Functional estimation of realized instantaneous volatility in the presence of Lévy price jumps.

Wale Dare and Matthias Fengler

January 26, 2017

Abstract

We propose functional estimators of the instantaneous volatility of asset prices. We introduce a Gabor frame estimator of spot volatility, and show that the estimator is convergent in probability and in the integrated mean squares error sense, when prices are without discontinuities. We also propose a second estimator for when prices cannot be assumed to be continuous, and show that the proposed estimator converges in probability in the presence of both finite and infinite activity price jumps.

Volatility estimation using discretely observed asset prices has received a great deal of attention recently, however, much of that effort has been focused on estimating the *integrated* volatility and, to a lesser extent, the *spot* volatility at a given point in time. Notable contributions to this literature include the papers by Foster & Nelson (1996), Fan & Wang (2008), Florens-Zmirou (1993), and Barndorff-Nielsen & Shephard (2004). In these studies, the object of interest is local in nature: spot volatility at a given point in time or integrated volatility up to a terminal point in time. In contrast, estimators which aim to obtain *spot* volatility estimates for entire time windows have received much less coverage. These are the so-called global spot volatility estimators. Here, the objects of interest are global: random elements whose realizations are sample paths, i.e. functions defined on nontrivial time intervals.

There are some potential benefits to obtaining global estimators of volatility. First, given an estimate of spot volatility σ^2 over an interval $[0, T]$, the integrated volatility at any point in the observation window can be obtained easily by integrating the estimated spot volatility up to that point in time.

In fact, by the continuous mapping theorem, any continuous function of the global spot volatility estimator is immediately available. Hence, integrated powers of spot volatility, $\int_0^t \sigma_s^p \, \mathrm{d}s$, $p > 0$, the running maximum of spot volatility, $\sigma_t^* := \sup_{s \le t} |\sigma_s|$, and volatility in excess of a given threshold, $\sigma_t^a := \sigma_t I_{\{|\sigma_t| > a\}}$, $a > 0$, to name just a few, are easily obtained via the obvious transformation of the estimated global spot volatility. This flexibility is the primary reason for the growing interest in this class of estimators.

Two estimators of global volatility have stood out in the recent literature: the Fourier-based estimator proposed by Malliavin & Mancino (2002) and the wavelet-based estimator proposed by Genon-Catalot et al. (1992) and later developed by Hoffmann et al. (2012). The Fourier-based estimator is built up by first obtaining an estimate of the Fourier series expansion of the price process. The estimated Fourier coefficients of the price process are then used to obtain estimates for the Fourier coefficients of the volatility function. While there is no doubt that the Fourier-based estimator works (converges) both in theory and in practice (See Malliavin et al., 2007; Malliavin & Mancino, 2009), the theoretical investigation of the estimator seems somewhat incomplete. For instance, Malliavin & Mancino show that estimates of the individual coefficients in the Fourier expansion converge in a mean square sense but stopped short of providing an explicit rate of convergence for the entire volatility function. On the other hand, a lot more is known about the wavelet-based estimator; for instance, explicit rates of convergence for the entire volatility function are well known.

In both the Fourier and the wavelet approaches, there is a reliance on orthonormal bases: the Fourier and wavelet orthonormal bases, respectively. Now the use of orthonormal bases in *practical* work is optimal if the individual coefficients in the orthonormal basis expansion can be estimated with good precision. A coefficient with a large estimation error may be expected to cause a proportional distortion in the overall estimate of the volatility function. In practical work, where we must rely on a finite number of data points to obtain estimates for the bases coefficients, it is clear that coefficient error can easily become an issue. The global spot volatility estimator we propose is aimed squarely at this problem; it employs a Gabor frame methodology to mitigate the effects of bases coefficient error. Frames are very flexible and yield robust estimates in practical situations where coefficients lack precision or have been entirely *erased*.

Perhaps, the principal contribution of the current paper besides bringing the flexibility of Gabor frames to volatility estimation, is the proposal of a global estimator of spot volatility that remains consistent, in a probabilistic sense, even if prices are subject to price jumps. The two estimators described above do not have this property, so that in situations where it is hard to justify the assumption that prices have no discontinuities, the current estimator may be the only reasonable option.

The rest of this paper is organized as follows: Section 1 gives a description of the dynamics of observed prices; Section 2 briefly reviews Gabor frames theory; Section 3 gives a specification of the Gabor frame based estimator in the case of continuous prices. We obtain a highly flexible mean integrated square error rate of convergence as a function of different bounds on the paths of both instantaneous mean and volatility. We also obtain an explicit convergence in probability result. Section 4 contains the main result of this paper; we apply the results of the previous section to establish the consistency result for a jump-robust estimator of global spot volatility. Section 5 provides further support for the various estimators via a simulation exercise. These simulations suggest potential directions for future research.

1 Prices

We follow the literature by assuming that the asset price is a semimartingale[1]. This assumption is motivated in part by a corollary of the fundamental theorem of asset pricing, which requires asset prices to be semimartingales as a neccessary condition for an arbitrage-free market. The class of semimartingales in its entirety is somewhat too broad and unwieldy. In fact, whether or not the notion of spot volatility makes sense for such a broad class is not immediately clear. We will instead confine our analysis to the class of stochastic processes known as *Itô semimartingales* with Lévy jumps.

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a filtered probability space satifying the *usual conditions*. We consider prices[2] that evolve over time according to:

$$X_t = X_0 + \int_0^t b(s)\,\mathrm{d}s + \int_0^t \sigma(s)\,\mathrm{d}W_s + J_t, \qquad t \in [0,1] \tag{1.1}$$

where J is a pure jump Lévy process; W is a standard Brownian motion; X_0 is a constant; b and σ are both progressively measurable processes. We follow the ususal practise of referring to either σ or σ^2 as *(spot) volatility*, with *(spot) variance* reserved exclusively for σ^2 when it is important to make a distinction between the two. In addition, to stress the connection with the arithmetic Brownian motion with drift case, b and σ will occasionally be

[1] Adapted processes which almost surely have right-continuous, left-limited paths and may be expressed as a sum of a local martingale and a finite variation process.

[2] Our use of the term *prices* is not intended to restrict the discussion to untransformed prices. X could represent log-transformed prices or asset returns. The only requirement is that the data being analyzed satisfies (1.1) and σ^2 is of interest to the analyst.

referred to, respectively, as the *drift coefficient* and the *diffusion coefficient*. The class of continuous Itô processes is large; it contains for example solutions of all stochastic differential equations.

We assume prices are observed in the fixed time interval $[0, 1]$ at discrete, equidistant times $t_i = i\Delta_n, i = 0, 1, \cdots, n$, where

$$\Delta_n = 1/n = t_{i+1} - t_i, \qquad i = 0, \cdots, n-1. \tag{1.2}$$

Given the finite sequence $\{X_{t_i}, i = 0, 1, 2, \cdots, n\}$, our aim is to estimate the spot variance σ^2 in the time interval $[0, 1]$ by nonparametric methods. Note that our objective is not an approximation of a point but rather the approximation of an entire function. Thus an estimator of the spot variance may be viewed as a random element (function), as opposed to a random variable, that must converge in some sense to the spot variance, which itself is a random element. We approach this task by estimating the expansion of the spot variance using finite collections of Gabor frame elements.

2 Frames

Frames generalize the notion of orthonormal bases in Hilbert spaces. If $\{f_k\}_{k\in\mathbb{N}}$ is a frame for a separable Hilbert space $\mathcal{H}$ then every vector $f \in \mathcal{H}$ may be expressed as a linear combination of the frame elements, i.e.

$$f = \sum_{k\in\mathbb{N}} c_k f_k. \tag{2.3}$$

This is similar to how elements in a Hilbert space may be expressed in terms of orthonormal basis; but unlike orthonormal basis, the representation in (2.3) need not be unique, and the frame elements need not be orthogonal. Loosely speaking, frames contain redundant elements. The absence of uniqueness in the frame representation is by no means a shortcoming; on the contrary, we are afforded a great deal of flexibility and stability as a result. In fact, given a finite data sample, the estimated basis expansion coefficients are likely to be imprecise. This lack of precision can create significant distortions when using an orthonormal basis. These distortions are somewhat mitigated when using frames because of the built-in redundancy of frame elements.

Furthermore, if $\{f_k\}_{k\in\mathbb{N}}$ is a frame for $\mathcal{H}$, then surjective, bounded transformations of $\{f_k\}_{k\in\mathbb{N}}$ also constitute frames for $\mathcal{H}$, e.g. $\{f_k + f_{k+1}\}_{k\in\mathbb{N}}$ is a frame. So, once we have a frame, we can generate an arbitrary number of them very easily. We may then obtain estimates using each frame and compare results. If our results using the different frames fall within a tight band, then we are afforded some indication of the robustness of our computations.

Our discussion of frame theory will be rather brief; we only mention concepts needed for our specification of the volatility estimator. For a more detailed treatment see the book by Christensen (2008). In the sequel if z is a complex number then we shall denote respectively by $\bar{z}$ and $|z|$ the complex conjugate and magnitude of z. Let $L^2(\mathbb{R})$ denote the space of complex-valued functions defined on the real line with finite norm given by

$$\|f\| := \left(\int_{\mathbb{R}} f(t)\overline{f(t)}\, \mathrm{d}t \right)^{1/2} < \infty, \qquad f \in L^2(\mathbb{R}).$$

Define the inner product of two elements f and g in $L^2(\mathbb{R})$ as $\langle f, g \rangle := \int_{\mathbb{R}} f(t)\overline{g(t)}\, \mathrm{d}t$.

Denote by $\ell^2(\mathbb{N})$ the set of complex-valued sequences defined on the set of natural numbers $\mathbb{N}$ with finite norm given by

$$\|c\| := \left(\sum_{k \in \mathbb{N}} c_k \overline{c_k} \right)^{1/2} < \infty, \qquad c \in \ell^2(\mathbb{N}),$$

where c_k is the k-th component of c. The inner product of two sequences c and d in $\ell^2(\mathbb{N})$ is $\langle c, d \rangle := \sum_{k \in \mathbb{N}} c_k \overline{d_k}$. Now we may give a definition for frames:

2.1 Definition *A sequence $\{f_k\}_{k \in \mathbb{N}} \subset L^2(\mathbb{R})$ is a frame if there exists positive constants C_1 and C_2 such that*

$$C_1 \|f\|^2 \leq \sum_{k \in \mathbb{N}} |\langle f, f_k \rangle|^2 \leq C_2 \|f\|^2, \qquad f \in L^2(\mathbb{R}).$$

The constants C_1 and C_2 are called *frame bounds.* If $C_1 = C_2$ then $\{f_k\}_{k \in \mathbb{N}}$ is said to be *tight.* Because an orthonormal basis satisfies Parseval's equality[3], it follows that an orthonormal basis is a tight frame with frame bounds identically equal to 1, i.e. $C_1 = C_2 = 1$. Now if $\{f_k\}$ is a frame, we may associate with it a bounded operator $\mathcal{A}$ that maps every function f in $L^2(\mathbb{R})$ to a sequence c in $\ell^2(\mathbb{N})$ in the following way:

$$\mathcal{A}f = c \qquad \text{where} \qquad c_k = \langle f, f_k \rangle, \qquad k \in \mathbb{N}. \tag{2.4}$$

[3]Parseval's equality states that if $\{f_k\}_{k \in \mathbb{N}}$ is an orthonormal basis for $\mathcal{H}$, a separable Hilbert space, then

$$\|f\|^2 = \sum_{k \in \mathbb{N}} |\langle f, f_k \rangle|^2 = \|\hat{f}\|^2, \qquad f \in \mathcal{H},$$

where $\hat{f}$ is the Fourier transform of f.

Because $\mathcal{A}$ takes a function defined on a continuum ($\mathbb{R}$) to a sequence, which is a function defined on the discrete set $\mathbb{N}$, $\mathcal{A}$ is known as the *analysis* operator associated with the frame $\{f_k\}_{k\in\mathbb{N}}$. The boundedness of the analysis operator follows from the frame bounds in Definition (2.1). Now $\mathcal{A}^*$, the adjoint of $\mathcal{A}$, is well-defined and takes sequences in $\ell^2(\mathbb{N})$ to functions in $L^2(\mathbb{R})$. Using the fact that $\mathcal{A}^*$ must satisfy the equality $\langle \mathcal{A}f, c\rangle = \langle f, \mathcal{A}^*c\rangle$ for all $f \in L^2(\mathbb{R})$ and $c \in \ell^2(\mathbb{N})$, it may be deduced that

$$\mathcal{A}^*c = \sum_{k\in\mathbb{N}} c_k f_k, \qquad c \in \ell^2(\mathbb{N}),$$

where c_k is the k-th component of the sequence c. The adjoint, $\mathcal{A}^*$, may be thought of as reversing the operation or effect of the analysis operator; for this reason it is known as the *synthesis* operator.

Now an application of the operator $(\mathcal{A}^*\mathcal{A})^{-1}$ to every frame element f_k yields a sequence $\{\tilde{f}_k := (\mathcal{A}^*\mathcal{A})^{-1}f_k\}_{k\in\mathbb{N}}$, which is yet another frame for $L^2(\mathbb{R})$. The frame $\{\tilde{f}_k\}_{k\in\mathbb{N}}$ is known as the *canonical dual* of $\{f_k\}_{k\in\mathbb{N}}$. Denoting the analysis operator associated with the canonical dual by $\tilde{\mathcal{A}}$, it may be shown[4] that

$$\mathcal{A}^*\tilde{\mathcal{A}} = \tilde{\mathcal{A}}^*\mathcal{A} = \mathcal{I}, \tag{2.5}$$

where $\mathcal{I}$ is the identity operator and $\tilde{\mathcal{A}}^*$ is the adjoint of the analysis operator of the canonical dual. Furthermore, Proposition 3.2.3 of Daubechies (1992) shows that $\tilde{\mathcal{A}}$ satisfies

$$\tilde{\mathcal{A}} = \mathcal{A}(\mathcal{A}^*\mathcal{A})^{-1}, \tag{2.6}$$

so that the analysis operator of the canonical dual frame is fully characterized by $\mathcal{A}$ and its adjoint. It is easily seen that (2.5) yields a representation result since if $f \in L^2(\mathbb{R})$ then

$$f = \tilde{\mathcal{A}}^*\mathcal{A}f = \mathcal{A}^*\tilde{\mathcal{A}}f = \sum_{k\in\mathbb{N}}\langle f, \tilde{f}_k\rangle f_k. \tag{2.7}$$

Thus, in a manner reminiscent of orthonormal basis representations, every function in $L^2(\mathbb{R})$ is expressible as a linear combination of the frame elements, with the frame coefficients given by $\langle f, \tilde{f}_k\rangle$, the correlation between the function and the elements of the dual frame. It follows from the first equality in (2.5) and the commutativity of the duality relationship that functions in $L^2(\mathbb{R})$ may also be written as linear combinations of the elements in $\{\tilde{f}_k\}_{k\in\mathbb{N}}$, with coefficients given by $\langle f, f_k\rangle$, i.e. $f = \sum_{k\in\mathbb{N}}\langle f, f_k\rangle \tilde{f}_k$.

[4] See for example (Daubechies, 1992, Proposition 3.2.3)

Now let $\mathcal{P} := \tilde{\mathcal{A}}\mathcal{A}^* = \mathcal{A}\tilde{\mathcal{A}}^*$. A consequence of the noncommutativity[5] of the composition operator is that even though $\mathcal{A}^*\tilde{\mathcal{A}} = \mathcal{I}$ (2.5), the operator $\mathcal{P}$ need not be equal to $\mathcal{I}$. In fact, Proposition 3.2.3 in (Daubechies, 1992) shows that in the general case $\mathcal{P}$ is the orthogonal projection operator of $\ell^2(\mathbb{N})$ onto $R(\mathcal{A})$, the range space of $\mathcal{A}$, that is, $R(\mathcal{A}) := \{c \in \ell^2(\mathbb{N}) : c = \mathcal{A}f, \text{ for some } f \in L^2(\mathbb{R})\}$. That $R(\mathcal{A})$ in general may not coincide with $\ell^2(\mathbb{N})$ is a consequence of the fact that frames may be redundant, i.e. that they may contain "more" elements than is required, and thus may be linearly dependent. The level of redundancy is inversely related to the "size" of $R(\mathcal{A})$. So that the more redundancy the frame contains the smaller the range space of $\mathcal{A}$. To see this, let $\{f_k\}_{k\in\mathbb{N}}$ be a frame such that $f_1 = f_2 + f_3$; so, $\{f_k\}_{k\in\mathbb{N}}$ is redundant and linearly dependent. Now if c is in the range space of $\mathcal{A}$, the associated analysis operator, then the k-th coordinate of c must satisfy $c_k = \langle f, f_k \rangle$ for some $f \in L^2(\mathbb{R})$. By the linearity of the inner product, $c_1 = c_2 + c_3$. Thus every element of $R(\mathcal{A})$ must satisfy the restriction that its first component must equal the sum of the second and the third. Since not every element of $\ell^2(\mathbb{N})$ is subject to this restriction, $R(\mathcal{A})$ must be a proper subset of $\ell^2(\mathbb{N})$. So, dependence relationships among the frame elements serve to restrict the range of $\mathcal{A}$; the greater the number of dependence relationships there are, the smaller $R(\mathcal{A})$ becomes. As we shall see shortly, this fact has important consequences for how coefficient error affects the precision of the synthesis operator. Of course if the frame has no redundancy, then $R(\mathcal{A})$ will coincide with $\ell^2(\mathbb{N})$ and $\mathcal{P}$ will just be the identity operator.

2.1 Why use frames?

The main reason we might be interested in frame methods for estimating volatility is robustness to coefficient noise. By this we mean the imprecision that may result by virtue of the fact that in practice the frame coefficients may not be known with precision and must be estimated. Coefficient error has many sources: error resulting from using a finite data sample, rounding or quantization error, and error arising from the use of data contaminated with market microstructure noise.

The robustness of redundant frames to coefficient error is well-documented. For instance, Munch (1992) report noise reduction that is directly proportional to the degree of redundancy of the frame. Cvetković & Vetterli (1998) consider coefficient error due to quantization, and report an even high degree of robustness to this type of coefficient errors. That redundant frames ex-

[5] A binary operation $\star$ is noncommutative if $A \star B$ is generally not equal to $B \star A$. A classic example is matrix multiplication.

hibit this kind of robustness is not entirely unexpected. Redundant frames in essence include near-duplicates of frame elements; so that, any error arising from a given frame coefficient is easily made up for by the presence of other frame elements with similar informational content.

Daubechies (1992) provides the following heuristic explanation in terms of the size of the range space of the analysis operator. Let $\{f_k\}_{k\in\mathbb{N}}$ be a redundant frame in $L^2(\mathbb{R})$, and $\mathcal{A} : L^2(\mathbb{R}) \to \ell^2(\mathbb{N})$ be the associated analysis operator defined as in (2.4). Now, it follows from (2.5) and (2.6) that

$$\mathcal{I} = \tilde{\mathcal{A}}^*\mathcal{A} = \tilde{\mathcal{A}}^*\tilde{\mathcal{A}}\mathcal{A}^*\mathcal{A}.$$

From the discussion in the previous section, the orthogonal projector of $\ell^2(\mathbb{N})$ onto $R(\mathcal{A})$ is $\tilde{\mathcal{A}}\mathcal{A}^* = \mathcal{P}$. Combining this with the equation above, we have $\mathcal{I} = \tilde{\mathcal{A}}^*\mathcal{P}\mathcal{A}$. So the representation result in (2.7) can be expressed as

$$f = \tilde{\mathcal{A}}^*\mathcal{P}\mathcal{A}f.$$

Now assuming the coefficients of f under the operation of the analysis operator were contaminated by white noise sequence ε, we would have at our disposal $\mathcal{A}f+\varepsilon$ instead of simply $\mathcal{A}f$. Further assume that ε is decomposable as follows: $\varepsilon = \varepsilon_{\mathcal{A}^\perp} + \varepsilon_A$, where ε_A resides in the range space of $\mathcal{A}$, and $\varepsilon_{\mathcal{A}^\perp}$ resides in the orthogonal complement of $R(\mathcal{A})$. So, by definition, $\mathcal{P}\varepsilon_{\mathcal{A}^\perp} = 0$. The operation of reconstructing a function from the noisy coefficients may now be expressed as

$$f_\varepsilon = \tilde{\mathcal{A}}^*\mathcal{P}(\mathcal{A}f + \varepsilon) = \tilde{\mathcal{A}}^*\mathcal{P}(\mathcal{A}f + \varepsilon_A + \varepsilon_{\mathcal{A}^\perp}) = \tilde{\mathcal{A}}^*\mathcal{P}(\mathcal{A}f + \varepsilon_A).$$

It is thus clear that the deviation of the approximation, $\|f - f_\varepsilon\|$, equals $\|\tilde{\mathcal{A}}^*\mathcal{P}\varepsilon_A\|$, which should be lower than $\|\varepsilon\|$ to the extent that the range space of $\mathcal{A}$ is small; this is another way of saying that the approximation error is reduced to the extent that the frame is redundant. As noted on (Daubechies, 1992, pp. 98), this explanation is heuristic and probably accounts for only a small portion of the noise reduction observed in practical work. Nevertheless, it provides a starting point for thinking about the source of the robustness of frames.

2.2 Gabor frames

Next, we specialize the discussion to Gabor frames. The analysis of Gabor frames involves two operators: the *translation* operator $\mathcal{T}$ and the *modulation* operator $\mathcal{M}$ defined as follows:

$$\mathcal{T}_b f(t) := f(t-b), \qquad b \in \mathbb{R}, f \in L^2(\mathbb{R}), \tag{2.8}$$
$$\mathcal{M}_a f(t) := \mathrm{e}^{2\pi i a t} f(t), \qquad a \in \mathbb{R}, f \in L^2(\mathbb{R}), \tag{2.9}$$

where i is the imaginary number, i.e. $\mathrm{i} = \sqrt{-1}$. Both $\mathcal{T}$ and $\mathcal{M}$ are shift operators: $\mathcal{T}$ is a shift or translation operator on the time axis, whereas $\mathcal{M}$ performs shifts on the frequency axis. A Gabor system is constructed by fixing $a, b \in \mathbb{R}$, and performing shifts of a single nontrivial function $g \in L^2(\mathbb{R})$ in time-frequency space. For example, if a and b are real numbers then the sequence of functions

$$\{\mathcal{M}_{ha}\mathcal{T}_{kb}g\}_{h,k\in\mathbb{Z}},$$

constitutes a Gabor system.

2.2 Definition *Let $g \in L^2(\mathbb{R})$, and let $a > 0$, $b > 0$ be positive real numbers. Define for $t \in \mathbb{R}$*

$$g_{h,k}(t) := \mathrm{e}^{2\pi\mathrm{i}hat} g(t - kb), \qquad h, k \in \mathbb{Z}.$$

If the sequence $\{g_{h,k}\}_{h,k\in\mathbb{Z}}$ constitutes a frame for $L^2(\mathbb{R})$, then it is called a Gabor frame[6]*.*

The fixed function g is known as the *Gabor frame generator*[7]; a is known as the *modulation parameter*; and b is known as the *translation parameter*. In order to obtain sharp asymptotic rates, we require g and its dual $\tilde{g}$ (see (2.7)) to be continuous and compactly supported. The following result, stated in (Christensen, 2006, Lemma 1.2) and in (Zhang, 2008, Proposition 2.4), tells us exactly how to construct such dual pairs.

2.1 Lemma *Let $[r, s]$ be a finite interval, let $a > 0$, $b > 0$ be positive constants, and let g be a continuous function. If $g(t) \neq 0$ when $t \in (r, s)$; $g(t) = 0$ when $t \notin (r, s)$; and a, b satisfy: $a < 1/(s - r)$, $0 < b < s - r$; then $\{g, \tilde{g}\}$ is a pair of dual Gabor frame generators, with the dual Gabor generator given by*

$$\tilde{g}(t) := g(t)/G(t), \ \text{where} \tag{2.10}$$

$$G(t) := \sum_{k\in\mathbb{Z}} |g(t - kb)|^2/a. \tag{2.11}$$

Furthermore,

$$\tilde{g}_{h,k}(t) := \mathrm{e}^{2\pi\mathrm{i}hat}\tilde{g}(t - kb), \qquad h, k \in \mathbb{Z} \tag{2.12}$$

is compactly supported.

---

[6]It is also sometimes referred to as a *Weyl-Heisenberg* frame.

[7]It is referred to elsewhere as the *window function.*

Next, we state here and prove in the Appendix that the dual generator $\tilde{g}$ inherits the continuity properties of g.

2.2 Lemma *Let the dual Gabor frame generator $\tilde{g}$ be constructed as in (2.10). If $\bar{\omega}(g, \delta)$ denotes the modulus of continuity of g, i.e. $\bar{\omega}(g, \delta) := \sup\{|g(t) - g(t')| : t, t' \in \mathbb{R}$ and $|t - t'| < \delta\}$, then*

$$\bar{\omega}(\tilde{g}_{j,k}, \delta) \leq C\bar{\omega}(g, \delta) \qquad h, k \in \mathbb{Z},$$

where C is a positive constant.

Proof. See Appendix A. □

In the sequel, we assume the Gabor frame setup in Lemma 2.1.

3 Volatility estimation: continuous prices

In this section we specify a consistent estimator of the spot volatility within a framework of continuous prices. That is $J_t = 0$ and (1.1) reduces to:

$$X_t = X_0 + \int_0^t b_s \,\mathrm{d}s + \int_0^t \sigma_s \,\mathrm{d}W_s \qquad t \in [0, 1], \tag{3.13}$$

where X_0 is a constant, both b and σ are adapted, b is càdlàg, and σ is continuous. We restate these assumptions for easy reference.

3.1 Assumption

1. Spot volatility σ is a strictly positive and adapted stochastic process with continuous paths on $[0, 1]$.
2. The drift coefficient b is adapted and càdlàg on $[0, 1]$.

The asymptotic results we shall establish rely heavily on the modulus of continuity of realized spot variance, so, it is not immediately clear that the continuity assumption can be further weakened. The right-continuity assumption (càdlàg is a French abbreviation for right continuous, with finite left limits) on the drift ensures that the drift coefficient is pathwise bounded on $[0, 1]$. Likewise, the spot variance is pathwise bounded as a result of the continuity assumption. Let $\{g, \tilde{g}\}$ be a pair of dual Gabor frame generators constructed as in Lemma 2.1, then σ^2 admits a Gabor frame expansion given by:

$$\sigma^2(t) = \sum_{h,k \in \mathbb{Z}} c_{h,k} \, g_{h,k}(t), \text{ where} \tag{3.14}$$

$$c_{h,k} = \langle \sigma^2, \tilde{g}_{h,k} \rangle. \tag{3.15}$$

Note that both σ^2 and $\tilde{g}$ have compact support. Indeed σ^2 has support in $[0,1]$, whereas $\tilde{g}$ has support in $[s,r]$. So, $c_{h,k} \neq 0$ only if the supports of σ^2 and $\tilde{g}_{h,k}$ overlap. Furthermore, we note from (2.12) that $\tilde{g}_{h,k+1}$ is simply $\tilde{g}_{h,k}$ shifted by b units; so, $c_{h,k} = 0$ if $|k| \geq K_0$ with

$$K_0 := \lceil (1 + |s| + |r|)/b \rceil, \tag{3.16}$$

where $\lceil x \rceil$, $x \in \mathbb{R}$, is the least integer that is greater than or equal to x. Thus σ^2 admits a representation of the form:

$$\sigma^2(t) = \sum_{\substack{(h,k)\in\mathbb{Z}^2 \\ |k|\leq K_0}} c_{h,k}\; g_{h,k}(t),$$

and for sufficiently large positive integer H,

$$\sigma^2(t) \approx \sum_{\substack{|h|\leq H \\ |k|\leq K_0}} c_{h,k}\; g_{h,k}(t).$$

Now, suppose n observations of the price process are available, and let

$$\Theta_n := \{(h,k) \in \mathbb{Z}^2 : |h| \leq H_n \text{ and } |k| \leq K_0\}, \tag{3.17}$$

where H_n is an increasing sequence in n. We propose the following estimator of the volatility coefficient:

$$v_n(X,t) := \sum_{(h,k)\in\Theta_n} \hat{c}_{h,k}\; g_{h,k}(t), \qquad t \in [0,1], \text{ where} \tag{3.18}$$

$$\hat{c}_{h,k} := \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}(X_{t_{i+1}} - X_{t_i})^2. \tag{3.19}$$

So $|\Theta_n|$ is the number of frame elements included in the expansion. Specifically, $|\Theta_n| = (2K_0+1)(2H_n+1)$; and since K_0 is a finite quantity, it follows that $|\Theta_n| = O(H_n)$, i.e. the number of estimated coefficients is proportional to H_n, and therefore, will grow with the number of observations, n. In the next section we show that the estimator converges to σ^2 on $[0,1]$ in a mean integrated square error sense.

3.1 Asymptotic properties

In this section we obtain an estimate of the rate of convergence of the Gabor frame estimator on compact intervals of the real line. We will take $[0,1]$ as the prototype for such intervals. The usual way to think about σ^2 is as a stochastic process, but it is just as natural to think of it as a random element. A random element is an extension of the familiar concept of a random variable to include situations where the state space can be any metric space E. Because our interest is in studying the convergence of the estimator on $[0,1]$, E will be set equal to $L^2[0,1]$, the set of real-valued, square integrable function on $[0,1]$. Now, due to the path continuity assumption on σ^2, we will restrict our attention to $C^0[0,1]$, the set of real-valued, continuous functions on $[0,1]$ equipped with the $L^2[0,1]$ norm. So, given an outcome ω in the sample space Ω, the restriction of realized volatility $\sigma^2(\omega)$ to $[0,1]$ is a continuous function defined on $[0,1]$. Now, with regards smoothness, $C^0[0,1]$ is a very diverse class, comprising functions that are infinitely differentiable, those that are nowhere differentiable, and everything in-between. For a random volatility coefficient it may very well be the case that for some outcome ω, realized volatility $\sigma^2(\omega)$ is very smooth with finite derivatives of all orders, whereas for outcome ω', $\sigma^2(\omega')$ is very rough with kinks everywhere. For instance, with probability one, the one-dimesional Brownian motion maps outcomes ω to continuous functions in $C^0[0,1]$, but it is almost surely the case that none of these functions will be differentiable.

Now, a good estimator should yield successively better approximations with increasing observation frequency, regarless of the degree of smoothness of the realized volatility function. More realistically, the *rate* of convergence of the approximation would depend on the smoothness of the realized volatility, with faster rates achieved for smooth functions. That is, for outcomes ω, ω' in Ω, if $\sigma^2(\omega)$ is smoother as a function of time than $\sigma^2(\omega')$ then the number of observations required to achieve a given level of accuracy when $\sigma^2(\omega)$ is realized should not exceed the required number when $\sigma^2(\omega')$ is realized. So, while an estimator might eventually converge regardless of the regularity of volatility, the convergence rate may be outcome- or state-dependent.

To develop an asymptotic theory for the Gabor frame estimator which can account for state-dependency in convergence rates, we characterize the effective state space of the volatility coefficient, viewed as a random element in $C^0[0,1]$, according to a smoothness criterion. A simple way to achieve this characterization is via the Hölder continuity criterion. Let $0 < \alpha \leq 1$, a function f in $C^0[0,1]$ is said to be Hölder continuous with exponent α if there is a finite constant K such that whenever x and y are distinct numbers in $[0,1]$ then

$$|f|_\alpha := \frac{|f(x) - f(y)|}{|x - y|^\alpha} \leq K. \tag{3.20}$$

The set of Hölder continuous functions with exponent α is denoted by $C^{0,\alpha}[0,1]$.

The Hölder class with $\alpha = 1$ is the familiar class of Lipschitz continuous functions. Hölder classes admit a natural ordering relation whereby if α is larger than β then every function that is Hölder-continuous with exponent α is also Hölder-continuous with exponent β. Note that with regards regularity (smoothness), the ordering is reversed: the smoother the function the larger the Hölder exponent. Consequently, the Lipschitz class ($\alpha = 1$) is contained in every Hölder class, and it is also the class with the smoothest (most regular) functions. As mentioned earlier, Brownian paths are nowhere differentiable, but using Hölder classes the regularity of Brownian paths can be further qualified; this is a consequence of the well-known Lévy's Modulus of Continuity Theorem (Rogers & Williams, 2000, Theorem I.10.2), which states that, with probability one, no Brownian path is Hölder-continuous with exponent less than 1/2 on $[0, 1]$. Now for a fixed exponent α, the following norm may be defined on $C^{0,\alpha}[0, 1]$:

$$\|f\|_\alpha := \sup_{t\in[0,1]} |f(t)| + |f|_\alpha,$$

where $|f|_\alpha$ is defined in (3.20). The norm is obviously well-defined since f is a continuous function defined on a compact set. Now using $\|\cdot\|_\alpha$ the state space may be characterized in a such a way that functions with similar regularity propoerties can be grouped together. We accomplish this via Hölder balls: A Hölder ball of radius $c > 0$ is given by:

$$\mathcal{H}(\alpha, c) := \{f \in C^{0,\alpha}[0, 1] : \|f\|_\alpha \leq c\}.$$

With this device it is possible to obtain convergence rates that take into account the regularity of realized volatility. While this is already quiet satisfying, it is also of some interest to achieve flexibility with regards the drift coefficient. Where the drift is concerned, regularity or even continuity for that matter is irrelevant; what is key is pathwise boundedness. The natural way to achieve flexibility in this respect is to group the realized drift according to membership in balls of radius c in $L^\infty[0, 1]$, that is,

$$\mathcal{U}(c) := \{f \in L^\infty[0, 1] : \|f\|_\infty \leq c\}. \tag{3.21}$$

In this way we are able to characterize the sample space according to the regularity of realized volatility and the boundedness of realized drift. This leads to the consideration of the asymptotic behavior of the estimator when $b \in \mathcal{U}(c)$ and $\sigma^2 \in \mathcal{H}(\alpha, c)$ for $c, \alpha > 0$[8]. We denote such events by $\mathcal{E}(\alpha, c)$, i.e.,

$$\mathcal{E}(\alpha, c) := \{\omega : b(\omega) \in \mathcal{U}(c)\} \cap \{\omega : \sigma^2(\omega) \in \mathcal{H}(\alpha, c)\}.$$

[8]It is not essential to use different c's for b and σ^2.

So, an outcome ω is in $\mathcal{E}(\alpha, c)$ if the realized drift is caught between $-c$ and c on $[0, 1]$, and if realized volatility $\sigma^2(\omega)$ is Hölder continuous with exponent α, and $\|\sigma^2(\omega)\|_\alpha \leq c$. Note that the implication of the last statement is that there is some $c' \leq c$ such that realized volatility is cought between 0 and c'.

We are now only left with the task of making the obvious modification to the usual integrated mean square error criterion :

$$R_n(\alpha, c) := E[\|v_n - \sigma^2\|^2 I_{\mathcal{E}(\alpha,c)}],$$

where $I_{\mathcal{E}(\alpha,c)}$ is the indicator function of $\mathcal{E}(\alpha, c)$, $\|\cdot\|$ is the $L^2[0, 1]$ norm, and n is the number of observations. Note that if $\mathcal{E}(\alpha, c) = \Omega$, the expression above will just be the usual integrated mean square error criterion. By restricting the volatility and the drift according to events $\mathcal{E}(\alpha, c)$, the asymptotic properties of the estimator may be studied with full flexibility. That is we may obtain results of the form $\limsup_{n\to\infty} \tilde{n}_{n,\alpha,c} R_n(\alpha, c) < \infty$, where $\tilde{n}_{n,\alpha,c}$ may vary for different values of α and c.

Much like the usual integrated mean square error risk, $R_n(\alpha, c)$ admits a decomposition in terms of an integrated square bias component and an integrated variance component. To see this, note the following:

$$\begin{aligned}
R_n(\alpha, c) &= E \int_0^1 \{(v_n(X, t) - \sigma^2(t)) I_{\mathcal{E}(\alpha,c)}\}^2 \, \mathrm{d}t && (3.22)\\
&= \int_0^1 E[\{(v_n(X, t) - \sigma^2(t)) I_{\mathcal{E}(\alpha,c)}\}^2] \, \mathrm{d}t && (3.23)\\
&= \int_0^1 E[(v_n(X, t) - \sigma^2(t)) I_{\mathcal{E}(\alpha,c)}]^2 \, \mathrm{d}t \\
&\quad + \int_0^1 \mathrm{var}[v_n(X, t) I_{\mathcal{E}(\alpha,c)}] \, \mathrm{d}t. && (3.24)
\end{aligned}$$

The equality in (3.23) results from an interchange of the expectectation and integration operators justified by Fubini's theorem. The decomposition in line (3.24) results from the usual mean square error decomposition into a square bias and a variance component for each $t \in [0, 1]$. The two summands in the last line are the bias and variance components and will be denoted $B_n^2(\alpha, c)$ and $V_n(\alpha, c)$, respectively; we obtain estimates for their rates of convergence below. First, we recall Theorem 4.1 from Zhang (2008), which plays a crucial role in our demonstrations.

3.1 Lemma (Theorem 4.1, Zhang (2008)) *Let $\{g_{h,k}, \tilde{g}_{h,k}\}$ be a pair of compactly supported dual Gabor frames satisfying the conditions of Lemma (2.1). If $f \in L^2(\mathbb{R})$ is bounded and continuous on $\mathbb{R}$ then for any positive* *integer l, if $K > (l + |s| + |r|)/b$ and H is a positive integer then the partial sums*

$$S_{H,K}(t) := \sum_{|h|\le H} \sum_{|k|\le K} \langle f, g_{h,k}\rangle \tilde{g}_{h,k}(t)$$

satisfies

$$\begin{aligned} &S_{H,K}(t) - f(t) \\ &\qquad = O(\|g\|_\infty/ab)\{\|g\|_\infty \bar{\omega}(f, 1/aH) + \|f\|_\infty \bar{\omega}(g, 1/aH)\} \log H, \end{aligned}$$

for all t such that $|t| \le l$.

The following two propositions are the main results of this section.

3.1 Proposition *Let $\{g, \tilde{g}\}$ be pair of dual Gabor generators constructed as in Lemma* (2.1)*. Suppose the conditions in Assumption 3.1 hold. If g is Lipschitz continuous and $H_n \uparrow \infty$ satisfies*

$$H_n^2 \Delta_n = o(1)$$

then $R_n(\alpha, c)$ converges to 0, with

$$\begin{aligned} &B_n^2(\alpha, c) = O(H_n^2 \Delta_n + H_n^{-2\alpha} \log^2 H_n) \\ &V_n(\alpha, c) = O(H_n^2 \Delta_n), \end{aligned} \tag{3.25}$$

where

$$\Delta_n := 1/n = t_{i+1} - t_i, \qquad 0 \le i \le n-1, \tag{3.26}$$

is the step size, and H_n is the order of magnitude of the number of estimated frame coefficients.

Proof. See the appendix.

3.1 Remark The above bounds are remarkably similar to those achievable using an orthonormal basis such as wavelets (Genon-Catalot et al., 1992). The variance component is slower by a factor of H_n. This comes about because the vectors in a frame need not be orthogonal. The bias term is also slower by a logarithmic factor. Intuitively, the logarithmic term shows up because frames unlike orthonormal basis may contain redundant terms. In practical implementations, this differences are likely to be insignificant.

3.2 Proposition *Suppose the price process is specified as in* (3.13)*. Let* $\{g, \tilde{g}\}$ *be pair of dual Gabor generators satisfying the conditions of Lemma 2.1 such that* g *is Lipschitz continuous on the unit interval. If* $H_n \uparrow \infty$ *satisfies*

$$H_n(n^{-1}\log(n))^{1/2} = o(1),$$

then $v_n(X,t)$*, defined in* (3.18)*, converges in* $L^2[0,1]$ *to* σ^2 *in probability.*

Proof. We begin by noting that

$$\begin{aligned} v_n(X,t) - \sigma^2(t) = &\sum_{(h,k)\in\Theta_n} (\hat{c}_{h,k} - c_{h,k})\, g_{h,k}(t) \\ &- \sum_{(h,k)\notin\Theta_n} c_{h,k}\, g_{h,k}(t), \end{aligned} \tag{3.27}$$

where

$$\begin{aligned} \hat{c}_{h,k} &= \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}(X_{t_{i+1}} - X_{t_i})^2 \text{ and} \\ c_{h,k} &= \int_0^1 \overline{\tilde{g}_{h,k}(s)}\sigma^2(s)\,\mathrm{d}s. \end{aligned}$$

We tackle the summands in (3.27) in turn starting with the first one. But first let

$$M_i := \int_{t_i}^{t_{i+1}} b(s)\,\mathrm{d}s, \quad \text{and} \quad S_i := \int_{t_i}^{t_{i+1}} \sigma(s)\,\mathrm{d}W_s,$$

and note that since $X_{t_{i+1}} - X_{t_i} = M_i + S_i$, it follows that

$$(X_{t_{i+1}} - X_{t_i})^2 = M_i^2 + 2M_iS_i + S_i^2.$$

So, (3.27) may be written as

$$v_n(X,t) - \sigma^2(t) = B_{1,n}(t) + B_{2,n}(t) + B_{3,n}(t) + B_{4,n}(t),$$

where

$$B_{1,n}(t) := \sum_{(h,k)\in\Theta_n} g_{h,k}(t)\left(\sum_{i=0}^{n-1}\overline{\tilde{g}_{h,k}(t_i)}S_i^2 - c_{h,k}\right),$$

$$B_{2,n}(t) := 2\sum_{(h,k)\in\Theta_n} g_{h,k}(t)\left(\sum_{i=0}^{n-1}\overline{\tilde{g}_{h,k}(t_i)}S_iM_i\right),$$

$$B_{3,n}(t) := \sum_{(h,k)\in\Theta_n} g_{h,k}(t)\left(\sum_{i=0}^{n-1}\overline{\tilde{g}_{h,k}(t_i)}M_i^2\right),$$

$$B_{4,n}(t) := -\sum_{(h,k)\notin\Theta_n} g_{h,k}(t)c_{h,k}. \tag{3.28}$$

We will estimate the summands starting with $B_{4,n}(t)$. Note the following:

$$\begin{aligned}\sum_{(h,k)\notin\Theta_n} g_{h,k}(t)c_{h,k} &= \sum_{(h,k)\notin\Theta_n} g_{h,k}(t)\int_0^1 \sigma^2(s)\overline{\tilde{g}_{h,k}(s)}\,\mathrm{d}s\\ &\leq c\bar{\omega}(\tilde{g}, 1/H_n)\log H_n + c\bar{\omega}(\sigma^2, 1/H_n)\log H_n,\end{aligned}$$

where the last line follows from Lemma 3.1 and Lemma 2.2. It follows from the Hölder continuity of σ^2 that $\bar{\omega}(\sigma^2, 1/H_n) \leq cH_n^{-\alpha}$. Furthermore, by Lemma 2.2 and the Lipschitz continuity of g we have $\bar{\omega}(\tilde{g}_{h,k}, 1/H_n) \leq cH_n^{-1}$. So,

$$B_{4,n}(t) = O(H_n^{-\alpha}\log H_n). \tag{3.29}$$

Note the generic use of the constant c. In the sequel, we will use c to denote the amalgamation of various constants resulting from multiple steps.

We now obtain an estimate for $B_{3,n}(t)$. Note the following:

$$\begin{aligned}M_i^2 &= \left(\int_{t_i}^{t_{i+1}} b(s)\,\mathrm{d}s\right)^2\\ &\leq ((\sup_{s\in[0,1]} b_s)n^{-1})^2.\end{aligned} \tag{3.30}$$

Now since $g_{h,k}$ and $\tilde{g}_{h,k}$ are bounded independently of h and k, and $n\Delta_n = 1$, it is almost surely the case that

$$B_{3,n}(t) = O(H_n\Delta_n). \tag{3.31}$$

Now consider the following,

$$\begin{aligned}S_i &= \int_{t_i}^{t_{i+1}} \sigma(s)\,\mathrm{d}W_s \\ &= O(\sqrt{n^{-1}\log(n)}),\end{aligned} \tag{3.32}$$

for sufficiently large n by a simple corollary to Lévy's modulus of continuity theorem (See Lemma A.1). Hence

$$M_i S_i = O(\sqrt{n^{-3}\log(n)}). \tag{3.33}$$

Since $g_{h,k}$ and $\tilde{g}_{h,k}$ are bounded independently of h and k, and $n\Delta_n = 1$, we have

$$B_{2,n}(t) = O(H_n\sqrt{n^{-1}\log(n)}). \tag{3.34}$$

Now we tackle the final piece $B_{1,n}(t)$. Let

$$A := \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} S_i^2 - \int_0^1 \sigma^2(s)\overline{\tilde{g}_{h,k}(s)}\,\mathrm{d}s. \tag{3.35}$$

We will first obtain an upper bound for A; we proceed by adding and subtracting $\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} \overline{\tilde{g}_{h,k}(t_i)}\sigma^2(s)\,\mathrm{d}s$ from A to yield:

$$\begin{aligned}A &= \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}\left(S_i^2 - \int_{t_i}^{t_{i+1}} \sigma^2(s)\,\mathrm{d}s\right) \\ &\quad + \sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}} \sigma^2(s)\{\overline{\tilde{g}_{h,k}(t_i)} - \overline{\tilde{g}_{h,k}(s)}\}\,\mathrm{d}s\right) \\ &=: A_1 + A_2.\end{aligned}$$

We obtain estimates in turn for the summands. By linearity of expectation

$$\begin{aligned}A_2 &= \sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} \sigma^2(s)\{\overline{\tilde{g}_{h,k}(t_i)} - \overline{\tilde{g}_{h,k}(s)}\}\,\mathrm{d}s \\ &\le c\bar{\omega}(\tilde{g}_{h,k}, \Delta_n),\end{aligned}$$

where $\bar{\omega}(\tilde{g}_{h,k}, \Delta_n)$ is the modulus of continuity of $\tilde{g}_{h,k}$ on an interval of length Δ_n. By Lemma (2.2) and the Lipschitz continuity of g we have,

$$A_2 \le c\bar{\omega}(g, \Delta_n) \le c\Delta_n.$$

Now, we obtain an estimate for A_1. First, let

$$T_c = \inf\{t > 0 : \sigma^2(t) > c\} \wedge 1.$$

By the regularity assumption on σ, $P(T_c < 1)$ can be made arbitrarily small by taking c sufficiently large. Now, let $D_i : \Omega \times [0,1] \to \mathbb{R}$ for $i = 0, \cdots, n-1$ be defined as follows:

$$D_i(t) := \overline{\tilde{g}_{h,k}(t_i)} \left(\int_{t_i}^{t} \sigma(u) \, \mathrm{d}W_u \right) \mathbb{1}_{(t_i, t_{i+1} \wedge T_c]}(t). \tag{3.36}$$

$$D_0(0) := 0. \tag{3.37}$$

So, $D_i(t)$ is 0 on $[0,1]$ except when t is in $(t_i, t_{i+1} \wedge T_c]$.

Now, using the integration by parts formula for semimartingales, we may write

$$\begin{aligned}\left(\int_{t_i}^{t_{i+1} \wedge T_c} \sigma(s) dW_s \right)^2 &- \int_{t_i}^{t_{i+1} \wedge T_c} \sigma^2(s) \, \mathrm{d}_s \\ &= 2 \int_{t_i}^{t_{i+1} \wedge T_c} \left(\int_{t_i}^{s} \sigma(u) \, \mathrm{d}W_u \right) \sigma(s) \, \mathrm{d}W_s\end{aligned}$$

so that

$$\begin{aligned}E(|A_1^{T_c}|) &= 2\, E \left[\sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1} \wedge T_c} \overline{\tilde{g}_{h,k}(t_i)} \left(\int_{t_i}^{s} \sigma(u) \, \mathrm{d}W_u \right) \sigma(s) \, \mathrm{d}W_s \right] \\ &\le 2\, E \left[\left| \int_0^{1 \wedge T_c} \sum_{i=0}^{n-1} D_i(s) \sigma(s) \, \mathrm{d}W_s \right| \right].\end{aligned}$$

Now using the fact that $\int_0^{t \wedge T_c} \sum_{i=0}^{n-1} D_i(s) \sigma(s) \, \mathrm{d}W_s$ is a martingale, we may make an appeal to the Burkholder-Davis-Gundy (BDG) inequality (Theorem 10.36 He et al., 1992) to yield:

$$\begin{aligned}E(|A_1^{T_c}|) &\le cE \left[\left| \int_0^{1 \wedge T_c} \left(\sum_{i=0}^{n-1} D_i(s) \sigma(s) \right)^2 \mathrm{d}s \right|^{1/2} \right] \\ &\le cE \left[\left| \int_0^{1 \wedge T_c} \sum_{i=0}^{n-1} \{D_i(s) \sigma(s)\}^2 \, \mathrm{d}s \right|^{1/2} \right],\end{aligned}$$

where the last line follows because $D_i(s)D_j(s) = 0$ whenever $i \neq j$. Now if we define $D_i^* := \sup_{t_i < s \le t_{i+1} \wedge T_c} D_i(s)$, and use the fact that σ^2 is less than c before T_c then

$$
\begin{aligned}
E(|A_1^{T_c}|) &\le cE\left[\left|\sum_{i=0}^{n-1} \Delta_n (D_i^*)^2\right|^{1/2}\right] \\
&\le cE\left[\left|\sum_{i=0}^{n-1} \Delta_n (D_i^*)^2\right|\right] \\
&\le c\Delta_n \sum_{i=0}^{n-1} E[(D_i^*)^2] \qquad (3.38)
\end{aligned}
$$

where c is a generic constant representing the bound on σ^2 and the BDG constant. Note from the definition of D_i that it is itself a martingale, so we may bound D_i^* with yet another application of the BDG inequality. That is

$$
\begin{aligned}
E((D_i^*)^2) &\le E\left(\int_{t_i}^{t_{i+1}\wedge T_c} \sigma^2(s)\,\mathrm{d}s\right) \\
&\le c\Delta_n. \qquad (3.39)
\end{aligned}
$$

Hence, given any $\varepsilon > 0$, $P(\sup_{t\in[0,1]} |B_{1,n}(t)| > \varepsilon) = O(H_n n^{-1}) + P(T_c < 1)$.

Collecting the estimates for $B_{j,n}(t)$ for $j = 1, \cdots, 4$, it is easily seen that $|v_n(X,t) - \sigma^2(t)|$ tends to zero in probability uniformly for all $t \in [0,1]$. $\square$

4 Volatility estimation: discontinuous prices

In this section we specify a global spot volatility estimator within a framework of prices X evolving in time as Itô semimatingales with continuous diffusion coefficients. Let $\tau : \mathbb{R} \to \mathbb{R}$ be bounded and satisfy $\tau(x) = x$ in a neighborhood of 0. Let ι be the identity function on the real line, i.e. $\iota(x) = x$ for $x \in \mathbb{R}$. The price process X admits the following representation:

$$
X_t = X_0 + \int_0^t b_s\,\mathrm{d}s + \int_0^t \sigma_s\,\mathrm{d}W_s + \tau(x) * (\mu - \nu)_t + (\iota - \tau)(x) * \mu_t, \qquad (4.40)
$$

for $t \geq 0$ where W is a standard Brownian motion; X_0 is either known or observable at time 0; both b and σ are adapted; b is càdlàg, and σ is continuous; μ is a Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ with intensity ν, where ν is a σ-finite Lévy measure on $\mathbb{R}_+ \times \mathbb{R}$. Note that because μ is a Poisson measure, if A and B are disjoint Borel sets on $\mathbb{R}_+ \times \mathbb{R}$, then the random measures $\mu(A)$ and $\mu(B)$ are Poisson distributed, independent, and and have intensity, $\nu(A)$ and $\nu(B)$, respectively. Moreover, because of the Lévy assumption on ν, it is the case that ν does not charge 0 and

$$(x^2 \wedge 1) * \nu_t < \infty, \qquad t \in [0, 1],$$

where $a \wedge b$, with $a, b \in \mathbb{R}$, denotes the minimum of a and b. The notation "$*$" denotes integration with respect to a random measure. So that

$$\begin{aligned} J_t^l &:= \tau * (\mu - \nu)_t = \int_0^t \int_{\mathbb{R}} \tau(x)[\mu(\mathrm{d}s, \mathrm{d}x) - \nu(\mathrm{d}s, \mathrm{d}x)], \\ J_t^s &:= (\iota - \tau) * \mu_t = \int_0^t \int_{\mathbb{R}} [\iota(x) - \tau(x)]\mu(\mathrm{d}s, \mathrm{d}x), \end{aligned}$$

for $t \geq 0$. Both J^l and J^s are purely discontinuous in the sense that they are orthogonal to all continuous semimartingales. J^s accounts for small jumps; it is a square-integrable martingale with possibly infinite activity. J^l accounts for large jumps, i.e. jumps with magnitude exceeding the bound on τ; it neccessarily has finite activity so it is a process with finite variation. In the sequel, we will specify τ as follows:

$$\tau(x) = xI_{\{|x| \leq 1\}}, \qquad x \in \mathbb{R}.$$

As in the preceeding section, we observe a realization of the price process at $n + 1$ equidistant points t_i, $i = 0, 1, \cdots, n$. The observation interval is normalized to $[0, 1]$ with practically no loss of generality. The estimator proposed in the previous section, where there is no jump activity, will not do here. It is inconsistent on account of the presence of jumps; its quality deteriorates as a function of how active the jumps of X are. We will counter this phenomenon with a modified spot variance estimator, but first we introduce the following notation. Let $\Delta_i X$ denote $X_{t_{i+1}} - X_{t_i}$ for $i = 0, 1, \cdots, n - 1$, and let u_n be a positive decreasing sequence such that

$$u_n(n^{-1} \log(n))^{-1} \tag{4.41}$$

diverges to infinity with n. We specify the jump-robust global estimator of spot volatility as follows:

$$V_n(X,t)(t) := \sum_{(h,k)\in\Theta_n} \hat{a}_{h,k}\, g_{h,k}(t), \qquad \forall t \in [0,1], \text{ where} \tag{4.42}$$

$$\hat{a}_{h,k} := \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}(\Delta_i X)^2 I_{\{(\Delta_i X)^2 \leq u_n\}}, \tag{4.43}$$

where $\{g_{h,k}, \tilde{g}_{h,k}\}$ is a pair of dual Gabor frames constructed as in Lemma (2.1); Θ_n retains its meaning from (3.17); and $I_{\{(\Delta_i X)^2 \leq u_n\}}$ is one if $(\Delta_i X)^2$ is less than or equal to u_n and zero otherwise.

There are obvious similarities between $v_n(X,t)$, defined at (3.18), and $V_n(X,t)$ with the key difference being that $V_n(X,t)$ discards realized squared increments over intervals that likely contain jumps; u_n determines the threshold for what is included in the computation and what is not. This determination becomes more accurate as the observation interval becomes infinitessimally small. Clearly it makes sense to use $v_n(X,t)$ if we have reason to believe that the price process is not subject to jumps; $v_n(X,t)$ will always employ all available data and therefore may be assumed to produce more accurate results.

4.1 Finite activity Lévy jumps

In order to demonstrate that the global estimator of spot volatility is consistent, we will proceed in stages. First suppose the price process specified in all generality in (4.40) experiences at most a finite number of Lévy jumps in any finite time interval. That is we assume that X has finite activity Lévy jumps, which is equivalent to ν being finite on the complement of $\{0\}$. The finite activity assumption also implies that the price process may be expressed as

$$X_t := \int_0^t b(s)\,\mathrm{d}s + \int_0^t \sigma(s)\,\mathrm{d}W_s + \sum_{i=1}^{N_t} Y_i, \qquad t \in [0,1], \tag{4.44}$$

where the Y_i's are i.i.d. jump sizes; N is a Poisson process with intensity λ, independent of each Y_i. Under this conditions, we have the following:

4.1 Proposition *Suppose the price process is specified as in* (4.44) *with σ and b satisfying Assumption 3.1. Let $\{g, \tilde{g}\}$ be pair of dual Gabor generators satisfying the conditions of Lemma* (2.1) *such that g is Lipschitz continuous on the unit interval. If $u_n \downarrow 0$ is a sequence in n such that*

$$n^{-1}\log(n)u_n^{-1} = o(1),$$

and the sequence $H_n \uparrow \infty$ satisfies

$$H_n(n^{-1}\log(n))^{1/2} = o(1),$$

then $V_n(X,t)$ as defined in (4.42) *converges in $L^2[0,1]$ in probability to σ^2.*

Proof. Let X^c denote the continuous part of X so that $X = X^c + J$, and

$$X_t^c = \int_0^t b(s)\,\mathrm{d}s + \int_0^t \sigma(s)\,\mathrm{d}W_s, \tag{4.45}$$

for t in $[0,1]$. Denote

$$v_n(X^c,t) := \sum_{(h,k)\in\Theta_n} \hat{d}_{h,k}\, g_{h,k}(t), \qquad t\in[0,1], \text{ where} \tag{4.46}$$

$$\hat{d}_{h,k} := \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}(\Delta_i X^c)^2.$$

We have

$$\begin{aligned}\int_0^1 &(V_n(X,t)-\sigma^2(t))^2\,\mathrm{d}t \\ &\le 2\int_0^1 (V_n(X,t)-v_n(X^c,t))^2\,\mathrm{d}t + 2\int_0^1 (v_n(X^c,t)-\sigma^2(t))^2\,\mathrm{d}t.\end{aligned} \tag{4.47}$$

That the second summand on the right converges to 0 in probability is a result of Proposition (1.1). Now note that

$$V_n(X,t)X - v_n(X^c,t)(t) = \sum_{(h,k)\in\Theta_n} (\hat{a}_{h,k} - \hat{d}_{h,k})\, g_{h,k}(t),$$

and

$$\hat{a}_{h,k} - \hat{d}_{h,k} = \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}\{(\Delta_i X)^2 I_{\{(\Delta_i X)^2\le u_n\}} - (\Delta_i X^c)^2\}.$$

By Corrollary 1 of Mancini (2009), for almost all outcomes, there is n' such that for all $n \ge n'$

$$I_{\{(\Delta_i X)^2\le u_n\}} = I_{\{\Delta_i N=0\}}.$$

Hence for almost all outcomes and sufficiently large n

$$\begin{aligned}\hat{a}_{h,k} - \hat{d}_{h,k} &\le \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} (\Delta_i X^c)^2 I_{\{\Delta_i N \neq 0\}} \\ &= O_{a.s}(\Delta_n \log(1/\Delta_n) \sum_{i=0}^{n-1} I_{\{\Delta_i N \neq 0\}}),\end{aligned}$$

where the last line follows from (4.45), the pathwise boundedness of b, and Lemma A.1. Hence for almost all outcomes,

$$\begin{aligned}\int_0^1 \{V_n(X,t) - v_n(X^c,t)\}^2 \,\mathrm{d}t &= O_{a.s}(H_n^2 \Delta_n^2 \log^2(1/\Delta_n)(N_1)^2) \\ &= O_{a.s}(H_n^2 \Delta_n^2 \log^2(1/\Delta_n)) \\ &\to 0, \end{aligned} \tag{4.48}$$

almost surely. □

4.2 Infinite activity Lévy jumps

We now turn to the case of a price process specified in full generality by (4.40), that is the price process is a sum of a continuous and a discontinuous process with possibly infinite activity. The infinite activity assumption is equivalent to the statement that ν assigns infinite measure to the complement of the singleton containing zero. The following is the consistency Proposition in this more general framework:

4.2 Proposition *Let the price process X be specified as in* (4.40) *with σ and b satisfying Assumption 3.1. Let $\{g, \tilde{g}\}$ be pair of dual Gabor generators satisfying the conditions of Lemma* (2.1) *such that g is Lipschitz continuous on the unit interval. Suppose the sequences $u_n \downarrow 0$ and $H_n \uparrow \infty$ satisfy*

$$H_n(n^{-1}\log(n))u_n^{-1} = o(1)$$

and that ν satisfies

$$(x^2 \wedge u_n^{1/2}) * \nu_1 = o(H_n^{-2}). \tag{4.49}$$

Then $V_n(X,t)$, defined in (4.42), *converges in $L^2[0,1]$ in probability to σ^2.*

Proof. We wish to show that the random variable $\int_0^1 (V_n(X,t) - \sigma^2(t))^2 \, \mathrm{d}t$ tends to zero in probability. The regularity conditions on X and σ^2 imply that $\sup_{t\in[0,1]}(V_n(X,t) - \sigma^2(t))^2$ is a random variable and that the previous claim would follow as soon as $\sup_{t\in[0,1]}(V_n(X,t) - \sigma^2(t))^2$ is shown to converge to zero in probability. To that end consider the following decomposition of the process X:

$$X = X^f + J^s, \tag{4.50}$$

$$X^f = X^c + J^l, \tag{4.51}$$

where $X^c = \int_0^t b_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}W_s$, $J^l = xI_{|x|>1} * \mu$, and $J^s = xI_{|x|\le 1} * (\mu - \nu)$. Let t be a point in the unit interval, then

$$\begin{aligned} V_n(X,t) - \sigma^2(t) = &\sum_{(h,k)\in\Theta_n} (\hat{a}_{h,k} - c_{h,k}) g_{h,k}(t) \\ &- \sum_{(h,k)\notin\Theta_n} c_{h,k} g_{h,k}(t), \end{aligned} \tag{4.52}$$

with $\hat{a}_{h,k}$ and $c_{h,k}$ defined by (4.42) and (3.15), respectively. By Theorem 4.1 of Zhang (2008) (See Lemma 3.1), the last term on the right converges uniformly on the unit interval to zero, almost surely, as $n \to \infty$.

To obtain a bound on the first item on the right of (4.52), we may use (4.50) to write

$$\begin{aligned} &\sum_{(h,k)\in\Theta_n} (\hat{a}_{h,k} - c_{h,k}) g_{h,k}(t) \\ &\qquad = \sum_{(h,k)\in\Theta_n} (w_{h,k} + x_{h,k} + y_{h,k} + z_{h,k}) g_{h,k}(t), \end{aligned} \tag{4.53}$$

where

$$\begin{aligned} w_{h,k} &:= \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} (\Delta_i X^f)^2 I_{\{(\Delta_i X^f)^2 \le 4u_n\}} - \int_0^1 \sigma^2(s) \overline{\tilde{g}_{h,k}(s)} \, \mathrm{d}s \\ x_{h,k} &:= \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} (\Delta_i X^f)^2 (I_{\{(\Delta_i X)^2 \le u_n\}} - I_{\{(\Delta_i X^f)^2 \le 4u_n\}}) \\ y_{h,k} &:= 2 \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} \Delta_i X^f \Delta_i J^s I_{\{(\Delta_i X)^2 \le u_n\}} \\ z_{h,k} &:= \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} (\Delta_i J^s)^2 I_{\{(\Delta_i X)^2 \le u_n\}}. \end{aligned} \tag{4.54}$$

By Proposition (4.1), if $\delta > 0$ then

$$P(\sup_{t\in[0,1]} |\sum_{(h,k)\in\Theta_n} w_{h,k} g_{h,k}(t)| > \delta) \le c\delta^{-1} H_n n^{-1}\log(n), \tag{4.55}$$

for some $c \in \mathbb{R}$. It remains to show that the last three terms on the right of (4.53) converge to zero in probability. Starting with the second summand, denote $A_i := \{(\Delta_i X)^2 \le u_n\}$, $B_i := \{(\Delta_i X^f)^2 \le 4u_n\}$ and note that $I_{A_i} - I_{B_i} = I_{A_i\cap B_i^c} - I_{A_i^c\cap B_i}$. Now for each outcome in $A_i \cap B_i^c$, it is the case that $2u_n^{1/2} - |\Delta_i J^s| < |\Delta_i X^f| - |\Delta_i J^s| \le |\Delta_i X^f + \Delta_i J^s| \le u_n^{1/2}$, so that $|\Delta_i J^s| > u_n^{1/2}$ and

$$\begin{aligned}
&\sum_{(h,k)\in\Theta_n} \left(\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} (\Delta_i X^f)^2 I_{A_i\cap B_i^c} \right) g_{h,k}(t) \\
&\quad \le \sum_{(h,k)\in\Theta_n} \left(\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} (\Delta_i X^f)^2 I_{\{(\Delta_i J^s)^2 > u_n\}} \right) g_{h,k}(t) \\
&\quad \le \sum_{(h,k)\in\Theta_n} \left(\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} (\Delta_i X^c)^2 I_{\{(\Delta_i J^s)^2 > u_n\}} \right) g_{h,k}(t) \\
&\qquad + \sum_{(h,k)\in\Theta_n} \left(\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} (\Delta_i J^l)^2 I_{\{(\Delta_i J^s)^2 > u_n\}} \right) g_{h,k}(t) \\
&\quad \le v_n + w_n,
\end{aligned} \tag{4.56}$$

where

$$v_n := 2cH_n\Lambda n^{-1}\log(n) \sum_{i=0}^{n-1} I_{\{(\Delta_i J^s)^2 > u_n\}}$$

$$w_n := cH_n \sum_{i=0}^{n-1} (\Delta_i J^l)^2 I_{\{(\Delta_i J^s)^2 > u_n\}}$$

where c is a sufficienty large constant, and Λ is a finite-valued random variable satisfying $\Lambda \ge \sup_{t\in[0,1]} |b(t)| + C$, where $C^{1/2}$ is the finite-valued random variable from Lemma A.1. Let $\delta > 0$ be given, put $x_n(t) := \sum_{(h,k)\in\Theta_n} \left(\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} (\Delta_i X^f)^2 I_{A_i\cap B_i^c}) \right) g_{h,k}(t)$ and note that

$$P\left(\sup_{t\in[0,1]} |x_n(t)| > \delta \right) \le P(v_n > \delta/2) + P(w_n > \delta/2).$$

Now let $\varepsilon > 0$ be given and note that because Λ is almost surely finite, there is a sufficiently large $K > 0$ such that $P(\Lambda > K) \leq \varepsilon/2$. Hence,

$$\begin{aligned} P(v_n > \delta/2) &\leq 2cKH_n\delta^{-1}E(n^{-1}\log(n)\sum_{i=0}^{n-1} I_{\{(\Delta_i J^s)^2 > u_n\}}) + P(\Lambda > K) \\ &= 2cKH_n\delta^{-1}\log(n)P((\Delta_i J^s)^2 > u_n) + \varepsilon/2 \\ &\leq 2cKH_n\delta^{-1}\log(n)E((\Delta_1 J^s)^2))u_n^{-1} + \varepsilon/2 \\ &\leq 2cKH_n\delta^{-1}\log(n)n^{-1}\kappa u_n^{-1} + \varepsilon/2 \end{aligned} \tag{4.57}$$

where $\kappa := E((\Delta_1 J^s)^2)) < \infty$. Obviously there is a large enough n such that the first expression above is less than or equal to $\varepsilon/2$. Moreover, because $\delta > 0$,

$$\begin{aligned} P(w_n > \delta/2) &\leq P\left(\cup_i\{I_{\{|x|>1\}} * \mu((t_i, t_{i+1}] \times \mathbb{R}) > 0, (\Delta_i J^s)^2 > u_n\}\right) \\ &\leq nP(\mu([0, 1/n] \times \{|x| > 1\}) > 0)E((\Delta_1 J^s)^2)u_n^{-1} \\ &\leq cn^{-1}\kappa u_n^{-1}, \end{aligned}$$

which clearly tends to zero in n.

Now, by Theorem 3.1 of Mancini (2009), if an outcome is in $A_i^c \cap B_i$ then there is a sufficiently small u_n such that $I_{\{|x|>1\}} * \mu((t_i, t_{i+1}] \times \mathbb{R}) = 0$. Hence, for such an outcome, it is the case that $(\Delta_i X)^2 > u_n$ if and only if $(\Delta_i X^c + \Delta_i J^s)^2 > u_n$, which in turn would hold if either $(\Delta_i X^c)^2 > u_n/4$ or $(\Delta_i J^s)^2 > u_n/4$. However, by theorem 3.1 of Mancini (2009), by taking n large enough $\{(\Delta_i X^c)^2 > u_n/4\} = \emptyset$. Let $\delta > 0$ be a given positive number; put $y_n(t) := \sum_{(h,k)\in\Theta_n}\left(\sum_{i=0}^{n-1}\overline{\tilde{g}_{h,k}(t_i)}(\Delta_i X^f)^2 I_{A_i^c\cap B_i})\right) g_{h,k}(t)$ and note that

$$P(\sup_{t\in[0,1]} |y_n(t)| > \delta) \leq P\left(cH_n\Lambda^2 n\left(I_{\{(\Delta_i J^s)^2 > u_n/4\}}\right)) > \delta\right),$$

which tends to zero exactly as in (4.57). Hence,

$$P(\sup_{t\in[0,1]} |\sum_{(h,k)\in\Theta_n} x_{h,k}g_{h,k}(t)| > \delta) \to 0. \tag{4.58}$$

Now we obtain a bound for the third summand in (4.53). First, denote $C_i := \{(\Delta_i J^s)^2 \leq 4u_n\}$, $p_{h,k} := 2\sum_{i=0}^{n-1}\overline{\tilde{g}_{h,k}(t_i)}\Delta_i X^f\Delta_i J^s I_{A_i\cap C_i}$, and $q_{h,k} := 2\sum_{i=0}^{n-1}\overline{\tilde{g}_{h,k}(t_i)}\Delta_i X^f\Delta_i J^s I_{A_i\cap C_i^c}$. Clearly, $y_{h,k} = p_{h,k} + q_{h,k}$. Now note that on $A_i \cap C_i^c$, it is the case that $2u_n^{1/2} - |\Delta_i X^f| < |\Delta_i J^s| - |\Delta_i X^f| \leq |\Delta_i X| \leq u_n^{1/2}$, so that $u_n^{1/2} < |\Delta_i X^f| < |\Delta_i J^l| + |\Delta_i X^c|$. In turn, the last inequlity implies that either $|\Delta_i J^l| > u_n^{1/2}/2$ or $|\Delta_i X^c| > u_n^{1/2}/2$. Now, for sufficiently large n, it is almost surely never the case that $|\Delta_i X^c| > u_n^{1/2}/2$ for some i, $0 \le i \le n-1$. Hence, for positive δ,

$$\begin{aligned}
&P(|\sum_{(h,k)\in\Theta_n} q_{h,k} g_{h,k}(t)| > \delta/2) \\
&\quad \le P(\cup_i\{\mu((t_i, t_{i+1}] \times \{|x| > 1\}) > 0, (\Delta_i J^s)^2 > u_n\}) \\
&\quad \le cn^{-1}\kappa u_n^{-1}. \qquad (4.59)
\end{aligned}$$

Meanwhile, on $A_i \cap C_i$, it is easily seen that $|\Delta_i J^l| - |\Delta_i X^c + \Delta_i J^s| < |\Delta_i X| \le u_n^{1/2}$, so that $|\Delta_i J^l| \le u_n^{1/2} + |\Delta_i X^c| + |\Delta_i J^s|$. On the other hand, $|\Delta_i J^l| < u_n^{1/2} + \Lambda n^{-1/2}\log^{1/2}(n) + 2u_n^{1/2} = O(u_n^{1/2})$. Let $r_{h,k} := 2\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}\Delta_i X^c \Delta_i J^s I_{A_i\cap C_i}$ and $s_{h,k} := 2cu_n^{1/2}\sum_{i=0}^{n-1}\overline{\tilde{g}_{h,k}(t_i)}\Delta_i J^s I_{A_i\cap C_i}$. Then

$$\begin{aligned}
&P(|\sum_{(h,k)\in\Theta_n} q_{h,k} g_{h,k}(t)| > \delta/2) \\
&\quad \le P(|\sum_{(h,k)\in\Theta_n} r_{h,k} g_{h,k}(t)| > \delta/4) + P(|\sum_{(h,k)\in\Theta_n} s_{h,k} g_{h,k}(t)| > \delta/4).
\end{aligned}$$

Now consider that $\sum_{(h,k)\in\Theta_n} r_{h,k} g_{h,k}(t) \le cH_n \sum_{i=0}^{n-1} \Delta_i X^c \Delta_i J^s I_{A_i\cap C_i}$, which implies that

$$\begin{aligned}
&P(|\sum_{(h,k)\in\Theta_n} r_{h,k} g_{h,k}(t)| > \delta/4) \le P(cH_n|\sum_{i=0}^{n-1} \Delta_i X^c \Delta_i J^s I_{A_i\cap C_i}| > \delta/4) \\
&\quad \le P\left(\left(\sum_{i=0}^{n-1} (\Delta_i X^c)^2\right)^{1/2} \left(\sum_{i=0}^{n-1} (\Delta_i J^s I_{A_i\cap C_i})^2\right)^{1/2} > \delta(4H_n c)^{-1}\right).
\end{aligned}$$

It is a well known fact that $\sum_{i=0}^{n-1}(\Delta_i X^c)^2(t)$ converges to $\int_0^t \sigma^2(s)\,\mathrm{d}s$ in probability uniformy on the unit interval. Hence, there is a sufficiently large N such that if $n > N$ then $P(|(\sum_{i=0}^{n-1}(\Delta_i X^c)^2)^{1/2} - (\int_0^1 \sigma^2(s)\,\mathrm{d}s)^{1/2}| > \delta) \le \varepsilon/4$, and because integrated volatility is almost surely finite, there is a sufficiently large K satisfying $K/2 > \delta$ such that $P(\int_0^1 \sigma^2(s)\,\mathrm{d}s > K/2) \le \varepsilon/4$. Hence, we may write

$$
\begin{aligned}
&P(|\sum_{(h,k)\in\Theta_n} r_{h,k}g_{h,k}(t)| > \delta/4)\\
&\le P\left(\sum_{i=0}^{n-1}(\Delta_i J^s I_{A_i\cap C_i})^2 > \delta^2(4KH_nc)^{-2}\right) + \varepsilon/2\\
&\le P\left((x^2 I_{\{|x|\le 1\wedge 2u_n^{1/2}\}} * \mu)_1 > \delta^2(4KH_nc)^{-2}\right) + \varepsilon/2\\
&\le \delta^{-2}(4KH_nc)^2 E\left((x^2 I_{\{|x|\le 1\wedge 2u_n^{1/2}\}} * \mu)_1\right) + \varepsilon/2\\
&\le \delta^{-2}(4KH_nc)^2 (x^2 I_{\{|x|\le 1\wedge 2u_n^{1/2}\}} * \nu)_1 + \varepsilon/2
\end{aligned}
$$

which for sufficiently large n is less than ε by (4.49).

Now it is easily seen that for sufficiently large c

$$
\begin{aligned}
&P(|\sum_{(h,k)\in\Theta_n} s_{h,k}g_{h,k}(t)| > \delta/4) \le P(\sum_{i=0}^{n-1}\Delta_i J^s I_{A_i\cap C_i} > (8cu_n^{1/2})^{-1}\delta)\\
&\le (64c^2u_n)\delta^{-2}E\left((x^2 I_{\{|x|\le 1\wedge 2u_n^{1/2}\}} * \mu)_1\right)\\
&\le (64c^2u_n)\delta^{-2}(x^2 I_{\{|x|\le 1\wedge 2u_n^{1/2}\}} * \nu)_1 \qquad (4.60)
\end{aligned}
$$

which, as above, is less than $\varepsilon/4$ for sufficiently large n. Hence,

$$
P(\sup_{t\in[0,1]}|\sum_{(h,k)\in\Theta_n} y_{h,k}g_{h,k}(t)| > \delta) \to 0.
$$

Next, write $z_{h,k} = a_{h,k}+b_{h,k}$ where $a_{h,k} := \sum_{i=0}^{n-1}\overline{\tilde g_{h,k}(t_i)}(\Delta_i J^s)^2 I_{A_i\cap C_i}$ and $b_{h,k} := \sum_{i=0}^{n-1}\overline{\tilde g_{h,k}(t_i)}(\Delta_i J^s)^2 I_{A_i\cap C_i^c}$. Then

$$
\begin{aligned}
&P(|\sum_{(h,k)\in\Theta_n} z_{h,k}g_{h,k}(t)| > \delta)\\
&\le P(|\sum_{(h,k)\in\Theta_n} a_{h,k}g_{h,k}(t)| > \delta/2) + P(|\sum_{(h,k)\in\Theta_n} b_{h,k}g_{h,k}(t)| > \delta/2).
\end{aligned}
$$

In the first instance,

$$
\begin{aligned}
&P(|\sum_{(h,k)\in\Theta_n} b_{h,k}g_{h,k}(t)| > \delta/2)\\
&\le P\left(\cup_i\{I_{\{|x|>1\}} * \mu((t_i,t_{i+1}]\times\mathbb{R}) > 0, (\Delta_i J^s)^2 > 4u_n\}\right)\\
&\le nP(I_{\{|x|>1\}} * \mu([0,1/n]\times\mathbb{R}) > 0)E((\Delta_i J^s)^2)(4u_n)^{-1}\\
&\le cn^{-1}\kappa u_n^{-1}. \qquad (4.61)
\end{aligned}
$$

which can be made as small as desired. Now consider

$$
\begin{aligned}
&P(|\sum_{(h,k)\in\Theta_n} a_{h,k}g_{h,k}(t)| > \delta/2)\\
&\quad \leq P(\sum_{i=0}^{n-1}(\Delta_i J^s)^2 I_{\{|\Delta_i J^s|\leq 2u_n^{1/2}\}} > \delta(2cH_n)^{-1})\\
&\quad \leq \delta^{-1}(2cH_n)E\left(x^2 I_{\{|x|\leq 1\wedge 2u_n^{1/2}\}} * \mu_1\right)\\
&\quad \leq \delta^{-1}(2cH_n)(x^2 I_{\{|x|\leq 1\wedge 2u_n^{1/2}\}} * \nu)_1
\end{aligned}
$$

which can be made arbitrarily small. Hence,

$$
P(\sup_{t\in[0,1]} |\sum_{(h,k)\in\Theta_n} z_{h,k}g_{h,k}(t)| > \delta) \to 0. \tag{4.62}
$$

The result follows from (4.55),(4.58),(4.61), and (4.62). □

5 Simulation

5.1 Continuous prices

In this section, we confirm via simulations the results established analytically. We will first focus on the continuous case to mirror Proposition (3.1). Specifically, we will demonstrate that the mean integrated square error (MISE), the square bias, and the variance of the frame-based estimator tends to zero as the number of obervations increases. We use prices generated by 4 commonly used models of asset prices and returns, namely, the arithmetic Brownian motion (ABM), the Ornstein-Uhlenbeck process (OU), the geometric Brownian motion (GBM), and the Cox-Ingersoll-Ross (CIR) process. Eventhough, we have limited ourselves to these four processes, it should be clear from the proof of Proposition (3.1) that any continuous Itô semimatingale with continuous volatility process and càdlàg drift process would yield similar results.

We simulate prices using the following stochastic differential equations:

$$X_t = 0.8 + 0.5t + 0.2W_t, \tag{ABM}$$

$$X_t = 0.8 - \int_0^t 4X_s \,\mathrm{d}s + \int_0^t 0.2 \,\mathrm{d}W_s, \tag{OU}$$

$$X_t = 0.8 + \int_0^t 0.5X_s \,\mathrm{d}s + \int_0^t 0.2X_s \,\mathrm{d}W_s, \tag{GBM}$$

$$X_t = 0.8 + \int_0^t (0.1 - 0.5X_s) \,\mathrm{d}s + \int_0^t 0.2\sqrt{X_s} \,\mathrm{d}W_s, \tag{CIR}$$

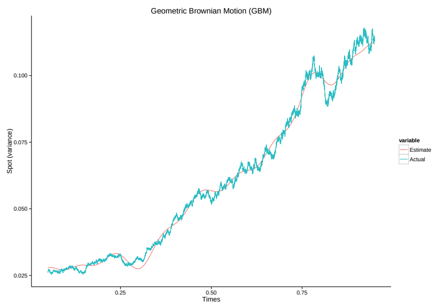
where W_t is a standard Brownian motion. For convenience, the observation interval is set to the unit interval $[0, 1]$. In all 4 cases, $X_0 = 0.8$. For each price model, we obtain estimates for the MISE, the square bias, and the variance of the estimator when the number of observations are 500, 5000, and 50000, respectively. In a high-frequency framework, 500 observations for an actively traded stock is likely too small; 5,000 is about right, but 50,000 is not entirely unheard of. At any rate, our objective is not to capture the average number of trades of any particular security, but rather, to obtain support for our asymptotic results by showing an inverse relationship between the number of observations and the MISE, and thereby gain a better understanding of the finite sample behavior of the estimator.

The starting point for constructing the estimator is to fix a generator for the Gabor frame. We have denoted the generator and its dual by g and $\tilde{g}$, respectively. For our purposes, any continuous and compactly supported function will work. In fact, part of the appeal of the frame method is this flexibility. We may chose the frame generator to match our prior assumptions about the smoothness of the latent volatility function. In this regard, a suitably *scaled*[9] member of the family of B-splines is particularly suited to the task of a Gabor frame *generator*. B-splines are piecewise polynomials, so, by varying their order or degree we may achieve any level of smoothness. Furthermore, the order of B-splines is directly related to the decay of their Fourier transforms. In fact the Fourier transform of a B-spline of order $p \geq 1$ decays like an $(p-1)$-th degree polynomial. This is important for the rate of decay of the MISE, and therefore, directly impacts the optimal choice of coefficients H_n to estimate. The upshot is: the higher the order of the B-spline, the smaller the number of coefficients needed to achieve a given level of accuracy.

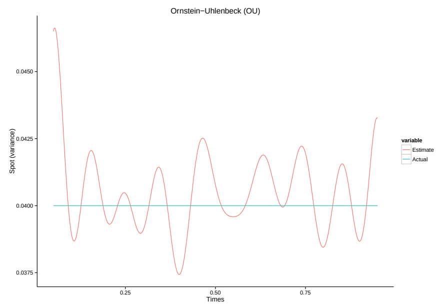
[9] More precisely *dilated*. See (5.63) for definition.

Figure 1: Estimated vs. actual spot volatility

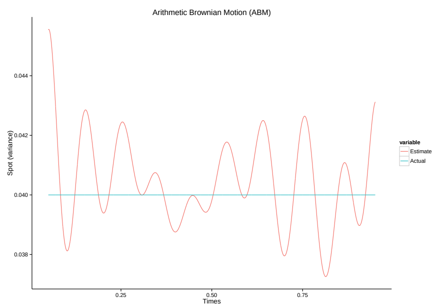
(a) GBM



(b) OU



(c) ABM



(d) CIR

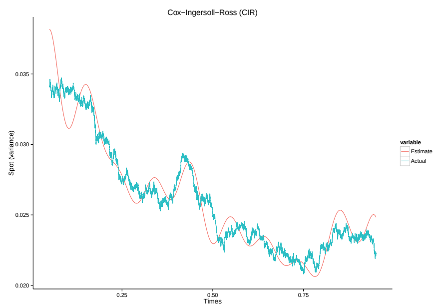


Table 1: Mean integrated square error (MISE) of $v_n(X,t)$.

	ABM			OU		
n	MISE	Sq. Bias	Var	MISE	Sq. Bias	Var
500	1.30×10^{-4}	2.86×10^{-6}	1.27×10^{-4}	1.43×10^{-4}	1.19×10^{-5}	1.31×10^{-4}
5000	1.41×10^{-5}	1.11×10^{-6}	1.30×10^{-5}	1.45×10^{-5}	1.62×10^{-6}	1.28×10^{-5}
50000	2.32×10^{-6}	1.02×10^{-6}	1.30×10^{-6}	2.36×10^{-6}	1.12×10^{-6}	1.23×10^{-6}

	GBM			CIR		
n	MISE	Sq. Bias	Var	MISE	Sq. Bias	Var
500	2.18×10^{-4}	4.18×10^{-6}	2.14×10^{-4}	6.26×10^{-5}	8.51×10^{-7}	6.17×10^{-5}
5000	2.33×10^{-5}	1.58×10^{-6}	2.17×10^{-5}	6.82×10^{-6}	6.00×10^{-7}	6.22×10^{-6}
50000	4.66×10^{-6}	1.02×10^{-6}	3.64×10^{-6}	1.46×10^{-6}	6.06×10^{-7}	8.52×10^{-7}

Note: The mean of the integrated square errors are obtained by taking an average over 100 sample paths generated for each model/number of observations pair.

From an implementation perspective, using a B-spline makes the construction of a *dual* frame generator a trivial matter. This is a consequence of Theorems 2.2 and 2.7 in Christensen (2006), which together specify a very simple rule for constructing dual pairs: Let $a > 0$ and $b > 0$ denote translation and modulation parameters, and let h be a B-spline of order p. Define the dilation operator $\mathcal{D}_c$ as follows:

$$\mathcal{D}_c f(x) = c^{-1/2} f(x/c). \tag{5.63}$$

If $0 < ab \leq 1/(2p-1)$ then $\{\mathcal{D}_a h, \mathcal{D}_a \tilde{h}\}$, where

$$\tilde{h}(x) = abh(x) + 2ab \sum_{n=1}^{p-1} h(x+n), \qquad x \in \mathbb{R}, \tag{5.64}$$

is a pair of dual Gabor frame generators. So if we start with a B-spline h then the dual generator will be a finite linear combination of scaled translates of h; consequently, the dual generator will be a spline, with similar regularity properties. For our simulation, we used a third-order B-spline. Our choice of the third order B-spline is motivated by a desire for a generator with a Fourier transform that decays like a quadratic polynomial. Specifically, we set

$$h(x) = \begin{cases} x^2/2 & x \in (1,0] \\ (-2x^2+6x-3)/2 & x \in (2,1] \\ (3-x^2)/2 & x \in (3,2] \\ 0 & x \notin (3,0] \end{cases}, \tag{5.65}$$

with $\tilde{h}$ computed as in (5.64) above. Our choice of the modulation and translation parameters is rather arbitrary. The only constraint is that $0 < ab \leq 1/(2p-1) = 1/5$; from our experimentation with different values, performance seems to be about the same for different choices satisfying the inequality; we settled on $a = 1/5$ and $b = 1/3$. Ideally H_n, the order of the number of frequency domain shifts, would be selected optimally to minimize MISE while balancing integrated variance and integrated square bias; we will turn our attention to this problem in future work. For the time being we set H_n naively equal to 50.

The simulation results indicate that the Gabor frame estimator performs satisfactorily. Figure 1 displays, for each of the 4 price models (ABM, OU, GBM, and CIR), simulated spot variance sample paths plotted against spot variance paths produced by the Gabor frame estimator. A visual inspection shows that the estimator produces a relatively good fit even with the naive selection of H_n. This claim is further corroborated by the analysis of the the integrated mean aquare error (MISE), the integrated square bias, and the integated variance summarized in Table 1. The figures in the table are arrived at in the following manner: first, 100 price histories are simulated for each observation frequency and model pair. So, each history is the result of sampling n price observations from distribution F, where n is the specified observation frequency and F is the distribution implied by the stochastic differential equation. The resulting data is a matrix with 100 rows and n columns. Each row represents a price history from which integrated quatities may be obatined, and each column indexes an orbservation time. Going down a column, average quantities may be computed. For instance, to arrive at the integrated square bias figures, average spot variances were computed for each observation times; the figures were then squared, weighted by Δ_n, and summed up. The integrated mean square error is computed similarly. We found that the variance, estimated in the foregoing manner, is only approximately the difference between the MISE and the integrated square bias. The reported figures for variance are in fact the difference between the MISE and the integrated square bias. The discrepancy is rather slight and does not materially change the result. In all 4 model, an inverse relation between MISE, square bias, and variance may be read off from the table. As was established mathematically, we expect MISE to vanish if the number of price observations were made to grow without bound.

5.2 Prices with jumps

In Propositions (4.1) and (4.2), we demonstrate analytically that $V_n(X,t)$ is consistent in term of the $L^2[0,1]$ distance when prices contain jumps of finite and infinite activity, respectivey. These results, however, do not assert the convergence of the mean integrated squared error of the estimator. Nevertherless, the consistency results lead us to suspect that the MISE of the estimator is also convergent. We investigate this analysis by simulating for price prosesses with jumps:

$$X_t = 0.8 + 0.5t + 0.2W_t + \sum_{i=1}^{N} Y_i, \tag{ABM + JMP}$$

$$X_t = 0.8 - \int_0^t 4X_s \,\mathrm{d}s + \int_0^t 0.2 \,\mathrm{d}W_s + \sum_{i=1}^{N} Y_i, \tag{OU + JMP}$$

$$X_t = 0.8 + \int_0^t 0.5X_s \,\mathrm{d}s + \int_0^t 0.2X_s \,\mathrm{d}W_s + \sum_{i=1}^{N} Y_i, \tag{GBM + JMP}$$

$$X_t = 0.8 + \int_0^t (0.1 - 0.5X_s) \,\mathrm{d}s + \int_0^t 0.2\sqrt{X_s} \,\mathrm{d}W_s + \sum_{i=1}^{N} Y_i, \tag{CIR + JMP}$$

wher N is a Poisson random variable with intensity 5 and Y_i, $1 \leq i \leq N$, is a normal random variable with mean zero and standard deviation 0.4.

We construct the dual Gabor frames as in the previous subsection using the third order B-Spline specified in (5.65). With the introduction of jumps into the simulation, we found out that better results may be obtained by varying tha parameters a, b, and H_n. We settled on $a = 1/7$, $b = 1/25$, and $H_n = 50$. The jump threshold is obtained by setting $u_n = n^{\alpha}$, where $\alpha = -0.45$. The results of the simulations are recorded in Table 2. We also produce a graph of a single observations (paths) in Figure 2. It is apparent from this simulation study that, eventhough the analytical part of the analysis assumed that volatility is continuous, the estimator $V_n(X, t)$ works fine in cases where volatiity is càdlàg. This is the case when we add jumps to the geometric Brownian and the CIR processes. We defer an analytical study of this more general situation to future work.

Figure 2: Estimated vs. actual spot volatility

(a) GBM + JMP **(b)** OU + JMP

(c) ABM + JMP **(d)** CIR + JUMP

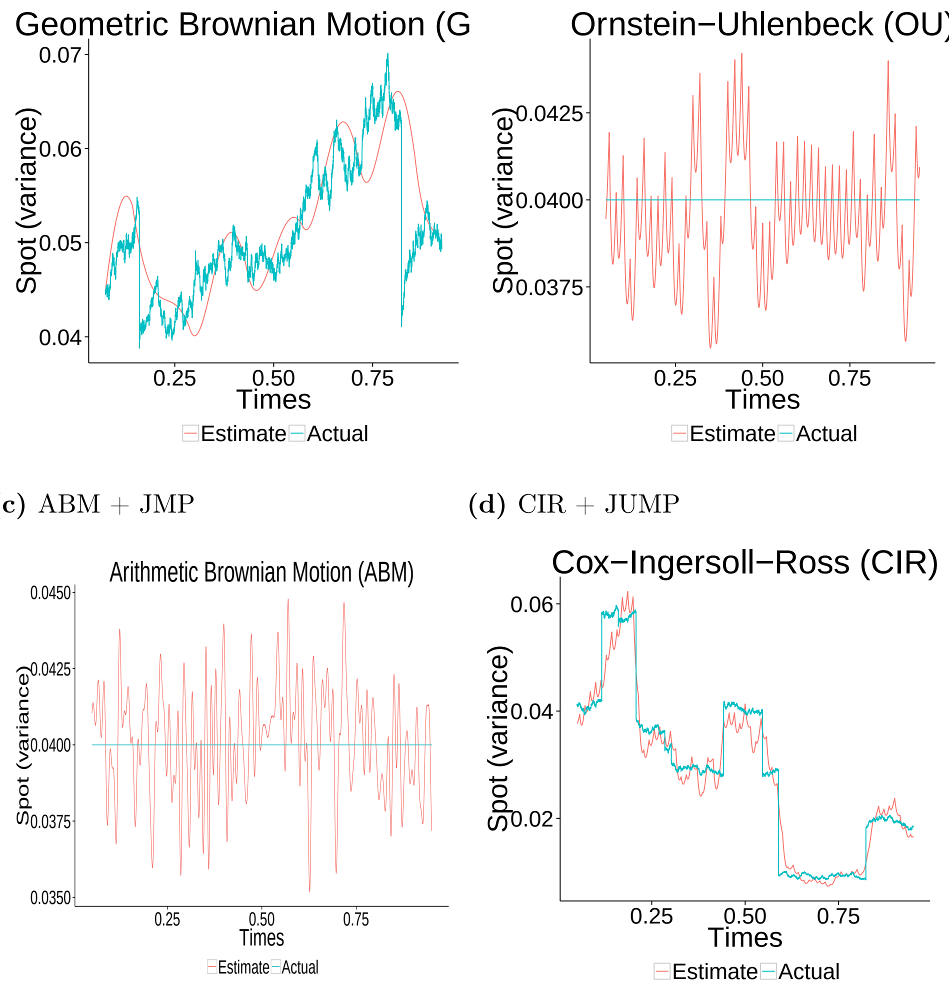


Table 2: Mean integrated square error (MISE) of $V_n(X,t)$.

	ABM + JMP			OU + JMP		
n	MISE	Sq. Bias	Var	MISE	Sq. Bias	Var
500	1.53×10^{-4}	8.95×10^{-6}	1.44×10^{-4}	8.51×10^{-4}	1.31×10^{-4}	7.20×10^{-4}
5000	2.19×10^{-5}	2.27×10^{-6}	1.96×10^{-5}	5.48×10^{-5}	9.76×10^{-6}	4.50×10^{-5}
50000	2.13×10^{-6}	9.00×10^{-8}	2.04×10^{-6}	6.61×10^{-6}	2.65×10^{-6}	3.97×10^{-6}

	GBM + JMP			CIR + JMP		
n	MISE	Sq. Bias	Var	MISE	Sq. Bias	Var
500	6.13×10^{-3}	8.70×10^{-4}	5.26×10^{-3}	3.74×10^{-4}	2.32×10^{-4}	1.43×10^{-4}
5000	3.42×10^{-4}	4.07×10^{-5}	3.02×10^{-4}	1.12×10^{-5}	8.29×10^{-6}	2.95×10^{-6}
50000	7.11×10^{-5}	6.36×10^{-6}	6.47×10^{-5}	7.05×10^{-6}	5.64×10^{-6}	1.40×10^{-6}

Note: The mean of the integrated square errors are obtained by taking an average over 50 sample paths generated for each model/number of observations pair.

6 Conclusion

We have investigated two types of estimators of the instantaneous volatility of asset prices. These estimators provide estimates of spot volatility not at a single point in time but for an entire time window. The main practical advantage of this type of estimator is their versatility. Once an estimate obtained various functionals of instantaneous volatility such as the ubiquitous integrated volatility are obtained immediately. Our contribution is two fold: we obtain estimators using frames, which have documented advantages over orthonormal basis such as wavelets and Fourier series; secondly, by modifying the basic estimator, we obtain an estimator of global spot volatility that remains consistent even in the presence of price jumps.

Appendices

A Proofs

We now give the proof of Lemma (2.2).

Proof. G is bounded away from zero. To see this, note that since g has support in $[r, s]$, the series on the left hand side of (2.11) has finitely many terms for each t. In addition, it is straight forward to verify that $G(t) = G(t + b)$ for all t; so, G is periodic with period b. It is also clear that because g is continuous, so is G. It follows that G attains its min and max on any interval of length b. Let I_b denote the interval $[(s + r - b)/2, (s + r + b)/2]$, then

$$\begin{aligned}\min_{t\in\mathbb{R}} G(t) &= \min_{t\in I_b} G(t) \\ &\geq a^{-1} \min_{t\in I_b} |g(t)|^2.\end{aligned}$$

Because g is continuous and g doesn't vanish in (r, s), we conclude that $G_* := \min_{t\in\mathbb{R}} G(t) > 0$. It is also straight forward that $G^* := \max_{t\in\mathbb{R}} G(t) < \infty$. Now, let $t, t' \in \mathbb{R}$, $t > t'$, such that $|t - t'| \leq \delta$, then

$$\begin{aligned}|\tilde{g}(t) - \tilde{g}(t')| &= |(G(t)G(t'))^{-1}(g(t)G(t') - g(t')G(t))| \\ &\leq (G_*^{-2})\{|g(t)||G(t) - G(t')| + |G(t)||g(t) - g(t')|\}.\end{aligned} \tag{A.66}$$

For a real number x, denote $\lfloor x\rfloor$ the largest integer less than or equal to x and $\lceil x\rceil$ the smallest integer that is greater than or equal to x. Now, Let A denote the set of integers i such that $r < t - ib < s$. By definition of g, $g(t - jb) = 0$, whenever $j \notin A$. Since $b > 0$, A contains at most $\lceil(1 + |s| + |r|)/b\rceil$ number of elements. Let $\tau := \min\{t - ib : i \in A\}$, i.e. τ is the smallest $t - ib$ such that $i \in A$. Because A contains at most a finite number of elements, there exists an integer k such that $\tau = t - kb$. Set $\tau' := t' - kb$.

It is straight forward to verify that $|\tau - \tau'| \leq \delta$ and

$$\begin{aligned}a|G(t) - G(t')| &\leq \sum_{j=0}^{\lceil(1+|s|+|r|)/b\rceil} |g(\tau + jb)^2 - g(\tau' + jb)^2| \\ &\leq \sum_{j=0}^{\lceil(1+|s|+|r|)/b\rceil} |g(\tau + jb) - g(\tau' + jb)||g(\tau + jb) + g(\tau' + jb)| \\ &\leq 2\lceil(1 + |s| + |r|)/b\rceil g^*\bar{\omega}(g, \delta),\end{aligned} \tag{A.67}$$

where $g^* := \max_{t\in\mathbb{R}} |g(t)|$. Returning to (A.66), we see that

$$|\tilde{g}(t) - \tilde{g}(t')| \le C_{\tilde{g}} \bar{\omega}(g, \delta),$$

where $C_{\tilde{g}} = G_*^2(2a(\lceil(1 + |s| + |r|)/b\rceil)(g^*)^2 + G^*)$. Now let $h, k \in \mathbb{Z}$, then

$$\begin{aligned} |\tilde{g}_{h,k}(t) - \tilde{g}_{h,k}(t')| &= |\mathrm{e}^{2\pi \mathrm{i} hat}(\tilde{g}(t - kb) - \tilde{g}(t' - kb))| \\ &\le |\tilde{g}(t - kb) - \tilde{g}(t' - kb)| \le C_{\tilde{g}} \bar{\omega}(g, \delta). \end{aligned} \tag{A.68}$$

The last inequality follows because translating a function leaves its modulus of continuity unchanged.

□

The following is a simple corollary to Lévy's modulus of continuity theorem.

A.1 Lemma *Suppose $\int_0^1 \sigma^2(s)\,\mathrm{d}s < \infty$, almost surely, where σ is adapted, strictly positive, and continuous. If W is an $\mathcal{F}_t$-Brownian motion, then for sufficiently large n, it is almost surely the case that*

$$\sup_{0\le i\le n-1} \left| \int_{t_i}^{t_{i+1}} \sigma(s)\,\mathrm{d}W_s \right| \le C(n^{-1}\log(n))^{1/2},$$

where C is a finite-valued random variable.

Proof. Let $T_t := \inf\{s > 0 : \int_0^s \sigma^2(u)\,\mathrm{d}u > t\}$ and $\mathcal{G}_t := \mathcal{F}_{T_t}$. Then, by Theorem 42 of Protter (2004), it is almost surely the case that

$$\int_0^t \sigma(s)\,\mathrm{d}W_s = B_{T_t},$$

where B is a $\mathcal{G}_t$-Brownian motion. Applying Lévy's modulus of continuity result (Rogers & Williams, 1994, Theorem 10.32) to B_{T_t}, we have, almost surely, for sufficiently large n

$$\sup_{0\le i\le n-1} \left| \int_{t_i}^{t_{i+1}} \sigma(s)\,\mathrm{d}W_s \right| (2\delta_i \log(1/\delta_i))^{-1/2} \le 1, \tag{A.69}$$

where $\delta_i := T_{t_{i+1}} - T_{t_i} = \int_{t_i}^{t_{i+1}} \sigma^2(s)\,\mathrm{d}s$. Because σ is continuous and strictly positive on $[0, 1]$, there are finite-valued random variables $c^* \ge c_* > 0$, such that $c^* \ge \sigma^2(s) \ge c_*$, $s \in [0, 1]$. Hence,

$$\begin{aligned} (2\delta_i \log(1/\delta_i))^{1/2} &\le (2c^* n^{-1} \log(n/c_*))^{1/2} \\ &= (2c^* n^{-1} \log(n) - 2c^* n^{-1} \log(c_*))^{1/2} \\ &\le (2c^* n^{-1} \log(n))^{1/2}. \end{aligned}$$

$\square$

Now we prove Proposition (3.1).

Proof. Without loss of generality let $X_0 = 0$, and take $\alpha \in (0, 1]$ and $c > 0$ as given. We begin with $B_n^2(\alpha, c)$, the integrated square bias component of $R_n(\alpha, c)$, which is defined as:

$$B_n^2(\alpha, c) = \int_0^1 E[(v_n((,t)t) - \sigma^2(t))I_{\mathcal{E}(\alpha,c)}]^2 \,\mathrm{d}t. \tag{A.70}$$

We make the following notational simplification:

$$E_{\mathcal{E}(\alpha,c)}[X] := E[XI_{\mathcal{E}(\alpha,c)}],$$

for all random variables X. We proceed by first obtaining an upper bound for the integrand in (A.70), i.e. the square bias at each fixed point t. To that end, let $t \in [0, 1]$, and note that

$$\begin{aligned} E_{\mathcal{E}(\alpha,c)}[v_n((,t)t) - \sigma^2(t)] = &\sum_{(h,k)\in\Theta_n} E_{\mathcal{E}(\alpha,c)}[\hat{c}_{h,k} \ - c_{h,k}] \, g_{h,k}(t) \\ &- \sum_{(h,k)\notin\Theta_n} E_{\mathcal{E}(\alpha,c)}[c_{h,k}] \, g_{h,k}(t), \end{aligned} \tag{A.71}$$

where

$$\begin{aligned} \hat{c}_{h,k} &= \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}(X_{t_{i+1}} - X_{t_i})^2 \text{ and} \\ c_{h,k} &= \int_0^1 \overline{\tilde{g}_{h,k}(s)}\sigma^2(s) \,\mathrm{d}s. \end{aligned}$$

We tackle the summands in (A.71) in turn starting with the first one. But first let

$$M_i := \int_{t_i}^{t_{i+1}} b(s) \,\mathrm{d}s, \quad \text{and} \quad S_i := \int_{t_i}^{t_{i+1}} \sigma(s) \,\mathrm{d}W_s,$$

and note that since $X_{t_{i+1}} - X_{t_i} = M_i + S_i$, it follows that

$$\begin{aligned} E_{\mathcal{E}(\alpha,c)}[(X_{t_{i+1}} - X_{t_i})^2] = &E_{\mathcal{E}(\alpha,c)}[M_i^2] + 2E_{\mathcal{E}(\alpha,c)}[M_iS_i] \\ &+ E_{\mathcal{E}(\alpha,c)}[S_i^2]. \end{aligned}$$

So, (A.71) may be written as

$$E_{\mathcal{E}(\alpha,c)}[v_n((,t)t) - \sigma^2(t)] = B_{1,n}(t) + B_{2,n}(t) + B_{3,n}(t) + B_{4,n}(t),$$

where

$$\begin{aligned}
B_{1,n}(t) &:= \sum_{(h,k)\in\Theta_n} g_{h,k}(t) \left(E_{\mathcal{E}(\alpha,c)} \left[\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} S_i^2 - c_{h,k} \right] \right), \\
B_{2,n}(t) &:= 2 \sum_{(h,k)\in\Theta_n} g_{h,k}(t) \left(\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} E_{\mathcal{E}(\alpha,c)} \left[S_i M_i \right] \right), \\
B_{3,n}(t) &:= \sum_{(h,k)\in\Theta_n} g_{h,k}(t) \left(\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} E_{\mathcal{E}(\alpha,c)} \left[M_i^2 \right] \right), \\
B_{4,n}(t) &:= - \sum_{(h,k)\notin\Theta_n} g_{h,k}(t) E_{\mathcal{E}(\alpha,c)}[c_{h,k}].
\end{aligned} \tag{A.72}$$

We will estimate the summands starting with $B_{4,n}(t)$. Note the following:

$$\begin{aligned}
\sum_{(h,k)\notin\Theta_n} g_{h,k}(t) E_{\mathcal{E}(\alpha,c)}[c_{h,k}] &= \sum_{(h,k)\notin\Theta_n} g_{h,k}(t) E[\langle \sigma^2, \tilde{g}_{h,k} \rangle I_{\mathcal{E}(\alpha,c)}] \\
&\leq E \left| \sum_{(h,k)\notin\Theta_n} g_{h,k}(t) \langle \sigma^2 I_{\mathcal{E}(\alpha,c)}, \tilde{g}_{h,k} \rangle \right| \\
&\leq cE[\bar{\omega}(\tilde{g}_{h,k}, 1/H_n) \log H_n] \\
&\quad + cE[\bar{\omega}(\sigma^2 I_{\mathcal{E}(\alpha,c)}, 1/H_n) \log H_n],
\end{aligned}$$

where the last line follows from Lemma 3.1. But since $\sigma^2 I_{\mathcal{E}(\alpha,c)}$ is bounded by c if it is in the Hölder ball $\mathcal{H}(\alpha, c)$ and 0 otherwise, it follows that $\bar{\omega}(\sigma^2 I_{\mathcal{E}(\alpha,c)}, 1/H_n) \leq cH_n^{-\alpha}$. Furthermore, by Lemma (2.2) and the Lipschitz continuity of g we have $\bar{\omega}(\tilde{g}_{h,k}, 1/H_n) \leq cH_n^{-1}$. So,

$$B_{4,n}(t) = O(H_n^{-\alpha} \log H_n). \tag{A.73}$$

Note the generic use of the constant c. In the sequel, we will use c to denote the amalgamation of various constants resulting from multiple steps; this should be harmless since constants are not asymptotically relevant.

We now obtain an estimate for $B_{3,n}(t)$. Note the following:

$$\begin{aligned}
E_{\mathcal{E}(\alpha,c)}[M_i^2] &= E\left[\left(\int_{t_i}^{t_{i+1}} b(s)\,\mathrm{d}s\right)^2 I_{\mathcal{E}(\alpha,c)}\right] \\
&= E\left[\left(\int_{t_i}^{t_{i+1}} b(s) I_{\mathcal{E}(\alpha,c)}\,\mathrm{d}s\right)^2\right] \\
&\leq E\left[\left(\int_{t_i}^{t_{i+1}} |b(s) I_{\mathcal{E}(\alpha,c)}|\,\mathrm{d}s\right)^2\right]
\end{aligned}$$

Note that $|b(s)I_{\mathcal{E}(\alpha,c)}|$ is either 0 or less than c, so

$$E_{\mathcal{E}(\alpha,c)}[M_i^2] \leq c\Delta_n^2. \tag{A.74}$$

Now since $g_{h,k}$ and $\tilde{g}_{h,k}$ are bounded independently of h and k, and $n\Delta_n = 1$, we have

$$B_{3,n}(t) = O(H_n \Delta_n). \tag{A.75}$$

We obtain an estimate for $B_{2,n}(t)$ next, but first let

$$\begin{aligned}
&T_{\sigma^2}(c) := \inf\{t \in (0,1] : \sigma^2(t) > c\}, \\
&T_b(c) := \inf\{t \in (0,1] : b(t) > c\}.
\end{aligned} \tag{A.76}$$

So, $T_{\sigma^2}(c)$ and $T_b(c)$ are the hitting times of the open set (c,∞) by b and σ^2, respectively; and they record the instant just before these coefficients exceed c. Because both processes are adapted and at least càdlàg, both hitting times are stopping times. Now, set

$$T_i(c) := T_{\sigma^2}(c) \wedge T_b(c) \wedge t_i, \qquad i = 1, \cdots, n.$$

Because $T_{\sigma^2}(c)$ and $T_b(c)$ are stopping times, so are the $T_i(c)$'s. The important thing to note is that at all times before $T_i(c)$, both b and σ^2 are bounded by c.

Returning to $B_{2,n}(t)$ note that $I^2_{\mathcal{E}(\alpha,c)} = I_{\mathcal{E}(\alpha,c)}$ so that $E_{\mathcal{E}(\alpha,c)}[M_i S_i] = E[(M_i I_{\mathcal{E}(\alpha,c)})(S_i I_{\mathcal{E}(\alpha,c)})]$. By the Cauchy-Schwarz inequality,

$$E[(M_i I_{\mathcal{E}(\alpha,c)})(S_i I_{\mathcal{E}(\alpha,c)})] \leq E[(M_i I_{\mathcal{E}(\alpha,c)})^2]^{1/2} E[(S_i I_{\mathcal{E}(\alpha,c)})^2]^{1/2} \tag{A.77}$$

Now, by repeating the same steps as in the case of $B_{3,n}(t)$ above, we may conclude that

$$E[(M_i I_{\mathcal{E}(\alpha,c)})^2]^{1/2} \leq c\Delta_n. \tag{A.78}$$

Now consider the following: if $\omega \in \mathcal{E}(\alpha, c)$ then $\sigma^2(\omega) \in \mathcal{H}(\alpha, c)$ and $\|\sigma^2(\omega)\|_\infty \leq \|\sigma^2(\omega)\|_\alpha \leq c$ so that $T_{\sigma^2(\omega)}(c) = 1$. Similarly $T_{b(\omega)}(c) = 1$ so that $T_i(c) = t_i$. Hence,

$$\begin{aligned} E[(S_i I_{\mathcal{E}(\alpha,c)})^2] &= E\left[\left(\int_{t_i}^{t_{i+1}} \sigma(s)\,\mathrm{d}W_s\right)^2 I_{\mathcal{E}(\alpha,c)}\right] \\ &= E\left[\left(\int_{t_i}^{T_{i+1}(c)} \sigma(s)\,\mathrm{d}W_s\right)^2 I_{\mathcal{E}(\alpha,c)}\right] \\ &\leq E\left[\left(\int_{t_i}^{T_{i+1}(c)} \sigma(s)\,\mathrm{d}W_s\right)^2\right], \end{aligned} \tag{A.79}$$

Note the role played by the $T_i(c)$'s; they serve to eliminate the factor $I_{\mathcal{E}(\alpha,c)}$ from the computations. An application of the Burkholder-Davis-Gundy (BDG) inequality now has the effect of eliminating the Wiener process W. So,

$$\begin{aligned} E[(S_i I_{\mathcal{E}(\alpha,c)})^2]^{1/2} &\leq cE\left[\left(\int_{t_i}^{T_{i+1}(c)} \sigma^2(s)\,\mathrm{d}s\right)\right]^{1/2} \\ &\leq (c\Delta_n)^{1/2}. \end{aligned} \tag{A.80}$$

Now, substituting (A.78) and (A.80) into (A.77) yields the estimate

$$E_{\mathcal{E}(\alpha,c)}[M_i S_i] \leq (c\Delta_n)^{3/2}. \tag{A.81}$$

Since $g_{h,k}$ and $\tilde{g}_{h,k}$ are bounded independently of h and k, and $n\Delta_n = 1$, we have

$$B_{2,n}(t) = O(H_n \Delta_n^{1/2}). \tag{A.82}$$

Now we tackle the final piece $B_{1,n}(t)$. Let

$$A := E_{\mathcal{E}(\alpha,c)}\left[\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} S_i^2 - \int_0^1 \sigma^2(s)\overline{\tilde{g}_{h,k}(s)}\,\mathrm{d}s\right]. \tag{A.83}$$

We will first obtain an upper bound for A; we proceed by adding and subtracting $E_{\mathcal{E}(\alpha,c)}[\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}}\overline{\tilde{g}_{h,k}(t_i)}\sigma^2(s)\,\mathrm{d}s]$ from A to yield:

$$\begin{aligned}A &= E_{\mathcal{E}(\alpha,c)}\left[\sum_{i=0}^{n-1}\overline{\tilde{g}_{h,k}(t_i)}\left(S_i^2 - \int_{t_i}^{t_{i+1}}\sigma^2(s)\,\mathrm{d}s\right)\right] \\ &\quad + E_{\mathcal{E}(\alpha,c)}\left[\sum_{i=0}^{n-1}\left(\int_{t_i}^{t_{i+1}}\sigma^2(s)\{\overline{\tilde{g}_{h,k}(t_i)} - \overline{\tilde{g}_{h,k}(s)}\}\,\mathrm{d}s\right)\right] \\ &=: A_1 + A_2.\end{aligned}$$

We obtain estimates in turn for the summands. By linearity of expectation

$$\begin{aligned}A_2 &= \sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} E[\sigma^2(s)I_{\mathcal{E}(\alpha,c)}]\{\overline{\tilde{g}_{h,k}(t_i)} - \overline{\tilde{g}_{h,k}(s)}\}\,\mathrm{d}s \\ &\leq c\bar{\omega}(\tilde{g}_{h,k},\Delta_n),\end{aligned}$$

where $\bar{\omega}(\tilde{g}_{h,k},\Delta_n)$ is the modulus of continuity of $\tilde{g}_{h,k}$ on an interval of length Δ_n. By Lemma (2.2) and the Lipschitz continuity of g we have,

$$A_2 \leq c\bar{\omega}(g,\Delta_n) \leq c\Delta_n. \tag{A.84}$$

Now, we obtain an estimate for A_1. First, let $D_i : \Omega \times [0,1] \to \mathbb{R}$ for $i = 0,\cdots,n-1$ be defined as follows:

$$D_i(t) := \overline{\tilde{g}_{h,k}(t_i)}\left(\int_{t_i}^{t}\sigma(u)\,\mathrm{d}W_u\right)I_{(t_i,t_{i+1}]}(t). \tag{A.85}$$

$$D_0(0) := 0. \tag{A.86}$$

So, $D_i(t)$ is 0 on $[0,1]$ except when t is in $(t_i,t_{i+1}]$. Now, using the integration by parts formula for semimartingales, we may write

$$S_i^2 - \int_{t_i}^{t_{i+1}}\sigma^2(s)ds = 2\int_{t_i}^{t_{i+1}}\left(\int_{t_i}^{s}\sigma(u)\,\mathrm{d}W_u\right)\sigma(s)\,\mathrm{d}W_s$$

so that

$$\begin{aligned}
A_1 &= 2\,E_{\mathcal{E}(\alpha,c)}\left[\sum_{i=0}^{n-1}\int_{t_i}^{t_{i+1}} \overline{\tilde{g}_{h,k}(t_i)}\left(\int_{t_i}^{s}\sigma(u)\,\mathrm{d}W_u\right)\sigma(s)\,\mathrm{d}W_s\right] \\
&= 2\,E_{\mathcal{E}(\alpha,c)}\left[\int_0^1\sum_{i=0}^{n-1} D_i(s)\sigma(s)\,\mathrm{d}W_s\right] \\
&= 2\,E\left[\left(\int_0^1\sum_{i=0}^{n-1} D_i(s)\sigma(s)\,\mathrm{d}W_s\right) I_{\mathcal{E}(\alpha,c)}\right].
\end{aligned}$$

Using the same stopping time argument as above, we may replace the upper limit of integration with $T_n(c)$ so that

$$\begin{aligned}
A_1 &\le 2\,E\left[\left(\int_0^{T_n(c)}\sum_{i=0}^{n-1} D_i(s)\sigma(s)\,\mathrm{d}W_s\right) I_{\mathcal{E}(\alpha,c)}\right] \\
&\le 2\,E\left[\left|\int_0^{T_n(c)}\sum_{i=0}^{n-1} D_i(s)\sigma(s)\,\mathrm{d}W_s\right|\right].
\end{aligned}$$

Now using the fact that $\int_0^{T_n(c)}\sum_{i=0}^{n-1} D_i(s)\sigma(s)\,\mathrm{d}W_s$ is a martingale, we may make another appeal to the BDG inequality to yield:

$$\begin{aligned}
A_1 &\le cE\left[\left|\int_0^{T_n(c)}\left(\sum_{i=0}^{n-1} D_i(s)\sigma(s)\right)^2\mathrm{d}s\right|^{1/2}\right] \\
&\le cE\left[\left|\int_0^{T_n(c)}\sum_{i=0}^{n-1}\{D_i(s)\sigma(s)\}^2\,\mathrm{d}s\right|^{1/2}\right],
\end{aligned}$$

where the last line follows because $D_i(s)D_j(s) = 0$ whenever $i \ne j$. Now if we define $D_i^* := \sup_{t_i<s\le T_n(c)} D_i(s)$, and use the fact that σ is less than c

before $T_n(c)$ then

$$\begin{aligned}
A_1 &\leq cE\left[\left|\sum_{i=0}^{n-1}\Delta_n(D_i^*)^2\right|^{1/2}\right] \\
&\leq cE\left[\left|\sum_{i=0}^{n-1}\Delta_n(D_i^*)^2\right|\right] \\
&\leq c\Delta_n\sum_{i=0}^{n-1}E[(D_i^*)^2] \qquad \text{(A.87)}
\end{aligned}$$

where c is a generic constant representing the bound on σ^2 and the BDG constant. Note from the definition of D_i (A.85) that it is itself a martingale, so we may bound D_i^* with yet another application of the BDG inequality. That is

$$\begin{aligned}
D_i^* &\leq \left(\int_{t_i}^{T_{i+1}(c)}\sigma^2(s)\,\mathrm{d}s\right)^{1/2} \\
&\leq c\Delta_n^{1/2}. \qquad \text{(A.88)}
\end{aligned}$$

Plugging the above into the estimate in (A.87) yields: $A_1 \leq c\Delta_n$. Combining the estimates for A_1 and A_2, it may be seen that

$$B_{1,n}(t) = O(H_n\Delta_n). \qquad \text{(A.89)}$$

Collecting the estimates for $B_{j,n}(t)$ for $j = 1, \cdots, 4$, it is easily seen that $E[v_n((,t)t) - \sigma^2(t)] = O(H_n\Delta_n^{1/2} + H_n^{-\alpha}\log H_n)$ for all $t \in [0,1]$. So that

$$B^2(\alpha, c) = O(H_n^2\Delta_n + H_n^{-2\alpha}\log^2 H_n).$$

Next, we obtain a bound for the variance term $V_n(\alpha, c)$. Recall that

$$V_n(\alpha, c) = \int_0^1 E_{\mathcal{E}(\alpha,c)}[\{v_n((,t)t) - E_{\mathcal{E}(\alpha,c)}[v_n((,t)t)]\}^2]\,\mathrm{d}t.$$

So that by the definition of the estimator, we may write

$$\begin{aligned}
V_n(\alpha, c) &= \int_0^1 E_{\mathcal{E}(\alpha,c)}\left[\left\{\sum_{(h,k)\in\Theta_n}(c_{h,k} - E_{\mathcal{E}(\alpha,c)}[c_{h,k}])g_{h,k}(t)\right\}^2\right] \mathrm{d}t \\
&= \sum_{(h,k)\in\Theta_n} \mathrm{var}_{\mathcal{E}(\alpha,c)}[\hat{c}_{h,k}]\left(\int_0^1 g_{h,k}^2(t)\,\mathrm{d}t\right) \\
&\quad + \sum_{(h,k)\neq(h',k')\in\Theta_n} \mathrm{cov}_{\mathcal{E}(\alpha,c)}[\hat{c}_{h,k}, \hat{c}_{h',k'}]\left(\int_0^1 g_{h,k}(t)g_{h',k'}(t)\,\mathrm{d}t\right) \\
&=: V_1 + V_2,
\end{aligned} \tag{A.90}$$

We will estimate these quantities in turn starting with V_1, but first let

$$\begin{aligned}
Y_i &:= \left(\int_{t_i}^{t_{i+1}} \sigma(s)\,\mathrm{d}W_s\right)^2, \\
Z_i &:= \left(\int_{t_i}^{t_{i+1}} b(s)\,\mathrm{d}s\right)^2 + 2\left(\int_{t_i}^{t_{i+1}} b(s)\,\mathrm{d}s\right)\left(\int_{t_i}^{t_{i+1}} \sigma(s)\,\mathrm{d}W_s\right), \\
\beta_{1,i} &:= \sum_{(h,k)\in\Theta_n} g_{h,k}^2(t_i)\left(\int_0^1 g_{h,k}^2(t)\,\mathrm{d}t\right),
\end{aligned}$$

for $i = 0, \cdots, n-1$. Now note that

$$\hat{c}_{h,k} = \sum_{i=0}^{n-1} g_{h,k}(t_i)(X_{t_{i=1}} - X_{t_i})^2 = \sum_{i=0}^{n-1} g_{h,k}(t_i)(Y_i + Z_i),$$

and since increments of the Brownian motion are independent, we have

$$V_1 = \sum_{i=0}^{n-1} \beta_{1,i}(\mathrm{var}_{\mathcal{E}(\alpha,c)}[Y_i] + \mathrm{var}_{\mathcal{E}(\alpha,c)}[Z_i] + 2\mathrm{cov}_{\mathcal{E}(\alpha,c)}[Y_i, Z_i]).$$

We will estimate the first two moments of Y_i and Z_i in turn. Note that

$$\begin{aligned}
E_{\mathcal{E}(\alpha,c)}[Y_i] = E[Y_i I_{\mathcal{E}(\alpha,c)}] &= E\left[\left(\int_{t_i}^{t_{i+1}} \sigma(s)\,\mathrm{d}W_s\right)^2 I_{\mathcal{E}(\alpha,c)}\right] \\
&= E\left[\left(\int_{t_i}^{T_i(c)} \sigma(s)\,\mathrm{d}W_s\right)^2 I_{\mathcal{E}(\alpha,c)}\right] \\
&\leq E\left[\left(\int_{t_i}^{T_i(c)} \sigma(s)\,\mathrm{d}W_s\right)^2\right] \\
&\leq cE\left[\left(\int_{t_i}^{T_i(c)} \sigma^2(s)\,\mathrm{d}s\right)\right] \\
&\leq c\Delta_n. \qquad \text{(A.91)}
\end{aligned}$$

where the fourth line results from an application of the BDG inequality. Repeating the exact same steps, it may be seen that $E_{\mathcal{E}(\alpha,c)}[Y^2] \leq c\Delta_n^2$. Thus,

$$\mathrm{var}_{\mathcal{E}(\alpha,c)}[Y_i] = E_{\mathcal{E}(\alpha,c)}[Y_i^2] - E_{\mathcal{E}(\alpha,c)}[Y_i]^2 \leq c\Delta_n^2. \qquad \text{(A.92)}$$

Next we obtain estimates for Z_i. From (A.74) and (A.81) we may conclude

$$E_{\mathcal{E}(\alpha,c)}[Z_i] = E_{\mathcal{E}(\alpha,c)}[M_i^2] + 2E_{\mathcal{E}(\alpha,c)}[M_i S_i] \leq c\Delta_n^{3/2}.$$

Using similar computations as above, it may be seen that $E_{\mathcal{E}(\alpha,c)}[Z_i^2] \leq c\Delta_n^3$ so that

$$\mathrm{var}_{\mathcal{E}(\alpha,c)}[Z_i] \leq c\Delta_n^3.$$

Now by the Cauchy-Schwarz inequality we may write

$$\mathrm{cov}_{\mathcal{E}(\alpha,c)}[Y_i, Z_i] \leq (\mathrm{var}_{\mathcal{E}(\alpha,c)}[Z_i]\mathrm{var}_{\mathcal{E}(\alpha,c)}[Y_i])^{1/2} \leq c\Delta_n^{5/2}.$$

Now because $g_{h,k}$ is bounded, it follows that $\beta_{1,i} = O(H_n)$ so

$$V_1 = O(H_n\Delta_n)$$

It is straight forward to see that V_2 may be estimated in a similar fashion. Indeed, let

$$\beta_{2,i} := \sum_{(h,k)\neq(h',k')\in\Theta_n} g_{h',k'}(t_i) g_{h,k}(t_i) \left(\int_0^1 g_{h,k}(t) g_{h',k'}(t)\,\mathrm{d}t\right),$$

then we may write

$$V_2 = \sum_{i=0}^{n-1} \beta_{2,i} (\mathrm{var}_{\mathcal{E}(\alpha,c)}[Y_i] + \mathrm{var}_{\mathcal{E}(\alpha,c)}[Z_i] + 2\mathrm{cov}_{\mathcal{E}(\alpha,c)}[Y_i, Z_i]),$$

the computations will then proceed identically as before from this point. Again, by the boundedness of $g_{h,k}$, it follows that $\beta_{2,i} = O(H_n^2)$ so that

$$V_2 = O(H_n^2 \Delta_n).$$

Therefore,

$$V(\alpha, c) = O(H_n^2 \Delta_n).$$

□

References

Barndorff-Nielsen, Ole E. and Shephard, Neil (2004) "Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics", *Econometrica*, Vol. 72, No. 3, pp. 885–925.

Christensen, Ole (2006) "Pairs of dual Gabor frame generators with compact support and desired frequency localization", *Applied Computational Harmonic Analysis*, Vol. 20, pp. 403–410.

——— (2008) *Frames and bases: An introductory course*, Boston: Birkhauser.

Cvetković, Zoran and Vetterli, Martin (1998) "Overcomplete expansions and robustness", *Wavelet Analysis and Its Applications*, Vol. 7, pp. 301–338.

Daubechies, Ingrid (1992) *Ten lectures on wavelets*: CBMS-NSF Series in Applied Mathematics, SIAM.

Fan, Jianqing and Wang, Yazhen (2008) "Spot volatility estimation for high-frequency data", *Statistics and its interface*, Vol. 1, pp. 279–288.

Florens-Zmirou, Danielle (1993) "On estimating the diffusion coefficicient from discrete observations", *Journal of Applied Probability*, Vol. 30, No. 4, pp. 790–804.

Foster, Dean P. and Nelson, Dan B. (1996) "Continuous record asymptotics for rolling sample variance estimators", *Econometrica*, Vol. 64, No. 1, pp. 139–174.

Genon-Catalot, V., Laredo, C., and Picard, D. (1992) "Non-parametric estimation of the diffusion coefficient by wavelets methods", *Scandinavian Journal of Statistics*, Vol. 19, No. 4, pp. 317–335.

He, Sheng-wu, Wang, Jia-gang, and Yan, Jia-an (1992) *Semimartingale theory and stochastic calculus*: CRC Press Inc.

Hoffmann, M., Munk, A., and Schmidt-Hieber (2012) "Adaptive wavelet estimation of the diffusion coefficient under additive error measurements", *Annales de l'institut Henry Poincaré, Probabilités et Statistiques*, Vol. 48, No. 4, pp. 1186–1216.

Malliavin, Paul and Mancino, Maria Elvira (2002) “Fourier series methods for measurement of multivariate volatilities”, *Finance and Stochastics*, Vol. 6, No. 1, pp. 49–61.

——— (2009) “A Fourier transform method for nonparametric estimation of multivariate volatility”, *Annals of Statistics*, Vol. 37, No. 4, pp. 1983 – 2010.

Malliavin, Paul, Mancino, Maria Elvira, and Recchioni, Maria Cristina (2007) “A non-parametric calibration of the hjm geometry: an application of itô calculus to financial statistics”, *Japanese Journal of Mathematics*, Vol. 2, pp. 55–77.

Mancini, Cecilia (2009) “Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps”, *Scandinavian Journal of Statistics*, Vol. 36, pp. 270 – 296.

Munch, N.J. (1992) “Noise reduction in tight weyl-heisenberg frames”, *IEEE transaction of information theory*, Vol. 38, pp. 608–619.

Protter, Philip E. (2004) *Stochastic Integration and Differential Equations*: Springer, 2nd edition.

Rogers, L.C.G. and Williams, D. (2000) *Diffusions, Markov Processes and Martingales: Volume 1: Foundations*: Cambridge University Press, 2nd edition.

Rogers, L.C.G. and Williams, Dvid (1994) *Diffusions, Markov Processes, and Martingales. Volume1: Foundations*: John Wiley & Sons, 2nd edition.

Zhang, Zhihua (2008) “Convergence of Weyl-Heisenberg frame series”, *Indian Journal of Pure and Applied Mathematics*, Vol. 39, No. 2, pp. 167–175.