Statistica Sinica **20** (2010), 1011-1024

SHARP MINIMAX ESTIMATION OF THE VARIANCE OF BROWNIAN MOTION CORRUPTED WITH GAUSSIAN NOISE

T. Tony Cai[1], A. Munk[2] and J. Schmidt-Hieber[2]

[1]*University of Pennsylvania and* [2]*Universität Göttingen*

Abstract: Let W_t be a Brownian motion with $\epsilon_{in} \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$, $i = 1, \ldots, n$, independent of W_t. $\sigma, \tau > 0$ are real, unknown parameters. Suppose we observe $Y_{i,n} = \sigma W_{i/n} + \tau \epsilon_{in}$. In this paper we establish sharp estimators for σ^2 and τ^2 in minimax sense, i.e. they attain the minimax constant asymptotically. A short and direct proof for the minimax lower bound is given. These estimators are based on a spectral decomposition of the underlying process $Y_{i,n}$ and can be computed explicitly taking $O(n \log n)$ operations. We outline how these estimators can be generalized from Brownian motion to processes with independent increments. Further we show that the spectral estimators presented are asymptotically normal.

Key words and phrases: Asymptotic normality, Brownian motion, deconvolution, minimax, spectral estimators, statistical inverse problems, variance estimation, oracle estimator.

1. Introduction

Suppose we observe

$$Y_{i,n} = \sigma W_{i/n} + \tau \epsilon_{in}, \tag{1.1}$$

where W_t, $t \in [0,1]$ denotes a standard Brownian motion and $\epsilon_{in} \stackrel{i.i.d.}{\sim} \mathcal{N}(0,1)$. W_t and the ϵ_{in} are assumed to be independent processes. We can think of the observed process as a linear combination of $W_{i/n}$ and ϵ_{in}, weighted with σ and τ, respectively. In this paper we analyze estimation of σ and τ from the viewpoint of a statistical inverse problem. From this perspective the process of interest, $\sigma W_{i/n}$, is additively corrupted by noise, $\tau\epsilon_{in}$, which reveals this problem as a particular deconvolution problem. In deconvolution it is often convenient to work in the spectral domain where convolution transforms to multiplication (see e.g., Mair and Ruymgaart (1996) for an early reference), and in this paper we adopt this point of view.

Model (1.1) has received much attention because it is the simplest model of high frequency financial data incorporating market microstructure noise, see e.g., Barndorff-Nielsen et al. (2008), Aït-Sahalia, Mykland, and Zhang (2005), or Huang, Liu, and Yu (2007) for further reading and more references. The aim is to estimate the parameters σ and τ. It is well known that σ can be estimated at a $n^{-1/4}$-rate and τ at a $n^{-1/2}$-rate, see e.g., Gloter and Jacod (2001a). In fact these are the minimax rates of convergence, i.e., the best possible rates of convergence of any estimator for τ and σ, respectively (see Tsybakov (2004) for a precise definition of a minimax rate). It is well-known that the Cramer-Rao lower bound is $2\tau^4 n^{-1}$ and $8\tau\sigma^3 n^{-1/2} + o(n^{-1/2})$ for estimation of τ^2 and σ^2, respectively (Gloter and Jacod (2001a,b)). These authors consider the more general model

$$\begin{aligned} Y_{in} &= X_{i/n} + \tau_n \epsilon_{in}, \quad \text{for} \quad i = 1, \ldots, n, \\ X_t &= \int_0^t \sigma(\theta, s) dW_s. \end{aligned}$$

Here, $\sigma(.,.)$ is a function satisfying some smoothness and identifiability assumptions, $\theta \in \Theta$ is the unknown parameter, and $\Theta \subset \mathbb{R}$ is compact. With constant τ and $\sigma(\theta, s)$ independent of s, we have model (1.1). If τ is known (although this is not a serious restriction), the authors obtain an estimator, based on minimizing a contrast functional, that is sharp with respect to Fisher information. They even establish LAN for their model, implying that Cramer-Rao lower bounds also provide the optimal constants in a minimax sense. In this paper we give an elementary and short proof for the sharp minimax lower bounds for estimation of τ^2 and σ^2 in model (1.1) without using LAN.

The common estimator for σ^2 in model (1.1) is the maximum likelihood estimator. It is asymptotically Cramer-Rao-efficient (Stein (1987) or Aït-Sahalia, Mykland, and Zhang (2005)) and hence a sharp minimax estimator. This estimator, however, requires numerical maximization of the likelihood function, which can be quite involved due to a flat likelihood function apart from its maximum. Further, the likelihood function is not convex albeit unimodal. Therefore, a good starting value for a Newton-type iteration (or any other optimization method) is of some importance. Hence, the second goal of this paper is to construct explicitly computable estimators that are minimax sharp as well. This is easy for τ^2, but not obvious for σ^2. In order to do this we transform the problem to the spectral domain. We split the spectrum of the covariance in an appropriate way and mimic the linear oracle estimator. The resulting estimator is explicitly computable and only depends on the precise spectral information of the covariance of the data. Hence no numerical minimization step is involved.

We believe that our spectral approach, combined with the viewpoint from nonparametric regression, sheds some new light on this problem and various important facts become immediately visible. For example we see that only $\sqrt{n}$

data in the transformed model can be used for efficient estimation of σ^2, immediately revealing $n^{-1/4}$ as the minimax rate, again. In Section 4 we also indicate how these estimators can be extended to more general models and that they are robust. For simplicity we restrict ourselves in this paper to model (1.1).

We briefly mention further related work on this subject. Another estimator was introduced in Aït-Sahalia, Mykland, and Zhang (2005). It does not require known τ and is asymptotically sharp in model (1.1). However, it is necessary to minimize a complicated expression in order to calculate the estimator.

In more general models, such as in Barndorff-Nielsen et al. (2008), Zhang, Mykland, and Aït-Sahalia (2005) and Jacod et al. (2009), σ is a smooth function (and possibly random) and $\int \sigma_s^2 ds$ (or as in Podolskij and Vetter (2009) $\int \sigma_s^p ds$) is to estimated. In this case the asymptotic variance for constant σ can be evaluated. So far, there is no known estimator that is efficient with respect to this case. In fact, to our knowledge, not even a sharp Cramer-Rao bound is known. The best constant $8.01\tau\sigma^3$ in this context is attained by the so called $Tuckey - Hanning_\infty$ estimator (Barndorff-Nielsen et al. (2008)), but to achieve this bound requires optimal choice of a bandwidth parameter depending on the unknown quantities σ and τ itself.

Another interesting generalization was considered by Gloter and Hoffmann (2007). These authors replaced the Brownian motion in model (1.1) by a fractional Brownian motion with unknown Hurst index H, $1/2 < H < 1$, and proved minimax rates for estimation of σ^2 under quite general assumptions on the noise term.

The paper is organized as follows. In Section 2 we present the spectral estimators and prove that they are sharp with respect to the optimal constants in minimax sense. Computational aspects are briefly discussed in Section 3, and we investigate robustness against violations of normality and indicate the extension to more general processes with independent increments in Section 4. To keep the work more readable, all technical proofs are deferred to the supplementary material [*SM*] (`http://www.stat.sinica.edu.tw/statistica`) that contains various additional lemmas, enumerated by S1.1, S1.2, ... Some further technicalities are postponed to Section S2.

Notation: Throughout this paper we suppress the index n and, for two sequences $(a_n)_n$ and $(b_n)_n$, we use the notation $a_n \ll b_n$ if $a_n = o(b_n)$.

2. Estimators and Sharp Minimax Bounds

Let

$$K := K_n := \text{Cov}\left[W_{i/n}, W_{j/n}\right]_{i,j=1,\ldots,n} = \left(\frac{i}{n} \wedge \frac{j}{n}\right)_{i,j=1,\ldots,n}.$$

Then $Y := (Y_{1,n}, \ldots, Y_{n,n})^t \sim \mathcal{N}(0, \sigma^2 K + \tau^2 I_n)$. We can write $K = D\Lambda D^t$, where

$$D := D_n := \left(\sqrt{\frac{4}{2n+1}} \sin\left(\frac{(2j-1)\, i\pi}{2n+1} \right) \right)_{i,j=1,\ldots,n} \tag{2.1}$$

is an orthogonal matrix, $D^t D = I_n$, and Λ is a diagonal matrix with diagonal elements

$$\lambda_i := \left[4n \sin^2\left(\frac{2i-1}{4n+2}\pi \right) \right]^{-1} = \frac{1}{n} \operatorname{Dir}_n^2 \left(\frac{(2i-1)\, \pi}{2n+1} \right), \tag{2.2}$$

where $\operatorname{Dir}_n(x)$ denotes the Dirichlet kernel $\operatorname{Dir}_n(x) = 1/2 + \sum_{i=1}^n \cos(i\pi x)$. This can be derived similarly as in Durbin and Knott (1972), and is based on solving a second order difference equation under given boundary conditions. Let $Z = (Z_1, \ldots, Z_n)^t = D^t Y$. Then,

$$Z_i \overset{ind.}{\sim} \mathcal{N}\left(0, \sigma^2 \lambda_i + \tau^2\right), \quad i = 1, \ldots, n. \tag{2.3}$$

Hence Z_i^2, $i = 1, \ldots, n$, are independent as well and have a scaled χ_1^2-distribution with expectation $\mathrm{E}\left(Z_i^2\right) = \sigma^2 \lambda_i + \tau^2$ and variance $\operatorname{Var}\left(Z_i^2\right) = 2\left(\sigma^2 \lambda_i + \tau^2\right)^2$. We work with the Z_i's from now on. Moreover, they form a sufficient statistic for model (1.1).

We give here a heuristic argument that from this representation the difficulty of estimating σ^2 becomes obvious. Because $\lambda_i \asymp n/i^2$ uniformly in $i = 1, \ldots, n$ (for a precise statement see Lemma B.1 in [*SM*]), only the variables Z_i^2/λ_i with $i = i(n) = O(\sqrt{n})$ have asymptotically bounded variances. Here we mean by $a_n \asymp b_n$ that $a_n = O(b_n)$ and $b_n = O(a_n)$. In contrast, for estimation of τ^2 only the "last" $n - \sqrt{n}$ variables Z_i^2 can be used. This observation is at the heart of our subsequent considerations.

2.1. Estimation of τ^2

First we consider the problem of estimating τ^2. There exist many alternative estimators for τ^2. For instance, scaled quadratic variation would work. However, in order to derive a sharp estimator of σ^2 we need some specific preliminary estimator of τ^2 that is independent of the random variables $Z_1, \ldots, Z_m$ for some $0 < m < n$. This motivates us to set

$$\hat{\tau}_m^2 := \frac{1}{n-m} \sum_{i=m+1}^{n} Z_i^2, \qquad 1 < m < n. \tag{2.4}$$

Theorem 1. *Assume model* (1.1) *holds, and let* $m = m(n)$ *be a sequence such that* $m/\sqrt{n} \to \infty$ *and* $m/n \to 0$ *for* $n \to \infty$. *Let further the estimator* $\hat{\tau}_m^2$ *of* τ^2 *be given in* (2.4). *Then*

(i) $\sup_{\sigma,\tau>0} \sigma^{-2} \left|\mathrm{E}\left(\hat{\tau}_m^2\right) - \tau^2\right| = o\left(n^{-1/2}\right)$,

(ii) $n^{1/2}(\hat{\tau}_m^2 - \tau^2) \xrightarrow{L} \mathcal{N}(0,\ 2\tau^4)$,

(iii) *and for any* $\epsilon > 0$,

$$\sup_{\sigma,\tau>\epsilon} (\sigma\tau)^{-4} \left|n\mathrm{Var}\left(\hat{\tau}_m^2\right) - 2\tau^4\right| = o\left(1\right), \sup_{\sigma,\tau>\epsilon} (\sigma\tau)^{-4}\, \mathrm{Var}\left(\hat{\tau}_m^2\right) = O\left(n^{-1}\right).$$

Proof. (i) Note that

$$\mathrm{E}\left(\hat{\tau}_m^2\right) = \frac{1}{n-m} \sum_{i=m+1}^{n} \left(\sigma^2\lambda_i + \tau^2\right) = \tau^2 + \sigma^2 \frac{1}{n-m} \sum_{i=m+1}^{n} \lambda_i.$$

By Lemma B.1 and the choice of m, (i) follows.
(iii) It holds that

$$\mathrm{Var}\left(\hat{\tau}_m^2\right) = \frac{2}{(n-m)^2} \sum_{i=m+1}^{n} \left(\tau^4 + 2\tau^2\sigma^2\lambda_i + \sigma^4\lambda_i^2\right)$$

and hence

$$\sup_{\sigma,\tau>\epsilon} \frac{1}{\sigma^4} \left|\frac{n}{2\tau^4}\mathrm{Var}\left(\hat{\tau}_m^2\right) - 1\right|$$
$$= \sup_{\sigma,\tau>\epsilon} \left|\frac{m}{\sigma^4\left(n-m\right)} + \frac{2n}{\sigma^2\tau^2\left(n-m\right)^2} \sum_{i=m+1}^{n} \lambda_i + \frac{n}{\tau^4\left(n-m\right)^2} \sum_{i=m+1}^{n} \lambda_i^2\right| = o\left(1\right).$$

The second statement in (iii) follows by triangle inequality.
(ii) Note that the estimator $\hat{\tau}_m^2$ can be written as

$$\hat{\tau}_m^2 = \sum_{i=m+1}^{n} \frac{\sqrt{2}(\sigma^2\lambda_i + \tau^2)}{n-m} \cdot \frac{X_i - 1}{\sqrt{2}} + \mathrm{E}\left(\hat{\tau}_m^2\right),$$

where $X_i \overset{i.i.d.}{\sim} \chi_1^2$. Set $c_i = \sqrt{2}(\sigma^2\lambda_i + \tau^2)n^{1/2}/\left(n-m\right)$ and $R_i = \left(X_i - 1\right)/\sqrt{2}$. Then the R_i are i.i.d. with mean zero and unit variance and

$$n^{1/2}(\hat{\tau}^2 - \tau^2) = \sum_{i=m+1}^{n} c_i R_i + n^{1/2}(\mathrm{E}\left(\hat{\tau}^2\right) - \tau^2).$$

Due to (i) and (iii), and since $\max_{i=m+1,\ldots,n} |c_i| = c_{m+1} \to 0$ and $\sum_{i=m+1}^{n} c_i^2 \to 2\tau^4 < \infty$ for $n \to \infty$, (ii) follows by using the CLT under Noether condition (see Theorem C.1 in [*SM*]).

Note that the theorem implies that

$$\begin{aligned}&\sup_{\sigma,\tau>\epsilon}(\sigma\tau)^{-4}\left|n\,\mathrm{MSE}\left(\hat{\tau}_m^2\right)-2\tau^4\right|\\&\leq\sup_{\sigma,\tau>\epsilon}(\sigma\tau)^{-4}\,n\,\mathrm{Bias}^2\left(\hat{\tau}_m^2\right)+\sup_{\sigma,\tau>\epsilon}(\sigma\tau)^{-4}\left|n\mathrm{Var}\left(\hat{\tau}_m^2\right)-2\tau^4\right|=o\left(1\right),\end{aligned}\tag{2.5}$$

and similarly

$$\sup_{\sigma,\tau>\epsilon}(\sigma\tau)^{-4}\,\mathrm{MSE}\left(\hat{\tau}_m^2\right)=O\left(n^{-1}\right).\tag{2.6}$$

Moreover, the constant $2\tau^4$ is sharp. More precisely, we have the following result.

Theorem 2. *Assume model* (1.1). *Then*

(i) *for any estimator* $\hat{\tau}^2$ *and any* $\sigma\geq 0$,

$$\underline{\lim}_{n\to\infty}\inf_{\hat{\tau}^2}\sup_{\tau>\epsilon}\frac{1}{2\tau^4}\mathrm{E}\left(n\left(\hat{\tau}^2-\tau^2\right)^2\right)\geq 1,\tag{2.7}$$

(ii) *for any* $\epsilon>0$,

$$\lim_n\inf_{\hat{\tau}^2}\sup_{\sigma,\tau>\epsilon}(\sigma\tau)^{-4}\left(\mathrm{E}\left(n\left(\hat{\tau}^2-\tau^2\right)^2\right)-2\tau^4\right)=0,$$

(iii) *for any* $0<\epsilon<c<\infty$,

$$\lim_n\inf_{\hat{\tau}^2}\sup_{\sigma,\tau>\epsilon,\ \sigma<c}\frac{1}{2\tau^4}\mathrm{E}\left(n\left(\hat{\tau}^2-\tau^2\right)^2\right)=1.$$

Proof. (i) We prove this by the Information Inequality Method (see Lehmann (1983, p.266)). Note that $Z_i\overset{ind.}{\sim}\mathcal{N}(0,\sigma^2\lambda_i+\tau^2)$, $i=1,\ldots,n$, can be written as $Z_i=U_i+V_i$, where $U_i\sim\mathcal{N}(0,\sigma^2\lambda_i)$, $V_i\sim\mathcal{N}(0,\tau^2)$, and $\{U_i,\ V_i,\ i=1,\ldots,n\}$ are mutually independent. Estimating τ^2 based on $Z_1,\ldots,Z_n$ is not easier than estimating τ^2 based on $V_1,\ldots,V_n$ since Z_i can be generated from V_i by adding random noise U_i and is thus less informative than V_i. Hence we may assume $\sigma=0$. The Fisher information for τ^2 is then $I\left(\tau^2\right)=n/\left(2\tau^4\right)$. Assume that (i) does not hold. Then there exists an estimator $\hat{\tau}^2$ and a subsequence $\{n_k\}$ such that

$$\lim_{k\to\infty}\sup_{\tau>\epsilon}\frac{1}{2\tau^4}\mathrm{E}\left(n_k(\hat{\tau}^2-\tau^2)^2\right)\leq(1-2\delta)^2$$

for some $0<\delta<1/2$. Hence there exists k_1 such that, for all $k\geq k_1$,

$$\mathrm{E}\,(\hat{\tau}^2-\tau^2)^2\leq(1-\delta)^2\,2\tau^4n_k^{-1},\quad\text{for all }\tau>\epsilon$$

and $n_k > 50/\delta^2$. For such an n_k, the Cramer-Rao information inequality yields

$$b^2\left(\tau^2\right) + \frac{\left(1 + b'\left(\tau^2\right)\right)^2}{I_{n_k}\left(\tau^2\right)} \leq (1-\delta)^2\, 2\tau^4 n_k^{-1} \quad \text{for all } \tau > \epsilon,$$

where $b\left(\tau^2\right)$ denotes the bias of $\hat{\tau}^2$. This implies both

$$b^2\left(\tau^2\right) \leq 2\tau^4 n_k^{-1} \quad \text{and} \quad b'\left(\tau^2\right) \leq -\delta, \quad \text{for all } \tau > \epsilon. \tag{2.8}$$

Integrating the second inequality yields $b\left(\tau^2\right) \leq -\delta\left(\tau^2 - 2\epsilon^2\right) + b\left(2\epsilon^2\right)$ for $\tau^2 \geq 2\epsilon^2$. This gives the contradiction

$$b\left(3\epsilon^2\right) \leq -\delta\epsilon^2 + b\left(2\epsilon^2\right) \leq -\delta\epsilon^2 + \sqrt{2}\left(2\epsilon^2\right) n_k^{-1/2} < -\sqrt{2}\left(3\epsilon^2\right) n_k^{-1/2} \leq b\left(3\epsilon^2\right),$$

where we used (2.8) in the second and last inequalities and $n_k > 50/\delta^2$ in the third one. Hence (i) holds.

In order to show (ii), note that by (i),

$$\begin{aligned}&\varliminf_n \inf_{\hat{\tau}^2} \sup_{\sigma,\tau>\epsilon} (\sigma\tau)^{-4}\left(\mathrm{E}\left(n\left(\hat{\tau}^2 - \tau^2\right)^2\right) - 2\tau^4\right)\\ &\qquad \geq 2\,(2\epsilon)^{-4} \varliminf_n \inf_{\hat{\tau}^2} \sup_{\tau>\epsilon,\ \sigma=2\epsilon} \left(\frac{1}{2\tau^4}\mathrm{E}\left(n\left(\hat{\tau}^2 - \tau^2\right)^2\right) - 1\right) \geq 0.\end{aligned}$$

On the other hand, we have due to (2.5),

$$\begin{aligned}&\varlimsup_n \inf_{\hat{\tau}^2} \sup_{\sigma,\tau>\epsilon} (\sigma\tau)^{-4}\left(\mathrm{E}\left(n\left(\hat{\tau}^2 - \tau^2\right)^2\right) - 2\tau^4\right)\\ &\qquad \leq \lim_n \sup_{\sigma,\tau>\epsilon} (\sigma\tau)^{-4}\left(n\,\mathrm{MSE}\left(\hat{\tau}_m^2\right) - 2\tau^4\right) = 0.\end{aligned}$$

Finally, (iii) follows from (ii) and (i) .

2.2. Estimation of σ^2

We now turn to the estimation of σ^2. Define the linear oracle "estimator"

$$\hat{\sigma}^2_{\text{oracle}} := C_n^{-1} \sum_{i=1}^n \frac{\lambda_i}{(\sigma^2\lambda_i + \tau^2)^2}(Z_i^2 - \tau^2), \tag{2.9}$$

where

$$C_n := \sum_{i=1}^n \frac{\lambda_i^2}{(\sigma^2\lambda_i + \tau^2)^2}. \tag{2.10}$$

It follows from Lemma A.1 that $\hat\sigma^2_{\mathrm{oracle}}$ attains the risk of $2C_n^{-1} = 8\tau\sigma^3 n^{-1/2}(1+o(1))$. Note that the oracle "estimator" $\hat\sigma^2_{\mathrm{oracle}}$ depends on the unknown parameters τ^2 and σ^2 and is thus not a statistical estimator.

We construct below a data-driven estimator of σ^2 that mimics the performance of the oracle. For $1 < k < m < n$, set

$$\bar\sigma^2_{k,m} = \frac{1}{k}\sum_{i=1}^{k} \lambda_i^{-1}\left(Z_i^2 - \hat\tau^2_m\right). \tag{2.11}$$

Then

$$\mathrm{E}\left(\bar\sigma^2_{k,m}\right) = \sigma^2 + \left(\tau^2 - \mathrm{E}\left(\hat\tau^2_m\right)\right)\frac{1}{k}\sum_{i=1}^{k}\lambda_i^{-1}, \tag{2.12}$$

$$\mathrm{Var}\left(\bar\sigma^2_{k,m}\right) = \frac{1}{k^2}\sum_{i=1}^{k} 2(\sigma^2+\tau^2\lambda_i^{-1})^2 + \mathrm{Var}\left(\hat\tau^2_m\right)\frac{1}{k^2}\left(\sum_{i=1}^{k}\lambda_i^{-1}\right)^2. \tag{2.13}$$

The idea is to divide the observations into three parts, using the observations $Z_1,\dots,Z_k$ and $Z_{m+1},\dots,Z_n$ to obtain estimates $\bar\sigma^2_{k,m}$ of σ^2 and $\hat\tau^2_m$ of τ^2, and using the middle part to construct an estimator $\hat\sigma^2$ by plugging $\bar\sigma^2_{k,m}$ and $\hat\tau^2_m$ in the oracle estimator of σ^2. The advantage of this procedure is that the estimates $\bar\sigma^2_{k,m}$ and $\hat\tau^2$ are independent of the observations used for estimating σ^2 in $\hat\sigma^2$. For $1 \le k \ll n^{1/2} \ll m \le n$, define in analogy to (2.9) the linear oracle estimator based on $Z_{k+1},\dots,Z_m$ by

$$\tilde\sigma^2_{k,m} := A_n^{-1}\sum_{i=k+1}^{m}\frac{\lambda_i}{(\sigma^2\lambda_i+\tau^2)^2}(Z_i^2-\tau^2),$$

where $A_n := A_n(k,m) := \sum_{i=k+1}^{m}\left(\sigma^2+\tau^2\lambda_i^{-1}\right)^{-2}$. Let $\hat\tau^2_m$ and $\bar\sigma^2_{k,m}$ be given as in (2.4) and (2.11), respectively and set $\hat A_n := \hat A_n(k,m) := \sum_{i=k+1}^{m}(\bar\sigma^2_{k,m} + \hat\tau^2_m\lambda_i^{-1})^{-2}$. Then for $1 \le k \ll n^{1/2} \ll m \le n$, define the estimator of σ^2 by

$$\hat\sigma^2 := \hat\sigma^2_{k,m} := \hat A_n^{-1}\sum_{i=k+1}^{m}\frac{\lambda_i}{(\bar\sigma^2_{k,m}\lambda_i+\hat\tau^2_m)^2}(Z_i^2-\hat\tau^2_m). \tag{2.14}$$

Theorem 3. *Let $k = [n^{1/2-b}]$ and $m = [n^{1/2+b}]$ with $0 < b < 1/20$. Let the estimator $\hat\sigma^2$ of σ^2 be given in* (2.14). *Then for any $\epsilon > 0$*

(i) $\sup_{\sigma,\tau>\epsilon}(\sigma\tau)^{-2}\left|\mathrm{E}\left(\hat\sigma^2-\sigma^2\right)\right| = o\left(n^{-1/4}\right)$,

(ii) $\sup_{\sigma,\tau>\epsilon}(\sigma\tau)^{-8}\left|\mathrm{Var}\left(\hat\sigma^2\right) - 8\tau\sigma^3 n^{-1/2}\right| = o\left(n^{-1/2}\right)$,

(iii) $n^{1/4}(\hat\sigma^2-\sigma^2) \xrightarrow{L} \mathcal{N}(0,\ 8\tau\sigma^3)$.

Proof. For ease of notation, we write $\tilde{\sigma}^2, \bar{\sigma}^2$ and $\hat{\tau}^2$ instead of $\tilde{\sigma}^2_{k,m}, \bar{\sigma}^2_{k,m}$ and $\hat{\tau}^2_m$, respectively. Introduce the oracle estimator

$$\hat{\sigma}^2_\tau := \hat{A}_n^{-1} \sum_{i=k+1}^{m} \frac{\lambda_i}{(\bar{\sigma}^2\lambda_i + \hat{\tau}^2)^2}(Z_i^2 - \tau^2).$$

(i) By construction we have that $\bar{\sigma}^2$ and Z_i as well as $\hat{\tau}^2$ and Z_i for $i = k+1, \dots, m$, are independent. Hence $\mathrm{E}\left(\hat{\sigma}^2_\tau\right) = \sigma^2$ and, due to

$$\left|\hat{\sigma}^2 - \hat{\sigma}^2_\tau\right| = \hat{A}_n^{-1} \sum_{i=k+1}^{m} \frac{\lambda_i}{\left(\bar{\sigma}^2\lambda_i + \hat{\tau}^2\right)^2} \left|\hat{\tau}^2 - \tau^2\right| \leq \lambda_m^{-1} \left|\hat{\tau}^2 - \tau^2\right|, \tag{2.15}$$

$$\sup_{\sigma,\tau>\epsilon} \frac{1}{\sigma^2\tau^2} \left|\mathrm{E}\left(\hat{\sigma}^2 - \sigma^2\right)\right| \leq \sup_{\sigma,\tau>\epsilon} \frac{1}{\sigma^2\tau^2} \mathrm{E}\left(\left|\hat{\sigma}^2 - \hat{\sigma}^2_\tau\right|\right) \leq \sup_{\sigma,\tau>\epsilon} \frac{\lambda_m^{-1}}{\sigma^2\tau^2} \mathrm{MSE}^{1/2}\left(\hat{\tau}^2\right).$$

By (2.6) this gives (i) .

(ii) We have the decomposition $\hat{\sigma}^2 - \tilde{\sigma}^2 = \left(\hat{\sigma}^2 - \hat{\sigma}^2_\tau\right) + \left(\hat{\sigma}^2_\tau - \tilde{\sigma}^2\right)$. In order to show that $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ have the same asymptotic variances, we bound the variance of the differences $\hat{\sigma}^2 - \hat{\sigma}^2_\tau$ and $\hat{\sigma}^2_\tau - \tilde{\sigma}^2$. Therefore, note that

$$\mathrm{Var}\left(\hat{\sigma}^2 - \tilde{\sigma}^2\right) \leq 2\mathrm{Var}\left(\hat{\sigma}^2 - \hat{\sigma}^2_\tau\right) + 2\mathrm{Var}\left(\hat{\sigma}^2_\tau - \tilde{\sigma}^2\right). \tag{2.16}$$

First we see that by (2.15),

$$\mathrm{Var}\left(\hat{\sigma}^2 - \hat{\sigma}^2_\tau\right) \leq \mathrm{E}\left(\hat{\sigma}^2 - \hat{\sigma}^2_\tau\right)^2 \leq \lambda_m^{-2}\,\mathrm{MSE}\left(\hat{\tau}^2\right). \tag{2.17}$$

Write $Z_i^2 = \left(\sigma^2\lambda_i + \tau^2\right)U_i$, where $U_i \sim \chi^2_1$, i.i.d. Let $w_{in} = A_n^{-1}\lambda_i/(\sigma^2\lambda_i + \tau^2)$ and $\hat{w}_{in} = \hat{A}_n^{-1}\lambda_i\left(\sigma^2\lambda_i + \tau^2\right)/(\bar{\sigma}^2\lambda_i + \hat{\tau}^2)^2$. Then

$$\hat{\sigma}^2_\tau - \tilde{\sigma}^2 = \hat{\sigma}^2_\tau - \sigma^2 + \sigma^2 - \tilde{\sigma}^2 = \sum_{i=k+1}^{m} \left(\hat{w}_{in} - w_{in}\right)\left(U_i - 1\right).$$

By construction we have that $\hat{w}_{in}$ and U_i, $i = k+1, \dots, m$ are independent. Therefore $\mathrm{E}\left(\hat{\sigma}^2_\tau - \tilde{\sigma}^2\right) = 0$ and, because $\mathrm{E}\left(\sum_{i=k+1}^m w_{in}\hat{w}_{in}\right) = A_n^{-1}$,

$$\mathrm{Var}\left(\hat{\sigma}^2_\tau - \tilde{\sigma}^2\right) = 2\mathrm{E}\left(\sum_{i=k+1}^{m} \left(\hat{w}_{in} - w_{in}\right)^2\right) = 2\mathrm{E}\left(\sum_{i=k+1}^{m} \hat{w}_{in}^2\right) - 2A_n^{-1}.$$

Furthermore, using the inequality

$$x^2 = y^2 + 2y\left(x-y\right) + \left(x-y\right)^2 \leq \left(1 + a^{-1}\right)y^2 + \left(1+a\right)\left(x-y\right)^2, \quad x, y \in \mathbb{R}, a > 0$$

we obtain

$$\left(\sigma^2\lambda_i+\tau^2\right)^2\leq\left(1+n^{-b}\right)\left(\bar{\sigma}^2\lambda_i+\hat{\tau}^2\right)^2+2\left(1+n^{b}\right)\left[(\sigma^2-\bar{\sigma}^2)^2\lambda_i^2+(\tau^2-\hat{\tau}^2)^2\right]. \tag{2.18}$$

With

$$\gamma_n:=\hat{A}_n^{-2}\sum_{i=k+1}^{m}\frac{\lambda_i^2}{(\bar{\sigma}^2\lambda_i+\hat{\tau}^2)^4}\left[(\sigma^2-\bar{\sigma}^2)^2\lambda_i^2+(\tau^2-\hat{\tau}^2)^2\right],$$

$$\begin{aligned}\sum_{i=k+1}^{m}\hat{w}_{in}^2&=\hat{A}_n^{-2}\sum_{i=k+1}^{m}\frac{\lambda_i^2}{(\bar{\sigma}^2\lambda_i+\hat{\tau}^2)^4}(\sigma^2\lambda_i+\tau^2)^2\\&\leq\hat{A}_n^{-1}\left(1+n^{-b}\right)+2\left(1+n^{b}\right)\gamma_n.\end{aligned} \tag{2.19}$$

It follows from (2.19) and Lemmas A.2 and A.3 in [*SM*] that

$$\sup_{\sigma,\tau>\epsilon}(\sigma\tau)^{-8}\operatorname{Var}\left(\hat{\sigma}_\tau^2-\tilde{\sigma}^2\right)=o\left(n^{-1/2}\right)$$

and hence with (2.6) and (2.17) this gives, for (2.16),

$$\begin{aligned}&\sup_{\sigma,\tau>\epsilon}\frac{1}{\sigma^8\tau^8}\operatorname{Var}\left(\hat{\sigma}^2-\tilde{\sigma}^2\right)\\&\quad\leq\sup_{\sigma,\tau>\epsilon}\frac{2}{\sigma^8\tau^8}\lambda_m^{-2}\operatorname{MSE}\left(\hat{\tau}^2\right)+o\left(n^{-1/2}\right)=o\left(n^{-1/2}\right).\end{aligned} \tag{2.20}$$

Therefore (ii) follows by Lemma A.1 and

$$\begin{aligned}&\left|\operatorname{Var}\left(\hat{\sigma}^2\right)-8\tau\sigma^3n^{-1/2}\right|\\&\quad\leq\operatorname{Var}\left(\hat{\sigma}^2-\tilde{\sigma}^2\right)+2^{3/2}\operatorname{Var}^{1/2}\left(\hat{\sigma}^2-\tilde{\sigma}^2\right)A_n^{-1/2}+2\left|A_n^{-1}-4\tau\sigma^3n^{-1/2}\right|,\end{aligned}$$

where we used $\operatorname{Var}\left(\tilde{\sigma}^2\right)=2A_n^{-1}$.
(iii) Since, by (i) and (2.20), $\operatorname{E}\left(\hat{\sigma}^2-\tilde{\sigma}^2\right)=\operatorname{E}\left(\hat{\sigma}^2-\sigma^2\right)=o\left(n^{-1/4}\right)$ and $\operatorname{Var}\left(\hat{\sigma}^2-\tilde{\sigma}^2\right)=o\left(n^{-1/2}\right)$, we have $\hat{\sigma}^2-\tilde{\sigma}^2=o_P(n^{-1/4})$. Therefore we can write $n^{1/4}\left(\hat{\sigma}^2-\sigma^2\right)=n^{1/4}\left(\tilde{\sigma}^2-\sigma^2\right)+o_P(1)$. For the asymptotic normality we apply again the CLT under Noether condition (Theorem C.1 in [*SM*]). We write $n^{1/4}\left(\tilde{\sigma}^2-\sigma^2\right)=n^{1/4}\sum_{i=k+1}^{m}w_{in}(U_i-1)$. Because of $\operatorname{E}\left(\tilde{\sigma}^2\right)=\sigma^2$ and $\operatorname{Var}\left(\tilde{\sigma}^2\right)=2A_n^{-1}$, we need only show that $\max_{i=k+1,\ldots,m}n^{1/4}w_{in}\to0$. To see this, note that

$$\max_{i=k+1,\ldots,m}n^{1/4}w_{in}\leq\frac{1}{\sigma^2}n^{1/4}A_n^{-1}\to0,$$

where we used Lemma A.1 (ii). This proves the asymptotic normality.

The constant $8\tau\sigma^3$ is sharp. As mentioned, the sharp minimax lower bound already follows by Theorem 12.1 in Ibragimov and Has'minskii (1981) from the LAN-property proved in Gloter and Jacod (2001a). However, we give a short and easily accessible proof that does not require the LAN property and, instead of assuming σ to be in a compact set, we may allow $\sigma, \tau \in [\epsilon, \infty)$, for some $\epsilon > 0$.

Theorem 4.
(i) *For any estimator* $\hat{\sigma}^2$,

$$\varliminf_{n\to\infty} \sup_{\tau,\sigma>\epsilon} \frac{1}{8\tau\sigma^3} \mathrm{E}\left(n^{1/2}(\hat{\sigma}^2 - \sigma^2)^2\right) \geq 1, \tag{2.21}$$

and equality holds if in addition $\sigma, \tau \leq K < \infty$.
(ii)

$$\lim_{n\to\infty} \inf_{\hat{\sigma}^2} \sup_{\tau,\sigma>\epsilon} (\sigma\tau)^{-8} \left(\mathrm{E}\left(n^{1/2}(\hat{\sigma}^2 - \sigma^2)^2\right) - 8\tau\sigma^3\right) = 0.$$

Proof. (i) The method of proof is similar to that of Theorem 2. Note that $Z_i \overset{ind.}{\sim} \mathcal{N}(0, \sigma^2\lambda_i + \tau^2)$, $i = 1, \ldots, n$. Straightforward calculations show that the Fisher information about σ^2 contained in $Z_1, \ldots, Z_n$ is $I_n(\sigma^2) = 1/2C_n = 1/2\sum_{i=1}^n 1/\left(\sigma^2 + \tau^2\lambda_i^{-1}\right)^2$, where C_n is as defined in (2.10). Suppose (2.21) does not hold. Then there exists an estimator $\hat{\sigma}^2$ such that, for a subsequence $\{n_k\}$,

$$\lim_{k\to\infty} \sup_{\tau,\sigma>\epsilon} \frac{1}{8\tau\sigma^3} \mathrm{E}\left(n_k^{1/2}(\hat{\sigma}^2 - \sigma^2)^2\right) \leq 1 - 4\delta$$

for some $0 < \delta \leq 1/4$. Hence there exists k_1 such that for all $k \geq k_1$,

$$\mathrm{E}\left(\hat{\sigma}^2 - \sigma^2\right)^2 \leq (1 - 3\delta) 8\tau\sigma^3 n_k^{-1/2}, \quad \text{for all } \tau, \sigma > \epsilon. \tag{2.22}$$

Let $\tau_0 > \epsilon$ be fixed. It follows from Lemma A.1 (i) that for all $\epsilon^2 < \sigma^2 \leq 3\epsilon^2$ and all sufficiently large n_k,

$$I_{n_k}(\sigma^2) \leq (1 + \delta) \frac{1}{8\tau_0\sigma^3} n_k^{1/2}. \tag{2.23}$$

Hence there exists an $n_0 > 0$ such that (2.22), (2.23), and

$$n_k > \frac{64\tau_0^2\epsilon^4\left(2^{3/2} + 3^{3/2}\right)^4}{\delta^4} \tag{2.24}$$

hold for all $n_k > n_0$ where (2.24) will be required later on. For such an n_k, the Cramer-Rao information inequality yields

$$b^2(\sigma^2) + \frac{(1 + b'(\sigma^2))^2}{I_{n_k}(\sigma^2)} \leq (1 - 3\delta) 8\tau_0\sigma^3 n_k^{-1/2} \quad \text{for all } \epsilon^2 < \sigma^2 \leq 3\epsilon^2,$$

where $b(\sigma^2)$ denotes the bias of $\hat{\sigma}^2$. This implies that

$$b^2(\sigma^2) \leq 8\tau_0\sigma^3 n_k^{-1/2} \quad \text{and} \quad b'(\sigma^2) \leq -\delta, \quad \text{for all } \epsilon^2 < \sigma^2 \leq 3\epsilon^2, \tag{2.25}$$

where the latter inequality follows from $(1 + b'(\theta))^2 \leq (1 - 3\delta)(1 + \delta)$. Further, $b'(\sigma^2) \leq -\delta$ gives

$$b(\sigma^2) \leq -\delta\left(\sigma^2 - 2\epsilon^2\right) + b(2\epsilon^2) \quad \text{for } 2\epsilon^2 \leq \sigma^2 \leq 3\epsilon^2. \tag{2.26}$$

Now with $\sigma^2 = 3\epsilon^2$ in (2.26), we obtain a contradiction for $n_k > n_0$ since

$$\begin{aligned} b\left(3\epsilon^2\right) &\leq -\delta\epsilon^2 + b\left(2\epsilon^2\right) \leq -\delta\epsilon^2 + \sqrt{8\tau_0}\left(2\epsilon^2\right)^{3/2} n_k^{-1/4} \\ &< -\sqrt{8\tau_0}\left(3\epsilon^2\right)^{3/2} n_k^{-1/4} \leq b\left(3\epsilon^2\right), \end{aligned}$$

where the second and the last inequality follow from (2.25), and the third one follows from (2.24). This proves the first part of (i) .

(ii) can be deduced in the same way as (ii) in Theorem 2 by using Theorem 3 and (i).

Combining (ii) and the first part of (i) gives equality in (i), if $\sigma, \tau \leq K < \infty$.

3. Computational Aspects

Finally, we discuss the computational complexity for calculating the spectral estimator. First we stress that this estimator can be implemented easily and in a straightforward manner. Note that the transform matrices D and D^t defined in (2.1) are discrete sine transforms (DST-IIo, DST-IIIo). For a reference see Britanak, Yip and Yao (2006) and Curci and Corsi (2006). Discrete sine transforms behave similar to Fourier transforms and fast algorithms are available. In fact, if we have n observations, performing the transformation requires $O(n \log n)$ operations. Additionally, computing $\hat{\tau}^2$, $\tilde{\sigma}^2$, and finally $\hat{\sigma}^2$ in the transformed model needs $O(n)$ steps. Hence the overall complexity is $O(n \log n)$.

Alternatively to our approach one could investigate numerically the performance of maximum likelihood methods in the difference model, where we have observations $(Y_{1,n}, Y_{2,n} - Y_{1,n}, \ldots, Y_{i,n} - Y_{i-1,n}, \ldots, Y_{n,n} - Y_{n-1,n})$, as well as in the transformed model (2.3), see Aït-Sahalia, Mykland, and Zhang (2005). This leads to maximum likelihood estimation of the parameters of an $AR(1)$ process. We mention that for computation of the maximum likelihood estimator a good starting value is of crucial importance due to the flat likelihood function in regions far from the maximum. One might use our estimator as a starting value and then iterate a few times to obtain the maximum likelihood estimator.

4. Discussion: Extension to Other Processes

Transforming the difference vector $(Y_{n,n} - Y_{n-1,n}, \ldots, Y_{2,n} - Y_{1,n}, Y_{1,n})$ by D^t, as defined in (2.1), gives us again a vector with independent observations, and we can follow the same arguments to obtain a sharp estimator. From the discussion so far it is not clear how this estimation method behaves if we consider more general models since, at a first glance, D^t seems to define a global transformation. However, suppose that τ and σ are sufficiently smooth functions, slight modifications of the proposed estimators in (2.4), (2.11) and (2.14) generalize to rate optimal estimators of $\int \tau_s^2 ds$ and $\int \sigma_s^2 ds$. Obviously our technique can be directly extended if we substitute the Brownian motion in model (1.1) by a centered Lévy-Process X with initial value $X_0 = 0$ a.s. and $\mathrm{E}\left(X_1^4\right) < \infty$ (for instance a compensated Poisson process), that is independent of $\epsilon_{i,n}$. As seen above, D^t defines a discrete sine transform. There is a strong connection between Karhunen-Loeve expansions and sine transforms. In fact, for wide classes of processes, D^t diagonalizes them approximately, for instance for general $\mathrm{MA}(q)$ processes. This gives us reason to believe that our approach is robust against various model misspecifications. However, our aim has been not to discuss these models in full generality rather to lay out these ideas as simply as possible.

Acknowledgement

The research of Tony Cai was supported in part by NSF Grant DMS-0604954. The research of Johannes Schmidt-Hieber and Axel Munk was supported by DFG Grant FOR 916. We are grateful to Holger Dette for bringing this problem to our attention and for many helpful discussions. Further we are grateful to Marc Hoffmann and Markus Reiß for their bibliographic suggestions and to two referees and an associate editor for valuable comments which improved the paper significantly.

References

Aït-Sahalia, Y., Mykland, P. and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Rev. Financial Studies* **18**, 351-416.

Barndorff-Nielsen, O., Hansen, P., Lunde, A. and Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica* **76**, 1481-1536.

Britanak, V., Yip, P. C. and Yao, K. R. (2006). *Discrete Cosine and Sine Transforms: General Properties, Fast Algorithms and Integer Approximations.* Academic Press, Oxford.

Curci, G. and Corsi, F. (2006). Discrete sine transform for multi-scales realized volatility measures. Working Paper. `papers.ssrn.com/sol3/papers.cfm?abstract_id=650504`.

Durbin, J. and Knott, M. (1972). Components of the Cramer-von Mises statistics I. *J. Roy. Statist. Soc. Ser. B* **34**, 290-307.

Gloter, A. and Jacod, J. (2001a). Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM: Probability and Statistics* **5**, 225-242.

Gloter, A. and Jacod, J. (2001b). Diffusions with measurement errors. II. Optimal estimators. *ESAIM: Probability and Statistics* **5**, 243-260.

Gloter, A. and Hoffmann, M. (2007). Estimation of the Hurst parameter from discrete noisy data. *Ann. Statist.* **35**, 1947-1974.

Huang, S., Liu, Q. and Yu, J. (2007). Realized daily variance of S&P 500 cash index: A revaluation of stylized facts. *Ann. Economics and Finance* **1**, 33-56.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation. Asymptotic Theory.* Springer, New York.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, P. and Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stoch. Processes and their Application* **119**, 2249-2276.

Lehmann, E. L. (1983). *Theory of Point Estimation.* Wiley, New York.

Mair, B. A. and Ruymgaart, F. H. (1996). Statistical inverse estimation in Hilbert scales, *SIAM J. Appl. Math.* **56**, 1424-1444.

Podolskij, M. and Vetter, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* **15**, 634-658.

Stein, M. (1987). Minimum norm quadratic estimation of spatial variograms. *Journal of the American Statistical Association* **82**, 765-772.

Tsybakov, A. (2004). *Introduction a l'Estimation Non-paramétrique.* Springer, New York.

Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* **472**, 1394-1411.

Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, U.S.A.

E-mail: tcai@wharton.upenn.edu

Institut für Mathematische Stochastik, Universität Göttingen, Maschmühlenweg 8-10, 37073 Göttingen, Germany.

E-mail: munk@math.uni-goettingen.de

Institut für Mathematische Stochastik, Universität Göttingen, Maschmühlenweg 8-10, 37073 Göttingen, Germany.

E-mail: schmidth@math.uni-goettingen.de

(Received January 2009; accepted April 2009)