Chapter 1

Nonparametric spot volatility estimation by Gabor frames methods

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 volatility estimates for entire time windows have received much less coverage. These are the so-called global estimators; the objects of interest are global: random elements whose realizations are sample paths, i.e. functions defined on nontrivial time intervals.

In the global volatility estimation literature, two estimators stand out: the Fourier-based estimator proposed by Malliavin & Mancino (2002) and the wavelet-based estimator proposed by Genon-Catalot et al. (1992) and later developped 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 estimators 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 of the volatility function converge in a mean square sense but stopped short of providing an explicit rate of convergence for the entire volatiltity function. On the other hand, a lot more is known about the wavelet-based estimator; for instance, uniform and integrated mean square convergence rates 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*. This robustness may be particularly pertinent in a high-frequency setting, where price measurements are subject to market microstructure noise. We elaborate on these points further below.

The rest of this paper is organized as follows: Section 1.1 gives a description of the dynamics of observed prices; Section 1.2 briefly reviews Gabor frames theory; Section 1.3 gives a specification of the Gabor frame based estimator; Section 1.4 discusses the asymptotic convergence of the frame-based estimator; Section 1.5 provides further support for the estimator via a simulation exercice; Section 1.6 provides a descriptive analysis of the diurnal pattern of intraday volatility in the bond markets; Section 1.7 proposes a multivariate estension; finally, Section 1.8 concludes and briefly discusses future work. The main technical arguments are contained in the Appendix.

1.1 Prices

We follow the literature by assuming that the asset price is a semimartingale[1]. This assumption is motivated in part by 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

[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.

processes known as *continuous Itô semimartingales.* Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]}, \mathrm{P})$ be a filtered probability satifying the *usual conditions.* A stochastic process X is a continuous Itô semimartingale if it is adapted, càlàg[2], and expressible in the form:

$$X_t = X_0 + \int_0^t \mu_s \mathrm{d}s + \int_0^t \sigma_s \mathrm{d}W_s, \qquad \forall t \in [0,T], \tag{1.1}$$

where W is a standard Brownian motion, X_0 is either known or observable at time 0, and both μ and σ are suitably restricted stochastic processes. We follow the ususal practise by 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 simple Brownian motion with drift case, μ and σ will occasionally be 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.

In the sequel, we will assume $T = 1$, without loss of generality. We assume prices are observed in the fixed time interval $[0,1]$ at discrete, equidistant times $t_i := i\Delta_n$, where $i = 0, 1, \cdots, n$ and $\Delta_n = 1/n$. 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.

1.2 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.

$$f = \sum_{k\in\mathbb{N}} c_k f_k. \tag{1.2}$$

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 (1.2)

[2]Processes that almost surely have sample paths which are right-continuous, with finite left-limits.

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 redundancies they contain. Of course, we end up computing more coefficients but there is no hard limit on the number of coefficients we should compute; we use the same n data points whether we compute k or $k+10$ coefficients.

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

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

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

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

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

1.1 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*

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

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 Parceval'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 A that maps every function f in $L^2(\mathbb{R})$ to a sequence c in $\ell^2(\mathbb{N})$ in the following way:

$$Af = c \qquad \text{where} \qquad c_k = \langle f, f_k \rangle. \tag{1.3}$$

On account of the fact that A takes a function defined on a continuum ($\mathbb{R}$) to a sequence, which is a function defined on the discrete set $\mathbb{N}$, A is known as the *analysis* operator associated with the frame $\{f_k\}_{k\in\mathbb{N}}$. That the analysis operator is bounded follows from the frame bounds in Definition (1.1). Now, the adjoint[4] of A, A^*, is well-defined and takes sequences in $\ell^2(\mathbb{N})$ to functions in $L^2(\mathbb{R})$. Using the fact that A^* must satisfy the equality $\langle Af, c\rangle = \langle f, A^*c\rangle$ for all $f \in L^2(\mathbb{R})$ and $c \in \ell^2(\mathbb{N})$, it may be deduced that

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

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

Now an application of the operator $(A^*A)^{-1}$ to every frame element f_k yields a sequence $\{\tilde{f}_k := (A^*A)^{-1} f_k\}_{k\in\mathbb{N}}$, 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 $\tilde{A}$, it may be shown[5] that

$$A^*\tilde{A} = \tilde{A}^*A = I, \tag{1.4}$$

where I is the identity operator and $\tilde{A}^*$ is the adjoint of the dual analysis operator $\tilde{A}$. The above yields a representation result since if $f \in L^2(\mathbb{R})$ then

$$f = A^*\tilde{A}f = \sum_{k\in\mathbb{N}} \langle f, \tilde{f}_k\rangle f_k. \tag{1.5}$$

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

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

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

[4]The adjoint is the functional counterpart of the transpose of a real matrix.

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

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 the correlation between the function and the dual frame elements, $\langle f, \tilde{f}_k \rangle$. It follows from the first equality in (1.4) and the commutativity of the duality relationship that functions in $L^2(\mathbb{R})$ are also expressible 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. $f = \sum_{k\in\mathbb{N}} \langle f, f_k \rangle \tilde{f}_k$.

Now a consequence of the noncommutativity of the composition operator is that $P := \tilde{A}A^* = A\tilde{A}^*$ need not be equal to the identity operator as in (1.4). In fact, (Daubechies, 1992, Proposition 3.2.3) shows that in the general case P is the orthogonal projection operator in $\ell^2(\mathbb{N})$ onto the range space of A, $R(A)$. That $R(A)$ is in general a proper subset of $\ell^2(\mathbb{N})$ is a consequence of the fact that, in general, frames are redundant, i.e. they contain "more" elements than is required in a basis, which is another way of saying that they form linearly dependent sets. The level of redundancy is inversely related to the "size" of the range space of A, $R(A)$. As we shall see shortly, this fact has important consequences for how coefficient error affects the precision of the reconstruction operator. Of course where the frame has no redundancy, i.e. the frame is in fact an orthonormal basis, the range space of A, $R(A)$, coincides with $\ell^2(\mathbb{N})$ and P is equal to the identity operator I.

1.2.1 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 exhibit 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 near-identical information 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 an

$L^2(\mathbb{R})$ frame and $A : L^2(\mathbb{R}) \to \ell^2(\mathbb{N})$ the associated analysis operator defined in (1.3). Now if the frame is redundant then it follows that it not a linearly independent set so that the range space of the analysis operator, $R(A)$, does not coincide exactly with $\ell^2(\mathbb{N})$. Now using propoerties of bounded linear operators, it may be shown that

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

From the discussion in the previous section, the orthogonal projector of $\ell^2(\mathbb{N})$ onto $R(A)$, P, is $\tilde{A}A^*$. Combining this with the equation above, we have $I = A^*P\tilde{A}$. So the representation result in (1.5) can be expressed as

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

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 $Af + \varepsilon$ instead of simply Af. Further assume that ε is decomposable as a sum of a component residing in the range space of A, ε_A, and a component $\varepsilon_{A^\perp}$ residing in the orthogonal complement of the range space of A, $A^\perp$, i.e. $\varepsilon = \varepsilon_A + \varepsilon_{A^\perp}$. So, by definition, $P\varepsilon_{A^\perp} = 0$. Now the operation of reconstructing a function from the noisy coefficients now yields

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

It is thus clear the deviation of the approximation $\|f - f_\varepsilon\|$ should be lower than $\|\varepsilon\|$ to the extent that the range space of A is small, which is another way of saying that the approximation error is reduced to the extent that the frame is redundant. As was noted by Daubechies (1992), this explanation is heuristic and probably accounts for only a small portion of the noise reduction gain observed in practical work. Nevertheless, it provides a starting point for starting to think about the source of the robustness of frames.

1.2.2 Gabor frames

Next, we specialize the discussion to Gabor frames. The analysis of Gabor frames involves two operators T and M, called translation and modulation operators, respectively. (T as used here will not be confused with the upper bound of the observation interval $[0, T]$, as the meaning of T will be clear from the context). If $f \in L^2(\mathbb{R})$ then

$$T_b f(t) := f(t - b), \tag{1.6}$$

$$M_a f(t) := e^{2\pi i a t} f(t), \tag{1.7}$$

for $a, b \in \mathbb{R}$, where $\mathrm{i}^2 = -1$. The parameters a and b are known, respectively, as the *modulation* and *translation* parameter. Both T and M are shift operators: T is a shift or translation operator on the time axis, whereas M performs shifts on the frequency axis. A Gabor system is constructed by performing time-frequency shifts on a single function $g \in L^2(\mathbb{R})$, i.e.

$$\{M_{ha}T_{kb}g\}_{h,k\in\mathbb{Z}},$$

for fixed a, b is a Gabor system. A Gabor system need not be a frame.

1.2 Definition *Let $g \in L^2(\mathbb{R})$ and $a, b > 0$ and, for all $t \in \mathbb{R}$, define*

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

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 or a Weyl-Heisenberg frame.

The fixed function g is called the *generator* or the *window function*. In order to obtain sharp asymptotic rates, we require g and its dual $\tilde{g}$ (see (1.5)) to be continuous and compactly supported. The following Lemma taken from Christensen (2006) and Zhang (2008) tells us exactly how to construct such dual pairs.

1.1 Lemma *Let $[r, s]$ be a finite interval, a and b positive constants, and g a continuous function. If $g(t) \neq 0$ when $t \in (r, s)$; $g(t) = 0$ when $t \notin (r, s)$; $a < 2\pi/(s - r)$; and $0 < b < s - r$; then $\{g, \tilde{g}\}$ is a pair of dual Gabor frame generators with*

$$\tilde{g}(t) := g(t)/G(t), \ \text{where} \tag{1.8}$$

$$G(t) := (2\pi/a)\sum_{k\in\mathbb{Z}}|g(t - kb)|^2. \tag{1.9}$$

Furthermore,

$$\tilde{g}_{h,k}(\cdot) := e^{\mathrm{i}hat}\tilde{g}(\cdot - kb), \qquad \forall h, k \in \mathbb{Z} \tag{1.10}$$

is compactly supported.

Next, we establish that the dual generator $\tilde{g}$ also inherits the continuity properties of g.

1.2 Lemma *Let the dual Gabor frame generator $\tilde{g}$ be constructed as in (1.8). If $\bar{\omega}(g, \delta)$ denotes the modulus of continuity of g, i.e. $\bar{\omega}(g, \delta) := \sup\{|g(t) - g(t')| : t, t' \in \mathbb{R} \text{ and } |t - t'| < \delta\}$, then*

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

where C is a positive constant.

Proof. See Appendix A. $\square$

In the sequel, we assume the Gabor frame setup in Lemma (1.1).

1.3 Volatility estimation

We make the following assumptions about the drift and volatility coefficients explicit.

1.1 Assumption

1. Spot volatility σ is a strictly positive and adapted stochastic process with continuous paths on $[0, 1]$.
2. The drift coefficient μ is adapted and càlàg 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 (càlàg 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 dual Gabor pair constructed as in Lemma (1.1), then σ^2 admits a Gabor frame expansion given by:

$$\sigma^2(t) = \sum_{h,k\in\mathbb{Z}} c_{h,k}\; g_{h,k}(t), \text{ where} \tag{1.11}$$

$$c_{h,k} = \langle \sigma^2, \tilde{g}_{h,k} \rangle. \tag{1.12}$$

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 (1.10) that $\tilde{g}_{h,k+1}$ is simply $\tilde{g}_{h,k}$ shifted by b units; so, $c_{h,k} = 0$ if $|k| > K_0$ with

$$K_0 := \lceil (1 + |s| + |r|)/b \rceil, \tag{1.13}$$

where $\lceil x \rceil$ is the smallest positive integer larger than $x \in \mathbb{R}$. Thus σ^2 admits a representation of the form:

$$\sigma^2(t) = \sum_{\substack{(h,k)\in\mathbb{Z}^2 \\ |k|\leq K_0}} c_{h,k}\; g_{h,k}(t). \tag{1.14}$$

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

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

where H_n is an increasing sequence[6] in n. We propose the following estimator of the volatility coefficient:

$$\hat{\sigma}_n^2(t) = \sum_{(h,k)\in\Theta_n} \hat{c}_{h,k}\, g_{h,k}(t), \qquad \forall t \in [0,1], \text{ where} \tag{1.15}$$

$$\hat{c}_{h,k} = \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}(X_{t_{i+1}} - X_{t_i})^2. \tag{1.16}$$

In the next section we show that the estimator converges to σ^2 on $[0,1]$ in a mean integrated square error sense.

1.4 Asymptotic properties

In this section we obtain an estimate of the rate of convergence of the Gabor frame estimator. The usual way to think about σ 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. In the current setting, E is the set of real-valued, continuous functions on $[0,1]$ equipped with the metric on $L^2[0,1]$ and denoted by $C^0[0,1]$. So, given an outcome ω in the sample space Ω, realized volatility $\sigma(\omega)$ is a real-valued, continuous function defined on $[0,1]$ with finite $L^2[0,1]$ norm. 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 $\sigma(\omega)$ is very smooth with finite derivatives of all orders, whereas for outcome ω', $\sigma(\omega')$ 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. But in all probability, the *rate* of convergence of the approximation would depend on the smoothness of the realized

[6]See Proposition (1.1) for restrictions on the growth of H_n.

volatility, with faster rates achieved for smooth functions. That is, for outcomes ω, ω' in Ω, if $\sigma(\omega)$ is smoother as a function of time than $\sigma(\omega')$ then the number of observations required to achieve a given level of accuracy when $\sigma(\omega)$ is realized should not exceed the required number when $\sigma(\omega')$ 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 partition 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 partition 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

$$|f|_\alpha := \frac{|f(x) - f(y)|}{|x-y|^\alpha} \leq K. \tag{1.17}$$

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. The 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 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 larger 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 (1.17). 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 partitioned 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 realized drift according to membership in balls of radius c in $L^\infty[0,1]$, that is,

$$\mathcal{U}(c) := \{f \in L^\infty[0,1] : \|f\|_\infty \leq c\}. \tag{1.18}$$

In this way we are able to partition 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 $\mu \in \mathcal{U}(c)$ and $\sigma^2 \in \mathcal{H}(\alpha, c)$ for $c, \alpha > 0$[7]. We denote such events by $\mathcal{E}(\alpha, c)$, i.e.,

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

So, an outcome ω is in $\mathcal{E}(\alpha, c)$ if realized drift is caught between $-c$ and c on $[0,1]$; 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 :

$$R_n(\alpha, c) := \mathrm{E}[\|\hat{\sigma}_n^2 - \sigma^2\|_2^2 \mathbb{1}_{\mathcal{E}(\alpha,c)}],$$

where $\mathbb{1}_{\mathcal{E}(\alpha,c)}$ is the indicator function of $\mathcal{E}(\alpha, c)$, $\|\cdot\|_2$ is the $L^2[0,1]$ norm, and n is the observation frequency. Note that with $\alpha = 0$ and $c = \infty$, the expression above would 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 the rate 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

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

integrated variance component. To see this, note the following:

$$
\begin{aligned}
R_n(\alpha, c) &= \mathrm{E} \int_0^1 \{(\hat{\sigma}_n^2(t) - \sigma^2(t))\mathbb{1}_{\mathcal{E}(\alpha,c)}\}^2 \mathrm{d}t && (1.19)\\
&= \int_0^1 \mathrm{E}[\{(\hat{\sigma}_n^2(t) - \sigma^2(t))\mathbb{1}_{\mathcal{E}(\alpha,c)}\}^2]\mathrm{d}t && (1.20)\\
&= \int_0^1 \mathrm{E}[(\hat{\sigma}_n^2(t) - \sigma^2(t))\mathbb{1}_{\mathcal{E}(\alpha,c)}]^2 \mathrm{d}t \\
&\quad + \int_0^1 \mathrm{var}[\hat{\sigma}_n^2(t)\mathbb{1}_{\mathcal{E}(\alpha,c)}]\mathrm{d}t. && (1.21)
\end{aligned}
$$

The equality in (1.20) results from an interchange of the expectectation and integration operators justified by Fubini's theorem. The decomposition in line (1.21) 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.

1.1 Proposition *Let $\{g, \tilde{g}\}$ be pair of dual Gabor generators constructed as in Lemma* (1.1)*. Suppose the conditions in Assumption* (1.1) *hold. If*

$$
H_n^2 \Delta_n + H_n \bar{\omega}(g, \Delta_n) + H_n^{-\alpha} \log H_n = o(1)
$$

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

$$
\begin{aligned}
B_n^2(\alpha, c) &= O(H_n^2 \Delta_n + \{H_n \bar{\omega}(g, \Delta_n)\}^2 + H_n^{-2\alpha} \log^2 H_n) \\
V_n(\alpha, c) &= O(H_n^2 \Delta_n^2), && (1.22)
\end{aligned}
$$

where $\bar{\omega}(g, \delta)$ is the modulus of continuity of function g, $\Delta_n = 1/n$ is the step size, and H_n is the order of the number of frame elements used in the estimation.

Proof. See the appendix.

1.1 Remark

1. First, 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 slower by a logarithmic factor. Intuitively, the logarithmic term shows up because we are expanding σ^2 using a frame, which may be thought

of as containing some redundant term. Overall, the rate of convergence of the integrated mean square error differ only by logarithmic term. In practical implementations, this is an insignificant price to pay for the added flexibility and robustness gained by using frames.

2. Second, this result shows that the variance component of the MISE does not depend on the smoothness properties of either σ^2 and g.

1.5 Simulation study

In this simulation study, we examine the impact of the observation frequency on the mean integrated square error (MISE), the square bias, and the variance of the frame-based estimator. We make this examination using prices generated by 4 commonly used models of asset prices, namely, the Arithmetic Brownian Motion (ABM), the Ornstein-Uhlenbeck process (OU), the Geometric Brownian Motion (GBM), and the Cox-Ingersoll-Ross (CIR) process.

Specifically, we simulate prices using the following stochastic differential equations:

$$\mathrm{d}X = 0.5\mathrm{d}t + 0.2\mathrm{d}W_t, \tag{ABM}$$
$$\mathrm{d}X = -4X_t\mathrm{d}t + 0.2\mathrm{d}W_t, \tag{OU}$$
$$\mathrm{d}X = 0.5X_t\mathrm{d}t + 0.2X_t\mathrm{d}W_t, \tag{GBM}$$
$$\mathrm{d}X = (0.1 - 0.5X_t)\mathrm{d}t + 0.2\sqrt{X_t}\mathrm{d}W_t, \tag{CIR}$$

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, 5,000, and 50,000, 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*[8] 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.

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 D_c as follows:

$$D_c f(x) = c^{-1/2} f(x/c). \tag{1.23}$$

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

$$\tilde{h}(x) = abh(x) + 2ab \sum_{n=1}^{p-1} h(x+n), \qquad \forall x \in \mathbb{R}, \tag{1.24}$$

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

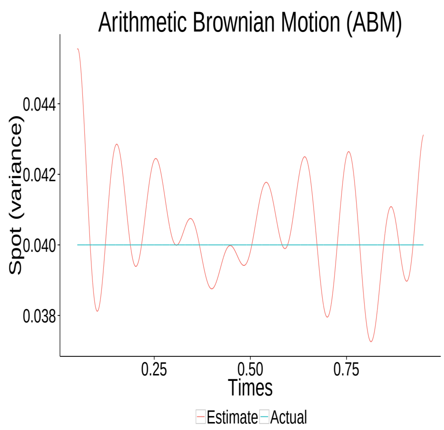
$$h(x) = \begin{cases} x^2/2 & x \in (1,0] \\ (-2x^2+6x-3)/2 & x \in (2,1] \\ (3-x^2)/2 & x \in (3,2] \\ 0 & x \notin (3,0] \end{cases},$$

with $\tilde{h}$ computed as in (1.24) above. Our choice of the modulation and translation parameters is rather arbitrary. The only constraint is that $0 <$

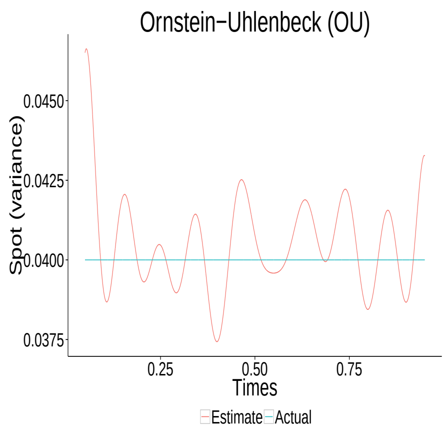
[8]More precisely *dilated*. See (1.23) for definition.

Figure 1.1: Estimated vs. actual spot volatility of common price models

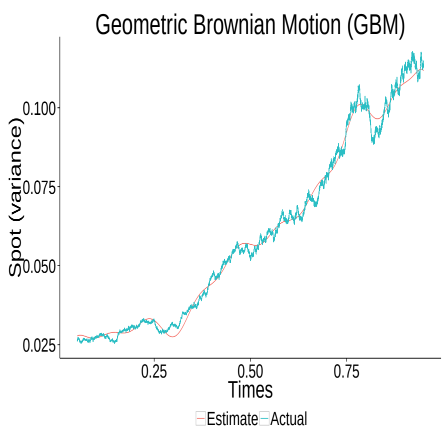
(a) ABM



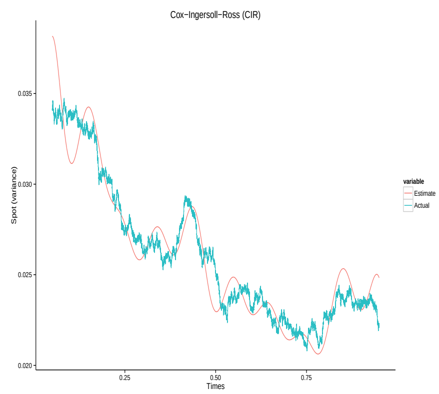
(b) OU



(c) GBM



(d) CIR



$ab \leq 1/(2p-1) = 1/5$; 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 Section (TBA). For the time being we set H_n naively equal to 50.

The simulation results indicate that the Gabor frame estimator performs satisfactorily. Figure 1.1 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 1.1. 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.

Table 1.1: Mean integrated square error (MISE) of the frame-based estimator $\hat{\sigma}_n^2(t)$ for popular price models.

	ABM			OU		
n	MISE	Sq. Bias	Var	MISE	Sq. Bias	Var
500	1.30E−4	2.86E−6	1.27E−4	1.43E−4	1.19E−5	1.31E−4
5000	1.41E−5	1.11E−6	1.30E−5	1.45E−5	1.62E−6	1.28E−5
50000	2.32E−6	1.02E−6	1.30E−6	2.36E−6	1.12E−6	1.23E−6

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

1.6 Volatility in the bond market

1.6.1 Balancing bias and variance

1.7 Multivariate Extension

We propose the following extension to $\hat{\sigma}_n^2$ from Chapter 1 to a multivariate setting. Let $\{X\}_{t\geq 0}$ be a d-dimensional vector of log-prices satisfying the stochastic integral equation:

$$X_t = X_0 + \int_0^t \mu(s, X_s)\mathrm{d}s + \int_0^t \sigma(s)\mathrm{d}W_s, \qquad t \geq 0,$$

where $\{W_s\}_{t\geq 0}$ is an m-dimensional standard Brownian motion; μ is $\mathbb{R}^d$-valued, continuous, and locally bounded in both variables; σ is $\mathbb{R}^d \times \mathbb{R}^m$-valued, continuous and locally bounded in time. Let

$$\Sigma(t) := \sigma(t)\sigma'(t), \qquad t \geq 0$$

where $\sigma'(t)$ is the transpose of $\sigma(t)$. Our aim is to obtain an estimate for Σ on the basis of n discretely and synchronously observed price vectors $\{X_1, \cdots, X_n\}$ in some fixed time interval $[0, 1]$. We assume that the prices are observed at equidistant intervals given by

$$\Delta_n := 1/n. \tag{1.25}$$

Let $\{g_{h,k}, \tilde{g}_{h,k}\}_{h,k\in\mathbb{Z}}$ denote a pair of dual Gabor frames generated under the assumptions of Lemma (1.1); let $\Theta_n := \{(h,k) \in \mathbb{Z}^2 : |h| \leq H_n \text{ and } |k| \leq K_0\}$, where H_n is an increasing sequence in n; and let $K_0 = \lceil (1+|s|+|r|)/b \rceil$ be constant. Intuitively, H_n represents the length of the frequency domain expansion for each time window; whereas, K_0, which is a constant because of the finite time domain support of the volatility function, captures the number of frame elements used to tile the time axis for each frequency domain shift. We propose to estimate the spot co-volatility matrix in $[0,1]$ using $\hat{\Sigma}_n$ defined component-wise for $1 \leq u, v \leq d$ as follows:

$$\hat{\Sigma}_n^{u,v}(t) = \sum_{(h,k)\in\Theta_n} \hat{c}_{h,k}^{u,v}\, g_{h,k}(t), \qquad \forall t \in [0,1], \tag{1.26}$$

where

$$\hat{c}_{h,k}^{u,v} = \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}(X_u(t_{i+1}) - X_u(t_i))(X_v(t_{i+1}) - X_v(t_i)). \tag{1.27}$$

We conjecture that $\hat{\Sigma}_n$ is consistent for Σ and that it converges in the mean integrated error sense at the same rate of convergence as $\hat{\sigma}_n^2$ (more precisely, the same order of convergence. So actual rate modulo a constant factor which we conjecture to be equal to the number m of driving Brownian motions). A rigorous proof of this conjecture will be given and further substantiated with simulations.

1.8 Conclusion

We proposed an estimator for the spot volatility function using Gabor frame methods. We showed that the estimator converges in a MISE sense and obtained an explicit convergence rate. The evidence for the validity of the proposed estimator will be further reinforced in a simulation study. We will also take the estimator to task using data from the Forex and bond market.

Bibliography

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.

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.

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

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

Appendices

Appendix A

Proofs

We now give the proof of Lemma (1.2).

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 (1.9) 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 $I_b := [(s + r - b)/2, (s + r + b)/2]$, then

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

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 $t, t' \in \mathbb{R}$ such that $|t - t'| \leq \delta$, then

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

Let $\tau := r + (t \bmod b)$, and $\tau' := r + (t' \bmod b)$. It is straight forward to verify that if $|\tau - \tau'| \leq \delta$, then

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

where $g^* := \max_{t\in\mathbb{R}} |g(t)|$. On the other hand, if $|\tau - \tau'| > \delta$, then

$$\begin{aligned}
&|G(t) - G(t')| \\
&\quad \leq |g(\tau')^2 - g(r)^2| + |g(s)^2 - g(\tau + c)^2| \\
&\qquad + \sum_{j=1}^{\lfloor (s+r)/b \rfloor} \{|g(\tau + (j-1)b)^2 - g(\tau' + jb)^2|\}.
\end{aligned}$$

where $c = \lfloor (s+r)/b \rfloor b$. It follows as above that

$$|G(t) - G(t')| \leq 2(\lceil (s+r)/b \rceil + 1) g^* \bar{\omega}(g, \delta). \tag{A.3}$$

Returning to (A.1), we see that

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

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

$$\begin{aligned}
|\tilde{g}_{h,k}(t) - \tilde{g}_{h,k}(t')| &= |e^{ihat}(\tilde{g}(t - kb) - \tilde{g}(t' - kb)| \\
&\leq |(\tilde{g}(t - kb) - \tilde{g}(t' - kb)| \leq C_{\tilde{g}} \bar{\omega}(g, \delta).
\end{aligned} \tag{A.4}$$

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

□

We establish Proposition (1.1) next.

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:

$$B_n^2(\alpha, c) = \int_0^1 \mathrm{E}[(\hat{\sigma}_n^2(t) - \sigma^2(t))\mathbb{1}_{\mathcal{E}(\alpha,c)}]^2 \mathrm{d}t. \tag{A.5}$$

We make the following notational simplification:

$$\mathrm{E}_{\mathcal{E}(\alpha,c)}[X] := \mathrm{E}[X \mathbb{1}_{\mathcal{E}(\alpha,c)}],$$

for all random variables X. We proceed by first obtaining an upper bound for the integrand in (A.5), i.e. the square bias at each fixed point t. To that end, let $t \in [0, 1]$, and note that

$$\begin{aligned}
\mathrm{E}_{\mathcal{E}(\alpha,c)}[\hat{\sigma}_n^2(t) - \sigma^2(t)] = &\sum_{(h,k)\in\Theta_n} \mathrm{E}_{\mathcal{E}(\alpha,c)}[\hat{c}_{h,k} - c_{h,k}]\; g_{h,k}(t) \\
&- \sum_{(h,k)\notin\Theta_n} \mathrm{E}_{\mathcal{E}(\alpha,c)}[c_{h,k}]\; g_{h,k}(t),
\end{aligned} \tag{A.6}$$

where

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

We tackle the summands in (A.6) in turn starting with the first one. But first let

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

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

$$\mathrm{E}_{\mathcal{E}(\alpha,c)}[(X_{t_{i+1}} - X_{t_i})^2] = \mathrm{E}_{\mathcal{E}(\alpha,c)}[M_i^2] + 2\mathrm{E}_{\mathcal{E}(\alpha,c)}[M_i S_i] + \mathrm{E}_{\mathcal{E}(\alpha,c)}[S_i^2].$$

So, (A.6) may be written as

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

where

$$\begin{aligned}
B_{1,n}(t) &:= \sum_{(h,k)\in\Theta_n} g_{h,k}(t) \left(\mathrm{E}_{\mathcal{E}(\alpha,c)} \left[\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} S_i^2 - c_{h,k} \right] \right), \\
B_{2,n}(t) &:= 2 \sum_{(h,k)\in\Theta_n} g_{h,k}(t) \left(\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} \mathrm{E}_{\mathcal{E}(\alpha,c)} \left[S_i M_i \right] \right), \\
B_{3,n}(t) &:= \sum_{(h,k)\in\Theta_n} g_{h,k}(t) \left(\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} \mathrm{E}_{\mathcal{E}(\alpha,c)} \left[M_i^2 \right] \right), \\
B_{4,n}(t) &:= - \sum_{(h,k)\notin\Theta_n} g_{h,k}(t) \mathrm{E}_{\mathcal{E}(\alpha,c)}[c_{h,k}].
\end{aligned} \tag{A.7}$$

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

$$\begin{aligned}
\sum_{(h,k)\notin\Theta_n} g_{h,k}(t) \mathrm{E}_{\mathcal{E}(\alpha,c)}[c_{h,k}] &= \sum_{(h,k)\notin\Theta_n} g_{h,k}(t) \mathrm{E}[\langle \sigma^2, g_{h,k} \rangle \mathbb{1}_{\mathcal{E}(\alpha,c)}] \\
&= \mathrm{E} \left| \sum_{(h,k)\notin\Theta_n} g_{h,k}(t) \langle \sigma^2 \mathbb{1}_{\mathcal{E}(\alpha,c)}, g_{h,k} \rangle \right| \\
&= O(\mathrm{E}[\bar{\omega}(\sigma^2 \mathbb{1}_{\mathcal{E}(\alpha,c)}, 1/H_n) \log H_n]),
\end{aligned}$$

where the last line follows from Theorem 4.1 of Zhang (2008). But since $\sigma^2 \mathbb{1}_{\mathcal{E}(\alpha,c)}$ is either in the Hölder ball $\mathcal{H}(\alpha, c)$ or 0, it follows that $\bar{\omega}(\sigma^2 \mathbb{1}_{\mathcal{E}(\alpha,c)}, 1/H_n) \leq cH_n^{-\alpha}$. So,

$$B_{4,n}(t) = O(H_n^{-\alpha} \log H_n). \tag{A.8}$$

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

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

Note that $|\mu(s)\mathbb{1}_{\mathcal{E}(\alpha,c)}|$ is either 0 or less than c, so

$$\mathrm{E}_{\mathcal{E}(\alpha,c)}[M_i^2] \leq c\Delta_n^2. \tag{A.9}$$

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 step; this should be harmless since constants are not asymptotically relevant. Now since $g_{h,k}$ and $\tilde{g}_{h,k}$ are bounded independently of h and k, and $n\Delta_n = 1$, we have

$$B_{3,n}(t) = O(H_n\Delta_n). \tag{A.10}$$

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

$$\begin{aligned}
&T_{\sigma^2}(c) := \inf\{t \in (0,1] : \sigma^2(t) > c\}, \\
&T_{\mu}(c) := \inf\{t \in (0,1] : \mu(t) > c\}.
\end{aligned} \tag{A.11}$$

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

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

Because $T_{\sigma^2}(c)$ and $T_\mu(c)$ 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 μ and σ^2 are bounded by c.

Returning to $B_{2,n}(t)$ note that $\mathbb{1}^2_{\mathcal{E}(\alpha,c)} = \mathbb{1}_{\mathcal{E}(\alpha,c)}$ so that $\mathrm{E}_{\mathcal{E}(\alpha,c)}[M_i S_i] = \mathrm{E}[(M_i \mathbb{1}_{\mathcal{E}(\alpha,c)})(S_i \mathbb{1}_{\mathcal{E}(\alpha,c)})]$. By the Cauchy-Schwarz inequality,

$$\mathrm{E}[(M_i \mathbb{1}_{\mathcal{E}(\alpha,c)})(S_i \mathbb{1}_{\mathcal{E}(\alpha,c)})] \leq \mathrm{E}[(M_i \mathbb{1}_{\mathcal{E}(\alpha,c)})^2]^{1/2} \mathrm{E}[(S_i \mathbb{1}_{\mathcal{E}(\alpha,c)})^2]^{1/2} \tag{A.12}$$

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

$$\mathrm{E}[(M_i \mathbb{1}_{\mathcal{E}(\alpha,c)})^2]^{1/2} \leq c\Delta_n. \tag{A.13}$$

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 $T_{\mu(\omega)}(c) = 1$ so that $T_i(c) = t_i$. Hence,

$$\begin{aligned}
\mathrm{E}[(S_i \mathbb{1}_{\mathcal{E}(\alpha,c)})^2] &= \mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}} \sigma(s)\mathrm{d}W_s\right)^2 \mathbb{1}_{\mathcal{E}(\alpha,c)}\right] \\
&= \mathrm{E}\left[\left(\int_{t_i}^{T_{i+1}(c)} \sigma(s)\mathrm{d}W_s\right)^2 \mathbb{1}_{\mathcal{E}(\alpha,c)}\right] \\
&\leq \mathrm{E}\left[\left(\int_{t_i}^{T_{i+1}(c)} \sigma(s)\mathrm{d}W_s\right)^2\right], \qquad \text{(A.14)}
\end{aligned}$$

Note the role played by the $T_i(c)$'s; they serve to eliminate the factor $\mathbb{1}_{\mathcal{E}(\alpha,c)}$ from the computations. An application of the Burkholder-Davis-Gundy (BDG) inequality now has the effect of eliminating the Weiner process W. So,

$$\begin{aligned}
\mathrm{E}[(S_i \mathbb{1}_{\mathcal{E}(\alpha,c)})^2]^{1/2} &\leq c\mathrm{E}\left[\left(\int_{t_i}^{T_{i+1}(c)} \sigma^2(s)\mathrm{d}s\right)\right]^{1/2} \\
&\leq (c\Delta_n)^{1/2}. \qquad \text{(A.15)}
\end{aligned}$$

Now, substituting (A.13) and (A.15) into (A.12) yields the estimate

$$\mathrm{E}_{\mathcal{E}(\alpha,c)}[M_i S_i] \leq (c\Delta_n)^{3/2}. \tag{A.16}$$

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

$$B_{2,n}(t) = O(H_n \Delta_n^{1/2}). \tag{A.17}$$

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

$$A := \mathrm{E}_{\mathcal{E}(\alpha,c)}\left[\sum_{i=0}^{n-1}\overline{\tilde{g}_{h,k}(t_i)}S_i^2 - \int_0^1 \sigma^2(s)\overline{\tilde{g}_{h,k}(s)}\mathrm{d}s\right]. \tag{A.18}$$

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

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

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

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

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 (1.2),

$$A_2 = O(\bar{\omega}(g,\Delta_n)).$$

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:

$$D_i(t) := \overline{\tilde{g}_{h,k}(t_i)}\left(\int_{t_i}^{t}\sigma(u)\mathrm{d}W_u\right)\mathbb{1}_{(t_i,t_{i+1}]}(t). \tag{A.19}$$

$$D_0(0) := 0. \tag{A.20}$$

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

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

so that

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

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

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

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:

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

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

$$
\begin{aligned}
A_1 &\leq c\mathrm{E}\left[\left|\sum_{i=0}^{n-1}\Delta_n(D_i^*)^2\right|^{1/2}\right] \\
&\leq c\mathrm{E}\left[\left|\sum_{i=0}^{n-1}\Delta_n(D_i^*)^2\right|\right] \\
&\leq c\Delta_n\sum_{i=0}^{n-1}\mathrm{E}[(D_i^*)^2] \qquad \text{(A.21)}
\end{aligned}
$$

where c is a generic constant representing the bound on σ^2 and the BDG constant. Note from the definition of D_i (A.19) that it is itself a martingale, so we may bound D_i^* with yet another application of the BDG inequality. That is

$$\begin{aligned} D_i^* &\le \left(\int_{t_i}^{T_{i+1}(c)} \sigma^2(s) \mathrm{d}s \right)^{1/2} \\ &\le c\Delta_n^{1/2}. \end{aligned} \tag{A.22}$$

Plugging the above into the estimate in (A.21) yields: $A_1 \le c\Delta_n$. Combining the estimates for A_1 and A_2, it may be seen that

$$B_{1,n}(t) = O(H_n\Delta_n + H_n\bar{\omega}(g,\Delta_n)). \tag{A.23}$$

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

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

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

$$\begin{aligned} V_n(\alpha, c) &= \int_0^1 \mathrm{E}_{\mathcal{E}(\alpha,c)}[\{\hat{\sigma}_n^2(t) - \mathrm{E}_{\mathcal{E}(\alpha,c)}[\hat{\sigma}_n^2(t)]\}^2]\mathrm{d}t \\ &= \int_0^1 \mathrm{E}_{\mathcal{E}(\alpha,c)} \left[\left\{ \sum_{(h,k)\in\Theta_n} (c_{h,k} - \mathrm{E}_{\mathcal{E}(\alpha,c)}[c_{h,k}]) g_{h,k}(t) \right\}^2 \right] \mathrm{d}t \\ &= \sum_{(h,k)\in\Theta_n} \mathrm{var}_{\mathcal{E}(\alpha,c)}[\hat{c}_{h,k}]\gamma_{h,k}^2 \\ &\quad + \sum_{(h,k)\neq(h',k')\in\Theta_n} \mathrm{cov}_{\mathcal{E}(\alpha,c)}[\hat{c}_{h,k}, \hat{c}_{h',k'}]\gamma_{h,k}\gamma_{h',k'} \\ &=: V_1 + V_2, \end{aligned} \tag{A.24}$$

where $\gamma_{h,k} := \int_0^1 g_{h,k}(t)\mathrm{d}t$. We will estimate these quantities in turn starting with V_1, but first let

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

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

$$\begin{aligned}
\mathrm{E}_{\mathcal{E}(\alpha,c)}[Y_i] = \mathrm{E}[Y_i \mathbb{1}_{\mathcal{E}(\alpha,c)}] &= \mathrm{E}\left[\left(\int_{t_i}^{t_{i+1}} \sigma(s)\mathrm{d}W_s\right)^2 \mathbb{1}_{\mathcal{E}(\alpha,c)}\right] \\
&= \mathrm{E}\left[\left(\int_{t_i}^{T_i(c)} \sigma(s)\mathrm{d}W_s\right)^2 \mathbb{1}_{\mathcal{E}(\alpha,c)}\right] \\
&\leq \mathrm{E}\left[\left(\int_{t_i}^{T_i(c)} \sigma(s)\mathrm{d}W_s\right)^2\right] \\
&\leq c\mathrm{E}\left[\left(\int_{t_i}^{T_i(c)} \sigma^2(s)\mathrm{d}s\right)\right] \\
&\leq c\Delta_n. \qquad \text{(A.25)}
\end{aligned}$$

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

$$\mathrm{var}_{\mathcal{E}(\alpha,c)}[Y_i] = \mathrm{E}_{\mathcal{E}(\alpha,c)}[Y_i^2] - \mathrm{E}_{\mathcal{E}(\alpha,c)}[Y_i]^2 \leq c\Delta_n^2. \qquad \text{(A.26)}$$

Next we obtain estimates for Z_i. From (A.9) and (A.16) we may conclude

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

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

$$V_1 = O(H_n \Delta_n^2)$$

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

$$\beta_{2,i} := \sum_{\substack{(h,k)\in\Theta_n \\ (h,k)\neq(h',k')}} g_{h',k'}(t_i) g_{h,k}(t_i) \gamma_{h',k'} \gamma_{h,k},$$

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

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

Therefore,

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

$\square$