h as a measure of IV. However the observation step $\hat{h}$ visually selected by means of the SP is not necessarily such that $RV_{\hat{h}}$ delivers a reliable estimate of IV, given that $RV_{\hat{h}}$ cannot disentangle the estimation error due to the choice of a too large h from the error induced by the presence of the noise. Moreover, in the presence of jumps in the DGP, RV_h undergoes a further source of estimation bias of IV, represented by the sum of the squared jumps. Another important criterion used to establish the limit frequency at which the noise can be neglected is theoretically minimizing the conditional mean square estimation error $RV_h - IV$, as described in [4,3,22]. However also in this case X is assumed to have continuous paths. Further, the selected h is optimal on average, along many paths of the price process, while it is possible that the optimal step for a given day is different from the step which is optimal in another day. This makes it useful to have a further tool allowing to establish, for a fixed path of a fixed asset and a given frequency, whether the noise is contaminating the asset returns in a non-negligible way or not. We are thus going to propose a test and to check its reliability on simulated data. The application to empirical financial data has been done in [15] and is also still in progress. Questions that we judge to be interesting are (1) checking whether, as stated by some authors (as in [20]), the mid-quotes are less affected by noise than the transaction prices and at which extent; (2) for a given high frequency, checking how much, when pre-averaging the data, the (normalized) pre-averaged time series has been decontaminated by the noise; (3) for an observation frequency at which the noise is judged to be negligible by our test, comparing the performances of TRV and pre-averaged TRV.

The paper is organized as follows: Section 2 draws the framework we are considering, Section 3 contains the main results allowing to construct out test, Section 4 illustrates how the test is constructed and how it works. In Section 5 we implement the test on simulated data in order to check whether its responses are reliable, meaning that when the test judges the noise to be relevant then the estimation error $\hat{IV}_h - IV$ is high while it is low otherwise. Appendix contains the proofs of all the results stated in this paper.

2. Model setup

For a fixed $T \in \mathbb{R}$ let us consider the filtered probability space $\mathcal{S}^0 = (\Omega^0, \mathcal{F}^0, \mathcal{F}^0_{t\in[0,T]}, P^0)$ generated by a Brownian motion W and a Poisson random measure μ (possibly allowing for infinite activity jumps), and let the log price of an asset be modeled as an Itô semimartingale X on $\mathcal{S}^0$. We can always arrange the different components of X so as $X = X_0 + J$, where $X_{0t} = x_0 + \int_0^t a_s ds + \int_0^t \sigma_s \, dW_s$, with cadlag integrands a and σ, and J has the following representation (see [11])

$$J_t = J_{1t} + \tilde{J}_{2t} = \int_0^t \int_{|\gamma(x,s,\omega)|>1} \gamma(x,s,\omega)\mu(dx,ds) + \int_0^t \int_{|\gamma(x,s,\omega)|\le 1} \gamma(x,s,\omega)[\mu(dx,ds) - dxds],$$

where $\int 1 \wedge \gamma^2(x,s,\omega)dx$ is a.s. finite. As such X_0 is called *Brownian semimartingale* (BSM) component of X and J *jump component*. Process J_1 is of *finite activity* of jump (i.e. almost all paths jump finitely many times in $[0,T]$), and it also has the representation

$$J_{1t}(\omega), = \sum_{s\le N_t(\omega)} \gamma(x_s,s,\omega)$$

where $N_t = \int_0^t \int_{|\gamma(x_s,s,\omega)|>1} 1\mu(dx,ds)$ is the counting measure of the jumps with size larger than 1 in absolute value and, for fixed ω if a jump occurs at time s then $x_s \in \mathbb{R}$ is the mark pointing at which jump size $\gamma(x_s,s,\omega)$ is realized. On the contrary, in general $\tilde{J}_2$ has *infinite activity* (some path can jump infinitely many times, even densely, in any finite time interval). Let ε be a *noise* process, defined on an extension $\mathcal{S} := (\Omega, \mathcal{F}, \mathcal{F}_{t\in[0,T]}, P)$ given as in [12]. We can only observe the noisy process $Y = X+\varepsilon$, which is the superposition of the *efficient* price process X with the contaminating noise. We have observations Y_{t_i} at discrete times $t_i = ih$, $i = 1..n$, for a given resolution $h = T/n$.

Define $r_h := h^\beta$, with $\beta \in (0,1)$, and $\hat{IV}_h := \sum_{i=1}^n (\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}$, where, for any process Z, $\Delta_i Z = Z_{t_i} - Z_{t_{i-1}}$. The following further notation is used throughout the paper: $\Delta_i Z_\star = \Delta_i Z I_{\{|\Delta_i Z|\le\sqrt{r_h}\}}$, for any process $Z, RV_h(Z) := \sum_{i=1}^n (\Delta_i Z)^2; RV_h = RV_h(Y)$; $\hat{IV}_h(Z) = \sum_{i=1}^n (\Delta_i Z_\star)^2$; for any Itô SM X as above, $\mathrm{QV}(X) = \int_0^T \sigma_s^2 ds + \sum_{s\le T} (\Delta J_s)^2$, where $\Delta J_s = J_s - J_{s-}$; $\varepsilon_i := \varepsilon_{t_i}$, $\mathcal{N}(0,b^2)$ denotes the Gaussian law having mean 0 and variance b^2; U denotes a standard Gaussian r.v.; c indicates a constant which does not depend on i, nor on n, and which keeps the same name even if it can change from line to line; $\approx$ denotes approximation of numerical results in computations; the asymptotic theory is conducted for $n \to \infty$, i.e. $h = T/n \to 0$. In view of the one to one correspondence between n and h, if f is written as a function of h (or alternatively of n) we indifferently indicate either $\lim_h f(h)$ or $\lim_n f(h)$; given two real functions f, g (possibly the paths of a stochastic process for a fixed ω), $f(h) \sim g(h)$ means asymptotic equivalence as $h \to 0$, i.e. there exist constants c_1, c_2 such that $c_1 \le \frac{f(h)}{g(h)} \le c_2$ keeps true when $h \to 0$, meaning that if f and g converge (or diverge), they do at the same speed; $f(h) \ll g(h)$ means that $f(h) = o(g(h))$.

Assumption 1. $\forall i = 1..n$, $P\{|\Delta_i \varepsilon| \le c\sqrt{r_h}\} = O(\sqrt{r_h})$.

Assumption 2. (i) $\int_{|\gamma|>1} 1dx$ is locally bounded in (t,ω);
(ii) there exists $\alpha \in]0,2[$ such that $\int_{|\gamma|\le\varepsilon} \gamma^2 dx \le c\varepsilon^{2-\alpha}$.

Remarks. (1) Assumption 1 is verified if e.g. all the ε_i are normally $\mathcal{N}(0,c^2)$ or uniformly $\mathcal{U}[-c,c]$ distributed for $i = 1..n$, for all n (which are the typical examples of additive i.i.d.

noise). More generally it is verified each time that the law $\Delta_i\varepsilon(P)$ does not depend on i and n, $\sigma_\varepsilon^2 \doteq \mathrm{Var}(\varepsilon_i) \neq 0$ and $\Delta_i\varepsilon(P)$ has a density f which is continuous at 0. Note that the assumption requires that the probability that an increment of the noise process be small tends to 0 when the observation step $h \to 0$. This is consistent with the idea that when h is very small, while the efficient price increments tend to zero in probability, because X is a semimartingale, the noise increments keep comparatively large, which gives an explanation of why the SP of RV_h increases when $h \to 0$.

(2) Assumption 2(i) is technical and is standard when proving CLTs (e.g. Assumption (K) in [10] implies our condition (i)). Assumption 2(ii) is satisfied when J is Lévy with Blumenthal–Getoor index α or is a semimartingale (with constant Blumenthal–Getoor index α) satisfying e.g. Assumption 2 in [2] (with β there playing the role of α here). The condition is needed to ensure that, for all n, $P\{|\Delta_i X| > \sqrt{r_h}\}$ and $P\{|\Delta_i \tilde{J}_2| > \sqrt{r_h}\}$ keep bounded by $ch^{1-\frac{\alpha\beta}{2}}$, uniformly in $i = 1..n$ (see Lemma A.1 in the Appendix), which is needed in the proof of Theorem 3.1.

3. Main results

Our first important result is showing that in the presence of noises the threshold estimator of IV tends to zero rather than to IV.

Theorem 3.1. *Let $Y = X + \varepsilon$ and take $\beta > 2/3$. Under Assumptions 1 and 2 we have*

$$\hat{IV}_h \xrightarrow{P} 0.$$

The intuition is the following. The increments $\Delta_i\varepsilon$ have the peculiarity that their variance keeps high even when $h \to 0$, which makes process ε to fall outside the semimartingales class. Microstructure noises typically satisfy Assumption 1, because they tend to keep large when $h \to 0$. On the contrary, as previously said, $\Delta_i X$ tends to be small for each i (in particular, under Assumption 2 we have $P\{|\Delta_i X| > \sqrt{r_h}\} \le ch^{1-\alpha\beta/2} \to 0$). It follows that when $h \to 0$ the increment $\Delta_i\varepsilon$ tends to predominate on $\Delta_i X$ and makes $\Delta_i Y$ large for all i, and all $I_{\{(\Delta_i Y)^2 \le r_h\}}$ will turn out to be zero.

Condition $\beta > 2/3$ is used in the proof to show that $E[I_1] \to 0$ (I_1 depends on h, which is omitted). If it was $\beta \le 2/3$, the threshold function could be too high in that too many noise increments would remain below it together with the ones of the Brownian part of Y, and the limit of $\hat{IV}_h$ could not be zero. In fact, in order to include all the frameworks, condition $\beta > 2/3$ is necessary, as if e.g. $Y = X + \varepsilon$ with $X \equiv 0$ and ε_i i.i.d. Gaussian with non-zero variance, then, by using (11) (which does not require $\beta > 2/3$) to prove conditions (a) to (c) described within the proof of Theorem 3.2, with $\phi_i = \Big((\Delta_i\varepsilon_\star)^2 - E[(\Delta_i\varepsilon_\star)^2]\Big)/\sqrt{n\mathrm{Var}\big((\Delta_i\varepsilon_\star)^2\big)}$, we reach that

$$\frac{\sum\limits_i (\Delta_i\varepsilon_\star)^2 - nE[(\Delta_i\varepsilon_\star)^2]}{c\sqrt{n}\, r_h^{\frac{5}{4}}} \xrightarrow{d} U,$$

for all $\beta \in (0, 1)$, meaning that $\sum_i (\Delta_i\varepsilon_\star)^2 \sim nE[(\Delta_i\varepsilon_\star)^2]$ which is of the same order as $nr_h^{3/2}$, and so $\hat{IV}_h(Y) = \hat{IV}_h(\varepsilon)$ tends to zero in probability iff $\beta > 2/3$.

Theorem 3.1 allows us to distinguish whether the observed process is contaminated by (a relevant) noise or not. In fact if $\sigma \not\equiv 0$, when X is not contaminated then $\hat{IV}_h(X) \stackrel{P}{\to} \int_0^T \sigma_s^2 ds > 0$ [16], while if X is contaminated then $\hat{IV}_h \stackrel{P}{\to} 0$. The next theorem enables us to establish confidence intervals for $\hat{IV}_h$ being significantly far form 0. When X models the log-price of a financial asset, its observation is always affected by some microstructure noises, however if $\hat{IV}_h$ turns out to be far from zero, the impact of the noise is as if it was absent, meaning that it is present but negligible, not relevant. This is the logic under which the test we propose in the next section works. In order to construct the mentioned confidence intervals (in Section 4) we need to compute the speed at which $\hat{IV}_h$ tends to zero in the case where X is contaminated, which is exactly the objective of the next theorem. In case where σ is null the next theorem is still valid, as within the proof the condition $\sigma \not\equiv 0$ is never invoked.

Theorem 3.2 (*CLT in the Presence of Additive i.i.d. Noise*)**.** *Assume that for all h the r.v.s $\varepsilon_{t_i}, i = 1..n, n \in \mathbb{N}$, are i.i.d. with zero mean and $0 < \sigma_\varepsilon^2 < \infty$, and are independent on X. Further assume that the law of each ε_{t_i} has (the same) Lipschitz and bounded density g. Then when X is contaminated by the noise and $\beta > 2/3$ we have*

(i)

$$\frac{E[\hat{IV}_h]}{nr_h^{\frac{3}{2}}} \stackrel{P}{\to} \frac{2}{3} E[g(\varepsilon_1)] = \frac{2}{3} \int_{\mathbb{R}} g^2(x) dx$$

(ii)

$$\mathcal{NB}_h := \frac{\hat{IV}_h - nr_h^{3/2} \frac{2}{3} E[g(\varepsilon_1)]}{\sqrt{n}\, r_h^{\frac{5}{4}} \sqrt{\frac{2}{5} E[g(\varepsilon_1)]}} \stackrel{\mathcal{F}_0\text{-stable}}{\to} U,$$

where U is a random variable on an extension $\mathcal{S}' := (\Omega', \mathcal{F}', \mathcal{F}_s', P')$ of $\mathcal{S}$, having standard Gaussian law, and is independent on $\mathcal{S}$.

Remarks. (1) We recognize that assuming i.i.d. noises when $\{Y_{t_i}\}_{i=1..n}$ represent the observed prices of a financial asset is not completely realistic, however, as in many other works following [23], this represents a starting point to understand what one can substantially do.

(2) The above assumptions on ε imply that Assumption 1 is satisfied (see point (5) below and Remark (1) after Assumption 2). The above assumptions on ε are satisfied if e.g. the noise is additive i.i.d. with Gaussian ε_{t_i}.

(3) When ε_{t_i} are i.i.d. with uniform laws, the density g is not Lipschitz over the whole $\mathbb{R}$, however the results (i) and (ii) are proved by using the specific features of the uniform density (see just after the proof of Theorem 3.2).

(4) Condition $\beta > 2/3$ implies, from (i), that $E[\hat{IV}_h] \to 0$.

(5) By the i.i.d. property of the r.v.s ε_i, also the differences $u_i = \varepsilon_i - \varepsilon_{i-1}$ have a common density f, for $i = 1..n$, for all n, and the relation between f and the density g of ε_i is

$$f(z) = \int_{\mathbb{R}} g(z+y) g(y) dy.$$

Consequently $E[g(\varepsilon_1)] = \int_{\mathbb{R}} g^2(y) dy = f(0)$, so we can estimate $E[g(\varepsilon_1)]$ by either making assumptions on the noise density (e.g. Gaussian or uniform) and then using parametric methods (e.g. deducing the value $f(0)$ from estimates of the variance of the noise increments given e.g. in [6, p. 20]) or using non-parametric methods (as kernel-type estimators of $f(0)$). Therefore we can implement a feasible version of $\mathcal{NB}_h$. $\mathrm{Var}(\varepsilon_1) \neq 0$ implies that $f(0) = E[g(\varepsilon_1)] \neq 0$, since $f(0) = E[g(\varepsilon_1)] = \int_{\mathbb{R}} g^2(y)dy$ and g cannot be null.

(6) In [21] a power variation based statistic is proposed to study which kind of noise could realistically affect a given record of observations. The statistic also serves to select an observation frequency at which the impact of the noise can be considered negligible. The theory however is developed for a noised Gaussian process ($X_t = \sigma W_t$).

4. Application: measuring the relevance of the noise in finite samples

In the previous section we obtained that in the presence of noises in the data, if we choose β close to one, then $\sum_{i=1}^n (\Delta_i Y)_\star^2 \xrightarrow{P} 0$, in a way such that Theorem 3.2(ii) holds true, and the feasible version of $\mathcal{NB}_h$, where $E[g(\varepsilon_1)] = f(0)$ is replaced by an estimate $\hat{f}(0)$, tends to a standard Gaussian r.v. On the contrary, in the absence of the noise, since the econometrician believes that some noises always affect the data, he still implements the same feasible version of $\mathcal{NB}_h$, but now we have $\sum_{i=1}^n (\Delta_i Y)_\star^2 \xrightarrow{P} IV \geq 0$ (see [16]) and $\hat{f}_n(0) \to +\infty$ in both the following cases: the case where we use kernel estimation

$$\hat{f}_n(0) = \frac{1}{ns} \sum_{i=1}^n I_{\{|\Delta_i Y| < s\}}$$

with $s = \vartheta\sqrt{h}$, for some constants ϑ, and the case where we believed that ε_i are Gaussian $\mathcal{N}(0, \sigma_\varepsilon^2)$ (or uniform) and estimated $f_n(0)$ through the empirical variance of the increments $\Delta_i Y$ taken at the highest available frequency, by using

$$\hat{f}_n(0) = \frac{1}{\sqrt{2\pi\hat{\sigma}_u^2}}, \qquad \hat{\sigma}_u^2 = \frac{1}{n}\sum_{i=1}^n (\Delta_i Y)^2 - \left(\frac{1}{n}\sum_{i=1}^n \Delta_i Y\right)^2. \tag{1}$$

($\hat{f}_n(0) = 1/\sqrt{6\hat{\sigma}_u^2}$ in the case of uniform noises ε_i). In fact, in both cases we have $\hat{f}_n(0) \sim h^{-1/2}$, which we checked under $X \equiv Y \equiv \sigma W$, by using the Lindeberg–Feller CLT for a 1-dependent sequence forming a triangular array. Therefore $nr_h^{3/2}\hat{f}_n(0) \to +\infty$, while $nr_h^{5/2}\hat{f}_n(0) \to 0$, so $S_h \to -\infty$.

As a consequence, with β close to 1, the following statistic

$$\mathcal{S}_h \doteq \frac{\hat{IV}_h - nr_h^{\frac{3}{2}}\frac{2}{3}\hat{E}[g(\varepsilon_1)]}{\sqrt{n}r_h^{\frac{5}{4}}\sqrt{\frac{2}{5}\hat{E}[g(\varepsilon_1)]}} = \frac{\hat{IV}_h - nr_h^{\frac{3}{2}}\frac{2}{3}\hat{f}_n(0)}{\sqrt{n}r_h^{\frac{5}{4}}\sqrt{\frac{2}{5}\hat{f}_n(0)}} \tag{2}$$

allows us to construct a formal test of the hypotheses

(H_0) presence of the noise, $\qquad (H_1)$ absence of the noise.

In fact, as soon as $\sqrt{n}r_h^{1/4}(f(0) - \hat{f}_n(0)) \to 0$ we have

$$\mathcal{S}_h \begin{cases} \overset{\mathcal{F}_0\text{-stable}}{\to} U & \text{if the noise is present, i.e. under } (H_0) \\ \overset{\text{a.s.}}{\to} -\infty & \text{if the noise is absent, i.e. under } (H_1). \end{cases} \tag{3}$$

Note that if e.g. we use (1) then within the specialized model $Y = \sigma W + \varepsilon$ we have $f(0) - \hat{f}_n(0) \sim h$ so the requirement $\sqrt{n} r_h^{1/4}(f(0) - \hat{f}_n(0)) \to 0$ is fulfilled.

The importance of this test stems from indicating us whether, for a given mesh h, we can rely or not on TRV in order to estimate the IV of X. In practice, financial data are always affected by some microstructure noises, so it is a bit delicate to be willing to test whether the noise is *present* or not. However on finite samples the contamination can be high or low and then it is meaningful to ask whether the noise can be neglected or not in order to estimate IV by using $\hat{IV}$. To give an answer to this last question is exactly our intent, and is made possible by looking at the behavior of $\mathcal{S}_h$: when, given an observation step h, $|\mathcal{S}_h|$ assumes a very large value we are led to think that the data behave like as if the noise was absent, meaning that the effect of the noise is sufficiently low to allow us to estimate IV through $\hat{IV}$. If on the contrary the value assumed by $|\mathcal{S}_h|$ is compatible with a standard Gaussian law, then the noise has to be judged to be relevant and $\hat{IV}$ has to be considered not reliable. The simulations experiments below substantially confirm that when the noise affecting the data has small variance or the observation frequency is low then $|\mathcal{S}_h|$ assumes large values, while it assumes small values otherwise. Thus we can use the magnitude of $|\mathcal{S}_h|$ as an indicator of how negligible is the present noise. The negligibility of the noise contribution to $\hat{IV}$ is measured below by the performance MEE $\doteq 100(\hat{IV} - IV)/IV$ of $\hat{IV}$ in estimating IV.

Note that our test is formulated in a not conventional way, as our hypothesis (H_0) is "presence of noise" rather than "absence of the noise".

The confidence intervals for our test statistic are given using that $P\{|U| > 1.96\} = 5\%$, so that S_h is compatible at the 95% confidence level with a standard normal r.v. if its assumed absolute value is below 1.96, and in such a case (H_0) is accepted and the noise has to be considered relevant. Otherwise, for large values of $|S_h|$ formally (H_0) would be rejected, however in practice we have an indication of the negligibility of the noise.

The test procedure we propose here summarizes as follows:

- estimate $f(0)$ (using a kernel or assuming a distribution for $\Delta_i\varepsilon$ and using the empirical variance of the $\Delta_i Y$ at the highest available frequency)
- RULE: consider the noise relevant at 5% level iff $|\mathcal{S}_h| \leq 1.96$.

We remark that using $\hat{IV}_h$ when possible, rather than applying estimators specifically accounting for the presence of the noise, has an advantage in efficiency. In fact $\hat{IV}_h$ converges at rate $n^{1/2}$, in the absence of the noise, when the jump component J of X has finite variation (see e.g. [17]), while the best rate of an estimator of IV accounting for the noise is $n^{1/4}$. This can make an important difference in finite samples.

By implementing $\mathcal{S}_h$ for different values of h, we can select optimally the observation mesh $\hat{h}$ to be used in order to estimate IV by $\hat{IV}_h$ in the presence of noise. In fact when the observation frequency h is low, the estimation error $\hat{IV}_h - IV$ can be high (even in the absence of the noises), because the theory asserts that $\hat{IV}_h \to IV$ when $h \to 0$. On the contrary, when the frequency is very high, $\hat{IV}_h$ tends to zero and not to IV, in fact RV_h would explode to infinity. We are thus proposing an alternative criterion to the ones proposed so far in the financial econometrics literature, namely the visual inspection of the SP of RV_h [8] or the minimization of the conditional (on σ) mean squared error (MSE) of $RV_h - IV$ [4,3,22]. SP is not necessarily such that $RV_{\hat{h}}$ delivers a reliable estimate of IV, given that $RV_{\hat{h}}$ cannot disentangle the estimation error due to choice of a too large h from the error induced by the presence of the noise. Furthermore both the SP and the MSE criterions are designed under the assumption that X has continuous paths, while in the presence of jumps RV_h undergoes a further source of estimation bias of IV, represented by the sum of the squared jumps. Moreover, for the MSE criterion the selected h is optimal on average, i.e. along many paths of the price process, while it is possible that the optimal frequency for a given day is different from the frequency which is optimal in another day. Our approach allows to establish the optimal observation mesh for any fixed path of a fixed asset and also in the presence of jumps in X.

In practice, for financial data realistically the noise variance can be different for different sampling frequencies. In a further paper (see in the meanwhile [18]) we study in simple frameworks how the test response changes in this case. Indicated by $\varepsilon_i^{(n)}$ the noise which is involved in the observations sampled at frequency $h = T/n$, we allow that when n changes the noise variance can change, but, for fixed n, $\mathrm{Var}(\varepsilon_i^{(n)})$ is the same for all $i = 1..n$. We separately tackle the case where $\rho_n := \mathrm{Var}(\varepsilon_1^{(n)}) \to \rho > 0$ and the one where $\rho_n \to 0$. The first case is probably the most realistic, and the test has theoretically the same asymptotic behavior as when ρ_n is the same for all n, while the second case serves to measure how reliable is the application of the test when we stress the difficulty in identifying the noise characteristics, i.e. when the hypotheses $(H_0^{(n)})$ and (H_1) get closer and closer while $n \to \infty$.

5. Reliability check on simulations

We check here the reliability of our proposed procedure in recognizing whether $\hat{IV}_h$ is a good estimate of IV by looking at the magnitude of $\mathcal{S}_h$. Through $\mathcal{S}_h$ we then select the optimal observation frequency to estimate IV. Further analysis on both simulated and empirical data is conducted in [15]. Here we conduct four different kinds of check. Our DGP is given by

$$Y = X + \tau\varepsilon$$

where X can follow one of the three models (5), (6) or (7) described below, the noise is additive and given by a process ε which is independent on X. The r.v.s ε_{t_i} are i.i.d. uniform and centered with different possible values of the variance parameter in the different experiments: $\mathrm{Var}(\varepsilon_i) \doteq \sigma_\varepsilon^2 = 2 \times 10^{-7}$ (low level) or $\sigma_\varepsilon^2 = 8 \times 10^{-6}$ (medium level) or $\sigma_\varepsilon^2 = 8 \times 10^{-5}$ (high level). In any case σ_ε^2 is constant as h varies, as assumed in [14, p. 16]. We discriminate the presence or the absence of the noise in the simulated data through the variable τ, which takes value 1 in the first case, and 0 in the second case. In the simulation experiments either we consider $f(0) = 1/\sqrt{12\sigma_\varepsilon^2}$ as known and plug its value directly into $\mathcal{S}_h$ or we estimate $f(0)$ by means of the empirical variance of the n observations $\Delta_i Y$, where n is the same as in $\hat{IV}$:

$$\hat{f}(0) = 1/\sqrt{6\hat{\sigma}_u^2}. \tag{4}$$

Applications of the test where $f(0)$ is estimated by the non-parametric kernel method is done in [15]. In all the three proposed models the values of σ keep around 0.4, as it is realistic for financial data, and the threshold r_h has to be such that about all the squared variations $(\sigma_{t_i}\Delta_i W)^2$ are below it, so we implement our test using $r_h = 0.95 \times h^{0.999}$, where 0.95 is about 6 times 0.4^2. We take different values of n and of the observation steps h in the different experiments, then $T = nh$. For instance when we consider $1''$ observations over a whole day with a 7 h open market ($T = 0.004$ years), then $h = 1/(252 \times 7 \times 60 \times 60)$ and $n = 25\,200$, while if we consider $5'$ observations over a day then $h = 1/(252 \times 7 \times 12)$ and $n = 84$.

We recall that we are interested in establishing whether for a given h the noise is too relevant or not in order to rely on the fact that $\hat{IV}_h$ correctly estimates IV, and such a relevance is measured by the discrepancy between the behavior of our test statistic $\mathcal{S}_h$ and the standard Gaussian law. Our formal hypothesis is

$$(H_0)\ \tau = 1$$

and we judge that the noise is relevant iff we have $|\mathcal{S}_h| < 1.96$, meaning that we cannot rely that $\hat{IV}_h$ correctly estimates IV, while we judge that the noise is negligible otherwise. In this last case we more or less rely on $\hat{IV}_h$.

In the following, when we take $\tau = 1$ we simulate H paths of Y, for each path we implement $\mathcal{S}_h$ and compute the following empirical quantile of $|\mathcal{S}_h|$

$$\text{pct} \doteq \frac{\#\{|\mathcal{S}_h| > 1.96\}}{H},$$

which we use as a test on the distribution of $|\mathcal{S}_h|$. More precisely, as the CLT we gave states an $\mathcal{F}_0$-stable convergence of $\mathcal{S}_h$, we operate *conditionally* on X, i.e. for a given h the H paths of Y are obtained by generating one path of X and by adding to it H different paths of ε. When $\tau = 0$ we implement $|\mathcal{S}_h|$ only once.

MODEL GP: Gauss–Poisson process. Here the efficient price X has constant volatility and compound Poisson jumps. However we condition process J to have one jump within $[0, T]$, in which case the jump time ν is a r.v. uniformly distributed on $[0, T]$, and J can be written as

$$J_t \equiv J_{1t} = Z\ I_{\{t \ge \nu\}},$$

with Z a r.v. independent on ν and that we choose to be Gaussian with law $\mathcal{N}(0, 0.6^2)$. Then we have

$$dX = 0.4dW + dJ, \tag{5}$$

the parameters are chosen as in [1] and are expressed in annual unit of measure.

MODEL SV-PJ: Stochastic volatility and Poisson jumps. The dynamics of σ is as in [9] and J is as above:

$$\begin{aligned} &dX = -\sigma_t^2/2dt + \sigma_t dW_t + dJ_t,\\ &d\log\sigma_t = -0.09 \times (\log\sigma_t - \log(0.25))dt + 0.05 \times dW_t^{(2)}, \quad J_t = Z\ I_{\{t\ge\nu\}}. \end{aligned} \tag{6}$$

The σ parameters $\log\sigma_0 = \log(0.4)$, $\rho = \text{corr}(W_t, W_t^{(2)}) = -0.7\ \forall t$, produce similar σ paths as in [9].

MODEL G-CGMY: constant volatility and CGMY jumps.

$$dX = 0.3815\, dW + dJ, \tag{7}$$

where J is a CGMY process as proposed in [5] with, for the Lévy density of J, scale parameter $C = 280.11$, tail decay parameter for the negative jump sizes $G = 102.84$, tail decay parameter for the positive jump sizes $M = 102.53$ and jump activity index $Y = 0.1191$. The parameter values have been estimated for MSFT asset prices in [5] (Table 2).

FIRST CHECK. We show the empirical density of the values assumed by our test statistic when implemented on $H = 1000$ paths of Model GP in the case $\tau = 1$, medium $\sigma_\varepsilon^2 = 8 \times 10^{-6}$, using $n = 1000$ observations. Consistently with our common sense, we observe two radically different behaviors when h is $20'$ (left panel of Fig. 1) and when h is $1''$ (right panel): according to the values assumed by $\mathcal{S}_h$ in the first case the noise is judged to be negligible, while in the second case it is judged to be relevant. And in fact pct $= 1$,

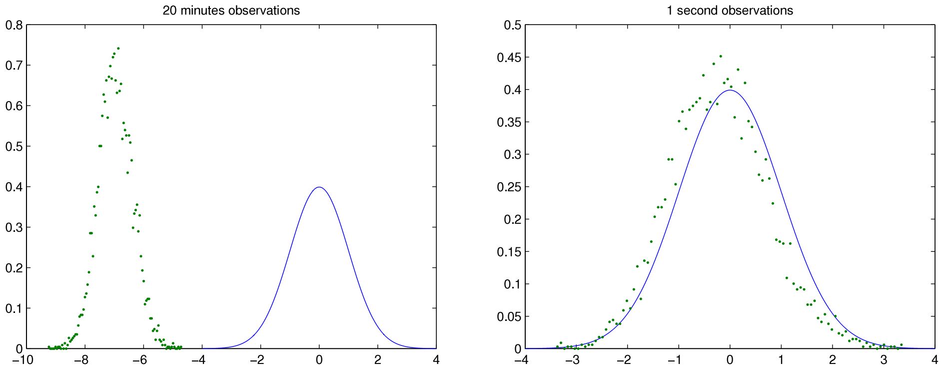

Fig. 1. Empirical density of the test statistic $\mathcal{S}_h$ under the simulated Model GP plus additive i.i.d. uniform noise with $\sigma_\varepsilon^2 = 8 \times 10^{-6}$ and $n = 1000$ observations, $h = 1/(252 \times 21)$ i.e. $20'$ (left), $h = 1/(252 \times 7 \times 60 \times 60)$ i.e. $1''$ (right), pct = 1 (left), pct = 0.045 (right).

Table 1
Performance of our test $\mathcal{S}_h$ under Model GP: $\tau = 1$ is when the simulated data contain the noise component, $\tau = 0$ otherwise; pct $= \#\{|\mathcal{S}_h| > 1.96|X\}/H$; low noise means $\sigma_\varepsilon^2 = 2 \times 10^{-7}$, medium noise $\sigma_\varepsilon^2 = 8 \times 10^{-6}$, high noise $\sigma_\varepsilon^2 = 8 \times 10^{-5}$. As we condition on the X path, when $\tau = 1$ then $\mathcal{S}_h$ is implemented on $H = 1000$ different simulated paths of $Y = X + \tau\varepsilon$, while when $\tau = 0$ then $\mathcal{S}_h$ is implemented only once.

Model GP	h	pct	MEE	SEE	pct	MEE	SEE	pct	MEE	SEE
		Low noise			Med noise			High noise		
Size	$5'$	1	−16.34	5.9	0.88	30.34	17.48	0.035	−27.32	15.90
$\tau = 1$	$1''$	1	−18.45	0.94	0.17	−84.50	0.46	0.046	−94.97	0.27
Power	$5'$	1	−9.059	0						
$\tau = 0$	$1''$	1	−9.91	0						

MEE $=$ mean$(100(\hat{IV} - IV)/IV) = 10.3434$, SEE $\doteq \sqrt{\mathrm{Var}(100(\hat{IV} - IV)/IV)} = 3.9709$ in the first case, while pct $= 0.0452$, MEE $= -84.5334$, SEE $= 2.3202$, in the second one.

SECOND CHECK. Under Model GP we check size ($\tau = 1$) and power ($\tau = 0$) of our test, for fixed $T = 0.004$ years, either with $h = 1$ second ($n = 25\,200$), or 5 min ($n = 84$), in the four cases of absence of noise, low, medium or high level of noise. Recall that when $\tau = 0$ only one path of Y is available and the value $100 \times (\hat{IV} - IV)/IV$ is computed only once, and that $\hat{f}(0)$ is still as in (4). When $\tau = 1$ then $H = 1000$ paths are generated in each scenario. The produced results are as in Table 1. Substantially the statistic behaves as one would expect: for instance, when the variance is low and we sample each $5'$, for the 100% of the paths of Y the statistic $|\mathcal{S}_h|$ assumes values above 1.96, indicating negligibility of the noise, and in fact the mean estimation error of IV by $\hat{IV}$ is not so high, about 16%; if we sample at $1''$ the noise is still classified as negligible by $\mathcal{S}_h$, in fact MEE is about 18%; when the noise has high variance and we sample at $1''$, for about the 95% of the samples $|\mathcal{S}_h|$ is below 1.96, so according to it the noise has to be considered relevant, and in fact the mean estimation error is high (about 95%). Consistently with our common sense, when the noise is at an intermediate level the statistic indicates that it is much more relevant when sampling at $1''$ than at $5'$. In the absence of the noise things go as expected, as according to $\mathcal{S}_h$ the noise is always correctly judged to be negligible.

We now change the volatility component in the simulated DGP, assuming Model SV-PJ, and repeat the previous experiment. Table 2 confirms the previous results.

Table 2
Performance $\mathcal{S}_h$ under Model SV-PJ.

Model SV-PJ	h	pct	MEE	SEE	pct	MEE	SEE	pct	MEE	SEE
		Low noise			Med noise			High noise		
Size	$5'$	1	-1.84	6.74	0.97	30.30	16.60	0.038	-27.70	15.17
$\tau = 1$	$1''$	1	-18.48	0.92	0.15	-84.48	0.47	0.035	-95.00	0.25
Power	$5'$	1	-10.91	0						
$\tau = 0$	$1''$	1	-11.10	0						

Table 3
Performance $\mathcal{S}_h$ under Model G-CGMY.

Model G-CGMY	h	pct	MEE	SEE	pct	MEE	SEE	pct	MEE	SEE
		Low noise			Med noise			High noise		
Size	$5'$	0.008	-64.45	6.12	0.043	-62.89	10.10	0.037	-68.74	10.91
$\tau = 1$	$1''$	1	-25.19	0.90	0.25	-84.95	0.45	0.05	-95.04	0.26
Power	$5'$	0	-76.27	0						
$\tau = 0$	$1''$	1	-19.09	0						

Table 4
Performance $\mathcal{S}_h$ under Model G-CGMY, $n = 2000$.

Model G-CGMY	h	pct	MEE	SEE	pct	MEE	SEE	pct	MEE	SEE
		Low noise			Med noise			High noise		
$\tau = 1$	$5'$	1	-2.15	1.56	0.97	15.08	3.52	0.56	-32.04	3.32
$\tau = 0$	$5'$	1	-3.94	0						

We finally consider Model G-CGMY. The outcomes for pct are given in Table 3.

In this G-CGMY framework in fact $\hat{IV}_h$ is almost always considered unreliable by the test based on the magnitude of $\mathcal{S}_h$, and in fact the mean estimation error MEE is high in all cases but when the noise is absent or low and we observe every $1''$. Now the many small jumps of the GCMY process are confused by the test with the noise process increments, in fact this confusion is higher for lower observation frequency when the jumps are not well disentangled of IV, with the result that the noise is perceived much higher than it is. When we sample each 5 min in fact $n = 84$ observations are not sufficient to disentangle IV from the jumps. If we implement the test with $n = 2000$ 5-min observations, things go much more as we would expect, as shown in Table 4.

THIRD CHECK. We now check the sensitivity of the proposed test to the noise variance σ_ε^2. For this, we simulate a GP model as in (5). Given an observation step h we vary σ_ε^2 and compute the resulting pct value. Fig. 2 displays the plots of pct as a function of σ_ε^2 in the two cases of $h = 5'$ (left panel) and $h = 1''$ (right panel).

Recall that the test classifies the noise as relevant iff pct ≤ 0.05, so we can see that for $h = 5'$ noises with variance less than or equal to about 10^{-8} are judged to be negligible, while with $h = 1''$ noises with variance between $10^{-8.5}$ and 10^{-8} are already judged to be relevant, as one would expect, because for the same level of noise the impact on the returns is higher at lower frequencies.

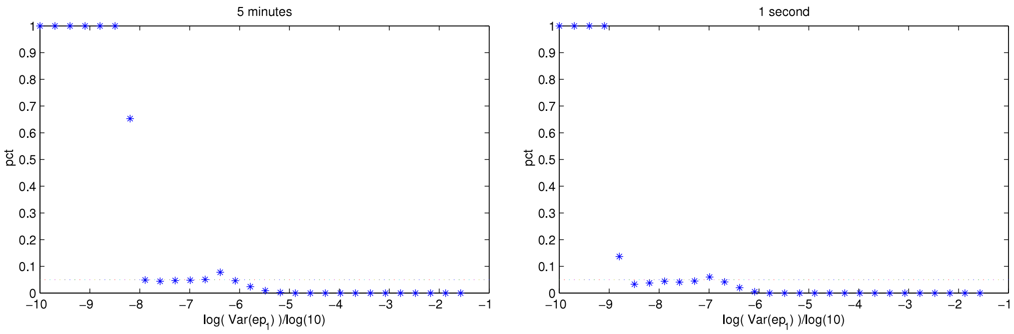

Fig. 2. Plot of pct $= \frac{\#\{|\mathcal{S}_h|>1.96\}}{H}$ as a function of σ_ε^2. From Model GP plus additive i.i.d. uniform noise $H = 1000$ paths were generated. Each path is observed at $n = 1000$ points in time and the observation step is either of five minutes (left panel) or one second (right panel). Recall that the noise is judged by the test to be negligible iff pct $\gg 5\%$.

FOURTH CHECK. We finally compare the response we obtain using our test with the responses given on one hand by visualizing the signature plot (SP) of RV_h and on the other hand by using the criterion of minimizing the conditional (on σ) mean square estimation error $RV_h - IV$ (MSE). We now simulate only one path of the DGP, with $n_{\max} = 33\,600$ observations with minimum observation frequency $h_{\min} = 1''$, then for each $h = h_{\min} \times k, k \in \{1, 2, 5, 10, 15, 20, 30, 60, 120, 300, 600, 900, 1200, 1800, 2400, 3000, 3600\}$, we aggregate the available data to reach $n = n_{\max}/k$ observations with frequency h and we jointly plot RV_h and $\mathcal{S}_h$ as functions of h. We also report the values of h obtained in [4,22,3], which give an approximately optimal MSE. In [4, p. 348], for a BSM model X with i.i.d. additive noise, the observation step minimizing MSE is $\hat{h} = T/\hat{n}$ where $\hat{n}$ minimizes

$$2\frac{T}{n}(IQ + R_n) + 4nE[\varepsilon_1^4] + 4n^2\sigma_\varepsilon^4 + 8IV\sigma_\varepsilon^2 + 2\sigma_\varepsilon^4 - E[\varepsilon_1^4],$$

where $R_n = o(1)$ as $n \to \infty$, and $IQ := \int_0^T \sigma_t^4 dt$. Because $T/n \to 0$, we computed the n minimizing

$$2\frac{T}{n}IQ + 4nE[\varepsilon_1^4] + 4n^2\sigma_\varepsilon^4 + 8IV\sigma_\varepsilon^2 + 2\sigma_\varepsilon^4 - E[\varepsilon_1^4],$$

which is unique and exactly given by $n_{\mathrm{BR}} \doteq y - a/3$, where

$$y = \sqrt[3]{-q/2 + \sqrt{q^2/4 + p^3/27}} + \sqrt[3]{-q/2 - \sqrt{q^2/4 + p^3/27}};$$
$$p = -a^2/3;\; q = 2a^3/27 - T \times IQ/(4\sigma_\varepsilon^4),$$

$a = E[\varepsilon_1^4]/2\sigma_\varepsilon^4$. We then set $h_{\mathrm{BR}} = T/n_{\mathrm{BR}}$. The authors also suggest that, when the number of used observations is sufficiently large, then $\hat{n}$ is well approximated by $\tilde{n}_{\mathrm{BR}} \doteq \sqrt[3]{T \times IQ/(4\sigma_\varepsilon^4)}$, so that

$$\tilde{h}_{\mathrm{BR}} \doteq \sqrt[3]{4T^2\sigma_\varepsilon^4/IQ}.$$

In [22, p. 1399], for a BSM model X, an analogous minimization of MSE is conducted and, in the framework of equally spaced observations, it gives the same approximate optimal observation step as h_{BR}. Note that in [3, p. 361] the same observation step value as h_{BR} is again selected for a parametric Gaussian model $X = \sigma W$ where σ is estimated by maximum likelihood and MSE is

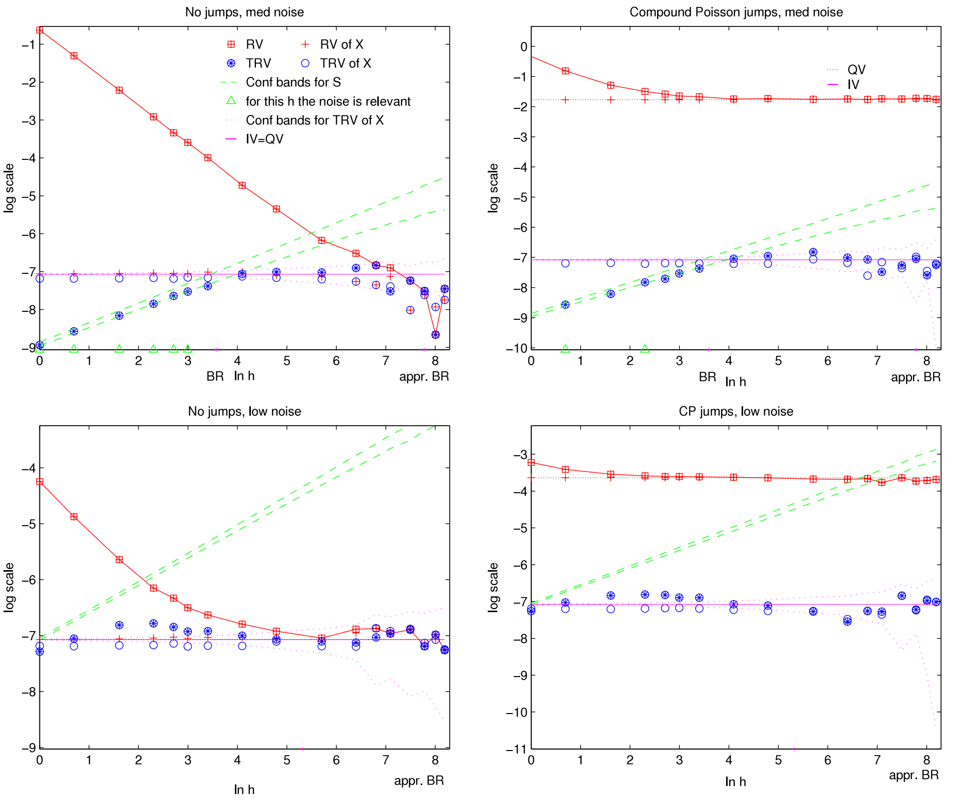

Fig. 3. Optimal choices of h to estimate IV, using the different criteria of $\mathcal{S}_h$, SP, h_{BR}, $\tilde{h}_{\mathrm{BR}}$, under Model SV-PJ. The noise is additive i.i.d. uniform, $n_{\max} = 33\,600$, $T = 0.0053$, $r_h = h^{0.999}$. The figures in the top row are characterized by $\sigma_\varepsilon^2 = 8 \times 10^{-6}$ (medium level of noise), while in the second row we have $\sigma_\varepsilon^2 = 2 \times 10^{-7}$ (low level of noise). The figures in the left column are characterized by the fact that we imposed $J \equiv 0$, while to generate the figures of the right column we conditioned to the occurrence of one jump. *TRV* stands for Threshold Realized Variance and coincides with $\hat{IV}_h$.

minimized. The value h_{BR} is still an approximation of the optimal h, this time the approximation error is small for large T. The coincidence of the selected observation steps in [3,4] is explained by the fact that in the framework of [3] the ML estimator coincides with RV_h/T.

We firstly assume Model SV-PJ and uniform noise. Fig. 3 visualizes a comparison of the different answers given by the different four criteria $\mathcal{S}_h$, SP, h_{BR}, $\tilde{h}_{\mathrm{BR}}$ within 4 different scenarios. The squares with a cross inside and connected by a line represent the SP of RV_h as h varies on the horizontal axis. The step h is expressed in seconds and the x-axis reports $\ln h$. In order to be able to clearly read the figure, also on the vertical axis we reported log values, such as $\log(RV_h)$, $\log(\hat{IV}_h)$ and so on. In the top left panel J is set equal to 0 and the noise level is medium with $\sigma_\varepsilon^2 = 8 \times 10^{-6}$. According to the plotted SP, as the minimal value is obtained with $h \approx e^8$ (corresponding to about $50'$), this is the limit frequency under which not to go in order to consider RV_h as a reliable estimate of IV. Note that in fact e^8 is close to the value $\tilde{h}_{\mathrm{BR}}$ ("APPR. BR", point in the x-axis). However note that, in this case, with $50'$ observations, RV_h does not approximate IV (continuous line) nicely. We also reported the unobservable $\log(RV_h(X))$ (crosses), $\log(\hat{IV}_h(X))$ (circles) and the log of the 95% confidence band (dotted lines) indicating when $\hat{IV}_h(X)$ is an acceptable estimate of IV in the absence of noise. Such confidence band is computed on the basis of the CLT for $\hat{IV}_h(X)$ given in [16]. We see that in the absence of the jumps and of the noise $RV_h(X)$ and $\hat{IV}_h(X)$ in fact coincide, but they give very accurate estimates of IV only for values of h less than or equal to e^7 seconds (about 18′), while for larger values of h they fall outside the confidence interval, meaning that $h \approx e^8$ is too large.

On the other hand, the minimal MSE criterion (h_{BR} and $\tilde{h}_{\text{BR}}$, points on the x-axis which are centered respectively with respect to the labels "BR" and "APPR. BR") would suggest that on average it is safe to use RV_h with $h_{\text{BR}} \approx e^{3.6}$ (about 36″) or $\tilde{h}_{\text{BR}} = e^{7.8}$ (41′). However, as we can directly check, for the realized path of Y we are analyzing, the estimation error $RV_h - IV$ is not really acceptable, especially at the frequency h_{BR}, as RV_h is outside and quite far from the dotted confidence range (in this framework of no jumps $P \lim_h RV_h = P \lim_h \hat{IV}_h$ and the same CLT holds for both the estimators).

On the contrary, if we use the threshold estimator of IV, we can take an even lower frequency than h_{BR}, such as about $h = e^{3.4}$ (about 30″) and still correctly approximating IV. In fact the stars surrounded by circles represent the values assumed by $\log(\hat{IV}_h)$. The dashed lines represent the log of the 95% confidence interval for S_h behaving like a standard Gaussian r.v., thus indicating relevance of the noise. A triangle on a given value h on the x-axis indicates that for that observation step our test accepts (H_0) (meaning relevance of the noise). As soon as $\log(\hat{IV}_h)$ enters the dashed lines confidence interval, we are aware that we cannot rely anymore on our estimator because the noise becomes too important. Note that, as h decreases, for a while $\hat{IV}_h$ follows the shape of RV_h, but then the threshold begins to truncate and $\hat{IV}_h$ is smoothed. Our test indicates not to use a frequency below about $h = e^{3.4}$ (30″) to estimate IV through $\hat{IV}_h$, considering acceptable the percentage estimation error $(\hat{IV}_h - IV)/IV$ of about 20% when $h = e^{3.4}$.

Since in this path no jumps occurred, QV equals IV, and we see that $\log(RV_h)$ and $\log(\hat{IV}_h)$ nearly coincide for $h \geq e^{6.8}$ (10′). However, if some jumps occur, as in the top right panel of Fig. 3, we know that it is forbidden to use RV_h to estimate IV, because RV_h tends to $\text{QV} = IV + \sum_{t \leq T} (\Delta J_t)^2$. So in this second panel it is even more evident the problem that the optimal h values for the SP and for the minimum MSE criterions is not necessarily such that the estimation error of IV is small.

On the other hand the bottom left panel of Fig. 3 shows the comparison among the illustrated optimal frequency selection criteria when the noise variance is decreased to $\sigma_\varepsilon^2 = 2 \times 10^{-7}$. In this case it turns out that $h_{\text{BR}} = 5.7 \times 10^{-6} \approx e^{-12}$ (which falls outside the x-axis range, and corresponds to about 0″), indicating that the noise is so low that we can use all the available 1″ data and rely on RV_h to estimate IV, which however is not the case from our picture. The threshold based test response on the optimal frequency selection is similar, because no triangles appear on the x-axis, indicating us to neglect the noise even when using 1″ observations if adopting $\hat{IV}_h$. In fact our picture clearly suggests that, with data at 1″, we have to estimate IV by $\hat{IV}_h$ and not by RV_h. This is even more so when X undergoes some jumps (bottom right panel).

We now repeat the comparison on two simulated paths of Model G-CGMY added with uniform noises. We have similar pictures (Fig. 4) and conclusions as before, for the noise variance levels of $\sigma_\varepsilon^2 = 8 \times 10^{-6}$ (left panel) and $\sigma_\varepsilon^2 = 2 \times 10^{-7}$ (right panel). Note that in this case QV always differs from IV, because J has infinite activity of jump and on $[0, T]$ it realizes countably many very small jumps.

Acknowledgments

I am grateful to Roberto Renò, Jean Jacod and Nour Meddahi for their very stimulating comments.

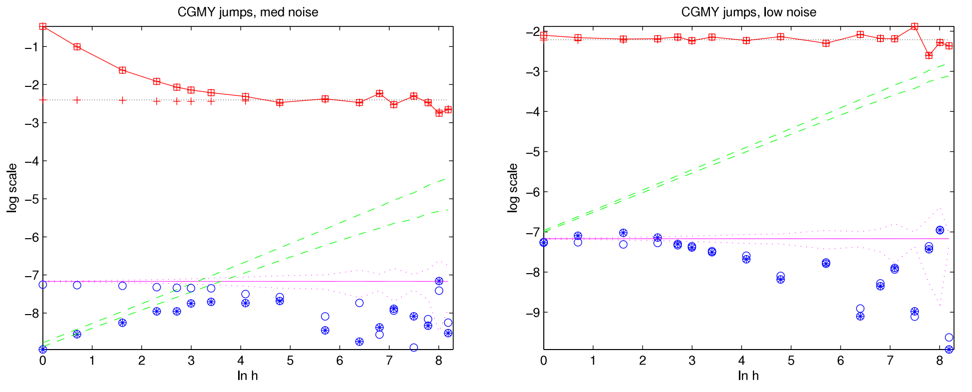

Fig. 4. Optimal choices of h to estimate IV, using the different criteria of $\mathcal{S}_h$, SP, h_{BR}, $\tilde{h}_{\mathrm{BR}}$, under Model G-CGMY. The noise is additive i.i.d. uniform, with medium level of noise $\sigma_\varepsilon^2 = 8 \times 10^{-6}$ in the left panel and $\sigma_\varepsilon^2 = 2 \times 10^{-7}$ in the right one. $n_{\max} = 33\,600$, $T = 0.0053$, $r_h = h^{0.999}$.

Appendix. Proofs of the results

Lemma A.1. *Under Assumption 2 we have, for all n,*

$$P\{|\Delta_i X| > \sqrt{r_h}\} \le ch^{1-\frac{\alpha\beta}{2}}, \qquad P\{|\Delta_i \tilde{J}_2| > \sqrt{r_h}\} \le ch^{1-\frac{\alpha\beta}{2}},$$

uniformly in $i = 1..n$.

Proof. Exactly as in Lemma 8.2(iii) in [7]. □

We remark that the càdlàg property of the paths of a, σ, X entails that the three processes are locally bounded. By a localization procedure similar to the one in [10, Section 5.4, p. 549], we can assume wlog that they are bounded (as (ω, t) vary within $\Omega \times [0, T]$).

Proof of Theorem 3.1. We have what follows.

$$0 \le \sum_{i=1}^{n} (\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}} \le 2 \sum_{i=1}^{n} [(\Delta_i X_0)^2 + (\Delta_i J + \Delta_i \varepsilon)^2]$$
$$\times [I_{\{\Delta_i N \ne 0, (\Delta_i Y)^2 \le r_h\}} + I_{\{\Delta_i N = 0, (\Delta_i Y)^2 \le r_h\}}]$$

note that for sufficiently small h on $(\Delta_i Y)^2 \le r_h$ we have $|\Delta_i J + \Delta_i \varepsilon| \le 2\sqrt{r_h}$, since $\sqrt{r_h} \ge |\Delta_i Y| \ge |\Delta_i J + \Delta_i \varepsilon| - |\Delta_i X_0|$ implies that $|\Delta_i J + \Delta_i \varepsilon| \le \sqrt{r_h} + |\Delta_i X_0| \le 2\sqrt{r_h}$ by (14) in [16], therefore

$$\sum_{i=1}^{n} [(\Delta_i X_0)^2 + (\Delta_i J + \Delta_i \varepsilon)^2] I_{\{\Delta_i N \ne 0, (\Delta_i Y)^2 \le r_h\}}$$
$$\le \sum_{i=1}^{n} [(\Delta_i X_0)^2 + (\Delta_i J + \Delta_i \varepsilon)^2] I_{\{\Delta_i N \ne 0, |\Delta_i J + \Delta_i \varepsilon| \le 2\sqrt{r_h}\}}$$
$$\le c\left(h \ln \frac{1}{h} + r_h\right) N_T \overset{\text{a.s.}}{\to} 0.$$

On the other hand

$$\sum_{i=1}^{n}(\Delta_i J+\Delta_i\varepsilon)^2 I_{\{\Delta_i N=0,(\Delta_i Y)^2\le r_h\}} \le \sum_{i=1}^{n}(\Delta_i J+\Delta_i\varepsilon)^2 I_{\{\Delta_i N=0,(\Delta_i Y)^2\le r_h,|\Delta_i J+\Delta_i\varepsilon|\le 2\sqrt{r_h}\}}$$

and this last term can be split in

$$I_1+I_2 := \sum_{i=1}^{n}(\Delta_i J+\Delta_i\varepsilon)^2 I_{\{\Delta_i N=0,(\Delta_i Y)^2\le r_h,|\Delta_i J+\Delta_i\varepsilon|\le 2\sqrt{r_h},|\Delta_i \tilde J_2|\le\sqrt{r_h}\}} + \sum_{i=1}^{n}(\Delta_i J+\Delta_i\varepsilon)^2 I_{\{\Delta_i N=0,(\Delta_i Y)^2\le r_h,|\Delta_i J+\Delta_i\varepsilon|\le 2\sqrt{r_h},|\Delta_i \tilde J_2|>\sqrt{r_h}\}}.$$

By Assumption 2 we have, uniformly on i, $P\{|\Delta_i\tilde J_2|>\sqrt{r_h}\}=O_P(h^{1-\frac{\alpha\beta}{2}})$, so in probability

$$I_2=O_P\left(r_h h^{-\frac{\alpha\beta}{2}}\right)=O_P\left(h^{\beta(1-\frac{\alpha}{2})}\right)\to 0.$$

As for I_1, on $\{\Delta_i N=0, |\Delta_i J+\Delta_i\varepsilon|\le 2\sqrt{r_h}, |\Delta_i\tilde J_2|\le\sqrt{r_h}\}$ we have $2\sqrt{r_h}\ge|\Delta_i J+\Delta_i\varepsilon|\ge |\Delta_i\varepsilon|-|\Delta_i\tilde J_2|$ then $|\Delta_i\varepsilon|\le 2\sqrt{r_h}+|\Delta_i\tilde J_2|\le 3\sqrt{r_h}$ and by Assumption 1 we reach that

$$E[I_1]\le c r_h n\sqrt{r_h}=h^{\frac{3}{2}\beta-1}\to 0.$$

Finally we consider $\sum_{i=1}^{n}(\Delta_i X_0)^2 I_{\{\Delta_i N=0,(\Delta_i Y)^2\le r_h\}}$ and we write it as

$$I_3+I_4 := \sum_{i=1}^{n}(\Delta_i X_0)^2\left[I_{\{\Delta_i N=0,(\Delta_i Y)^2\le r_h\}}-I_{\{(\Delta_i X)^2\le A r_h\}}\right]+\sum_{i=1}^{n}(\Delta_i X_0)^2 I_{\{(\Delta_i X)^2\le A r_h\}},$$

with $A>1$ any constant. We have

$$I_3=\sum_{i=1}^{n}(\Delta_i X_0)^2\left[I_{\{\Delta_i N=0,(\Delta_i Y)^2\le r_h,(\Delta_i X)^2>A r_h\}}-I_{\{(\Delta_i X)^2\le A r_h\}\cap(\{\Delta_i N\ne 0\}\cup\{(\Delta_i Y)^2>r_h\})}\right]:$$

we now show that on $\{\Delta_i N=0, (\Delta_i Y)^2\le r_h, (\Delta_i X)^2>A r_h\}$ we have $(\Delta_i\tilde J_2)^2>c r_h$, for a suitable constant c. In fact, given any constant $\delta>0$, analogously as for the statement in the fourth line of this proof, a.s. for sufficiently small h if $(\Delta_i Y)^2\le r_h$ then

$$|\Delta_i\tilde J_2+\Delta_i\varepsilon|\le(1+\delta)\sqrt{r_h}; \tag{8}$$

moreover if $(\Delta_i X)^2>A r_h$ and $(\Delta_i Y)^2\le r_h$ then $|\Delta_i\varepsilon|>(\sqrt{A}-1)\sqrt{r_h}$, since

$$|\Delta_i\varepsilon|=|\Delta_i Y-\Delta_i X|>|\Delta_i X|-|\Delta_i Y|\ge\sqrt{A r_h}-\sqrt{r_h}=(\sqrt{A}-1)\sqrt{r_h} \tag{9}$$

putting together (8) and (9) we reach

$$|\Delta_i\tilde J_2|=|\Delta_i\tilde J_2+\Delta_i\varepsilon-\Delta_i\varepsilon|>|\Delta_i\varepsilon|-|\Delta_i\tilde J_2+\Delta_i\varepsilon|\ge((\sqrt{A}-1)-1-\delta)\sqrt{r_h}$$

and $\sqrt{A}-2-\delta>0$ as soon as we choose $A>(2+\delta)^2$, as we wanted. Now a.s., for sufficiently small h, with $\sqrt{c}:=\sqrt{A}-2-\delta$

$$\sum_{i=1}^{n}(\Delta_i X_0)^2 I_{\{\Delta_i N=0,(\Delta_i Y)^2\le r_h,(\Delta_i X)^2>Ar_h\}} \le \sum_{i=1}^{n}(\Delta_i X_0)^2 I_{\{|\Delta_i \tilde{J}_2|>\sqrt{c}\sqrt{r_h}\}}$$
$$\le ch\ln\frac{1}{h}h^{-\frac{\alpha\beta}{2}} \to 0$$

and the almost sure limit of $I_3 + I_4$ is the same as

$$-\sum_{i=1}^{n}(\Delta_i X_0)^2 I_{\{(\Delta_i X)^2\le Ar_h\}\cap\left(\{\Delta_i N\ne 0\}\cup\{(\Delta_i Y)^2>r_h\}\right)} + \sum_{i=1}^{n}(\Delta_i X_0)^2 I_{\{(\Delta_i X)^2\le Ar_h\}}.$$

Note that a.s., for sufficiently small h, $\sum_{i=1}^{n}(\Delta_i X_0)^2 I_{\{(\Delta_i X)^2\le Ar_h\}\cap\{\Delta_i N\ne 0\}} \le h\ln\frac{1}{h}N_T$ is negligible, so we are left with

$$\sum_{i=1}^{n}(\Delta_i X_0)^2\left[-I_{\{(\Delta_i X)^2\le Ar_h\}\cap\{(\Delta_i Y)^2>r_h\}} + I_{\{(\Delta_i X)^2\le Ar_h\}}\right]$$
$$= \sum_{i=1}^{n}(\Delta_i X_0)^2 I_{\{(\Delta_i X)^2\le Ar_h,(\Delta_i Y)^2\le r_h\}}. \tag{10}$$

However on $\{(\Delta_i X)^2 \le Ar_h, (\Delta_i Y)^2 \le r_h\}$ we have $|\Delta_i\varepsilon| = |\Delta_i Y - \Delta_i X| \le |\Delta_i Y| + |\Delta_i X| \le \sqrt{r_h} + \sqrt{Ar_h}$, so that almost surely (10) is bounded by $h\ln\frac{1}{h}\sum_{i=1}^{n} I_{\{|\Delta_i\varepsilon|\le(\sqrt{A}+1)\sqrt{r_h}\}}$, whose expectation is $O(nh\ln\frac{1}{h}\sqrt{r_h}) \to 0$. □

Lemma A.2. *Under the assumptions of Theorem 3.2, for any even integer $q > 0$ we have what follows.*

(1) *For fixed ω, for all n for all $i = 1..n$, define the r.v.*

$$H_i^q(r) \doteq \int_{-\sqrt{r}}^{\sqrt{r}} |u|^q g(u - \Delta_i X + \varepsilon_{i-1})du.$$

It holds that for fixed ω, for all n for all $i = 1..n \exists\ \xi_i = \xi_i^n(\omega) \in (0, r)$:

$$H_i^q(r) = \frac{r^{(q+1)/2}}{q+1}\left[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})\right]. \tag{11}$$

(2) $\forall n, \forall i$

$$E_{i-1}[(\Delta_i Y_\star)^q] = E_{i-1}[H_i^q(r_h)]$$
$$= \frac{r_h^{(q+1)/2}}{q+1}E_{i-1}\left[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})\right].$$

(3) $\frac{1}{n}\sum_{i=1}^{n} E_{i-1}[g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})] \xrightarrow{L^1} E[g(\varepsilon_1)]$ *for both cases $s = +1$, and $s = -1$.*

Proof. (1) Define $G^{(q)}(r) := r^{\frac{q+1}{2}}$ and note that $G^q(0) = H_i^{(q)}(0) = 0$. Using the Cauchy theorem, a.s. for all i there exist numbers $\xi_i \in]0, r[$ such that

$$H_i^{(q)}(r) = \frac{(H_i^{(q)})'(\xi_i)}{(G^{(q)})'(\xi_i)}G^{(q)}(r)$$
$$= \frac{g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})}{q+1}r^{\frac{q+1}{2}}. \tag{12}$$

(2) For fixed (h, i) we have $E_{i-1}[(\Delta_i Y_\star)^q] = E_{i-1}\big[(\Delta_i X + \Delta_i \varepsilon)^q I_{\{|\Delta_i X + \Delta_i \varepsilon| \le \sqrt{r_h}\}}\big]$, and by the independence of ε_i on (ε_{i-1}, X) and since q is even the above term equals

$$E_{i-1}\left[\int_{\mathbb{R}} (\Delta_i X + z - \varepsilon_{i-1})^q I_{\{|\Delta_i X + z - \varepsilon_{i-1}| \le \sqrt{r_h}\}} g(z) dz\right]$$
$$= E_{i-1}\left[\int_{-\sqrt{r_h}}^{\sqrt{r_h}} u^q g(u - \Delta_i X + \varepsilon_{i-1}) du\right] = E_{i-1}\big[H_i^{(q)}(r_h)\big],$$

having changed variable as $u = \Delta_i X + z - \varepsilon_{i-1}$. Now for fixed (i, h), for any fixed ω we have equality (11), so for fixed (i, h) the two terms in (11) are a.s. equal, therefore their expectations E_{i-1} coincide a.s., and the thesis follows.

(3) Firstly note that by the law of large numbers $\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(\varepsilon_{i-1})] = \frac{1}{n}\sum_{i=1}^n g(\varepsilon_{i-1}) \overset{L^2}{\to} E[g(\varepsilon_1)]$. Secondly we show that $\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})]$ behaves asymptotically in L^1 norm as $\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(\varepsilon_{i-1})]$. In fact by the Lipschitz property of g, denoting with L its Lipschitz constant,

$$E\left[\left|\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) - g(\varepsilon_{i-1})]\right|\right] \le \frac{L}{n}\sum_{i=1}^n E[|s\sqrt{\xi_i} - \Delta_i X|]. \quad (13)$$

Because $|s\sqrt{\xi_i}| \le \sqrt{r_h}$ and we assumed X bounded wlog, we have, for all i, for small h, $E[|\Delta_i X|] \le \sqrt{h} < \sqrt{r_h}$ and the last display above is dominated by $c(E[|s\sqrt{\xi_i}|] + E[|\Delta_i X|]) \le c\sqrt{r_h} \to 0$. □

Proof of Theorem 3.2. (i) We have

$$E[\hat{IV}_h] = E\left[\sum_{i=1}^n (\Delta_i Y_\star)^2\right] = E\left[\sum_i E_{i-1}[(\Delta_i Y_\star)^2]\right]$$

by Lemma A.2 part (2) the last expectation equals

$$\frac{r_h^{3/2}}{3} E\left[\sum_{i=1}^n E_{i-1}[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})]\right]$$
$$= n r_h^{3/2} \frac{1}{3} E\left[\frac{1}{n}\sum_i E_{i-1}[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})]\right],$$

and the thesis follows from Lemma A.2 part (3).

In order to prove (ii), we apply a classical theorem of convergence for sums of r.v.s belonging to a triangular array [11, Lemma 4.3 of the preprint draft] to show the convergence in law of the normalized bias $\mathcal{NB}_h$. We then refine the result to an $\mathcal{F}^0$-stable convergence. Recall that $h = T/n$ and define

$$\phi_i = \phi_i^n \doteq \frac{(\Delta_i Y_\star)^2 - r_h^{3/2}\frac{2}{3}E[g(\varepsilon_1)]}{\sqrt{n r_h^{5/2}\frac{2}{5}E[g(\varepsilon_1)]}}$$

and note that $\phi_i \in \mathcal{F}_{t_i}$. We are going to verify that

$$\text{(a)} \sum_{i=1}^{[t/h]} E_{i-1}[\phi_i] \overset{P}{\to} 0 \quad \text{(b)} \sum_{i=1}^{[t/h]} E_{i-1}[\phi_i^2] - E_{i-1}^2[\phi_i] \overset{L^1}{\to} C_T, \quad \text{(c)} \sum_{i=1}^{[t/h]} E_{i-1}[\phi_i^4] \overset{P}{\to} 0,$$

with C a deterministic increasing process with continuous paths. Such conditions imply the convergence in law of processes $\{\sum_{i=1}^{[t/h]} \phi_i, t \geq 0\}$ to a Gaussian process B with continuous paths, centered, with independent increments and such that $\forall t \geq 0$, $E[B_t^2] = C_t$.

As for (a),

$$\sum_{i=1}^{[t/h]} E_{i-1}[\phi_i] = \sum_{i=1}^{[t/h]} \left(E_{i-1}[\phi_i] \pm \frac{2r_h^{\frac{3}{2}}}{3} \frac{E_{i-1}[g(\varepsilon_{i-1})]}{\sqrt{n} r_h^{\frac{5}{4}} \sqrt{\frac{2}{5} E[g(\varepsilon_1)]}} \right)$$

$$= \frac{r_h^{\frac{3}{2}}}{3} \sum_{i=1}^{[t/h]} \frac{E_{i-1}\Big[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) - 2g(\varepsilon_{i-1})\Big]}{\sqrt{n} r_h^{\frac{5}{4}} \sqrt{\frac{2}{5} E[g(\varepsilon_1)]}}$$

$$+ \frac{r_h^{\frac{3}{2}}}{3} \sum_{i=1}^{[t/h]} \frac{2E_{i-1}[g(\varepsilon_{i-1})] - 2E[g(\varepsilon_1)]}{\sqrt{n} r_h^{\frac{5}{4}} \sqrt{\frac{2}{5} E[g(\varepsilon_1)]}}. \tag{14}$$

Using the Lipschitz property of g, the first term above has absolute value bounded by

$$c \frac{r_h^{\frac{1}{4}}}{\sqrt{n}} \sum_{i=1}^{[t/h]} E_{i-1}\Big[|g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) - 2g(\varepsilon_{i-1})|\Big]$$

$$\leq c \frac{r_h^{\frac{1}{4}}}{\sqrt{n}} L \sum_{i=1}^{[t/h]} E_{i-1}\Big[|\sqrt{\xi_i}| + |\Delta_i X|\Big].$$

As argued for (13), $E_{i-1}\Big[|\sqrt{\xi_i}| + |\Delta_i X|\Big] \leq \sqrt{r_h}$, further for $t \leq T$ we have $[t/h] \leq n$, so the above display is dominated by

$$c r_h^{\frac{1}{4}} \sqrt{n} \sqrt{r_h} = h^{\frac{3}{4}\beta - \frac{1}{2}}$$

which tends to zero by the assumption $\beta > 2/3$. As for the second term in (14), it coincides with

$$c \frac{r_h^{\frac{1}{4}}}{\sqrt{n}} \left[\sum_{i=1}^{[t/h]} g(\varepsilon_{i-1}) - \left[\frac{t}{h}\right] E[g(\varepsilon_1)] \right]$$

which, by the central limit theorem for a sequence of i.i.d. r.v.s with finite mean and variance, behaves asymptotically as $r_h^{1/4} \to 0$.

As for condition (b), using Lemma A.2 we have

$$\sum_{i=1}^{[t/h]} E_{i-1}^2[\phi_i] = \frac{\sum_{i=1}^{[t/h]} E_{i-1}^2 \left[(\Delta_i Y_\star)^2 - r_h^{3/2} \frac{2}{3} E[g(\varepsilon_1)] \right]}{n r_h^{\frac{5}{2}} \frac{2}{5} E[g(\varepsilon_1)]}$$

$$\leq c r_h^3 \frac{\sum_{i=1}^{[t/h]} E_{i-1}^2 \big[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) \big]}{n r_h^{\frac{5}{2}}}$$

$$+ c r_h^3 \left[\frac{t}{h}\right] \frac{E^2[g(\varepsilon_1)]}{n r_h^{\frac{5}{2}}}:$$

the last term has the same asymptotic behavior as $[t/h]r_h^3/(nr_h^{5/2}) \le r_h^{1/2} \to 0$, while the first term of the rhs above is dominated by

$$c\frac{r_h^{1/2}}{n}\sum_{i=1}^{[t/h]} E_{i-1}[g^2(\sqrt{\xi_i}-\Delta_i X+\varepsilon_{i-1})+g^2(-\sqrt{\xi_i}-\Delta_i X+\varepsilon_{i-1})].$$

By the boundedness of g this in turn is dominated by $cr_h^{1/2} \to 0$. We now compute

$$\sum_{i=1}^{[t/h]} E_{i-1}[\phi_i^2] = \sum_{i=1}^{[t/h]} E_{i-1}\left[\frac{(\Delta_i Y_\star)^4}{nr_h^{\frac{5}{2}}\frac{2}{5}E[g(\varepsilon_1)]}\right] - r_h^{\frac{3}{2}}\frac{4}{3}E[g(\varepsilon_1)]\sum_{i=1}^{[t/h]}\frac{E_{i-1}\big[(\Delta_i Y_\star)^2\big]}{nr_h^{\frac{5}{2}}\frac{2}{5}E[g(\varepsilon_1)]}$$
$$+\left[\frac{t}{h}\right]r_h^3\frac{4}{9}\frac{E^2[g(\varepsilon_1)]}{nr_h^{\frac{5}{2}}\frac{2}{5}E[g(\varepsilon_1)]}.$$

By Lemma A.2 part (2) and the analogous result as in part (3) with $[t/h]$ in place of n, the first term tends to t/T in probability and the second and the third terms above have both the same asymptotic behavior as $r_h^{1/2} \to 0$. We can conclude that condition (b) holds with $C_t = t/T$, so that the limit process $B = Z/\sqrt{T}$ has the same law of a standard Brownian motion Z_t divided by $\sqrt{T}$.

We now check condition (c). We have

$$\sum_{i=1}^{[t/h]} E_{i-1}[\phi_i^4] \le \frac{c}{n^2 r_h^5}\sum_{i=1}^{[t/h]} E_{i-1}\left[\left|(\Delta_i Y_\star)^2 - r_h^{3/2}\frac{2}{3}E[g(\varepsilon_1)]\right|^4\right]$$
$$\le \frac{c\sum_{i=1}^{[t/h]} E_{i-1}[(\Delta_i Y_\star)^8]}{n^2 r_h^5} + \frac{c}{n^2 r_h^5} n r_h^6.$$

The last term is of the same order as $r_h/n \to 0$, while, using again Lemma A.2, parts (2) and (3), the first term of the rhs above is dominated by

$$c\frac{r_h^{9/2}}{nr_h^5}\frac{[t/h]}{n}\frac{\sum_{i=1}^{[t/h]} E_{i-1}[g(\sqrt{\xi_i}-\Delta_i X+\varepsilon_{i-1})+g(-\sqrt{\xi_i}-\Delta_i X+\varepsilon_{i-1})]}{[t/h]} \sim \frac{1}{n\sqrt{r_h}} \to 0.$$

We now come to the $\mathcal{F}_0$-stable convergence of $\sum_{i=1}^n \phi_i$. By Proposition VIII.5.33 in [13], because $\sum_{i=1}^n \phi_i$ converges in law, then it is tight, it is thus sufficient to show that for all $A \in \mathcal{F}_0$ and all bounded continuous f the sequence $E[I_A f(\sum_{i=1}^n \phi_i)]$ converges. In fact

$$E\left[I_A f\left(\sum_{i=1}^n \phi_i\right)\right] = E\left[E_0\left[I_A f\left(\sum_{i=1}^n \phi_i\right)\right]\right] = E\left[I_A E_0\left[f\left(\sum_{i=1}^n \phi_i\right)\right]\right].$$

By the convergence in law of $\sum_{i=1}^n \phi_i$ we have that $E_0\big[f(\sum_{i=1}^n \phi_i)\big] \to \int f(x)\phi(x)dx$, where ϕ is the density of $B_T = Z_T/\sqrt{T}$, which is standard Gaussian, and by the dominated convergence theorem, the last term in the display above converges to $P(A)\int f(x)\phi(x)dx$, which concludes the proof of the stated stable convergence. □

Proof of results (i) *and* (ii) *in the statement of* Theorem 3.2 *when the noise has uniform law*.

We are now assuming that process ε is independent on X, ε_i are i.i.d. with uniform law, $\beta > 2/3$. We begin by checking the validity of Lemma A.2. In this framework we have $g(x) = C^{-1} I_{[-C/2,C/2]}(x)$, for some fixed constants C, thus for fixed h, ω, i, $H_i^{(q)}(r) = C^{-1} \int_{-\sqrt{r}}^{\sqrt{r}} u^q I_{|u-\Delta_i X+\varepsilon_{i-1}|\le C/2} du$ is not differentiable at the points $r = \Delta_i X - \varepsilon_{i-1} - C/2$, $\Delta_i X - \varepsilon_{i-1} + C/2$. We can apply the Cauchy theorem as in the proof of part (1) of the lemma only when $(-\sqrt{r}, \sqrt{r}) \subset (\Delta_i X - \varepsilon_{i-1} - C/2, \Delta_i X - \varepsilon_{i-1} + C/2)$ (or equivalently when $\sqrt{r} < C/2 - |\Delta_i X - \varepsilon_{i-1}|$), and the following results will be sufficient to prove the CLT of Theorem 3.2:

(1′) for fixed ω, n, i, for $r < C/2 - |\Delta_i X - \varepsilon_{i-1}|$ then $\exists\, \xi_i = \xi_i^n(\omega) \in (0, r)$ such that (11) holds true

(2′) for any fixed (n, i)

$$E_{i-1}[(\Delta_i Y_\star)^q I_{\sqrt{r_h}<C/2-|\Delta_i X-\varepsilon_{i-1}|}] = E_{i-1}[H_i^q(r_h) I_{\sqrt{r_h}<C/2-|\Delta_i X-\varepsilon_{i-1}|}]$$
$$= \frac{r_h^{(q+1)/2}}{q+1} E_{i-1}\Big[[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})] I_{\sqrt{r_h}<C/2-|\Delta_i X-\varepsilon_{i-1}|}\Big].$$

(3′) $\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})] \xrightarrow{L^1} E[g(\varepsilon_1)]$ for both cases $s = +1$, and $s = -1$.

Proof. Parts (1′), (2′) are proved analogously as for Lemma A.2. As for (3′) we only have to show that

$$\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) - g(\varepsilon_{i-1})] \xrightarrow{L^1} 0.$$

Using the expression for $g(x)$ and noting that with probability 1 we have $\varepsilon_{i-1} \in (-C/2, C/2)$, the rhs term of the last expression equals

$$\frac{-C^{-1}}{n}\sum_{i=1}^n E_{i-1}[I_{\{\varepsilon_{i-1}\in(-C/2,C/2), |\varepsilon_{i-1}-\Delta_i X+s\sqrt{\xi_i}|>C/2\}}]$$

which has absolute value

$$\frac{1}{nC}\sum_{i=1}^n \Big(P_{i-1}\{\varepsilon_{i-1} > C/2 + \Delta_i X - s\sqrt{\xi_i}\} + P_{i-1}\{\varepsilon_{i-1} < -C/2 + \Delta_i X - s\sqrt{\xi_i}\}\Big)$$
$$\le \frac{1}{nC}\sum_{i=1}^n \Big(P_{i-1}\{\varepsilon_{i-1} > C/2 + \Delta_i X - \sqrt{r_h}\} + P_{i-1}\{\varepsilon_{i-1} < -C/2 + \Delta_i X + \sqrt{r_h}\}\Big).$$

Thus

$$E\left|\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) - g(\varepsilon_{i-1})]\right|$$
$$\le \frac{1}{nC}\sum_{i=1}^n \Big(P\{\varepsilon_{i-1} > C/2 + \Delta_i X - \sqrt{r_h}\} + P\{\varepsilon_{i-1} < -C/2 + \Delta_i X + \sqrt{r_h}\}\Big)$$

$$= \frac{1}{nC}\sum_{i=1}^{n}\Big(E[P\{\varepsilon_{i-1} > C/2 + \Delta_i X - \sqrt{r_h}\}|\Delta_i X]$$
$$+ E[P\{\varepsilon_{i-1} < -C/2 + \Delta_i X + \sqrt{r_h}|\Delta_i X\}]\Big).$$

Noting that if $\Delta_i X - \sqrt{r_h} > 0$ the first term is 0 and if $\Delta_i X + \sqrt{r_h} > 0$ the second one is 0, the last display equals

$$\frac{1}{nC}\sum_{i=1}^{n}\left(E\left[\int_{C/2+\Delta_i X-\sqrt{r_h}}^{C/2} 1dz\; I_{\Delta_i X-\sqrt{r_h}<0}\right]\right.$$
$$\left. +\; E\left[\int_{-C/2}^{-C/2+\Delta_i X+\sqrt{r_h}} 1dz\; I_{\Delta_i X+\sqrt{r_h}>0}\right]\right)$$
$$\le \frac{c}{n}\sum_{i=1}^{n} E[\sqrt{r_h} + |\Delta_i X|] \le c\sup_i(E[|\Delta_i X|] + \sqrt{r_h}) \to 0. \quad \square$$

We now prove the result (i) and (ii) of Theorem 3.2 when ε_i are uniform.

For (i), we have

$$E[\hat{IV}] = E\Big[\sum_i E_{i-1}[(\Delta_i Y_\star)^2\, I_{\{(-\sqrt{r_h},\sqrt{r_h})\subset(\Delta_i X-\varepsilon_{i-1}-C/2,\Delta_i X-\varepsilon_{i-1}+C/2)\}}]\Big]$$
$$+ E\Big[\sum_i E_{i-1}[(\Delta_i Y_\star)^2\, I_{\{(-\sqrt{r_h},\sqrt{r_h})\subset(\Delta_i X-\varepsilon_{i-1}-C/2,\Delta_i X-\varepsilon_{i-1}+C/2)\}^c}]\Big]. \tag{15}$$

Firstly note that

$$P\{-\sqrt{r_h} > \Delta_i X - \varepsilon_{i-1} - C/2\} = P\{\sqrt{r_h} < -\Delta_i X + \varepsilon_{i-1} + C/2,\ \Delta_i X - \varepsilon_{i-1} > 0\}$$
$$+ P\{\sqrt{r_h} < |\Delta_i X - \varepsilon_{i-1}| + C/2,\ \Delta_i X - \varepsilon_{i-1} < 0\}:$$

for small h, we have $\sqrt{r_h} < C/2$, thus for any i the last term equals

$$P\{\Delta_i X - \varepsilon_{i-1} < 0\} = E[P\{\Delta_i X < \varepsilon_{i-1}|\Delta_i X\}] = C^{-1}E\left[\int_{\Delta_i X}^{C/2} 1dz\right]$$
$$= C^{-1}E[C/2 - \Delta_i X] = 1/2 - E[\Delta_i X]/(C),$$

and the first term equals

$$P\{\Delta_i X > \varepsilon_{i-1} > \sqrt{r_h} - C/2 + \Delta_i X\} = C^{-1}E\left[\int_{\sqrt{r_h}-C/2+\Delta_i X}^{\Delta_i X} 1dz\right]$$
$$= C^{-1}(C/2 - \sqrt{r_h}) = 1/2 - \sqrt{r_h}/C.$$

Moreover

$$P\{\sqrt{r_h} \ge \Delta_i X - \varepsilon_{i-1} + C/2\} = E[P\{\sqrt{r_h} \ge \Delta_i X - \varepsilon_{i-1} + C/2|\Delta_i X\}]$$
$$= E[C^{-1}\int_{\Delta_i X+C/2-\sqrt{r_h}}^{C/2} 1dz] = C^{-1}E[-\Delta_i X + \sqrt{r_h}].$$

Thus in (15) the second term is dominated by

$$nr_h \sup_i \Big(P\{-\sqrt{r_h} \le \Delta_i X - \varepsilon_{i-1} - C/2\} + P\{\sqrt{r_h} \ge \Delta_i X - \varepsilon_{i-1} + C/2\} \Big)$$
$$\le cnr_h \Big(\sup_i E[|\Delta_i X|] + \sqrt{r_h} \Big) \le cnr_h^{3/2} \to 0,$$

as $\beta > 2/3$ and $\sup_i E[|\Delta_i X|] \le c\sqrt{h} \le c\sqrt{r_h}$.

On the other hand, to the first term on the rhs of (15) we can apply result (2′) above, and obtain

$$nr_h^{3/2} \frac{1}{3} E\left[\frac{1}{n} \sum_i E_{i-1}[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})] \right]$$
$$- nr_h^{3/2} \frac{1}{3} E\Bigg[\frac{1}{n} \sum_i E_{i-1}[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})$$
$$+ g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})] I_{\{(-\sqrt{r_h}, \sqrt{r_h}) \subset (\Delta_i X - \varepsilon_{i-1} - C/2, \Delta_i X - \varepsilon_{i-1} + C/2)\}^c} \Bigg].$$

By the boundedness of g and the fact that $nr_h^{3/2} \to 0$, the second term above is negligible and by result (3′) we reach our thesis.

We now prove (ii). We can proceed almost in the same way as in the previous proof of (ii) conducted under the assumption that g was Lipschitz. It is sufficient to give an alternative treatment of the first term in (14), the only point where we used the Lipschitz property of g in the previous proof. We need to deal with two terms of kind

$$c \frac{r_h^{1/4}}{\sqrt{n}} \sum_{i=1}^{[t/h]} E_{i-1}\Big[|g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) - g(\varepsilon_{i-1})| \Big].$$

Using the above computations, the last term is given by

$$c \frac{r_h^{1/4}}{\sqrt{n}} \sum_{i=1}^{[t/h]} E_{i-1}[I_{\{\varepsilon_{i-1} \in (-C/2, C/2), |\varepsilon_{i-1} - \Delta_i X + s\sqrt{\xi_i}| > C/2\}}]$$
$$\le cr_h^{1/4} \sqrt{n} \sup_i (E[|\Delta_i X|] + \sqrt{r_h}) \le c\sqrt{r_h^{3/2} n} \to 0,$$

as for small h, $\sqrt{h} < \sqrt{r_h}$. □

References

[1] Y. Aït-Sahalia, Disentangling volatility from jumps, Journal of Financial Economics 74 (2004) 487–528.

[2] Y. Aït-Sahalia, J. Jacod, Estimating the degree of activity of jumps in high frequency data, Annals of Statistics 37 (2009) 2202–2244.

[3] Y. Ait-Sahalia, P. Mykland, L. Zhang, How often to sample a continuous-time process in the presence of market microstructure noise, Review of Financial Studies 18 (2005) 351–416.

[4] F.M. Bandi, J.R. Russell, Microstructure noise, realized volatility, and optimal sampling, Review of Economic Studies 75 (2) (2008) 339–369.

[5] P. Carr, H. Geman, D. Madan, M. Yor, The fine structure of asset returns: an empirical investigation, Journal of Business 75 (2) (2002) 305–332.

[6] K. Christensen, R. Oomen, M. Podolskij, Realised quantile-based estimation of the integrated variance, Journal of Econometrics 159 (2010) 74–98.
[7] R. Cont, C. Mancini, Non-parametric tests for pathwise properties of semimartingales, Bernoulli 17 (2) (2011) 781–813.
[8] P.R. Hansen, A. Lunde, Realized variance and market microstructure noise, Journal of Business and Economic Statistics 24 (2) (2006) 127–218.
[9] X. Huang, G. Tauchen, The relative contribution of jumps to total price variance, Journal of Financial Econometrics 3 (4) (2005) 456–499.
[10] J. Jacod, Asymptotic properties of realized power variations and associated functions of semimartingales, Stochastic Processes and their Applications 118 (2008) 517–559.
[11] J. Jacod, Statistics and high-frequency data, in: Statistical Methods for Stochastic Differential Equations, in: Chapman & Hall/CRC Monographs on Statistics & Applied Probability, 2012, Lecture Notes, SEMSTAT in La Manga (Spain).
[12] J. Jacod, Y. Li, P.A. Mykland, M. Podolskij, M. Vetter, Microstructure noise in the continuous case: the pre-averaging approach, Stochastic Processes and their Applications 119 (2009) 2249–2276.
[13] J. Jacod, A.N. Shiryaev, Limit Theorems for Stochastic Processes, second ed., Springer-Verlag, Berlin, 2003.
[14] G.J. Jiang, R.C.A. Oomen, Testing for jumps when asset prices are observed with noise. a swap variance approach, Journal of Econometrics 144 (2) (2008) 352–370.
[15] G. Lorenzini, Estimating integrated variance with noisy data, Master Thesis, University of Florence, 2012.
[16] C. Mancini, Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps, Scandinavian Journal of Statistics 36 (2009) 270–296.
[17] C. Mancini, The speed of convergence of the threshold estimator of integrated variance, Stochastic Processes and their Applications 121 (4) (2011) 845–855.
[18] C. Mancini, Measuring the relevance of the microstructure noise in financial data, Technical Report, 2012, Working Paper on www.ssrn.com.
[19] M. Podolskij, M. Vetter, Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps, Bernoulli 15 (3) (2009) 634–658.
[20] R. Roll, A simple implicit measure of the effective bid-ask spread in an efficient market, Journal of Finance XXXIX (4) (1984).
[21] M. Rosenbaum, A new microstructure noise index, Quantitative Finance 6 (2011) 883–899.
[22] L. Zhang, P. Mykland, Y. Ait-Sahalia, A tale of two time scales: determining integrated volatility with noisy high-frequency data, Journal of the American Statistical Association 100 (2005) 1394–1411.
[23] B. Zhou, High-frequency data and volatility in foreign exchange rates, Journal of Business and Economic Statistics 14 (1996) 45–52.