Measuring the relevance of the microstructure noise in financial data

Cecilia MANCINI*

November 16, 2012

Abstract

We show that the Truncated Realized Variance (TRV) of a semimartingale asset price converges to zero when observations are contaminated by microstructure noises. Under the additive iid noise assumption, a central limit theorem is also proved. In consequence it is possible to construct a feasible test allowing us to measure the relevance of the noise in the data of a given asset price at a given observation step. For a given observed price path we thus can optimally select the observation frequency at which we can "safely" use TRV to estimate the efficient price integrated variance IV. The *local size* of our test is investigated and its performance is verified on simulated data. A comparison conducted with Bandi and Russel (2008) and Ait-Sahalia, Mykland and Zhang (2005) mean square error criterions 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. The advantages of our method are at least two: on the one hand the underlying model for the efficient asset price is less restrictive in that any kind of Ito semimartingale (SM) jump component is allowed. 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 data is conducted in [15].

Key words. Semimartingales with jumps, integrated variance, threshold estimation, test to select optimal sampling frequency

Gel codes: C12, C14, C32, C60, G12

1 Introduction

When we want to estimate the integrated variance $IV \doteq \int_0^T \sigma_s^2 ds$ of an efficient data generating process (DGP) X based on observations $Y = X + \varepsilon$ affected by microstructure noise, we have to decide whether to pre-average the observations (as e.g. in [17]) 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 frequency at which we pick the observations. Given a time series of noisy Brownian semimartingale

*Dipartimento di Matematica per le Decisioni, Università di Firenze, Italy. Email: `cecilia.mancini@dmd.unifi.it`.

we can use the signature plot of the realized variance (RV_h) as a function of the observation step h to decide whether at a predetermined frequency the noise contamination is relevant or not; 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 signature plot (SP; [8]) is not necessarily such that $RV_{\hat{h}}$ leaves a reliable estimate of IV, given that $RV_{\hat{h}}$ cannot disentangle the estimation error from the error induced by the presence of noise. Moreover, in the presence of jumps in the DGP, RV 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 becomes relevant is theoretically minimizing the conditional mean square estimation error $RV_h - IV$, as described in [4], [3], [20]. However also there the efficient data generating process 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 (a test) 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 now going to propose such a test, to check its reliability on simulated data and it is in progress (in [15]) the application to empirical data. Questions that we judge to be interesting are 1) checking whether, as stated by some authors ([18]), the mid-quotes are less affected by noise than the transaction prices and at which extent; 2) for a given frequency, checking how much the pre-averaged time series has been decontaminated by the noise.

2 Model setup

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 Ito 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])

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

where $\int 1 \wedge \gamma^2(x,s,\omega)dx$ is a.s. finite. As such X_0 is called *Brownian semimartingale* 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

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

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 $\hat{IV}_h := \sum_{i=1}^n \Delta_i Y I_{\{(\Delta_i Y)^2 \le r_h\}}$, where, for any process Z, $\Delta_i Z = Z_{t_i} - Z_{t_{i-1}}$. The following further notation is used throughout the paper:

· $\Delta_i Z_\star = \Delta_i Z I_{\{|\Delta_i Z| \le \sqrt{r_h}\}}$, $\varepsilon_i := \varepsilon_{t_i}$

· 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 $\lim_n f(h)$.

· 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,

· given two real functions f, g (possibly the paths of a stochastic process for 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; $f(h) << g(h)$ means that $f(h) = o(g(h))$;

· $\approx$ denotes approximation of numerical results of computations;

· $\mathcal{N}(0, b^2)$ denotes a r.v. with Gaussian law having mean 0 and variance b^2; U denotes a standard Gaussian rv.

We also recall that

· for any process Y, $RV_h := \sum_{i=1}^n (\Delta_i Y)^2$;

· for any Ito SM X as above, $QV(X) = \int_0^T \sigma_s^2 ds + \sum_{s \le T} (\Delta J_s)^2$, where $\Delta J_s = J_s - J_{s-}$.

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

Assumption 2. *i) $\int_{|\gamma|>1} 1 dx$ 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 increments ε_i are normally $\mathcal{N}(0, c^2)$ or uniformly $\mathcal{U}[-c, c]$ distributed (which are the typical examples of additive i.i.d. noise). More generally it is verified each time that $Var(\varepsilon_i) \ne 0$ and the common law $\Delta_i \varepsilon(P)$, $i = 1..n$ has a density f continuous at 0. In fact the Assumption requests that the probability that an increment of the noise process be small is smaller and smaller 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 8.1 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 microstructure noises the threshold estimator of IV tends to zero rather than to IV.

Theorem 3.1. *Let $Y = X + \varepsilon$ be contaminated by microstructure noises 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.

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 $\hat{IV}_h \xrightarrow{P} \int_0^T \sigma_s^2 ds > 0$ ([16]), while if X is contaminated then $\hat{IV}_h \xrightarrow{P} 0$. Next theorem enables us to establish confidence intervals for $\hat{IV}_h$ being significantly far form 0. Observation of X 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 σ be 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 iid noise] Assume that for all h the r.v.s ε_{t_i}, $i = 1..n$, $n \in \mathbb{N}$, are IID with zero mean and finite strictly positive variance, 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*

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

ii)

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

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 iid noises is not completely realistic, however, as in many other works following [21], 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 7) below and Remark 1) after Ass. 2).

3) The above assumptions on ε are satisfied if e.g. the noise is additive iid with Gaussian ε_{t_i}.

4) When ε_{t_i} are i.i.d. with uniform laws, the density g is not Lipschitz over the whole $\mathbb{R}$, however, as shown in the proofs section (just after the proof of Theorem 3.2), the results i) and ii) above can be proved also for the uniform $\{\varepsilon_{t_i}\}$, by using the specific features of the uniform density.

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

6) By the iid property of the rvs ε_i, also the differences $u_i = \varepsilon_i - \varepsilon_{i-1}$ have a common density f, 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 $\mathcal{NB}_h$

7) $Var(\varepsilon_1) \neq 0$ 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.

8) In [19] 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 microstructure noises in the data, if we choose β close to one then $\sum_{i=1}^n (\Delta_i Y)^2_\star \xrightarrow{P} 0$ in a way such that Theorem 3.2, ii) holds true, and the feasible version of $\mathcal{NB}_h$ where $E[g(ep_1)] = f(0)$ is replaced by an estimate $\hat{f}(0)$ tends to a standard Gaussian rv. On the contrary, in the absence of the noise, since the econometrician believes that some noise affects the data anyway, he still implements the same feasible version of $\mathcal{NB}_h$, but now we have $\sum_{i=1}^n (\Delta_i Y)^2_\star \xrightarrow{P} IV \geq 0$ (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 $\Delta_i\varepsilon$ are Gaussian $\mathcal{N}(0, 2\sigma_\varepsilon^2)$ and estimated $f_n(0)$ through the empirical variance of the ultra high frequency returns $\Delta_i Y$ by

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

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

$$H_0) \text{ presence of the noise}, \qquad H_1) \text{ absence of the noise}.$$

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

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

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 the 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, than 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 rv 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 $\Delta_i Y$)
- 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. [**?**]), 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. 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 $\hat{IV}_h - IV$ tends to zero 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 would explode to infinity. We are thus proposing an alternative criterion to the ones proposed so far in the 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], [20]). 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 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 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 step for a given day is different from the step 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.

5 When the noise variance changes as $h \to 0$: the case $\rho_n \to \rho > 0$

In practice it is realistic to account for the fact that the noise variance can be different for different sampling frequencies. We are interested here in having the flavor of how the test response changes in this case. We now indicate by $\varepsilon_i^{(n)}$ the noise which is involved in the observations sampled at frequency $h = T/n$ and allow that when n changes the noise variance can change, but, for fixed n, $Var(\varepsilon_i^{(n)})$ is the same for all $i = 1..n$. We now denote $Var(\varepsilon_1^{(n)})$ by ρ_n^2 rather than by σ_ε^2, to recall that in this context such a variance is not constant. After some preliminary remarks, we separately tackle the case where $\rho_n \to \rho > 0$ 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 far 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 $H_1^{(n)}$) get closer and closer while $n \to \infty$.

If $\varepsilon = \lim_n \varepsilon^{(n)}$ was a SM then it would give a finite contribution to both $\lim_h RV_h(Y)$ and $\lim_h \hat{IV}_h$ rather than making $RV_h(Y)$ to explode and $\hat{IV}_h$ to tend to zero when $h \to 0$. However even if $\rho_n \to 0$ process ε is never a SM unless it has finite variation (FV). But in the last case the contribution to both $\lim_n RV_h(Y)$ and $\lim_h \hat{IV}_h$ would be null. We illustrate how the contribution of the noise process to $\lim_h RV_h(Y)$ and $\lim_h \hat{IV}_h$ depends on the behavior of ρ_n, in the specialized framework where for all n, for $i = 1..n$, $\varepsilon_i^{(n)}$ are iid with Gaussian law $\mathcal{N}(0, \rho_n^2)$. As for each fixed n the r.v.s $\Delta_{2k+1}\varepsilon^{(n)}$ are iid $\mathcal{N}(0, 2\rho_n^2)$, we apply the classical Lindeberg-Feller Central Limit Theorem (LF-CLT) for triangular arrays to $\sum_{k=1}^{(n-1)/2}(\Delta_{2k+1}\varepsilon^{(n)})^2$, and we reach that

$$\frac{\sum_{k=1}^{(n-1)/2}(\Delta_{2k+1}\varepsilon^{(n)})^2 - (n-1)\rho_n^2}{\sqrt{4(n-1)\rho_n^4}} \xrightarrow{d} \mathcal{N}(0,1). \tag{4}$$

Therefore in probability we have $\sum_{k=1}^{(n-1)/2} \Delta_{2k+1}(\varepsilon^{(n)})^2 \sim (n-1)\rho_n^2$ an the following cases are possible:
1) if $n\rho_n^2 \to +\infty$ the noise process ε cannot be a SM, and the contribution of the $\{\varepsilon_i^{(n)}\}_i$ can make $RV_h(Y)$ either to explode or not and $\hat{IV}_h$ either to tend to 0 or not (see remark 6.4);
2) if $n\rho_n^2 \to c \neq 0$ then the $\lim_h RV_h(\varepsilon^{(n)})$ is finite, but ε cannot be a non-trivial local martingale (if it was, we would have $\varepsilon_t = \int_0^t \alpha_s dW_s + \int_0^t \int_{\mathbb{R}} \gamma(x,s)[\mu(dx,ds) - dxds]$, for some processes α, γ but as $E[\varepsilon_t^2] = \lim_n E[(\varepsilon_t^{(n)})^2] = \lim \rho_n^2 = \lim 1/n = 0$ then $\alpha \equiv \gamma \equiv 0$ in probability), nor $\{\varepsilon_i^{(n)}\}_{i\in\mathbb{N}}$ is a local martingale for fixed n (if it was, as $\sup_i E[(\varepsilon_i^{(n)})^2] < \infty$ it would have to be a martingale, however this is not possible because $E_{i-1}[\varepsilon_i^{(n)}] = 0$ rather than $\varepsilon_{i-1}^{(n)}$). In this case the noise contribution to $\lim_h RV_h(Y)$ and $\lim_h \hat{IV}_h$ is not negligible but finite;
3) if $n\rho_n^2 \to 0$, then by using the LF-CLT it turns out that, in this Gaussian setting, ε is of FV iff $n\rho_n \to c \geq 0$, but in any case the noise process gives no contribution to $\lim_h RV_h(Y)$ nor to $\lim_h \hat{IV}_h$.

To verify the behavior of our test in the present framework we are interested in knowing the rate at which $\hat{IV}_n$ converges. The case where $\rho_n \to \rho > 0$ falls within case 1) above, while the case $\rho_n \to 0$, dealt with in the next section, can arise in any of the three cases above.

When $\rho_n \to \rho > 0$, Assumption 1 is still satisfied under the hypotheses of Theorem 5.1. For instance in the Gaussian case $\varepsilon_i^{(n)}(P) = \mathcal{N}(0, \rho_n^2)$ we have

$$P\{|\Delta_i \varepsilon^{(n)}| < c\sqrt{r_h}\} = \frac{\sqrt{2}}{\sqrt{\pi}} \int_0^{\frac{c\sqrt{r_h}}{\rho_n\sqrt{2}}} e^{-\frac{y^2}{2}} dy, \tag{5}$$

and using the Taylor approximation of the above integral we find that $P\{|\Delta_i \varepsilon^{(n)}| < c\sqrt{r_h}\} \sim \frac{\sqrt{r_h}}{\rho_n}$ which tends to zero like as $\sqrt{r_h}$.

Similarly as in the previous section, we require that for any n, the r.v.s $\varepsilon_i^{(n)}, i = 1..n$ are iid and that their law has a density g_n. Then also the first differences $u_i^{(n)} = \varepsilon_i^{(n)} - \varepsilon_{i-1}^{(n)}$ have a density f_n, and if we assume that there exists a bounded function $g(x,y)$ continuous in the variable y such that $g_n(x) = g(x, \rho_n)$ then

$$f_n(0) = \int g^2(x, \rho_n)dx \to_n \int g^2(x, \rho)dx \doteq f(0) > 0$$

(the last strict positivity follows from the fact that $\lim_n Var(\varepsilon_i^{(n)}) > 0$, implying that $g^2(x,\rho)$ cannot be null), and we show that an analogous version of Theorem 3.2 holds true in the present framework. It follows that with $\mathcal{S}_h = \frac{\hat{IV}_h - nr_h^{\frac{3}{2}} \frac{2}{3}\hat{f}_n(0)}{\sqrt{n} r_h^{\frac{5}{4}} \sqrt{\frac{2}{5}\hat{f}_n(0)}}$, as soon as $\sqrt{n} r_h^{1/4}(f(0) - \hat{f}_n(0)) \to 0$ we still have (3) and our test works in the same way as in the previous section.

Note that the densities of the Gaussian laws with zero mean and variances ρ_n^2, $n \in \mathbb{N}$, satisfy the requirements on the functional form of g_n. Further $\sqrt{n} r_h^{1/4}(f(0) - \hat{f}_n(0)) \to 0$ is fulfilled for instance in the specialized model $Y = \sigma W + \varepsilon$, if we use (1).

Theorem 5.1. *[CLT with changing noise] Assume that for any given n the r.v.s $\varepsilon_{t_i}^{(n)}$, $i = 1..n$, are IID with zero mean and finite variance ρ_n^2, and are independent on X. Further assume that there exists a bounded function $g(x,y)$, which is continuous in the variable y and Lipschitz in x in the following sense*

$$\forall x_1, x_2, y \in \mathbb{R}, \quad |g(x_1,y) - g(x_2,y)| \leq L \ \eta(y),$$

with L a constant and η a bounded function, and g is such that, for any fixed n, $\varepsilon_{t_i}^{(n)}(P)$ has density $g_n(x) = g(x,\rho_n)$, for all $i = 1..n$. Then if $\rho_n \to \rho > 0$, $w_n \doteq Var(g_n(\varepsilon_{t_i}^{(n)})) = \int g_n^3(x)dx - \left(\int g_n^2(x)dx\right)^2 \to \ell < \infty$ and $\beta > 2/3$ we have

i)

$$\frac{E[\hat{IV}_h]}{nr_h^{\frac{3}{2}}} \xrightarrow{P} \frac{2}{3}E[g(\varepsilon_1)] = \frac{2}{3}f(0);$$

ii)

$$\mathcal{NB}_h := \frac{\hat{IV}_h - nr_h^{3/2}\frac{2}{3}f(0)}{\sqrt{n}\ r_h^{\frac{5}{4}}\sqrt{\frac{2}{5}f(0)}} \xrightarrow{\mathcal{F}_0-stable} U$$

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 $\mathcal{S}$.

We remark that, for all n, $w_n > 0$ and if $\varepsilon_{t_i}^{(n)}(P) = \mathcal{N}(0,\rho_n^2)$ then $w_n = (1/\sqrt{3} - 1/2)/(2\pi\rho_n^2)$ and fulfill the above requirement.

6 Local size: when $\rho_n \to 0$

For conventional tests the *local power* is studied, where H_0) keeps fixed while the alternative H_1^n) moves towards the null as $n \to \infty$, and the local power of the test is defined as the limit of the probability of the (moving) critical region under the (moving) alternative. Here we only can move the null hypothesis, so we assume that under H_0^n) the noise has variance $\rho_n = Var(\varepsilon_1^{(n)})$ such that $\rho_n \to 0$ as $n \to \infty$, while we keep fixed the alternative of the absence of the noise. We call *local size* the quantity

$$\lim_n P\{|\mathcal{S}_h| > 1.96 | H_0^n\}.$$

The case $\rho_n \to 0$ splits into two cases where the local size of our test is different. We show the complete picture, in the chosen restricted framework specified below, after having made some preliminary remarks.

Firstly note that when $\rho_n \to 0$, Assumption 1 is never satisfied. More precisely, still in the case where $\varepsilon_i^{(n)}(P) = \mathcal{N}(0, \rho_n^2)$, by (5) we have:
1) if $\frac{\sqrt{r_h}}{\rho_n} \to \infty$ then $P\{|\Delta_i \varepsilon| < c\sqrt{r_h}\} \to 1$, meaning that when the noise variance tends to zero but is less than the squared threshold r_h then the increment $\Delta_i \varepsilon$ stays substantially always below $\sqrt{r_h}$.
2) If $\frac{\sqrt{r_h}}{\rho_n} \to c \neq 0$, then $P\{|\Delta_i \varepsilon| < c\sqrt{r_h}\} \to \ell \in (0,1)$ and still Assumption 1 is not satisfied.
3) If $\frac{\sqrt{r_h}}{\rho_n} \to 0$, then $P\{|\Delta_i \varepsilon| < c\sqrt{r_h}\} \sim \frac{\sqrt{r_h}}{\rho_n}$ which now tends to zero but more slowly than $\sqrt{r_h}$.

As Assumption 1 is not valid anymore, in order to find the rate of convergence of $\hat{IV}_h$ we cannot directly apply Theorem 3.2. In this paper we only are interested in understanding how substantially the size of our test changes when $\rho_n \to 0$, so we specialize here our framework as in the following:

Assumption 3. *$X \equiv \sigma W$; for all n, $\varepsilon_i^{(n)}$ are iid, $i = 1..n$, and independent on W, with law $\mathcal{N}(0, \rho_n^2)$, $\rho_n = h^\gamma, \gamma > 0$.*

Secondly, note that $\rho_n \to 0$ means that the densities g_n of $\{\varepsilon_i^{(n)}\}_{i=1..n}$ form a δ-sequence, because $\varepsilon_{1/n}^{(n)} \xrightarrow{P} 0$ so g_n tends a.s. to a delta function. It follows that $g_n(0) \to +\infty$ and $f_n(0) = E[g_n(\varepsilon_i^{(n)})] = \int g_n^2(x)dx \to +\infty$, which obliges us to properly reformulate the CLT for $\hat{IV}_h$. We obtain what follows.

Theorem 6.1. *[CLT for $\hat{IV}_h$ in the presence of vanishing noise] Under Assumption 3, taking $\beta \in (2/3, 1)$, we have*
a) if $\gamma > 1/2$ then
$$\frac{\hat{IV}_h - IV}{\sqrt{2hT\sigma^4}} \xrightarrow{d} U, \tag{6}$$
b) if $\gamma = 1/2$ then
$$\frac{\hat{IV}_h - (\sigma^2+2)T}{\sqrt{2h(\sigma^2+2)^2 T}} \xrightarrow{d} U, \tag{7}$$
c) if $\gamma \in (\beta/2, 1/2)$ then
$$\frac{\hat{IV}_h - 2nh^{2\gamma}}{\sqrt{8nh^{4\gamma}}} \xrightarrow{d} U, \tag{8}$$
d) if $\gamma = \beta/2$ then, with $\phi_2 \doteq \int_{-\infty}^{1/\sqrt{2}} y^2 e^{-\frac{y^2}{2}} dy/\sqrt{2\pi}$, $\phi_4 \doteq \int_{-\infty}^{1/\sqrt{2}} y^4 e^{-\frac{y^2}{2}} dy/\sqrt{2\pi}$ and $\psi \doteq 8\phi_4 - 16\phi_2^2 + 16\phi_2 - 16$,
$$\frac{\hat{IV}_h - 4nr_h(\phi_2 - 1/2)}{\sqrt{nr_h^2 \psi}} \xrightarrow{d} U, \tag{9}$$
e) if $\gamma \in (0, \beta/2)$ then
$$\frac{\hat{IV}_h - \frac{nh^{3\beta/2-\gamma}}{3\sqrt{\pi}}}{\sqrt{\frac{nh^{5\beta/2-\gamma}}{5\sqrt{\pi}}}} \xrightarrow{d} U. \tag{10}$$

As consequences in probability we have the following limits for $\hat{IV}_h$ in the listed different cases.
a) If $\gamma > 1/2$ then $\hat{IV}_h \to IV$, meaning that if $\rho_n << \sqrt{h} < \sqrt{r_h}$, even if the noise remains below the threshold (we are in case 1) of the list before Assumption 3 and in case 3) of the list after (4) and then enters within $\hat{IV}_h$, since it is of FV then its impact is negligible.

b) If $\gamma = 1/2$ then $\hat{IV}_h \to IV + 2T$, i.e. if $\rho_n \sim \sqrt{h} << \sqrt{r_h}$ then all the noise remains below the threshold and its magnitude is now such that its impact is non-negligible but finite.
c) and d): if $\gamma \in [\beta/2, 1/2)$ then $\lim_h \hat{IV}_h = +\infty$: the noise variance is now higher that $\sqrt{h}$ but still less (or equal) than $\sqrt{r_h}$, meaning that any $\Delta_i\varepsilon$ remains below the threshold but it is big and makes $\hat{IV}_h$ to explode.
e) If $\gamma \in (0, \beta/2)$ then $\lim_h \hat{IV}_h \in \{+\infty, c, 0\}$. Now $\rho_n >> \sqrt{r_h}$, and makes many terms $\Delta_i Y$ to go above the threshold and to be excluded by $\hat{IV}_h$. The higher ρ_n is the more terms are excluded thus making the $\lim_h \hat{IV}_h$ to decline again. In fact, since the denominator in (10) tends to zero, we have $\hat{IV}_h \sim nr_h^{3/2}/\rho_n = h^{3\beta/2-\gamma-1}$, which can either explode, or tend to a constant or to 0, depending on whether $\gamma >$ or $=$ or $< 3\beta/2 - 1$ respectively.

The above CLT allows us to understand the asymptotic behavior, and thus the local size, of our test when the noise is vanishing. In practice we do not know which kind of noise is in play and thus we implement in any case $\mathcal{S}_h$ as in the last term of (2).

Theorem 6.2. *[Asymptotic behavior of our test when the noise is vanishing] Under Assumption 3, taking $\beta \in (2/3, 1)$ and $\hat{f}_n(0)$ as in (1), we have what follows:*
in the cases a), b), c), d) above, i.e. if $\gamma \geq \beta/2$, then in probability $\mathcal{S}_h \to -\infty$;
in case e), i.e. if $\gamma \in (0, \beta/2)$, then in distribution $\mathcal{S}_h \to U$.

The local size immediately follows and is shown to depend on how $\rho_n \to 0$.

Corollary 6.3. *[Local size] Under Assumption 3, taking $\beta \in (2/3, 1)$:*
if $\gamma \geq \beta/2$ then for any fixed quantile $q_\alpha \geq 0$ of the law of $|U|$, we have $P\{|\mathcal{S}_h| > q_\alpha | H_0^n\} \to 1$;
if $\gamma \in (0, \beta/2)$ then $P\{|\mathcal{S}_h| > q_\alpha | H_0^n\} \to \alpha$.

In the first case above the asymptotic behavior of $P\{|\mathcal{S}_h| > q_\alpha | H_0^{(n)}\}$ is the same as for $P\{|\mathcal{S}_h| > q_\alpha | H_1\}$, and thus we are not able to disentangle the relevance of the noise by using our test. In fact we have $\rho_n \leq \sqrt{r_h}$, meaning that the noise is too small to be recognized: it would be judged to be negligible by the test because $|\mathcal{S}_h|$ under $H_0^{(n)}$ explodes. On the contrary whenever $\gamma < \beta/2$, i.e. $\rho_n > \sqrt{r_h}$ the local size of the test coincides with the size of the case where the noise does not vanish, meaning that it is sufficiently big that in principle we would have to be able to recognize its relevance. The watershed exponent of h within ρ_n for the two different cases is exactly $\beta/2$, the exponent within the threshold: the noise is judged to be negligible by our test when $\gamma \geq \beta/2$, which is exactly when it falls below the threshold, as $\lim_n \sqrt{r_h}/\rho_n \in (0, \infty]$ is when $P\{|\Delta_i\varepsilon| < c\sqrt{r_h}\} <\to \ell > 0$.

Remark 6.4. *If $\gamma < \beta/2$ we have $\frac{\sqrt{r_h}}{\rho_n} \to 0$ and the test for the relevance of the noise has the same size as in section 4, when σ_ε was fixed. However now the 95% confidence bands*

$$\hat{IV}_n \in \Big(2nr_h^{3/2}\hat{f}_n(0)/3 - 1.96\sqrt{2nr_h^{5/2}\hat{f}_n(0)/5},\ \ 2nr_h^{3/2}\hat{f}_n(0)/3 + 1.96\sqrt{2nr_h^{5/2}\hat{f}_n(0)/5}\Big),$$

that we use for deciding whether the noise is negligible or not, have amplitude of the order of $c\sqrt{nr_h^{5/2}h^{-\gamma}}$, which for large n is larger than in the case of fixed σ_ε, when the amplitude was $c\sqrt{nr_h^{5/2}}$. This can be *explained by the fact that a noise with vanishing variance confuses more with $\Delta_i X$ than when $\sigma_\varepsilon > 0$ keeps fixed. Note however that $\gamma < \beta/2$ means that $\rho_n >> \sqrt{r_h} >> \sqrt{h}$, i.e. $n\rho_n^2 \to \infty$, but the impact of the noise on $\hat{IV}_h$ can either be relevant or negligible, since the probability limit of $\hat{IV}_h \sim n r_h^{3/2} h^{-\gamma}/(3\sqrt{\pi})$ can either be 0, or a positive constant, or infinite.*

7 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 ((12), (13) or (14)) described below, the noise is additive and given by a process ε which is independent on X. The rvs ε_{t_i} are iid uniform and centered with different possible values of the variance parameter in the different experiments: $Var(\varepsilon_i) \doteq \sigma_\varepsilon^2 = 2 \times 10^{-7}$ (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 $\hat{IV}$:

$$\hat{f}(0) = 1/\sqrt{12\pi\hat{\sigma}_{ep}^2}. \tag{11}$$

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 realistically around 0.4, 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 hour open market (T=0.004 years), then $h = 1/(252 \times 7 \times 60 \times 60)$ and $n = 25200$, 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 $\hat{IV}_h$ correcly 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:

$$dX = 0.4dW + dJ, \quad J \equiv J_{1t} = \sum_{\ell \leq N_t} Y_\ell, \tag{12}$$

where J is a compound Poisson jump process, N is a simple Poisson process with intensity $\lambda = 5$, $\forall \ell,\ Y_\ell(P) = \mathcal{N}(0, 0.6^2)$, and the parameters are realistically 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:

$$dX = -\sigma_t^2/2dt + \sigma_t dW_t + dJ_t, \quad d\log\sigma_t = -k(\log\sigma_t - \theta)dt + \nu dW_t^{(2)}, \quad dJ_t = \sum_{\ell \leq N_t} Y_\ell. \tag{13}$$

The σ parameters $\log\sigma_0 = log(0.4), k = 0.09, \theta = \log(0.25), \nu = 0.05, \rho = corr(W_t, W_t^{(2)}) = -0.7\ \forall t$, produce similar σ paths as in [9].

MODEL G-CGMY: constant volatility and CGMY jumps.

$$dX = 0.3815\ dW + dJ, \tag{14}$$

where J is a CGMY process as proposed in [5] with scale parameter on the J Levy density $C = 280.11$, tail decay parameter of the density 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 Figure 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 $pct = 1, MEE = mean(100(\hat{IV} - IV)/IV) = 10.3434, SEE \doteq \sqrt{Var(100(\hat{IV} - IV)/IV)} = 3.9709$ in the first case, while $pct = 0.0452, MEE = -84.5334, SEE = 2.3202$, in the second one.

SECOND CHECK, Model GP. Under Model 1 we check size ($\tau = 1$) and power ($\tau = 0$) of our test, for fixed $T = 0.004$ years, either with h =1 second (n=25200), or 5 minutes (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

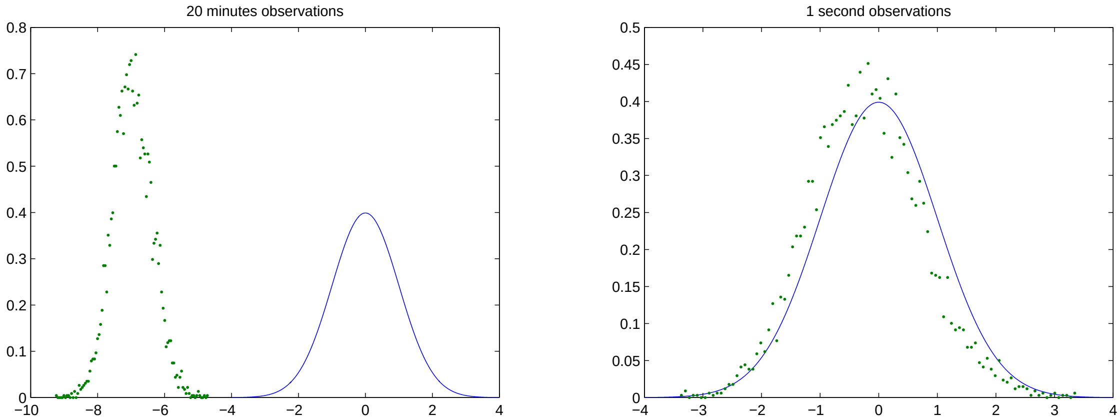

Figure 1: Empirical density of the test statistic $\mathcal{S}_h$ under the simulated model GP plus additive iid uniform noise with $\sigma_\varepsilon^2 = 8 \times 10^{-6}$ and $n = 1000$ observations, $h = 1/(252 \times 21)$ i.e. 20' (left), $h = 1/(252 \times 7 \times 60 \times 60)$ i.e. 1" (right),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

available and the value $100 \times (\hat{IV} - IV)/IV$ is computed only once, and that $\hat{f}(0)$ is still as in (11). When $\tau = 1$ then $H = 1000$ paths are generated in each scenario. The produced results are as in Table 1 below.

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						

TABLE 1. Performance of our test $\mathcal{S}_h$ under Model GP: τ 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 $Y = X + \tau\varepsilon$, while when $\tau = 0$ then $\mathcal{S}_h$ is implemented only once.

Substantially the statistic behaves as one would expect: for instance, when the variance is low and we sample at 20', for the 100% of the paths of Y $|\mathcal{S}_h|$ 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 UHF than at HF. 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. The following Table 2 confirms the previous results.

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

We finally consider Model G-CGMY. The outcomes for pct are the following.

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

In this 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". It is possible that 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.

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 (12). Given an observation step h we vary σ_ε^2 and compute the resulting pct value. Figure 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 $pct \leq 0.05$, so we can see that for $h = 5'$ noises with variance less than or equal to about 10^{-8} are negligible, while with $h = 1''$ noises with variance between $10^{-8.5}$ and 10^{-8} are already relevant, as one would expect, because for the same level of noise the impact on the returns is higher at lower frequency.

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

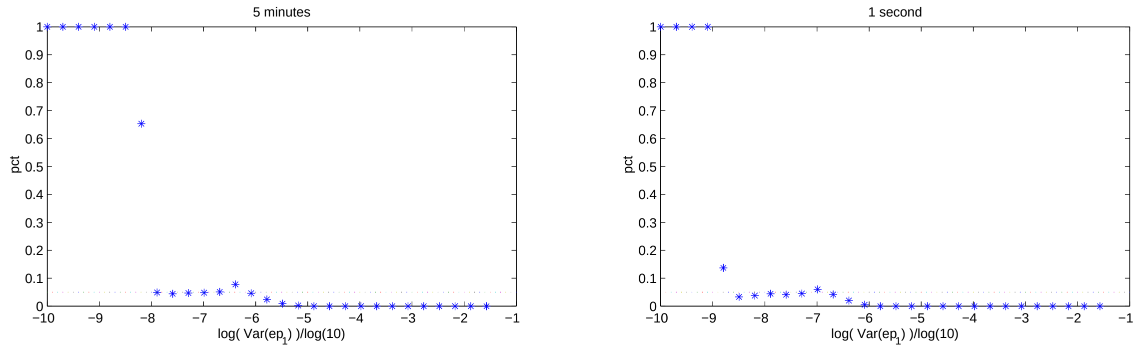

Figure 2: Plot of $pct = \frac{\#\{|\mathcal{S}_h|>1.96\}}{H}$ as a function of σ_ε^2. From model GP plus additive iid uniform noise $H = 1000$ paths were generated. Each path is observed at $n = 1000$ points in time and the observation step is either of five minutes (left panel) or one second (right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

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=33600 observations with minimum discretisation step $h_{min} = 1$", 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 observation step 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], [20] and [3], which give an approximately optimal MSE. In [4] (p.348), for a Brownian semimartingale model X (i.e. a semimartingale without jump part, hereafter indicated by BSM) with iid additive noise, the observation step minimizing MSE is $\hat{h} = T/\hat{n}$ where $\hat{n}$ minimizes

$$2\frac{T}{n}(IQ + o(1)) + 4nE[\varepsilon_1^4] + 4n^2\sigma_\varepsilon^4 + 8IV\sigma_\varepsilon^2 + 2\sigma_\varepsilon^4 - E[\varepsilon_1^4],$$

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

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

In [20] (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 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] and [4] is explained by the fact that the ML estimator coincides with RV_h/T.

We firstly assume Model SV-PJ and uniform noise. Figure 3 visualizes a comparison of the different answers given by the different four criteria $\mathcal{S}_h$, SP, h_{BR}, $\tilde{h}_{BR}$ within 4 different scenarios. The red squares with a red cross inside and connected by a red line represent the SP of RV_h (RV_h without further specification means $RV_h(Y)$, while later we use $RV_h(X)$ to indicate the realized variance of the efficient process X) 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, 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 no jumps occurred 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'), one would decide not to use observations with step below 50' in order to consider RV_h as a reliable estimate of IV. However note that, in this case, with 50' observations, RV_h does not approximate IV (pink continuous line) nicely. We also reported the unobservable $\log(RV_h(X))$ (red crosses), $\log(\hat{IV}_h(X))$ (blue circles) and the log of the 95% confidence band (pink dotted lines) indicating when $\hat{IV}_h(X)$ is an acceptable estimate of IV. Such confidence band is computed on the basis of the CLT

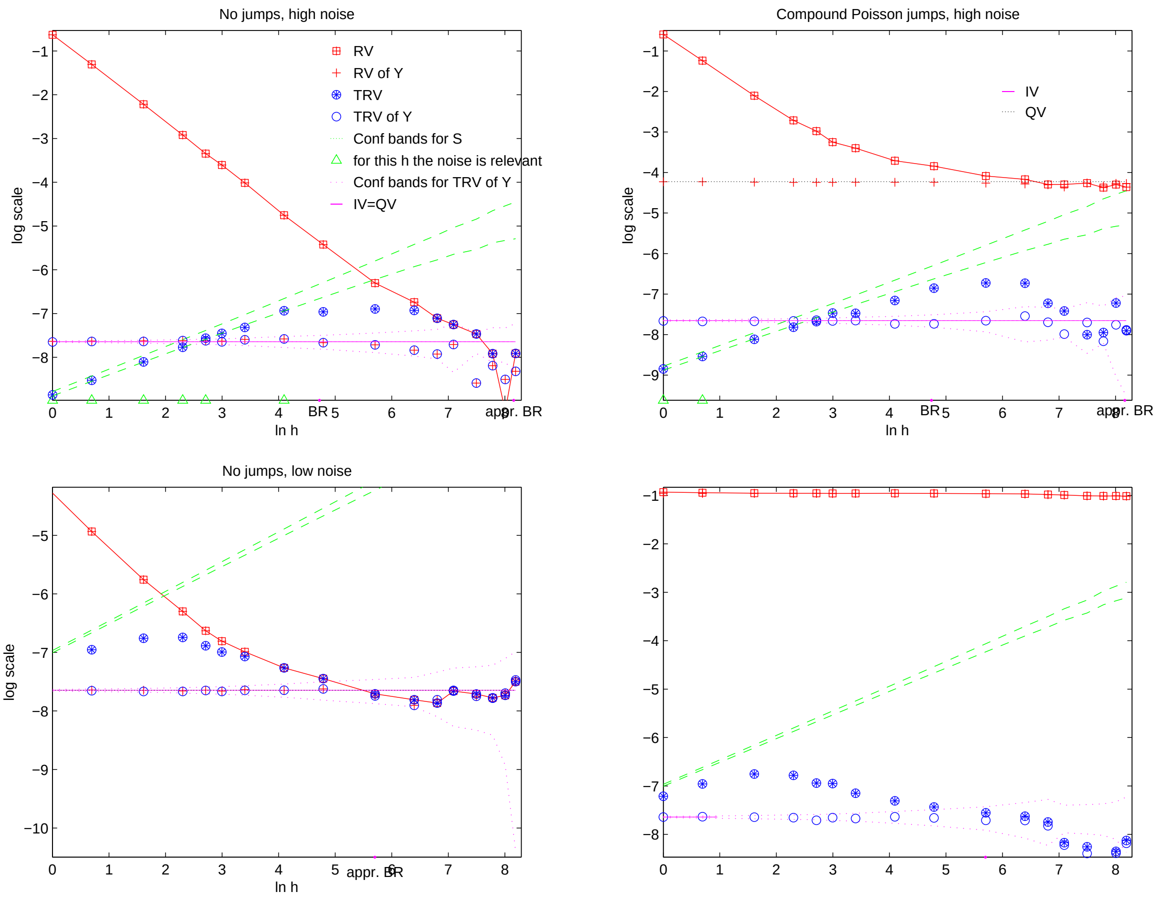

Figure 3: Optimal choices of h to estimate IV, using the different criteria of $\mathcal{S}_h$, SP, h_{BR}, $\tilde{h}_{BR}$, under model SV-PJ. The noise is additive iid uniform, $n = 33600$, $T = 0.0053$, $r_h = h^{0.999}$. The figures in the top row are characterized by $\sigma_\varepsilon^2 = 8 \times 10^{-6}$ (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 along the simulated path of Y no jumps occurred, 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 $\hat{IV}_h$.

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 $\tilde{h}_{BR}$, pink points on the x-axis) would suggest that on average it is safe to use RV_h with $h_{BR} \approx e^{4.7}$ (about 2') or $\tilde{h}_{BR} = e^{8.2}$ (61'). 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 at anyone of the two observation frequencies, as in both cases RV_h is outside the pink dotted confidence range (in this framework of no jumps $Plim_h RV_h = Plim_h \hat{IV}_h$ 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 step ($h \approx e^3$, i.e. about 33") and still safely to approximate IV. In fact the blue stars surrounded by blue circles represent the values assumed by $\log(\hat{IV}_h)$. The green dashed lines represent the log of the 95% confidence interval for S_h behaving like a standard Gaussian rv, thus indicating relevance of the noise. A green 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 green 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. Note that with $h = e^{2.2}$ the $\hat{IV}_h$ estimation error would not be that different than with $h = e^3$ (in terms of the distance from the pink line), however the test realizes that the noise is too relevant if $h = e^{2.2}$ and it is safer not to rely on the response of $\hat{IV}_h$. On the other hand, with e.g. $h = e^{5.8}$ the estimation error of $\hat{IV}_h$ is higher than with e.g. $h = e^{7.5}$, however $\hat{IV}_h - IV$ is of the magnitude order of about 6×10^{-4} in both cases, which is considered acceptable by the test.

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 $h \geq e^{6.4}$ (10'). However, if some jumps occur, as in the top right panel of Figure 3, we know that it is forbidden to use RV_h to estimate IV, because RV_h tends to $QV = IV + \sum_{t\leq T}(\Delta J_t)^2$. 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 in fact small.

On the other hand the bottom left panel of Figure 3 shows the comparison among the illustrated optimal frequency selection criterions when the noise variance is decreased to $\sigma_\varepsilon^2 = 2 \times 10^{-7}$. In this case it turns out that $h_{BR} \approx e^{-2.6}$ (which falls outside the x-axis range, and corresponds to about 0.07”), 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 green 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 UHF, 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 with uniform noises. We have similar pictures (Figure 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.

8 Appendix: proofs of the results

Lemma 8.1. *Under Assumption 2 we have, for all n,*

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

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

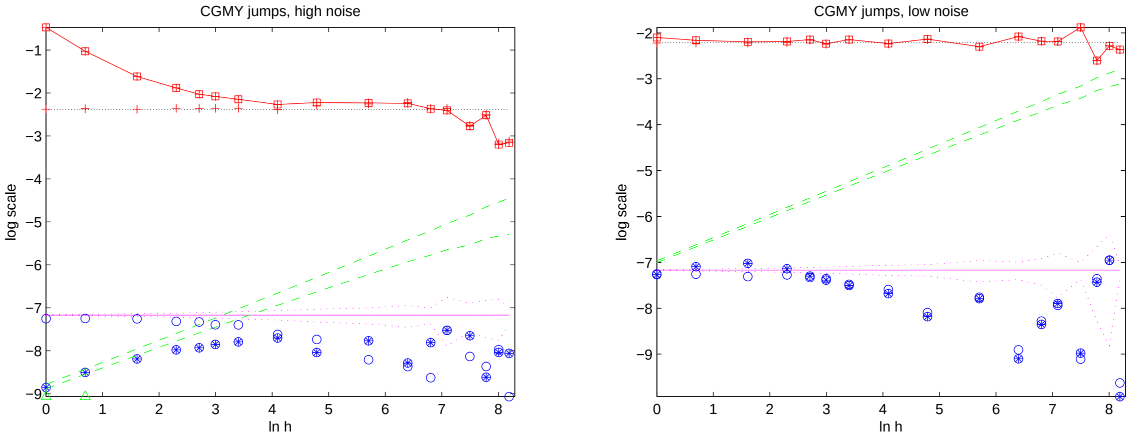

Figure 4: Optimal choices of h to estimate IV, using the different criteria of $\mathcal{S}_h$, SP, h_{BR}, $\tilde{h}_{BR}$, under Model G-CGMY. The noise is additive iid uniform, with $\sigma_\varepsilon^2 = 8 \times 10^{-6}$ in the left panel and $\sigma_\varepsilon^2 = 2 \times 10^{-7}$ in the right one. n = 33600, T = 0.0053, $r_h = h^{0.999}$.

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.

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

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

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

$$\le c\big(h \ln \frac{1}{h} + r_h\big) N_T \stackrel{a.s.}{\to} 0.$$

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

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

As for I_1, on $\{\Delta_i N = 0, |\Delta_i J + \Delta_i \varepsilon| \leq 2\sqrt{r_h}, |\Delta_i \tilde{J}_2| \leq \sqrt{r_h}\}$ we have $2\sqrt{r_h} \geq |\Delta_i J + \Delta_i \varepsilon| \geq |\Delta_i \varepsilon| - |\Delta_i \tilde{J}_2|$ then $|\Delta_i \varepsilon| \leq 2\sqrt{r_h} + |\Delta_i \tilde{J}_2| \leq 3\sqrt{r_h}$ and by assumption 1 we reach that

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

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

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

with $A > 1$ any constant. We have

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

we now show that on $\{\Delta_i N = 0, (\Delta_i Y)^2 \leq r_h, (\Delta_i X)^2 > A r_h\}$ we have $(\Delta_i \tilde{J}_2)^2 > c r_h$, for a suitable constant c. In fact, given any constant $\delta > 0$, as before, a.s. for sufficiently small h if $(\Delta_i Y)^2 \leq r_h$ then

$$|\Delta_i \tilde{J}_2 + \Delta_i \varepsilon| \leq (1 + \delta)\sqrt{r_h}; \tag{15}$$

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

$$|\Delta_i \varepsilon| = |\Delta_i Y - \Delta_i X| > |\Delta_i X| - |\Delta_i Y| \geq \sqrt{r_h} - \sqrt{r_h} = (\sqrt{} - 1)\sqrt{r_h}. \tag{16}$$

putting together (15) and (16) we reach

$$|\Delta_i \tilde{J}_2| = |\Delta_i \tilde{J}_2 + \Delta_i \varepsilon - \Delta_i \varepsilon| > |\Delta_i \varepsilon| - |\Delta_i \tilde{J}_2 + \Delta_i \varepsilon| \geq (\sqrt{} - 1 - 1 - \delta)\sqrt{r_h}$$

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,

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

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

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

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

$$\sum_{i=1}^n (\Delta_i X_0)^2 \big[-I_{\{(\Delta_i X)^2 \leq A r_h\} \cap \{(\Delta_i Y)^2 > r_h\}} + I_{\{(\Delta_i X)^2 \leq A r_h\}} \big] = \sum_{i=1}^n (\Delta_i X_0)^2 I_{\{(\Delta_i X)^2 \leq A r_h, (\Delta_i Y)^2 \leq r_h\}}. \tag{17}$$

However on $\{(\Delta_i X)^2 \leq A r_h, (\Delta_i Y)^2 \leq r_h\}$ we have $|\Delta_i \varepsilon| = |\Delta_i Y - \Delta_i X| \leq |\Delta_i Y| + |\Delta_i X| \leq \sqrt{r_h} + \sqrt{A r_h}$, so that almost surely (17) is bounded by $h \ln \frac{1}{h} \sum_{i=1}^n I_{\{|\Delta_i \varepsilon| \leq (\sqrt{A}+1)\sqrt{r_h}\}}$, whose expectation is $O(nh \ln \frac{1}{h} \sqrt{r_h}) \to 0$. $\square$

Lemma 8.2. *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 $i = 1..n$ $\exists$ $\xi_i = \xi_i^n(\omega) \in (0, r)$:

$$H_i^q(r) = \frac{r^{(q+1)/2}}{q+1}\big[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})\big]. \tag{18}$$

2) $\forall n, \forall i$

$$E_{i-1}[(\Delta_i Y_\star)^q] = E_{i-1}[H_i^q(r_h)]$$

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

3) $\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})] \stackrel{L^1}{\to} E[g(\varepsilon_1)]$ *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

$$H_i^{(q)}(r) = \frac{(H_i^{(q)})'(\xi_i)}{(G^{(q)})'(\xi_i)}\ G^{(q)}(r) = \frac{g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})}{q+1}\ r^{\frac{q+1}{2}}. \tag{19}$$

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

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

$$E_{i-1}\big[\int_{-\sqrt{r_h}}^{\sqrt{r_h}} u^q g(u - \Delta_i X + \varepsilon_{i-1}) du\big] = E_{i-1}\big[H_i^{(q)}(r_h)\big],$$

having changed variable as $u = \Delta_i X + z - \varepsilon_{i-1}$. Now for fixed (i, h), for any fixed ω we have equality [?], so for fixed (i, h) the two terms in [?] 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 $\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(\varepsilon_{i-1})] = \frac{1}{n}\sum_{i=1}^n g(\varepsilon_{i-1}) \stackrel{L^2}{\to} E[g(\varepsilon_1)]$. 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 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,

$$E\left[\left|\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) - g(\varepsilon_{i-1})]\right|\right] \le \frac{L}{n}\sum_{i=1}^n E[|s\sqrt{\xi_i} - \Delta_i X|]. \tag{20}$$

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 $c(E[|s\sqrt{\xi_i}|] + E[|\Delta_i X|]) \le c\sqrt{r_h} \to 0$. $\square$

Proof of theorem 3.2.

i) We have

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

by Lemma 8.2 part 2) the last expectation equals

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

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

and the thesis follows from Lemma 8.2 part 3).

In order to prove ii), we apply a classical theorem of convergence for sums of rvs belonging to a triangular array ([11], Lemma 4.3) 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

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

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

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

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),

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

$$\frac{r_h^{\frac{3}{2}}}{3}\sum_{i=1}^{[t/h]} \frac{E_{i-1}\Big[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) - 2g(\varepsilon_{i-1})\Big]}{\sqrt{n}r_h^{\frac{5}{4}}\sqrt{\frac{2}{5}E[g(\varepsilon_1)]}} + \tag{21}$$

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

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

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

which tends to zero by the assumption $\beta > 2/3$. As for the second term in (21), it coincides with

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

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

As for condition (b), using Lemma 8.2 we have

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

$$cr_h^3 \frac{\sum_{i=1}^{[t/h]} E_{i-1}^2\big[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})\big]}{nr_h^{\frac{5}{2}}} + cr_h^3 \left[\frac{t}{h}\right] \frac{E^2\big[g(\varepsilon_1)\big]}{nr_h^{\frac{5}{2}}} :$$

the last term has the same asymptotic behavior as $[t/h]r_h^3/(nr_h^{5/2}) \sim r_h^{1/2} \to 0$, 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 8.2 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

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

The last term is of the same order as $r_h/n \to 0$, while, using again Lemma 8.2, 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

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

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 iid with uniform law, $\beta > 2/3$. We begin by checking the validity of Lemma 8.2. 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 (18) 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') $\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})] \stackrel{L^1}{\to} E[g(\varepsilon_1)]$ for both cases $s = +1$, and $s = -1$.

Proof. Parts 1'), 2') are proved analogously as for Lemma 8.2. As for 3') we only have to show that

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

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

$$E\Big|\frac{1}{n}\sum_{i=1}^n E_{i-1}[g(s\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) - g(\varepsilon_{i-1})]\Big|$$

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

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

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

$$\frac{1}{nC}\sum_{i=1}^n \Big(E[\int_{C/2 + \Delta_i X - \sqrt{r_h}}^{C/2} 1dz\ I_{\Delta_i X - \sqrt{r_h} < 0}] + E[\int_{-C/2}^{-C/2 + \Delta_i X + \sqrt{r_h}} 1dz\ I_{\Delta_i X + \sqrt{r_h} > 0}]\Big)$$

$$\le \frac{c}{n}\sum_{i=1}^n E[\sqrt{r_h} + |\Delta_i X|] \le c \sup_i (E[|\Delta_i X|] + \sqrt{r_h}) \to 0.$$

$\square$

We now prove the result i) and ii) of Theorem 3.2 when ε_i are uniform.
For i), we have

$$E[\hat{IV}] = E\Big[\sum_i E_{i-1}[(\Delta_i Y_\star)^2\ I_{\{(-\sqrt{r_h},\sqrt{r_h})\subset(\Delta_i X-\varepsilon_{i-1}-C/2,\Delta_i X-\varepsilon_{i-1}+C/2)\}}]\Big]+ \\ E\Big[\sum_i E_{i-1}[(\Delta_i Y_\star)^2\ I_{\{(-\sqrt{r_h},\sqrt{r_h})\subset(\Delta_i X-\varepsilon_{i-1}-C/2,\Delta_i X-\varepsilon_{i-1}+C/2)\}^c}]\Big]. \tag{22}$$

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, for any i the last term equals

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

and the first term equals

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

thus in (22) the second term is dominated by

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

as $\beta > 2/3$ and $\sup_i E[|DX|] \leq c\sqrt{h}$.

On the other hand, to the first term on the rhs of (22) we can apply result 2') above, and obtain

$$nr_h^{3/2}\frac{1}{3}E\Big[\frac{1}{n}\sum_i E_{i-1}[g(\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1}) + g(-\sqrt{\xi_i} - \Delta_i X + \varepsilon_{i-1})\Big]-$$

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

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 (21), the only point where we used the Lipshitz 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

$$c\frac{r_h^{1/4}}{\sqrt{n}}\sum_{i=1}^{[t/h]} E_{i-1}[I_{\{\varepsilon_{i-1}\in(-C/2,C/2),|\varepsilon_{i-1}-\Delta_i X+s\sqrt{\xi_i}|>C/2\}}] \leq cr_h^{1/4}\sqrt{n}\sup_i(E[|\Delta_i X|] + \sqrt{r_h}) \leq cr_h^{5/4}\sqrt{n} \to 0,$$

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

Proof of Theorem 5.1 [CLT with changing noise] Firstly note that the analogous of Lemma 8.2 holds true with $\varepsilon_i^{(n)}$ and $\xi_i^{(n)}$ in place of ε_i and ξ_i, and, at point 3), with $f(0)$ in place of $E[g(\varepsilon_1)]$. In fact points 1) and 2) are obtained exactly as for Lemma 8.2, while for point 3) we consider the triangular array $\sum_{i=1}^n \phi_{i,n}$, with $\phi_{i,n} = g_n(\varepsilon_i^{(n)})/n$. Since $\sum_i E_{i-1}[\phi_i^{(n)}] \to f(0)$ and $\sum_i E_{i-1}[(\phi_i^{(n)})^2] \to 0$, we have $\sum_{i=1}^n \phi_{i,n} \stackrel{ucp}{\to} f(0)$. We then write $L_n \doteq L\eta(\rho_n)$ and remark that $L_n, n \in \mathbb{N}$ are bounded, thus exactly the same reasoning as in Lemma 8.2 can be applied here.

Secondly, the proof of i) of Theorem 5.1 follows exactly the same lines as for Theorem 3.2.

Third, remark that the speed of convergence of $\sum_{i=1}^n \phi_{i,n}$ is $\sqrt{w_n/n}$. In fact $\sum_i E_{i-1}[(\phi_i^{(n)})^3] = \int g_n^4(x)dx/n^2 \to 0$ and thus by the Lindeberg-Feller CLT the speed is $\sqrt{nVar(\phi_i^{(n)})} = \sqrt{w_n/n}$. It follows that, while following the proof of the convergence to zero of the second term in (21), we still obtain $\frac{r_h^{\frac{1}{4}}}{\sqrt{n}}\Big[\sum_{i=1}^{[t/h]} g_n(\varepsilon_{i-1}^{(n)}) - \big[\frac{t}{h}\big] f(0)\Big] \le\sim r_h^{1/4}\sqrt{w_n} \to 0$.

This said, also the proof of ii) follows the one for ii) given in Theorem 3.2. $\square$

Proof of Theorem 6.1 [CLT for $\hat{IV}_h$ in the presence of vanishing noise]. We prove a CLT separately for the sum $\sum_{i=2k+1} I_{\{(\Delta_i Y)^2 \le r_h\}}$ of the terms with odd index and the sum $\sum_{i=2k} I_{\{(\Delta_i Y)^2 \le r_h\}}$ of the terms with even index of $\hat{IV}_h$. As the two sums are independent, a CLT holds for the global sum with limit law being the sum of the two separate limits laws. In order to compute the mean and variance of each summand we need to evaluate $E[(\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}]$ and $E[(\Delta_i Y)^4 I_{\{(\Delta_i Y)^2 \le r_h\}}]$. Under Assumption 3 we have

$$E[(\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}] = \frac{2}{\sqrt{2\pi v_h}} \int_0^{h^{\beta/2}} x^2 e^{-\frac{x^2}{2v_h}} dx = \sqrt{\frac{2}{\pi}} v_h \int_0^{\frac{h^{\beta/2}}{\sqrt{v_h}}} y^2 e^{-\frac{y^2}{2}} dy,$$

where $v_h = \sigma^2 h + 2\rho_n^2$ and where the above equality is obtained by changing variable $y = x/\sqrt{v_h}$, and

$$E[(\Delta_i Y)^4 I_{\{(\Delta_i Y)^2 \le r_h\}}] = \sqrt{\frac{2}{\pi}} v_h^2 \int_0^{\frac{h^{\beta/2}}{\sqrt{v_h}}} y^4 e^{-\frac{y^2}{2}} dy.$$

According to the choice of γ we have 5 following different cases, where when we indicate asymptotic equivalence we also account for the correct constants. In any case the Lindeberg condition turns out to be verified.

a) If $\gamma > 1/2$ then $v_h \sim h$, $\frac{h^{\beta/2}}{\sqrt{v_h}} \to +\infty$, $E[(\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}] \sim \sigma^2 h$, $E[(\Delta_i Y)^4 I_{\{(\Delta_i Y)^2 \le r_h\}}] \sim 3\sigma^4 h^2$ and $Var((\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}) \sim 2\sigma^2 h^2$, so (6) follows. As a consequence, since the denominator tends to zero, $\hat{IV}_h \to IV$.

b) If $\gamma = 1/2$ then $v_h = (\sigma^2 + 2)h$, $\frac{h^{\beta/2}}{\sqrt{v_h}} \to +\infty$, $E[(\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}] \sim (\sigma^2 + 2)h$, $E[(\Delta_i Y)^4 I_{\{(\Delta_i Y)^2 \le r_h\}}] \sim 3(\sigma^2 + 2)^2 h^2$, $Var((\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}) \sim 3(\sigma^2 + 2)^2 h^2$ and (7) follows, entailing that $\hat{IV}_h \to IV + 2T$.

c) If $\gamma \in (\beta/2, 1/2)$ then $v_h \sim 2\rho_n^2$, $\frac{h^{\beta/2}}{\sqrt{v_h}} \to \ell \in \{+\infty, c\}$, $E[(\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}] \sim 2\rho_n^2$ is slower than h, $E[(\Delta_i Y)^4 I_{\{(\Delta_i Y)^2 \le r_h\}}] \sim 12h^{4\gamma}$, $Var((\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}) \sim 8h^{4\gamma}$ so (8) holds true, and implies that $\lim_h \hat{IV}_h = +\infty$.

d) If $\gamma = \beta/2$ then $v_h \sim 2\rho_n^2$, $\frac{h^{\beta/2}}{\sqrt{v_h}} \to 1/\sqrt{2}$, $E[(\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}] \sim 4(\phi_2 - 1/2)h^{2\gamma}$, $E[(\Delta_i Y)^4 I_{\{(\Delta_i Y)^2 \le r_h\}}] \sim 8r_h^2(\phi_4 - 3/2)$, $Var((\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \le r_h\}}) \sim r_h^2 \psi$, and (9) follows. Therefore we reach that $\hat{IV}_h \to \infty$.

e) If $\gamma < \beta/2$ then $v_h \sim 2\rho_n^2$, $\frac{h^{\beta/2}}{\sqrt{v_h}} \to 0$, and $F(h) \doteq \int_0^{\frac{h^{\beta/2}}{\sqrt{v_h}}} y^2 e^{-\frac{y^2}{2}} dy \to 0$. By applying the Cauchy theorem to $F(h)/G(h)$ with $G(h) = \frac{h^{3(\beta/2-\gamma)}}{6\sqrt{2}}$ we obtain that $E[(\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \leq r_h\}}] \sim h^{\frac{3}{2}\beta-\gamma}/(3\sqrt{\pi})$. Further still by using Cauchy theorem with $F(h) = \int_0^{h^{\beta/2-\gamma}/\sqrt{2}} y^4 e^{-\frac{y^2}{2}} dy$ and $G(h) = (\beta/2-\gamma)h^{5(\beta/2-\gamma)}/(20\sqrt{2}(\beta/2-\gamma))$, we have $E[(\Delta_i Y)^4 I_{\{(\Delta_i Y)^2 \leq r_h\}}] \sim h^{(5\beta/2-\gamma)}/(5\sqrt{\pi})$ and $Var((\Delta_i Y)^2 I_{\{(\Delta_i Y)^2 \leq r_h\}}) \sim h^{\beta/2-\gamma}$, so that (10) is proved. Since the denominator tends to 0, we obtain that $\hat{IV}_h \sim nh^{3\beta/2-\gamma}/(3\sqrt{\pi})$, which in fact can either explode, or tend to to a constant or to 0, depending on whether $r_h^{3/2} n/\rho_n = h^{3\beta/2-\gamma-1} \to +\infty$ or to a constant or to 0, respectively. □

Proof of Theorem 6.2 [Asymptotic behavior of our test when the noise is vanishing] By using the Lindeberg-Feller CLT we find that, with the correct constant,

$$\hat{f}_n(0) \sim \frac{1}{\sqrt{2\pi(\sigma^2 h + 2\rho_n^2)}} \sim \begin{cases} \frac{h^{-1/2}}{\sqrt{2\pi\sigma^2}} & \text{if } \gamma \geq 1/2 \\ \\ \frac{h^{-\gamma}}{\sqrt{4\pi}} & \text{if } \gamma < 1/2. \end{cases} \tag{23}$$

Then the statistic we consider behaves, with the right constants, as

$$\mathcal{S}_h \sim \frac{\hat{IV}_h - \frac{\sqrt{2}nh^{3\beta/2}}{3\sqrt{\pi(\sigma^2 h + 2\rho_n^2)}}}{\frac{\sqrt{ch}}{\sqrt{\frac{\sqrt{2}nh^{5\beta/2}}{5\sqrt{\pi(\sigma^2 h + 2\rho_n^2)}}}}}$$

Now putting together Theorem 6.1 and (23) we have what follows.

a) if $\gamma > 1/2$ we write $\mathcal{S}_h$ as

$$\mathcal{S}_h = \frac{\hat{IV}_h - IV}{\sqrt{ch}} \frac{\sqrt{ch}}{\sqrt{c'nh^{5\beta/2-1/2}}} + \frac{IV - c''nh^{3\beta/2-1/2}}{\sqrt{c'nh^{5\beta/2-1/2}}}.$$

Since the first factor is asymptotically Gaussian, the second one tends to 0 and the third term to $-\infty$ the result is immediate.

b) If $\gamma = 1/2$ then, analogously as before,

$$\mathcal{S}_h \sim \frac{\hat{IV}_h - ch^{3(\beta-1)/2}}{c'h^{(5\beta-3)/4}} = \frac{\hat{IV}_h - (\sigma^2+2)T}{c''\sqrt{h}} \frac{c''\sqrt{h}}{c'h^{(5\beta-3)/4}} + \frac{(\sigma^2+2)T - ch^{3(\beta-1)/2}}{c'h^{(5\beta-3)/4}} \to -\infty.$$

c) If $\gamma \in (\beta/2, 1/2)$ then similarly

$$\mathcal{S}_h \sim \frac{\hat{IV}_h - 2nh^{2\gamma}}{\sqrt{8nh^{4\gamma}}} \frac{\sqrt{8nh^{4\gamma}}}{\sqrt{cnh^{5\beta/2-\gamma}}} + \frac{2nh^{2\gamma} - cnh^{3\beta/2-\gamma}}{\sqrt{c'nh^{5\beta/2-\gamma}}} \to -\infty.$$

d) If $\gamma = \beta/2$ then

$$\mathcal{S}_h \sim \frac{\hat{IV}_h - \frac{nr_h}{3\sqrt{\pi}}}{\sqrt{\frac{nr_h^2}{5\sqrt{\pi}}}} = \frac{\hat{IV}_h - 4(\phi_2 - 1/2)r_h n}{\sqrt{nr_h^2\psi}} \sqrt{5\psi\sqrt{\pi}} + \frac{\left(4(\phi_2 - 1/2) - \frac{1}{3\sqrt{\pi}}\right) r_h n}{\sqrt{\frac{nr_h^2}{5\sqrt{\pi}}}} \to -\infty,$$

as the first factor is asymptotically Gaussian, the second factor is a constant and the third term has $4(\phi_2 - 1/2) - \frac{1}{3\sqrt{\pi}} < 0$ and thus tends to $-\infty$.

e) If $\gamma \in (0, \beta/2)$ then, with the right constants

$$\mathcal{S}_h \sim \frac{\hat{IV}_h - nh^{3\beta/2-\gamma}/(3\sqrt{\pi})}{\sqrt{nh^{5\beta/2-\gamma}/(5\sqrt{\pi})}},$$

which is exactly the same as in (10) and thus is asymptotically Gaussian. □

References

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

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

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

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

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

[6] K. Christensen, R. Oomen, and M. Podolskij. Realised quantile-based estimation of the integrated variance. *Journal of Econometrics*, 159:74–98, 2010.

[7] R. Cont and C. Mancini. Non-parametric tests for pathwise properties of semimartingales. *Bernoulli*, 17(2):781–813, 2011.

[8] P.R. Hansen and A. Lunde. Realized variance and market microstructure noise. *J. Bus. Econom. Statist.*, 24 (2):127–218, 2006.

[9] X. Huang and G. Tauchen. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics*, 3 (4):456–499, 2005.

[10] J. Jacod. Asymptotic properties of realized power variations and associated functions of semimartingales. *Stochastic Processes and Their Applications*, 118:517–559, 2008.

[11] J. Jacod. *Statistics and high-frequency data. Lecture Notes, SEMSTAT in La Manga (Spain) 2007.* in press.

[12] J. Jacod, Y. Li, P. A. Mykland, and M. Podolskij. Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Processes and Their Applications*, 119:2249–2276, 2009.

[13] J. Jacod and A.N. Shiryaev. *Limit Theorems for Stochastic Processes, 2nd ed.* Springer-Verlag, Berlin, 2003.

[14] G.J. Jiang and R.C.A. Oomen. Testing for jumps when asset prices are observed with noise. a swap variance approach. *Journal of Econometrics*, 144(2):352–370, 2008.

[15] G. Lorenzini. Testing on empirical data for the relevance of microstructure noise. *in progress*, 2012.

[16] C. Mancini. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36:270–296, 2009.

[17] M. Podolskij and M. Vetter. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15 (3):634–658, 2009.

[18] R. Roll. A simple implicit measure of the effective bid-ask spread in an efficient market. *Journal of Finance*, XXXIX(4), 1984.

[19] M. Rosenbaum. A new microstructure noise index. *Quantitative Finance*, 6:883–899, 2011.

[20] L. Zhang, P. Mykland, and Y. Ait-Sahalia. A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association.*, 100:1394–1411, 2005.

[21] B. Zhou. High-frequency data and volatility in foreign exchange rates. *Journal of Business and Economic Statistics*, 14:45–52, 1996.