Ph.D. research proposal:
On nonparametric spot volatility estimation, market microstructure noise, and fixed-income market stability.

Wale Dare

August 16, 2015

Contents

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 **?**, **?**, **?**, and **?**. 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 **?** and the wavelet-based estimator proposed by **?** and later developped by **?**. 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 **??**), the theoretical investigation of the estimator seems somewhat incomplete. For instance, **?** 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 Model

Let $\{X_t\}_{t\geq 0}$ be log prices with dynamics given by

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t)\mathrm{d}W_t, \qquad X_0 = x, \tag{1.1}$$

where $\{W_t\}$ is a standard Brownian motion with respect to the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}\}_t, \mathrm{P})$ satisfying the usual conditions; the initial price $x \in \mathbb{R}$ is known; the non-stochastic functions μ and σ are as yet unknown, but assumed to satisfy the Lipschitz and growth conditions sufficient for the existence of a strong solution. We assume prices are observed in the fixed time interval $[0, T]$ at discrete, equidistant times $t_i := i\Delta_n$, where $i = 0, 1, \cdots, n$ and $\Delta_n = T/n$. 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 $[0, T]$ by means of projection methods. We approach this task by estimating the projection of the spot variance in the finite dimensional subspace spanned by finite 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) 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 **?**. 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*

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

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[1], 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[2] 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

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

[1] 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.

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

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[3] 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

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

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, (**?**, 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

[3] See for example (**?**, Proposition 3.2.3)

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, **?** report noise reduction that is directly proportional to the degree of redundancy of the frame. **?** 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.

? 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 **?**, 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

$$T_b f(t) := f(t-b),$$
$$M_a f(t) := e^{2\pi \mathrm{i} at} f(t),$$

for $a, b \in \mathbb{R}$, where $\mathrm{i}^2 = -1$. 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.

$$\{M_h T_k g\}_{h,k\in\mathbb{Z}}$$

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 **?** and **?** 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*

$$\tilde{g}(t) := g(t)/G(t), \text{ where} \tag{1.6}$$
$$G(t) := (2\pi/a) \sum_{k\in\mathbb{Z}} |g(t-kb)|^2. \tag{1.7}$$

Furthermore,

$$\tilde{g}_{h,k}(\cdot) := e^{ihat}\tilde{g}(\cdot - kb), \qquad \forall h,k \in \mathbb{Z} \tag{1.8}$$

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* (1.6)*. If $\omega_g(\delta)$ denotes the modulus of continuity of g, i.e. $\omega_g(\delta) := \sup\{|g(t) - g(t')| : t, t' \in \mathbb{R}$ and $|t - t'| < \delta\}$, then*

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

where C is a positive constant.

Proof. See Appendix A.

□

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

1.3 Volatility estimation

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

1.1 Assumption

1. The volatility function is strictly positive, bounded, and continuous.
2. The modulus of continuity of the volatility coefficient, $\omega_\sigma(\Delta_n)$, is equal to $o(1/\log(1/\Delta_n))$ as $n \to \infty$.
3. There is $0 < C_T < \infty$ such that $|\mu(t,x)| \le C_T(1 + |x|)$, for all $t \in [0,T]$ and $x \in \mathbb{R}$.

Since σ^2 may not necessarily be defined on the entire real line, we proceed as in **?** by constructing an extension over the real line that has support in $[0,T]$; this way we are able to apply the Hilbert space machinery in $L^2(\mathbb{R})$. We let $\bar{\sigma}^2$ denote the extension of σ^2, i.e. $\bar{\sigma}^2$ coincides exactly with σ^2 on $[0,T]$. We summarize these notions as follows:

1.2 Assumption $\bar{\sigma}^2$ is in $L^2(\mathbb{R})$, has support in $[-\upsilon, T+\upsilon]$, where $\upsilon > 0$, and coincides with σ^2 on $[0,T]$.

With this substitution, we end up with a new process $\bar{X}$ coinciding with X on $[0, T]$ such that

$$\mathrm{d}\bar{X} = \mu(t, \bar{X}_t)\mathrm{d}t + \bar{\sigma}(t)\mathrm{d}W_t, \qquad \bar{X}_0 = x.$$

Now we may avail ourselves of the Gabor frame representation on $L^2(\mathbb{R})$. Let $\{g, \tilde{g}\}$ be a dual Gabor pair constructed as in Lemma (1.1), then $\bar{\sigma}^2$ admits a Gabor frame expansion given by:

$$\bar{\sigma}^2(t) = \sum_{h,k\in\mathbb{Z}} c_{h,k}\; g_{h,k}(t), \text{ where} \tag{1.9}$$

$$c_{h,k} = \langle \bar{\sigma}^2, \tilde{g}_{h,k} \rangle. \tag{1.10}$$

Note that both $\bar{\sigma}^2$ and $\tilde{g}$ have compact support. Indeed $\bar{\sigma}^2$ has support in $[0, T]$, whereas $\tilde{g}$ has support in $[s, r]$. So, $c_{h,k} \neq 0$ only if the supports of $\bar{\sigma}^2$ and $\tilde{g}_{h,k}$ overlap. Furthermore, we note from (1.8) 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

$$K_0 := \lceil (T + |s| + |r|)/b \rceil, \tag{1.11}$$

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

$$\bar{\sigma}^2(t) = \sum_{\substack{(h,k)\in\mathbb{Z}^2 \\ |k|\le K_0}} c_{h,k}\; g_{h,k}(t). \tag{1.12}$$

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

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

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

$$\hat{\sigma}_n^2(t) = \sum_{(h,k)\in\Theta_n} \hat{c}_{h,k}\; g_{h,k}(t), \qquad \forall t \in [0, T], \text{ where} \tag{1.13}$$

$$\hat{c}_{h,k} = \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} (X_{t_{i+1}} - X_{t_i})^2. \tag{1.14}$$

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

1.4 Asymptotic properties

Let R_n denote the average integrate square deviation of $\hat{\sigma}_n^2$ from $\bar{\sigma}^2$, i.e.

$$R_n = \mathrm{E}\int_{\mathbb{R}} \{\hat{\sigma}_n^2(t) - \bar{\sigma}^2(t)\}^2 \lambda(t)\mathrm{d}t, \tag{1.15}$$

where λ is a positive and continuous weight function with support in $(0, T)$. The weight function allows us to emphasis different time windows when estimating the volatility. For instance, we may wish to emphasize the recent past in real-time applications. We show that R_n tends to 0 as a function of the sample size, n. Note that R_n is the sum of a bias and a variance component, which we write as follows:

$$R_n = B_n^2 + V_n,$$

where

$$B_n^2 := \int_{\mathbb{R}} (\mathrm{E}[\hat{\sigma}_n^2(t) - \bar{\sigma}^2(t)])^2 \lambda(t)\mathrm{d}t \tag{1.16}$$

$$V_n := \int_{\mathbb{R}} \mathrm{E}[\{\hat{\sigma}_n^2(t) - \mathrm{E}[\hat{\sigma}_n^2(t)]\}^2]\lambda(t)\mathrm{d}t. \tag{1.17}$$

1.1 Proposition *Let $\{g, \tilde{g}\}$ be pair of dual Gabor generators constructed as in Lemma (1.1). Suppose the conditions in Assumptions (1.1) and (1.2) hold. If $H_n^2\Delta_n, H_n\omega_g(\Delta_n)$, and $\omega_{\sigma^2}(1/H_n)\log H_n \to 0$, then the mean integrated square error R_n tends to 0 as n tends to infinity, with*

$$\begin{aligned} B_n^2 &= O(H_n^2\Delta_n + \{H_n\omega_g(\Delta_n)\}^2 + \{\omega_{\sigma^2}(1/H_n)\log H_n\}^2) \\ V_n &= O(H_n^2\Delta_n), \end{aligned} \tag{1.18}$$

where $\omega_g(\delta) := \sup\{|g(t) - g(t')| : t, t' \in \mathbb{R}$ and $|t - t'| < \delta\}$.

Proof. See Appendix A.

□

1.1 Remark

1. First, the above bounds are remarkably similar to those achievable using an orthonormal basis such as wavelets (**?**). 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 $\bar{\sigma}^2$ using a frame, which may be thought of as containing some redundant term. In practical implementations, this may be a small 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

Obs	ABM	OU	GBM	CIR
500				
5000				
50000				

In practice, it is straight forward to obtain an estimator of the spot volatility using the approach of the previous section. The starting point is the selection of a suitable Gabor frame generator g. Any smooth, compactly supported function would do the trick, but using a piecewise polynomial such as a B-spline makes the task particularly simple because the dual generator $\tilde{g}$ is given by a finite linear combination of translates of g. This is the content of Theorem TBA in **?**.

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:

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

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 $[0, T]$, where $T < \infty$. With very little loss of generality we assume that the prices are observed at equidistant intervals given by

$$\Delta_n := T/n. \tag{1.19}$$

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 $K_0 = \lceil (T+|s|+|r|)/b \rceil$ 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 $[0, T]$ using $\hat{\Sigma}_n$ defined component-wise for $1 \leq u, v \leq d$ as follows:

$$\hat{\Sigma}_n^{u,v}(t) = \sum_{(h,k)\in\Theta_n} \hat{c}_{h,k}^{u,v}\, g_{h,k}(t), \qquad \forall t \in [0,T], \tag{1.20}$$

where

$$\hat{c}_{h,k}^{u,v} = \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}(X_u(t_{i+1}) - X_u(t_i))(X_v(t_{i+1}) - X_v(t_i)). \tag{1.21}$$

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.

Chapter 2

Market microstructure noise and spot volatility estimation

The need to estimate the volatility of an asset is pervasive in finance. Volatility is the key component in portfolio selection, option pricing, and risk management. Without further restriction on the price process the estimation of the volatility coefficient would be all but impossible. Thankfully financially theory provides some guidance in this regard. At relatively long observation intervals such as a week or a month, it is generally agreed that observed prices may be thought of as discrete realizations from a semimartigales. The theory of semimartingales provides a complete answer on what form we should expect the volatility coefficient of the price process to take. Furthermore, the theory tells us that the usual realized variance estimator converges to the quadratic variation process as the observation interval shrinks to zero (**?**, Theorem 23).

Unfortunately, this general concensus on the semimartingale status of observed prices breaks down when we consider prices sampled at higher frequencies. The problem is that, at high frequencies, it is hard to justify the assumption that the efficient price process, which may be a semimartingale, is directly observable. Instead, what is observable is part efficient price and part noise resulting from the established processes in the market. The noisy component of observed prices has several sources; these include the so-called bid-ask bound, the release of asynchronous information, and rounding error resulting from discrete prices etc. We elaborate on this sources below. The term *market microstructure noise* was coined by **?** to describe this type of price contamination. In short, due to the presence of market microstructure noice in high frequency prices, it may no longer be justifiable to assume that observed prices are the discrete realizations of a semimartigale.

There is a large literature dealing with nonparametric estimation of the volatility process in the presence of market microstructure noise. Here too the focus of these efforts has been on obtaining estimates for the integrated volatility. Some of the proposed approaches such as the two-time scale estimator of **?** and the pre-averging estimator of **?** have been extended to cover local spot volatility estimation; see for example **?**. On the other hand, the study of market-microstructure-robust global volatility estimators has thus far been rather scant. A notable contribution, is the wavelet-based estimator proposed by **?**, which works essentially by estimating the wavelet coefficient using pre-averaged market data. In this chapter, we propose an alternative estimator which combines the Gabor frame estimator from the previous chapter with the multi-time scale procedure popularized by **?**. We believe that frame-based estimators are very well-suited for dealing with high-frequency data. In a high-frequency setting, we may expect basis coefficient estimation error to be amplified by the presence of market microstructure noise in price data. As was explained in Chapter 1 (see subsection 1.2.1) , the redundancy inherent in frames can be leveraged to efficiently reduce the effects of market microstructure noise on overall volatility estimates.

The rest of this chapter is organized as follows: In Section 2.1 we specify the price process and the structure of the noise process. In Section 2.2 we give a specification of the noise robust spot volatility estimator. In Section 2.3 we conduct a simulation exercise to verify the validity of the proposed estimator. In section 2.4 we use the estimator to gain insight into the diurnal pattern of volatility in the bond market; Section 2.5 concludes the paper. A rigorous proof of the consistency of the estimator will be given in the Appendix.

2.1 Model

We consider the problem of making inference on the spot volatility of a security price process using market data sampled discretely in the presence of market microstructure noise. The presence of the noise component implies that the price process of interest is unobservable directly; instead what we have are discrete transaction or bid and ask price data with market microstructure noise contamination. The usual way this market setup is modeled in the literature is via the Additive Market Microstructure (AMN) model, which as the name implies, states that prices are affected additively by microstructure noise. That is, for $i = 0, 1, \cdots, n$ and $0 = t_0 \leq t_i \leq t_n = T$, the i-th observed price at time t_i may be modeled as:

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}, \tag{2.1}$$

where ε_{t_i} is the i-th coordinate in an i.i.d sequence of market microstructure noise. The noise component is assumed to be independent of the efficient price process. The unobserved efficient price of interest X is the unique solution to the following stochastic differential equation:

$$\mathrm{d}X_t = \mu(t, X_t)\mathrm{d}t + \sigma(t)\mathrm{d}W_t, \tag{2.2}$$

where W_t is a Brownian motion, σ is strictly positive, and both μ and σ are continuous and bounded from above. The functions μ and σ are referred to respectively as the drift and volatility functions. This setup is similar to the microstructure environment considered by **?** in their estimation of the integrated volatility.

2.1.1 Market microstructure noise

According to **?**, market microstructure noise arises from the moment-to-moment aggregate exchange behavior. Some major sources of market microstructure noise include:

1. *The bid-ask spread.* The price at which an investor can buy an asset, at any fixed point in time, is almost always greater than the price at which he may sell the asset. The *real* or efficient price of the asset is somewhere in between (in some cases, it could be outside the range if there is private information not available to the other participants in the market)

2. *The price impact of trade.* The idea is that each transaction releases information about the underlying asset. For instance, a buyer-initiated transaction tells the market that the asset is more valuable than its current price to somebody. Now, a really big buyer-initiated transaction tells the market that someone with a lot of money and, with no doubt a sophisticated knowledge of the market, thinks the asset is more valuable in the future than its current price. This type of information release can lead to a domino effect where the market goes through several rounds overbidding the price of the asset even though the fundamentals of the asset may not have changed. A pioneering work in the theoretical study of the price impact of trades is the paper by **?**; **?** provides an empirically-oriented treatment.

3. *Price round-off* Suppose the market valuation of IBM stock is CHF 19.95666, but because market prices are quoted up to a certain decimal place, the security may be exchanged at say CHF 19.95. Economically,

this seems like a small matter, but implementation-wise this a problem for any statistical procedure relying on the assumption that prices satisfy some form of *recurrence or mixing* property. This is because with prices rounded at 2 decimal places, it is no longer the case that any possible value in the continuous range of the asset price will eventually show up in the data given enough time. Thus the vast majority of price information will in fact *never* be reported.

4. *Human error* This is especially a problem for prices resulting from trading pits. The chaos of the trading pit almost surely guarantees the occurence of data entry errors throughout the trading day.

2.2 Noise-robust estimator

We start by highlighting the difficulty with which we are faced when trying to estimate the volatility function in the presence of market microstructure noise. The main difficulty in Chapter 1 is that we are attempting to estimate an unobservable or latent volatility function on the basis of discretely observable price process; in the present setting, not even the price process X is observable. What we have at our disposal is data Y, which is part efficient price X and part noise ε. We have no idea how much of the observed price Y is noise and how much is efficient price. It is thus safe to assume that we need to do something different in order to take care of the contamination. Motivated by the ideas proposed by **?** in the integrated volatility case, we propose to divide the sample n into R_n subsamples containing m_n data points so that $R_n m_n$ is approximately n or $R_n m_n \sim n$. For this approach to work, it is required that both R_n and m_n be large and tend asymptotically to infinity. A consequence of this assumption is that the current estimator does not merely collapse to the estimator proposed in Chapter 1 in the noiseless case. Once the subsamples are in place, we may compute coefficients of the Gabor frame expansion using each subsample (one set of basis coefficients for each subsample); next, we take an average of the basis coefficient estimates over the R_n subsamples; finally, we use the coefficient estimates based on the entire sample n to bias-correct the average coefficient estimates obtained in the previous step. In notation, we propose the following estimator of the spot volatility function in an environment with market microstructure noise:

$$\hat{\sigma}^2_{n,b}(t) = \sum_{(h,k)\in\Theta_n} \hat{c}^b_{h,k}\, g_{h,k}(t), \qquad \forall t \in [0,T], \tag{2.3}$$

where

$$\hat{c}^b_{h,k} = \hat{c}^R_{h,k} - (m_n/n)\hat{c}_{h,k} \tag{2.4}$$

$$\hat{c}^R_{h,k} = (1/R_n) \sum_{i=0}^{n-R_n} \overline{\tilde{g}_{h,k}(t_i)}(Y_{t_{i+R_n}} - Y_{t_i})^2 \tag{2.5}$$

$$\hat{c}_{h,k} = \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)}(Y_{t_{i+1}} - Y_{t_i})^2. \tag{2.6}$$

Here, as in Chapter 1, $\{g_{h,k}, \tilde{g}_{h,k}\}_{h,k\in\mathbb{Z}}$ denotes a pair of dual Gabor frames generated under the assumptions of Lemma (1.1); $\Theta_n := \{(h,k) \in \mathbb{Z}^2 : |h| \le H_n \text{ and } |k| \le K_0\}$; H_n is an increasing sequence in n representing the number of frame elements used along the frequency axis for each time axis shift; whereas $K_0 = \lceil (T + |s| + |r|)/b \rceil$ is a constant representing the number of frame elements used along the time axis for each frequency domain shift.

We would like to mention that the similarities between $\hat{\sigma}^2_n$ from Chapter 1 and $\hat{\sigma}^2_{n,b}$ are at best superficial. Note that $\hat{\sigma}^2_n$ is constructed using the actual efficient price process X, whereas in the present context we have to make do with corrupted market data Y. Also, note that the coefficients $\hat{c}_{h,k}$ play a secondary role here; they merely serve as a devise for bias correction. On the other hand there are strong similarities between the two time scale estimators of **?** and **?**. To see this note that we may express (2.5) as follows:

$$\hat{c}^R_{h,k} = (1/R_n) \sum_{i=1}^{R_n-1} \sum_{j=1}^{m_n} \overline{\tilde{g}_{h,k}(t_{i+(j-1)R_n\Delta_n})}(Y_{i+jR_n\Delta_n} - Y_{i+(j-1)R_n\Delta_n})^2.$$

Now it is clear that $\hat{c}^R_{h,k}$ is the average coefficient taken over the R_n coefficients estimates obtained using the R_n subsamples. In the sequel we show using a simulation study the validity of the proposed estimator. We also apply the estimator compute and study diurnal patterns in intra day volatility in the bond market. A rigorous proof of the consistency of the estimator will be given in the Appendix.

2.3 Simulation study

2.4 Diurnal pattern in the bond market revisited

2.5 Conclusion

Chapter 3

On measuring changes in bond market stability

It is a well-known tenet in empirical finance that a large portion of the variability in financial markets results from a handful of risk-factors. Accurate measurement and use of these principal factors has been shown to be economically worthwhile. For instance, **?**, working directly with spot rate yields, argue that a large portion ($> 98\%$) of the volatility in fixed-income returns can be explained by no more than three characteristics of the yield curve, with the most principal factor accounting for roughly 89% of return variance by itself. In addition, **?** show by way of a case study the superiority of a hedged portfolio immunized against these three principal factors over the traditional duration-matched portfolio.

Approaching the problem from a slightly different angle by using futures contracts to reconstract a proxy for the foreward rate curve (FRC), **?** report a very fast decay of the eigenvalues associated with the stochastic component of the foreward rate curve, implying that much of the observed fluctuation observed in FRC may be attributed to a small number of "modes". In addition, **?** demonstrate the statistical stability and significance of the principal factors driving the stochastic components of the the FRC via a simulation exercise.

Now, on account of the large contribution the first few principal factors, it is suggestive that these factors capture systematic risk. Furthermore, it is well-agreed that systematic risk is the main source of risk to which well-diversified portfolios are susceptible. It is thus clear that accurate measurements of these factors is economically worthwhile. Traditionally, factor measurement is achieved via a principal component analysis; our proposed contribution is within this tradition. Using nonparametric methods, we endeavour to obtain robust, accurate, real-time (intra-day) principal factor measurements in the form of eigen values and vectors of the spot covolatility matrix; from these measurements, we will study the time evolution and potential predictiveness of these principal factors with regards major market movements.

3.1 Forward rates and bond prices

We fix notation by going over the basics of continuous-time fixed-income pricing. For a detailed presentation see **?** or **?**. Let $P(t, \tau)$ denote the time-t price of a zero-coupon bond (bond) with maturity date τ and unit nominal value, where $0 \leq t \leq \tau < \infty$. Under the assumption that the bond price $P(t, \tau)$ is a smooth function of the maturity date τ, the forward rate $f(t, \tau)$ is given by the instantaneous rate of return on a loan contracted at time t to take effect starting at time τ; it is defined as:

$$f(t, \tau) = -\partial \log P(t, \tau)/\partial T. \tag{3.1}$$

The above definition yields a formula for the unit price of a bond at time t with maturity τ; this is achieved by integrating and then taking the exponential of both sides of (3.1) to yield:

$$P(t, \tau) = \exp\left(-\int_t^\tau f(t, s)\mathrm{d}s\right). \tag{3.2}$$

It is clear from (3.2) that the forward rate fully characterizes bond prices. Specializing to the case where the contract and effective dates coincides yields a characterization of the spot interest rate in terms of the forward rate function. Thus, denoting the spot interest rate by $r(t)$, we have

$$r(t) = f(t, t).$$

It is clear from the above equality that a discussion of the spot rate may be subsumed in a discussion of the forward rate. This is the point of departure of the model of fixed-income markets proposed by **?** (HJM). We highlight the salient point of the HJM perspective in the next section.

3.2 HJM model of fixed-income markets

The HJM model provides a model for the dynamics of the entire term structure. It is assumed that there is continuum of bonds indexed by maturity date τ taking values in some closed and bounded interval. Bond prices and forward rates are denoted by $P(t,\tau)$ and $f(t,\tau)$, respectively. The forward rate evolution in time is subject to uncertainty arising from m-dimensional standard Brownian vector $(W_1, \cdots, W_d)$, with $m < \infty$, defined on the filtered probability space $\{\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathrm{P}\}$, satisfying the usual conditions. To recap, while there are uncountably many bonds and consequently forward rates - one for each maturity date - there are only m risk factors affecting the time evolution of the entire term structure. Given a fixed maturity date τ, the forward rate $f(t,\tau)$ satisfies the following:

$$f(t,\tau) = f(0,\tau) + \int_0^t \mu(t,\tau)\mathrm{d}s + \int_0^t \sum_{j=1}^m \sigma_j(s,\tau)\mathrm{d}W_j(s), \tag{3.3}$$

where $0 \leq t \leq \tau$; the drift function is locally integrable and the volatility coefficient is locally square integrable. **?** showed that the no-arbitrage condition restricts the drift coefficient according to the following:

$$\mu(t,\tau) = \sum_{j=1}^m \sigma_j(t,\tau) \int_t^\tau \sigma_j(t,s)\mathrm{d}s.$$

So, the drift function is fully characterized by the volatility coefficient.

3.3 Realized Spectrum

Now, in practice we do not have market data for a continuum of maturities. Most fixed-income instruments are contracted for maturities that are multiples of 3 months, up to a mximum of 30 years; so, in practice we have about 120 maturities. Based on the empirical evidence (**??**), the principal risk factors, those accounting for the vast majority of the variability, say 98%, are less than 4; so, 120 maturities is probably more than we need to start teasing out what these principal risk factors are, and to monitor how they change or evolve over time. We start by fixing $1 < d < \infty$ maturities for which we have high-frequency data available. We let $[0,T]$ denote the observation interval. We assume the d forward rates are observed synchronously at equidistant intervals. Let $\Sigma(t)$ denote the time-t $d \times d$ spot co-volatility matrix of the d forward rates. Since, $\Sigma(t)$ is symmetric and positive definite, its spectral decomposition is given by:

$$\Sigma(t) = \sum_{i=1}^d \lambda_i(t)\{v_i(t) \otimes v_i(t)\}, \qquad t \geq 0,$$

where $\{\lambda(t)_i, v_i(t)\}$ are time-t eigen pairs, with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$, and $\otimes$ is the Kronecker product symbol. Given the empirical evidence aluded to previously, there is $d' << d$, i.e. there is a d' much smaller than d, such that

$$\Sigma(t) \approx \sum_{i=1}^{d'} \lambda_i(t)\{v_i(t) \otimes v_i(t)\}, \qquad t \geq 0.$$

Now, because Σ is not accessible, we propose using the Gabor frame co-volatility matrix estimator, $\hat{\Sigma}_n$, specified in Equation (1.20) of Chapter 1 to estimate Σ. Now given, $\hat{\Sigma}_n$, we propose to estimate the spectrum of Σ by means of the spectrum of $\hat{\Sigma}_n$; we denote this by $\{\hat{\lambda}_{n,j}, \hat{v}_{n,j}\}$ for j between 1 and d. Now for each j between 1 and d, let

$$\hat{\chi}_{n,j}(t_i) = \frac{\langle \hat{v}_{n,j}(t_i), \hat{v}_{n,j}(t_i - 1)\rangle}{\|\hat{v}_{n,j}(t_i)\|\|\hat{v}_{n,j}(t_i - 1)\|}, \qquad i = 1, \cdots, n, \tag{3.4}$$

where $\langle\cdot\rangle$ and $\|\cdot\|$ are, respectively, the Euclidean inner product and norm. The statistic $\hat{\chi}_{n,j}(t_i)$ is an empirical construct measuring the cosine between successive realizations of the eigenvectors $\hat{v}_{n,j}(t_i)$ and $\hat{v}_{n,j}(t_i - 1)$.

Now the empirical envidence (**?**) suggests that the principal or most important factor can account for a disproportionately portion of total variability; so, it makes sense to pay special attention to the most principal eigenvector. Since this eigenvector is a description of the composition of the most important or principal risk factor, we may view $\hat{\chi}_{n,1}(t_i)$ as telling us how the direction or composition of the principal risk factor changes over time. Knowledge of this time evolution, if tracked in real-time, could provide actionable information to traders about market direction and stability. Furthermore, **?** showed empirically that the Fourier-based counterpart of $\hat{\chi}_{n,1}(t_i)$ is fairly stable when it is business-as-usual in the market; in crisis periods this statistic can exhibit substantial fluctuations. Our aim is to replicate this empirical result using the robust Gabor frame co-volatility estimator and to answer a question that **?** have left unanswered, i.e. whether the realized velocity of the principal eigenvector is a leading indicator for major market movements? Intuitively, changes in volatility need not be reflected immediately in prices since the Brownian motions driving rates may offset the change in volatility for some time before they may begin to show in return data. We conjecture that an affirmative answer to this question may not be rejected with statistical significance. In the sequel we will conduct an empirical study to settle the matter. We propose the following simple model:

$$f(t_i, \tau) = \alpha + \sum_{j=0}^{q} \beta_j \hat{\chi}_{n,1}(t_{i-j}) + \sum_{j=0}^{p} \gamma_j f(t_{i-j}, \tau) + \eta_i, \tag{3.5}$$

where η_i is i.i.d error with constant variance and zero mean; p and q are positive integers; and α, $\{\beta_j\}_{j=0}^{q}$ and $\{\gamma_j\}_{j=0}^{p}$ are constants to be determined and tested for statistical significance.

3.4 Empirical study

3.5 Conclusion

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 (1.7) 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

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

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

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

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

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

Returning to (A.1), we see that

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

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

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

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

□

Next, we provide a proof of Proposition (1.1).

Proof. We begin with B_n^2, the bias component of the integrated mean square error. Using Itô's product formula (**?**, p. 257), we may express, using Itô's isometry theorem, the average deviation at time t as follows:

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

where

$$\begin{aligned}
B_{1,n}(t) &:= \sum_{(h,k)\in\Theta_n} g_{h,k}(t) \left\{ \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} \int_{t_i}^{t_{i+1}} \sigma^2(s) \mathrm{d}s - c_{h,k} \right\}, \\
B_{2,n}(t) &:= 2 \sum_{(h,k)\in\Theta_n} g_{h,k}(t) \left\{ \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} \mathrm{E}\left[\int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) \mu(s, X_s) \mathrm{d}s \right] \right\}, \\
B_{3,n}(t) &:= - \sum_{(j,k)\notin\Theta_n} g_{h,k}(t) c_{h,k}.
\end{aligned} \tag{A.5}$$

The first two components, $B_{1,n}(t)$ and $B_{2,n}(t)$, result from the fact that X is being observed discretely; whereas $B_{3,n}$ results because only a finite number of the frame elements are being used in the approximation. We refer to Theorem 4.1 in **?** for an estimate of $B_{3,n}$:

$$B_{3,n} = O(\omega_{\bar{\sigma}^2}(1/H_n)\log H_n).$$

So, the smoother the volatility coefficient the smaller the number of frame elements needed to obtain a decent approximation. We obtain a bound on $B_{1,n}$ by noting that

$$\begin{aligned}\sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} \int_{t_i}^{t_{i+1}} \sigma^2(s)\mathrm{d}s - c_{h,k} &= \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} \sigma^2(s)\{\overline{\tilde{g}_{h,k}(t_i)} - \overline{\tilde{g}_{h,k}(s)}\}\mathrm{d}s \\ &\leq C_\beta \omega_{\tilde{g}_{h,k}}(\Delta_n),\end{aligned}$$

where $C_\beta = \int_0^T \sigma^2(s)\mathrm{d}s < \infty$. It follows from Lemma (1.2) that

$$B_{1,n}(t) \leq C_B H_n \omega_g(\Delta_n)$$

with $C_B = 2(2K_0+1)g^*\tilde{g}^* C_\beta C_{\tilde{g}}$. Next, we estimate $B_{2,n}(t)$. Note that

$$X_s - X_{t_i} = \int_{t_i}^{s} \mu(u, X_u)\mathrm{d}u + \int_{t_i}^{s} \sigma(u)\mathrm{d}W_u.$$

So, we may write

$$\begin{aligned}&\mathrm{E}\left[\int_{t_i}^{t_{i+1}} (X_s - X_{t_i})\mu(s, X_s)\mathrm{d}s\right] \\ &\quad = \mathrm{E}\left[\int_{t_i}^{t_{i+1}} \left(\int_{t_i}^{s} \mu(u, X_u)\mathrm{d}u\right) \mu(s, X_s)\mathrm{d}s\right] \\ &\qquad + \mathrm{E}\left[\int_{t_i}^{t_{i+1}} \left(\int_{t_i}^{s} \sigma(u)\mathrm{d}W_u\right) \mu(s, X_s)\mathrm{d}s\right] \\ &\quad =: \beta_{2,i} + \beta_{3,i}.\end{aligned}$$

By Fubinni's theorem, the Cauchy-Schwarz inequality, Itô's isometry theorem, and the linear growth condition on the drift, we have

$$\begin{aligned}\beta_{3,i} &\leq C_T \int_{t_i}^{t_{i+1}} \left(\int_{t_i}^{s} \sigma^2(u)\mathrm{d}u\right)^{1/2} \mathrm{E}[(1+|X_s|)^2]^{1/2}\mathrm{d}s \\ &\leq C_3 \Delta_n^{3/2}. \qquad \text{(A.6)}\end{aligned}$$

Similarly, it may be verified that $\beta_{2,i} \le C_2\Delta_n^2$. Now we may write

$$\begin{aligned} B_{2,n}(t) &\le \Delta_n^{1/2} \sum_{(h,k)\in\Theta_n} g_{h,k}(t) \left\{ \sum_{i=0}^{n-1} \overline{\tilde{g}_{h,k}(t_i)} C_4 \Delta_n \right\} \\ &\le ((2K_0+1)C_4 g^* \tilde{g}^* T)(2H_n+1)\Delta_n^{1/2} \\ &= O(H_n \Delta_n^{1/2}). \end{aligned} \tag{A.7}$$

Thus the square bias is bounded as follows:

$$\begin{aligned} B_n^2 &= \int_R (B_{1,n}(t) + B_{2,n}(t) + B_{3,n}(t))^2 \lambda(t) \mathrm{d}t \\ &= O(H_n^2 \Delta_n + H_n^2 \omega_g^2(\Delta_n) + \omega_{\tilde{\sigma}^2}^2(1/H_n) \log^2 H_n). \end{aligned} \tag{A.8}$$

Next, we obtain a bound for the variance term V_n. Recall from (1.17) that

$$V_n := \int_{\mathbb{R}} \mathrm{E}[\{\hat{\sigma}_n^2(t) - \mathrm{E}[\hat{\sigma}_n^2(t)]\}^2] \lambda(t) \mathrm{d}t$$

and note that V_n may be expressed as follows:

$$V_n = V_{1,n} + V_{2,n}$$

with

$$\begin{aligned} V_{1,n} &:= \sum_{(h,k)\in\Theta_n} \mathrm{var}[\hat{c}_{h,k}] \gamma_{h,k}^2, \text{ and} \\ V_{2,n} &:= \sum_{\substack{(h,k),(h',k')\in\Theta_n \\ (h,k)\ne(h',k')}} \mathrm{cov}[\hat{c}_{h,k}, \hat{c}_{h',k'}] \gamma_{h,k} \gamma_{h',k'}, \end{aligned}$$

where $\gamma_{h,k} := \int_0^T g_{h,k}(t)\mathrm{d}t$. We start with $V_{1,n}$. If we set

$$\begin{aligned} Y_i &:= \left(\int_{t_i}^{t_{i+1}} \sigma(s) \mathrm{d}W_s \right)^2 \qquad \text{and} \\ Z_i &:= \left(\int_{t_i}^{t_{i+1}} \mu(s, X_s) \mathrm{d}s \right)^2 + 2 \left(\int_{t_i}^{t_{i+1}} \mu(s, X_s) \mathrm{d}s \right) \left(\int_{t_i}^{t_{i+1}} \sigma(s) \mathrm{d}W_s \right), \end{aligned}$$

then we may write

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

Furthermore, setting

$$\alpha_{1,i} := \sum_{(h,k)\in\Theta_n} g_{h,k}^2(t_i)\gamma_{h,k}^2,$$

allows us to write

$$V_{1,n} := \sum_{i=0}^{n-1} \alpha_{1,i}(\mathrm{var}[Y_i] + \mathrm{var}[Z_i] + 2\mathrm{cov}[Y_i, Z_i]).$$

Using the Cauchy-Schwarz inequality, we obtain

$$\mathrm{var}[Y_i] = \left(\int_{t_i}^{t_{i+1}} \sigma^2(t_i)\mathrm{d}s\right)^2 = O(\Delta_n^2); \tag{A.9}$$

$$\mathrm{var}[Z_i] \le \mathrm{E}[Z_i^2] = O(\Delta_n^3); \tag{A.10}$$

$$\mathrm{cov}[Y_i, Z_i] \le (\mathrm{var}[Z_i]\mathrm{var}[Y_i])^{1/2} = O(\Delta_n^{5/2}). \tag{A.11}$$

Since $\alpha_{1,i} = O(H_n)$, we conclude that $V_{1,n} = O(H_n\Delta_n)$. It is clear that $V_{2,n}$ may be bounded in a similar fashion. Indeed, let

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

then we may write

$$V_{2,n} := \sum_{i=0}^{n-1} \alpha_{2,i}(\mathrm{var}[Y_i] + \mathrm{var}[Z_i] + 2\mathrm{cov}[Y_i, Z_i]).$$

Since $\alpha_{2,i} = O(H_n^2)$, it follows that $V_{2,n} := O(H_n^2\Delta_n)$.

$\square$