Weierstraß-Institut
für Angewandte Analysis und Stochastik

im Forschungsverbund Berlin e.V.

Preprint ISSN 0946 – 8633

Nonparametric volatility estimation on the real line from low frequency data

Markus Reiß

Weierstraß Institute for Applied Analysis and Stochastics

Mohrenstraße 39, D-10117 Berlin, Germany

email: mreiss@wias-berlin.de

submitted: February 18, 2004

No. 911
Berlin 2004

1991 *Mathematics Subject Classification.* 62M05,60H10.

Key words and phrases. diffusion process, nonparametric inference, wavelet, spectral approximation, low-frequency observation.

Edited by
Weierstraß-Institut für Angewandte Analysis und Stochastik (WIAS)
Mohrenstraße 39
10117 Berlin
Germany

Fax: + 49 30 2044975
E-Mail: preprint@wias-berlin.de
World Wide Web: http://www.wias-berlin.de/

Abstract

We estimate the volatility function of a diffusion process on the real line on the basis of low frequency observations. The estimator is based on spectral properties of the estimated Markov transition operator of the embedded Markov chain. Asymptotic risk estimates for a growing number of observations are provided without assuming the observation distance to become small.

1 Introduction

Diffusion processes are widely used in physical, chemical or economical applications to model random fluctuations of some quantity over time. Especially in mathematical finance it has become very popular to model asset prices by diffusion processes because this allows the use of strong tools from stochastic analysis for option pricing or risk analysis. Removing seasonal effects and long-term growth results in time-homogeneous diffusion processes. A typical time-homogeneous scalar diffusion $(X_t,\, t \geq 0)$ solves the Itô stochastic differential equation

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \quad t \geq 0, \tag{1.1}$$

with drift coefficient $b(\bullet)$, volatility or diffusion coefficient $\sigma(\bullet)$ and with a one-dimensional Brownian motion $(W_t,\, t \geq 0)$.

Statistical inference for the volatility function has attracted a lot of interest recently, see the discussions in (Kleinow 2002) or (Gobet, Hoffmann, and Reiß 2002) for an overview. Especially, in (Kleinow 2002) it is argued on the basis of empirical data that common parametric assumptions on the coefficients are highly misspecified in models for financial markets. Moreover, statistical methods developed for high frequency observations, that is small observation distances, have been typically applied to daily asset price data over periods of several years, which should be qualified rather as low-frequency observations. Therefore, the work (Kessler and Sørensen 1999) on low-frequency statistical methods became a popular alternative, but remains restricted to certain parametric models.

Here, we consider the case of nonparametric inference for the volatility function $\sigma(\bullet)$ in the case of an unknown drift function $b(\bullet)$ and equidistant observations $(X_{n\Delta})_{0\leq n\leq N}$ with some fixed $\Delta > 0$. If a linear parametric form of the drift $b(\bullet)$ is imposed, then the nonparametric inference for $\sigma(\bullet)$ can be based mainly on the invariant density, which is easy to estimate (Aït-Sahalia 1996). In (Gobet, Hoffmann, and Reiß 2002) it was shown that for diffusions with reflections on a compact interval the nonparametric estimation problems for $b(\bullet)$ and $\sigma(\bullet)$ together can be solved using ideas in (Hansen, Scheinkman, and Touzi 1998), but it involves some ill-posedness such that the minimax rate of convergence is $N^{-s/(2s+3)}$ for $N \to \infty$ and regularity $s \geq 1$ of $\sigma(\bullet)$. Moreover, first numerical simulations in the reflected setting have shown that the spectral estimator outperforms the traditional quadratic variation estimator already for rather small observation distances Δ. We generalize this approach to cope also with diffusions on the entire real line.

The basic ideas are that (a) we can only draw inference on the law of the embedded Markov chain $(X_{n\Delta})_{n\geq 0}$, that (b) by spectral calculus its transition operator determines the infinitesimal generator of the diffusion process and that (c) this generator encodes rather explicitly the two unknown functions $b(\bullet)$ and $\sigma(\bullet)$. More specifically, the spectral estimator we propose is based on estimates of the invariant density and of one eigenfunction and its eigenvalue of the transition operator of $(X_{n\Delta})_{n\geq 0}$, see formula (2.3) below. Leaving the case of a compact state space, we face several new problems compared with the situation treated in (Gobet, Hoffmann, and Reiß 2002): (1) the observation design is degenerate, (2) the invariant densities are not uniformly comparable and (3) the eigenfunctions are unbounded. Point (1) is overcome by using warped wavelet functions or equivalently a suitable state transformation. To avoid problem (2) we work on parameter-dependent function spaces and problem (3) is treated by smoothing differently at the boundaries. By this approach we obtain that our spectral estimator also attains the rate $N^{-s/(2s+3)}$ as in the simpler case of reflected diffusions, provided the coefficients guarantee that the process is well mixing and the first eigenfunction exists and does not grow too fast to infinity. For the proof we assume the invariant law of the diffusion to be known. This is, of course, not realistic, but the estimation of the invariant density is standard and contributes less to the overall risk than the spectral estimations, as can also be seen from the lower bound proof in (Gobet, Hoffmann, and Reiß 2002).

Section 2 introduces the diffusion model and recalls some theory for diffusions, Section 3 presents and discusses the estimator and Section 4 provides the mathematical results. We adopted (hopefully) standard notation. In particular, $C^r(\mathbb{R})$ denotes the space of r-times continuously differentiable functions and $C^r_b(\mathbb{R})$ its subspace such that all derivatives are uniformly bounded including the function itself. The relation $A \lesssim B$ means that A is bounded by a multiple of B, independent of the quantities appearing in the expression B. The relation $A \sim B$ stands for $A \lesssim B$ and $B \lesssim A$. A sequence of random variables that is bounded in probability will be abbreviated by $O_P(1)$. Vectors and matrices are usually set in bold fonts.

2 The diffusion model

In this section fundamental results for one-dimensional diffusions are recalled, for more details and proofs see e.g. (Karlin and Taylor 1981) or (Bass 1998). We consider diffusion processes $(X_t, t \geq 0)$ solving (1.1). The drift $b(\bullet)$ and diffusion coefficient or volatility $\sigma(\bullet)$ are assumed to be Lipschitz continuous functions such that a strong solution exists. We shall henceforth assume the uniform ellipticity condition

$$\exists\, \sigma_0, \sigma_1 > 0 : \ \sigma_0 \leq \sigma(x) \leq \sigma_1 \text{ for all } x \in \mathbb{R} \tag{2.1}$$

and the mixing condition

$$\lim_{x\to+\infty} b(x) = -\infty \text{ and } \lim_{x\to-\infty} b(x) = +\infty. \tag{2.2}$$

These conditions imply the existence of a stationary solution X with invariant marginal density

$$\mu(x) = \frac{2C}{\sigma^2(x)} \exp\Big(\int_0^x \frac{2b(x)}{\sigma^2(x)}\,dx\Big), \quad x \in \mathbb{R},$$

where $C > 0$ is a suitable norming constant. Moreover, the solution process is time-reversible and β-mixing with exponential speed such that for statistical purposes the hypothesis of stationary observations is reasonable and will be assumed henceforth.

Diffusions are efficiently described by their Markov transition operators $(P_t)_{t\geq 0}$ with

$$P_t f(x) = \mathbb{E}[f(X_t)\,|\,X_0 = x] = \int_{-\infty}^{\infty} f(\xi)p_t(x,\xi)\,d\xi, \quad x \in \mathbb{R},\ f \in C_b(\mathbb{R}),$$

where $p_t(x,\xi)$ denotes the transition probability density. The operators $(P_t)_{t\geq 0}$ can be extended to the Hilbert space

$$L^2(\mu) = \Big\{f : \mathbb{R} \to \mathbb{R} \mid \int f^2(x)\mu(x)\,dx < \infty\Big\},$$

on which they form a strongly continuous, self-adjoint semigroup of contraction operators with infinitesimal generator

$$Lf(x) = \tfrac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x), \quad x \in \mathbb{R},$$

for functions f in the domain (with natural boundary conditions)

$$\mathcal{D}(L) = \{f \in L^2(\mu) \,|\, Lf \in L^2(\mu)\}.$$

L is a closed selfadjoint operator with spectrum on the negative real axis and the spectral mapping theorem asserts $P_t = \exp(tL)$. In particular, the eigenfunctions of P_t and L coincide and the eigenvalues are transformed like the operators. The Markov semigroup can be described equivalently by the invariant density $\mu(\bullet)$ and the inverse scale density $S(\bullet)$ given by

$$S(x) = \tfrac{1}{2}\sigma^2(x)\mu(x), \quad x \in \mathbb{R}\,.$$

Then the infinitesimal generator is given in divergence form by

$$Lf(x) = \mu^{-1}(x)(Sf')'(x), \quad x \in \mathbb{R}\,.$$

Any eigenfunction $u \in L^2(\mu)$ of L with eigenvalue ν satisfies

$$S(x)u'(x) = \nu \int_{-\infty}^{x} u(\xi)\mu(\xi)\,d\xi, \quad x \in \mathbb{R},$$

which yields

$$\sigma^2(x) = \frac{2\nu \int_{-\infty}^{x} u(\xi)\mu(\xi)\,d\xi}{u'(x)\mu(x)}, \quad x \in \mathbb{R}. \tag{2.3}$$

This identity allows to determine the volatility $\sigma(\bullet)$ from quantities accessible from the embedded Markov chain $(X_{n\Delta})_{n\geq 0}$, namely from the invariant density and a spectral pair $(u, e^{\nu\Delta})$ of the transition operator. This approach was first proposed by (Hansen, Scheinkman, and Touzi 1998) and statistically analyzed in (Gobet, Hoffmann, and Reiß 2002).

For this method to work we have to ensure that at least parts of the spectrum are discrete, that is proper eigenvalues exist. In the sequel we shall only need that the largest nontrivial (i.e., nonzero) spectral value is discrete, but to avoid any technicalities we assume $\sigma \in C^1(\mathbb{R})$ and

$$\lim_{|x|\to\infty} \Big(\sigma'(x) - \frac{2b(x)}{\sigma(x)}\Big)^2 = \infty, \tag{2.4}$$

which by Section 4.2 in (Hansen, Scheinkman, and Touzi 1998) ensures that the entire spectrum of L is discrete. In view of our previous assumptions this is already satisfied if $\sigma'(\bullet)$ is uniformly bounded.

The mathematical analysis of our proposed estimators relies on some additional growth restrictions for the first nontrivial eigenfunction u_1 of L, namely

$$u_1 \in L^p(\mu) \text{ and } u_1' \in L^p(\mu) \tag{2.5}$$

for some arbitrary $p > 2$. For $p = 2$ this condition is always satisfied because u_1 is in the domain of L and thus also of $(-L)^{1/2}$:

$$\|(-L)^{1/2}u_1\|_\mu^2 = \langle(-L)u_1, u_1\rangle_\mu = \langle Su_1', u_1'\rangle \sim \langle u_1', u_1'\rangle_\mu.$$

Observe the different norms and scalar products employed, where the index μ always refers to $L^2(\mu)$ and no index to L^2 with respect to the Lebesgue measure. For the canonical example of a stationary Ornstein-Uhlenbeck process all eigenfunctions satisfy condition (2.5) even for exponential moments. It is plausible that this behaviour remains the same whenever the tails of the invariant densities are equally small which is to say that the negative drift $-b(\bullet)$ grows linearly. A formal mathematical result in this direction still lacks and we can merely provide an example of a sufficient result under nonasymptotic conditions on the coefficients.

2.1 Proposition. *Condition* (2.5) *is satisfied for all* $p < \infty$ *if the coefficients* $\sigma^2 \in C^2(\mathbb{R})$, $b \in C^1(\mathbb{R})$ *of the diffusion satisfy*

$$\inf_{x\in\mathbb{R}} \Big(\frac{(\sigma^2(x))'}{\sigma^2(x)} \Big(b - \frac{(\sigma^2(x))'}{2\sigma^2(x)} \Big) + \frac{1}{2}\big(\sigma^2(x)\big)'' - 2b'(x) \Big) > 0.$$

For constant volatility this reduces to $\sup_x b'(x) < 0$.

Proof. By definition 1.2 in (Ledoux 1998) the diffusion process X satisfies the condition $CD(R, \infty)$ for some $R > 0$ under our assumption, which implies that the Markov semigroup is hypercontractive. It is proved in (Bakry 1994) that any eigenfunction u of a hypercontractive semigroup operator has exponential moments, that is satisfies $\int \exp(cu^\alpha(x))\mu(x)\,dx < \infty$ for some $c, \alpha > 0$.

By Lemma 1.3 in (Ledoux 1998) the condition $CD(R,\infty)$ is equivalent to

$$|\sigma(x)(P_t f)'(x)| \le e^{-Rt} P_t(|\sigma(x) f'(x)|), \quad x \in \mathbb{R},$$

for all sufficiently smooth f. For any eigenfunction u of L with eigenvalue ν we thus obtain

$$|\sigma(x)u'(x)|e^{t\nu} \le e^{-Rt} P_t(|\sigma(x)u'(x)|), \quad x \in \mathbb{R}\,.$$

The hypercontractivity of (P_t) and $\sigma u' \in L^2(\mu)$ therefore imply $\sigma u' \in L^p(\mu)$ for all $p < \infty$ and by ellipticity also $u' \in L^p(\mu)$. □

3 Construction of the estimators

We describe the spectral estimation procedure using the projection method in detail. The use of projection methods has the advantage of approximating the abstract operators by finite-dimensional matrices, for which the spectrum is easy to calculate numerically. In addition, mathematical results for spectral approximation by kernel-smoothed operators seem to be difficult to obtain. A projection approach was already suggested by (Chen, Hansen, and Scheinkman 1997) and adopted by (Gobet, Hoffmann, and Reiß 2002). More specifically, we make use of compactly supported wavelets on the interval $[0,1]$. For the notion of wavelet bases on compact intervals and their properties we refer to (Cohen 2000).

3.1 Definition. *Let (ψ_λ) with multi-indices $\lambda = (j,k)$ be a compactly supported orthonormal wavelet basis of $L^2(0,1)$ including the scaling function $\psi_{-1,0} = \mathbf{1}$. For $\lambda = (j,k)$ we set $|\lambda| := j$. The approximation spaces (V_Λ) are defined as the linear span of the wavelets indexed with Λ:*

$$V_\Lambda := \operatorname{span}\{\psi_\lambda \,|\, \lambda \in \Lambda\}.$$

The L^2-orthogonal projection onto V_Λ will be called Π_Λ. For a function $M : \mathbb{R} \to [0,1]$ we introduce the warped wavelets *$\psi_\lambda^M(x) := \psi_\lambda(M(x))$, $x \in \mathbb{R}$.*

Note that (ψ_λ^M) constitutes an orthonormal basis of $L^2(\mu)$ with $\mu(x) = M'(x)$, if M is (weakly) differentiable.

The first main idea is to use wavelets warped by the empirical stationary distribution function of the diffusion process X in order to obtain a regular autoregressive design, see (Kerkyacharian and Picard 2003) for a similar approach in classical regression with random design, but note that our density does not define a Muckenhoupt weight. An equivalent viewpoint is that we consider the data

$$\hat{Y}_n := \hat{M}(X_{n\Delta}), \qquad \text{where } \hat{M}(x) := \frac{1}{N+1}\sum_{n=0}^{N} \mathbf{1}_{(-\infty, X_{n\Delta}]}(x)$$

is the empirical stationary distribution function. The transformed observations $(\hat{Y}_n)_{0\le n\le N}$ form a permutation of the set $\{n/(N+1)\,|\,1 \le n \le N+1\}$.

For such equispaced data Mallat's pyramidal algorithm for computing wavelet coefficients is very efficient and widely available. Since the marginal density does not determine the diffusion process if drift and volatility are both unknown, we use the dependency structure in the data $(\hat{Y}_n)$ in order to draw further inference. By the Markov property of diffusion processes, it suffices to consider the empirical distribution of the transitions $X_{(n-1)\Delta} \mapsto X_{n\Delta}$ or $\hat{Y}_{n-1} \mapsto \hat{Y}_n$, respectively. Furthermore, the time reversibility asserts that the laws of $(X_{(n-1)\Delta}, X_{n\Delta})$ and $(X_{n\Delta}, X_{(n-1)\Delta})$ coincide such that we may symmetrize our estimators.

Under the stationarity assumption, $\hat{M}$ converges for $N \to \infty$ uniformly to the true distribution function $M(x) := \int_{-\infty}^{x} \mu(\xi)\, d\xi$, $x \in \mathbb{R}$. We are thus naturally lead to consider the diffusion process $Y_t = M(X_t)$, $t \geq 0$, with values in the open unit interval $(0,1)$, which has natural boundaries and satisfies by Itô's formula

$$dY_t = \mu_M(Y_t)\Big(b_M(Y_t) + \frac{1}{2}(\mu_M)'(Y_t)\sigma_M^2(Y_t)\Big)\,dt + \mu_M(Y_t)\sigma(Y_t)\,dW(t).$$

The process Y is equivalently described by the following quantities, where we write $f_M(y) := f(M^{-1}(y))$ for any function $f : \mathbb{R} \to \mathbb{R}$:

invariant measure:	$\mu_Y(y) = \mathbf{1}_{(0,1)}(y)$, (uniform),
scale density:	$S_Y^{-1}(y) = 2\mu_M^{-2}(y)\sigma_M^{-2}(y)$,
transition density:	$p_{t,Y}(y,\eta) = p_t(M^{-1}(y), M^{-1}(\eta))\mu_M(\eta)^{-1}$,
inf. generator:	$L_Y f(y) = \big(\frac{1}{2}\sigma_M^2\mu_M^2 f'\big)'(y)$,
domain of L_Y:	$\mathcal{D}(L_Y) = \{f \in L^2(0,1) \mid L_Y f \in L^2(0,1)\}$.

Note that quantities without index usually refer to X, whereas those related to Y carry an index. From the formula for the transition operator $(P_{t,Y} f_M)(M(x)) = (P_t f)(x)$ it follows that any eigenvalue ν_Y of L_Y with eigenfunction u_Y is also an eigenvalue of L, but with the rescaled eigenfunction $u = u_Y \circ M$ and vice versa.

We thus separate the estimation problem for the volatility function $\sigma(\bullet)$ of the original process X into the two subproblems of estimating the invariant density $\mu(\bullet)$ of X and of drawing inference on the Markov transitions of the transformed diffusion process Y. Of course, the latter is the much more demanding task, because the invariant density can be estimated classically under a suitable mixing hypothesis on X.

3.2 Example. *The stationary Ornstein-Uhlenbeck process with parameters α, $\sigma > 0$ satisfies the stochastic differential equation*

$$dX_t = -\alpha X_t\,dt + \sigma\,dW_t.$$

It is a Gaussian process with normal stationary law $N(0, \frac{\sigma^2}{2\alpha})$. Its generator L has discrete spectrum $\Sigma(L) = \{-\alpha n \mid n \geq 0\}$ and the eigenfunctions are given by Hermite-type polynomials.

The transformed process Y satisfies the stochastic differential equation

$$dY_t = -2\alpha\mu_M(Y_t)M^{-1}(Y_t)\,dt + \sigma\mu_M(Y_t)\,dW_t,$$

where by normality $\mu_M(y)$ is up to logarithmic terms of order y for y near zero and of order $(1-y)$ for y close to one. The eigenfunctions of L_Y are polynomials in $M^{-1}(y)$ such that they have logarithmic singularities and their derivatives of order r have polynomial singularities of order r at the boundary.

Recall that by formula (2.3) we can estimate the volatility function $\sigma(\bullet)$ by a plug-in from estimates of the invariant density $\mu(\bullet)$ and the inverse scale density $S(\bullet)$ of the process X. Hence, we make use of the transformation of this formula

$$\sigma_M^2(y) = \frac{2\nu_1 \int_0^y u_{1,Y}(\eta)\,d\eta}{(u_{1,Y})'(y)\mu_M^2(y)}, \tag{3.1}$$

where $u_{1,Y}$ denotes the eigenfunction of L_Y corresponding to the largest nontrivial eigenvalue ν_1. By the spectral mapping theorem $(e^{\Delta\nu_1}, u_{1,Y})$ is the corresponding spectral pair of the transition operator $P_{\Delta,Y}$.

Consequently, we are interested in obtaining spectral information about the transition operator $P_{\Delta,Y}$ of Y. Its expansion in the wavelet basis (ψ_λ) of $L^2(0,1)$ can be estimated by the symmetrized empirical operator coefficients

$$(\hat{\mathbf{P}}_\Delta)_{\lambda,\lambda'} := \frac{1}{2N}\sum_{n=1}^{N}\Big(\psi_\lambda(\hat{Y}_{n-1})\psi_{\lambda'}(\hat{Y}_n) + \psi_\lambda(\hat{Y}_n)\psi_{\lambda'}(\hat{Y}_{n-1})\Big).$$

Note that this is equivalent to estimating the transition operator P_Δ of X in terms of the empirically warped wavelet basis $(\psi_\lambda^{\hat{M}})$:

$$(\hat{\mathbf{P}}_\Delta)_{\lambda,\lambda'} = \frac{1}{2N}\sum_{n=1}^{N}\Big(\psi_\lambda^{\hat{M}}(X_{(n-1)\Delta})\psi_{\lambda'}^{\hat{M}}(X_{n\Delta}) + \psi_\lambda^{\hat{M}}(X_{n\Delta})\psi_{\lambda'}^{\hat{M}}(X_{(n-1)\Delta})\Big).$$

If we had $\hat{M} = M$, this would give an unbiased estimate because of

$$\begin{aligned}\mathbb{E}[\psi_\lambda(M(X_{(n-1)\Delta}))\psi_{\lambda'}(M(X_{n\Delta}))] &= \int_0^1\int_0^1 \psi_\lambda(y)\psi_{\lambda'}(\eta)p_{\Delta,Y}(y,\eta)\,d\eta\,dy \\ &= \langle P_{\Delta,Y}\psi'_\lambda, \psi_\lambda\rangle.\end{aligned}$$

The eigenfunction $u_1 \in L^2(\mu)$ of P_Δ with eigenvalue κ_1 satisfies for any multi-index λ the coefficient equation

$$\sum_{\lambda'}\langle P_\Delta\psi_\lambda^M, \psi_{\lambda'}^M\rangle_\mu\langle u_1, \psi_{\lambda'}^M\rangle_\mu = \kappa_1\langle u_1, \psi_\lambda^M\rangle_\mu.$$

Furthermore, we have $u_{1,Y} = \sum_\lambda\langle u_1, \psi_\lambda^M\rangle_\mu\psi_\lambda$. We therefore calculate the largest nontrivial eigenvalue $\hat{\kappa}_1$ (i.e. $\hat{\kappa}_1 < 1$) with eigenvector $\hat{\mathbf{u}}_1$ of the symmetric $|\Lambda|\times|\Lambda|$-matrix $\hat{\mathbf{P}}_{\Delta,\Lambda} := (\hat{\mathbf{P}}_\Delta)_{\lambda,\lambda'\in\Lambda}$ and use the estimators

$$\hat{u}_{1,Y}(x) := \sum_{\lambda\in\Lambda}(\hat{\mathbf{u}}_1)_\lambda\psi_\lambda(x), \quad \hat{\nu}_1 := \Delta^{-1}\log(\hat{\kappa}_1).$$

Observe that by construction $\hat{\mathbf{P}}_{\Delta,\Lambda}$ always has the eigenvector $\hat{\mathbf{u}}_0 = (1,0,\dots,0)$ corresponding to the constant scaling function $\psi_{-1,0}$ with eigenvalue 1.

Even though formula (3.1) is valid for any nontrivial spectral pair of L_Y, we prefer taking the first nontrivial eigenfunction $u_{1,Y}$ for two reasons: first, all other eigenfunctions oscillate such that the denominator vanishes at some point and the estimate in its neighbourhood is worthless. Second, the spectral estimation quality depends very much on the separation of the eigenvalue from the remaining spectrum (cf. Proposition 4.6) and the spectrum $\Sigma(P_{\Delta,Y}) = \{e^{\Delta\nu} \mid \nu \in \Sigma(L_Y)\}$ is such that it becomes rapidly very dense for smaller eigenvalues. Nevertheless, it might be reasonable to use the information about the other spectral pairs, compare also the embeddability discussion in (Hansen, Scheinkman, and Touzi 1998).

The usage of warped basis functions simplifies the design and thus the analysis of the stochastic error term, but does not overcome the complex structure of the deterministic approximation error. As proved later, the eigenfunctions of L_Y have logarithmic singularities at the boundary of the unit interval and its derivatives have even polynomial-type singularities. This is why, theoretically and in practice, the finite index set Λ employed in the construction of $\hat{\mathbf{P}}_{\Delta,\Lambda}$ has to be chosen carefully. On the one hand, we have the usual bias-variance balance that lets us choose the highest resolution level J in accordance with the smoothness s of the eigenfunction and the number N of observations. On the other hand we have to take into account the singular behaviour such that we shall refine more in the neighbourhood of the boundary points. We roughly choose a maximal frequency level $J(y)$ for wavelets with support in the point $y \in (0,1)$ that satisfies $2^{J(y)} \sim 2^J \min(y, 1-y)^{-1+\varepsilon}$ with some small $\varepsilon > 0$, see Proposition 4.3 for details.

It remains to estimate μ_M, which we propose to do by the – up to transformation – classical projection estimate

$$\hat{\mu}_M(y) := \sum_{|\lambda| \le J} \hat{\mu}_\lambda \psi_\lambda, \quad \hat{\mu}_\lambda := \frac{1}{N+1} \sum_{n=0}^{N} \psi_\lambda(\hat{Y}_n).$$

Equipped with these estimates we use formula (3.1) in order to derive an estimate $\hat{\sigma}_M^2$ of σ_M^2 and use the estimated invariant law to transform it to an estimator $\hat{\sigma}^2$ of σ^2, which is our proposed spectral estimator.

Let us summarize our estimation procedure:

1. Form the empirical distribution function $\hat{M}$ and the transformed observations $\hat{Y}_n = \hat{M}(X_{n\Delta})$, $n = 0, 1, \ldots, N$.

2. Estimate the transition operator by the matrix $\hat{\mathbf{P}}_{\Delta,\Lambda}$ of empirical wavelet coefficients.

3. Calculate the first nontrivial spectral pair $(\kappa_1, \mathbf{u}_1)$ of $\hat{\mathbf{P}}_{\Delta,\Lambda}$ and build the estimate $\hat{u}_{1,M}$ of the eigenfunction.

4. Estimate the invariant density μ by some classical method.

5. Derive the estimator $\hat{\sigma}_M^2$ by inserting the preceding estimates in formula (3.1) and transform it back to the real line.

As already mentioned in the introduction, we provide a proof in the case that the invariant law of X is known, that is M and μ are available exactly.

In this case our spectral estimator is given by

$$\hat{\sigma}^2(x) := \frac{2\hat{\nu}_1 \int_0^{M(x)} \hat{u}_{1,Y}(\eta)\,d\eta}{\hat{u}'_{1,Y}(M(x))\mu^2(x)}, \tag{3.2}$$

derived from formula (3.1) by plug-in and transformation. To avoid theoretical complications we must keep $\hat{\nu}_1$, $\|\hat{u}_{1,Y}\|_{L^2}$ and $\|\mu_M \hat{u}'_{1,Y}\|_{L^2}^2$ uniformly bounded, e.g. by applying a cut-off for unreasonably large values. Similarly, we guarantee that the a priori knowledge $\hat{\sigma}^2(x) \geq \sigma_0^2$ is fulfilled by changing the denominator if necessary. Then our main result is the following:

3.3 Theorem. *Let us assume that the invariant distribution function M and its derivative μ are known and that $s \geq 2$, $s \in \mathbb{N}$. Then for $\sigma \in C_b^s(\mathbb{R})$ and $b \in C^{s-1}(\mathbb{R})$ the spectral volatility estimator $\hat{\sigma}^2$ from (3.2) satisfies for any $\delta > 0$*

$$(2^{-2Js} + N^{-1}2^{3J})^{-1} \int_{M^{-1}(\delta)}^{M^{-1}(1-\delta)} |\hat{\sigma}^2(x) - \sigma^2(x)|^2 \mu(x)\,dx = O_P(1).$$

In particular, we obtain with the asymptotically optimal choice $2^J \sim N^{1/(2s+3)}$ that $N^{s/(2s+3)}\|\hat{\sigma}^2 - \sigma^2\|_{L^2(K)}$ is bounded in probability for any compact set $K \subset \mathbb{R}$.

3.4 Remark. *For true minimax results we should have a uniform constant for all parameters in some smoothness class. This might be feasible, although very technical, and requires also uniform estimates on the separation of the spectrum which are usually difficult to obtain. It is not clear whether it is possible to get rid of the restriction to bounded intervals. In the case of reflected diffusion a lower bound proof shows that estimation at the boundary is definitely more difficult, but whether this holds also in our situation with a μ-weighted loss function is an open question. Following the approach in (Gobet, Hoffmann, and Reiß 2002) we can extend our procedure to estimate also the drift coefficient $b(\bullet)$.*

4 Mathematical results

4.1 Lemma. *Suppose $\sigma \in C_b^s(\mathbb{R})$ and $b \in C^{s-1}(\mathbb{R})$ with $b' \in C_b^{s-2}(\mathbb{R})$ for some $s \geq 2$. Then the inverse scale density S_Y of Y is s-times differentiable and satisfies*

$$|S_Y^{(r)}(y)| \lesssim S_Y(y)\left|\frac{b_M(y)}{\mu_M(y)}\right|^r \quad 0 \leq r \leq s.$$

Proof. The derivatives $S^{(r)}(x)$ for $r \leq s$ are given by

$$S(x) = \tfrac{1}{2}\sigma^2(x)\mu(x) =: a(x)\mu(x), \qquad S^{(r)}(x) = \sum_{k=0}^{r}\binom{r}{k} a^{(k)}(x)\mu^{(r-k)}(x).$$

Applying iteratively the formula $\mu'(x) = 2(-\sigma'(x)+b(x))\mu(x)/\sigma(x)$ and using that $\sigma^{(r)}(\bullet)$, $0 \leq r \leq s$, and $b^{(r)}(\bullet)$, $1 \leq r \leq s$, are uniformly bounded, we obtain

$$|S^{(r)}(x)| \lesssim |b(x)|^r \mu(x).$$

If we now use $S_Y(y) = S_M(y)\mu_M(y)$ and thus $S_Y'(y) = (S')_M(y)$, we arrive at

$$|S_Y^{(r)}| \lesssim \left|(S^{(r)})_M(y)\mu_M^{-r+1}\right| + \left|(S')_M(y)(\mu^{(r-1)})_M\mu_M^{-r+1}\right| \lesssim b_M^r\mu_M^{-r+2}.$$

By the uniform ellipticity condition on $\sigma(\bullet)$ the assertion follows. $\square$

4.2 Proposition. *Suppose $\sigma \in C_b^s(\mathbb{R})$ and $b \in C^{s-1}(\mathbb{R})$ for some $s \geq 2$, and the eigenfunction u of L satisfies $u,\, u' \in L^p(\mu)$ for some $p \geq 2$. Then the derivatives of the corresponding eigenfunction u_Y of L_Y exist up to order $s+1$ and satisfy for any $1 \leq r \leq s+1$*

$$\mu_M w^{r-1} u_Y^{(r)} \in L^p(0,1) \qquad \textit{with the weight function } w(y) := \min(y, 1-y).$$

Proof. The $L^p(\mu)$-integrability of u and u' translates via $u_Y = u_M$ into $u_Y,\, u_Y'\mu_M \in L^p(0,1)$. We now apply the eigenfunction relation

$$S_Y u_Y'' + S_Y' u_Y' = (S_Y u_Y')' = \nu u_Y \in L^p(0,1).$$

From $S_Y'\mu_M^{-1} = (S'\mu^{-1})_M = b_M$ we conclude $\|(S_Y')^{-1}\mu_M\|_\infty < \infty$ and

$$S_Y(S_Y')^{-1}\mu_M u_Y'' = \mu_M u_Y' - \nu(S_Y')^{-1}\mu_M u_Y \in L^p(0,1).$$

Consequently the estimate $|S_Y'(y)| \lesssim |S_Y(y)b_M(y)\mu_M^{-1}(y)|$ from Lemma 4.1 shows that $\mu_M^2 b_M^{-1} u_Y'' \in L^p(0,1)$ holds. More generally, we use $(S_Y u_Y')^{(r)} = \nu u_Y^{(r-1)}$ and $\|(S_Y^{(r)})^{-1}\mu_M\|_\infty < \infty$ to obtain inductively over $0 \leq r \leq s$

$$S_Y(S_Y^{(r)})^{-1}\mu_M u_Y^{(r+1)} \in L^p(0,1).$$

Hence, Lemma 4.1 yields that $\mu_M^{r+1} b_M^{-r} u_Y^{(r+1)}$ lies in $L^p(0,1)$ for $0 \leq r \leq s$. While $b_M\mu_M^{-1}$ is obviously bounded on compact subintervals of $(0,1)$, we have by L'Hopital's rule

$$\lim_{y\to 0+} \frac{b_M(y)y}{\sigma_M^2(y)\mu_M(y)} = \lim_{x\to -\infty} \frac{b(x)M(x)}{\sigma^2(x)\mu(x)} = \lim_{x\to -\infty} \frac{b'(x)M(x) + b(x)\mu(x)}{2b(x)\mu(x)}.$$

Due to $M(x)\mu^{-1}(x) \to 0$ and $b'(x)/b(x) \to 0$ (μ decays faster than exponentially because of $|b(x)| \to \infty$ and b' is bounded) we obtain

$$b_M(y)y \sim \sigma_M^2(y)\mu_M(y) \sim \mu_M(y) \text{ for } y \to 0. \tag{4.1}$$

Together with the symmetric argument for $y \to 1$ we obtain the assertion. $\square$

4.3 Proposition. *The projection $\Pi_\Lambda u_Y$ of the eigenfunction u_Y of L_Y with*

$$\Lambda := \Lambda(J,\varepsilon) := \{(j,k) \mid j \leq J \text{ or } w(k2^{-j}) \in (2^{-J/\varepsilon}, 2^{(J-j)/(1-\varepsilon)})\}$$

satisfies $\|(\mathrm{Id} - L_Y)^{1/2}(\mathrm{Id} - \Pi_\Lambda)u_Y\|_{L^2(0,1)} \lesssim 2^{-Js}$ for any $J \in \mathbb{N}$, provided $\sigma \in C_b^s(\mathbb{R})$, $b' \in C_b^{s-2}(\mathbb{R})$ and $\varepsilon \in (0, (p-2)/2ps)$.

4.4 Remark. *By construction of Λ, we only use wavelet coefficients in $\Pi_\Lambda u_Y$ up to the maximal resolution level $2^{J/\varepsilon}$. Furthermore, the number of indices contained in Λ is of order*

$$|\Lambda| \sim 2^J + \sum_{j>J} |\{k \,|\, k2^{-j} \leq 2^{(J-j)/(1-\varepsilon)}\}| \sim 2^J + \sum_{j>J} 2^{J/(1-\varepsilon)} 2^{-\varepsilon j/(1-\varepsilon)} \sim 2^J$$

such that the variance term will behave as in the case of spatially homogeneous approximation.

Proof. Due to $\|(\mathrm{Id} - L)^{1/2} f\|_{L^2}^2 = \|f\|_{L^2}^2 + \|S_Y^{1/2} f'\|_{L^2}^2$ we can separately bound the norms of $(\mathrm{Id} - \Pi_\Lambda)u_Y$ and its derivative. Since the first norm bound is a much simpler version of the second, we only present the estimate for $\|S_Y^{1/2}((\mathrm{Id} - \Pi_\Lambda)u_Y)'\|_{L^2}$. For this note that due to inequalities of the type

$$\|S_Y^{1/2} u_Y' \mathbf{1}_{[0,\delta]}\|_{L^2} \leq \|S_Y^{1/2} u_Y'\|_{L^p} \delta^{(p-2)/2p}, \quad p > 2,$$

we only need to bound the $L^2(\delta, 1-\delta)$-norm with $\delta^{(p-2)/2p} \sim 2^{-Js}$, that is $\delta \sim 2^{-2Jsp/(p-2)}$ and thus $\delta/2^{-J/\varepsilon} \to \infty$.

We use the compact support and the vanishing moment property of the wavelet functions and its derivatives following the classical approximation estimates via Taylor expansion. Denoting the supporting interval of ψ_λ by $\mathfrak{S}_\lambda$, that is $\mathfrak{S}_{j,k} = [k2^{-j}, (s_0 + k)2^{-j}]$, its length by $|\mathfrak{S}_\lambda|$ and the L^2-Sobolev space of order 1 by H^1, we obtain

$$\begin{aligned}
\|S_Y^{1/2}((\mathrm{Id} - \Pi_\Lambda)u_Y)'\|_{L^2}^2 &= \int_0^1 S_Y(y) \Big(\sum_{\lambda \notin \Lambda} \langle u_Y, \psi_\lambda \rangle \psi_\lambda'(y) \Big)^2 \, dy \\
&\lesssim \int_0^1 \Big(\sum_{\lambda \notin \Lambda} |\mathfrak{S}_\lambda|^s \Big| \int_{\mathfrak{S}_\lambda} u_Y^{(s+1)} \Big| \|\psi_\lambda\|_{L^1} S_Y^{1/2}(y) \psi_\lambda'(y) \Big)^2 \, dy \\
&\lesssim \Big\| \sum_{\lambda \notin \Lambda} 2^{-(s+1)|\lambda|} \Big(\int_{\mathfrak{S}_\lambda} |u_Y^{(s+1)}|^2 \Big)^{1/2} \big(\max_{y \in \mathfrak{S}_\lambda} S_Y^{1/2}(y) \big) \psi_\lambda \Big\|_{H^1}^2 \\
&\sim \sum_{\lambda \notin \Lambda} 2^{-2s|\lambda|} \big(\max_{y \in \mathfrak{S}_\lambda} S_Y(y) \big) \int_{\mathfrak{S}_\lambda} |u_Y^{(s+1)}|^2.
\end{aligned}$$

Since we only need to consider wavelet coefficients $(j, k) \notin \Lambda$ satisfying additionally $w(k2^{-j}) \geq \delta$ with $\delta/2^{-J/\varepsilon} \to \infty$, the corresponding support intervals $\mathfrak{S}_{j,k}$ have a distance of at least $\max(\delta, 2^{(J-j)/(1-\varepsilon)}) - s_0 2^{-j} \gtrsim 2^{-j}$ from the boundary. The estimates $S_Y(y) \sim \mu_M^2(y)$ and $\mu_M(y) \gtrsim w(y)$ yield

$$\frac{\sup_{y \in \mathfrak{S}_{j,k}} S_Y(y)}{\inf_{y \in \mathfrak{S}_{j,k}} S_Y(y)} \lesssim \Big(1 + \frac{S2^{-j}}{2^{-j}} \Big)^2 \sim 1,$$

which gives the bound

$$\|S_Y^{1/2}((\mathrm{Id} - \Pi_\Lambda)u_Y)'\|_{L^2(\delta, 1-\delta)}^2 \lesssim \sum_{\lambda \notin \Lambda} 2^{-2s|\lambda|} \int_{\mathfrak{S}_\lambda} S_Y(\zeta) |u_Y^{(s+1)}(\zeta)|^2 \, d\zeta.$$

We apply the Hölder inequality with $\frac{p}{2}$ and $q = \frac{p}{p-2} > 1$ and obtain for $j \geq J$ by Proposition 4.2

$$\begin{aligned}
&\sum_{k\colon (j,k)\notin\Lambda} \int_{S_{j,k}} S_Y(\zeta)|u_Y^{(s+1)}(\zeta)|^2\,d\zeta \\
&\leq \Big(\sum_k \int_{S_{j,k}} S_Y(\zeta)^{p/2}|u_Y^{(s+1)}(\zeta)|^p w^{sp}(\zeta)\,d\zeta\Big)^{2/p}\Big(\sum_k \int_{S_{j,k}} w^{-2sq}(\zeta)\,d\zeta\Big)^{1/q} \\
&\lesssim \|S_Y^{1/2} w^s u_Y^{(s+1)}\|_{L^p}^2 \Big(\sum_k 2^{-j} w(k2^{-j})^{-2sq}\Big)^{1/q} \\
&\lesssim \big(2^{-j}2^{2jqs}\big(2^{(J-j)/(1-\varepsilon)}2^j\big)^{1-2qs}\big)^{1/q} \\
&= 2^{(J-j)(q^{-1}-2s)/(1-\varepsilon)}.
\end{aligned}$$

Consequently, $\|S_Y^{1/2}((\mathrm{Id}-\Pi_\Lambda)u)'\|_{L^2}$ is of order 2^{-Js}, provided

$$\sum_{j\geq J} 2^{-2(j-J)s}2^{(J-j)(q^{-1}-2s)/(1-\varepsilon)} = \sum_{j\geq 0} 2^{-j(2s+(q^{-1}-2s)/(1-\varepsilon))}$$

is finite, which is ensured for $\varepsilon < (p-2)/2ps$. □

4.5 Proposition. *For any function $v \in V_\Lambda$, $\|v\|_{L^2} = 1$, we have*

$$\mathbb{E}[\|(\mathrm{Id}-L_Y)^{1/2}(\hat{P}_{\Delta,\Lambda} - P_{\Delta,\Lambda})v\|_{L^2}^2] \lesssim N^{-1}2^{3J},$$

where we have introduced the operators

$$P_{\Delta,\Lambda} := \Pi_\Lambda P_{\Delta,Y} \text{ and } \hat{P}_{\Delta,\Lambda} v := \sum_{\lambda\in\Lambda}\Big(\hat{\mathbf{P}}_{\Delta,\Lambda}(\langle v,\psi_{\lambda'}\rangle)_{\lambda'\in\Lambda}\Big)_\lambda \psi_\lambda.$$

Proof. The bound on $\|(\hat{P}_{\Delta,\Lambda} - P_{\Delta,\Lambda})v\|_{L^2}$ is again easy and therefore omitted.

We obtain by the mixing properties of Y, cf. Lemma 5.2 in (Gobet, Hoffmann, and Reiß 2002):

$$\begin{aligned}
&\mathbb{E}\Big[\|S_Y^{1/2}((\hat{P}_{\Delta,\Lambda} - P_{\Delta,\Lambda})v)'\|_{L^2}^2\Big] \\
&= \int_0^1 S(y)\,\mathrm{Var}\Big[\sum_{\lambda\in\Lambda}\frac{1}{N}\sum_{n=1}^N \psi_\lambda(Y_{(n-1)\Delta})v(Y_{n\Delta})\psi_\lambda'(y)\Big]\,dy \\
&\lesssim N^{-1}\int_0^1 S_Y(y)\,\mathbb{E}\Big[\Big(\sum_{\lambda\in\Lambda}(\psi_\lambda(Y_0)v(Y_\Delta)\psi_\lambda'(y)\Big)^2\Big]\,dy \\
&= N^{-1}\sum_{\lambda,\lambda'\in\Lambda}\Big(\int_0^1 S_Y(y)\psi_\lambda'(y)\psi_{\lambda'}'(y)\,dy\Big)\,\mathbb{E}\Big[\psi_\lambda(Y_0)\psi_{\lambda'}(Y_0)v^2(Y_\Delta)\Big].
\end{aligned}$$

Because of (4.1) and the logarithmic growth bound on b_M we obtain for $\lambda,\lambda'\in\Lambda$ with $j' := |\lambda'| \geq |\lambda| =: j$ and $j' > J$

$$\begin{aligned}
\Big|\int_0^1 S_Y(y)\psi_\lambda'(y)\psi_{\lambda'}'(y)\,dy\Big| &\lesssim |S_{\lambda'}|\mu_M^2(2^{(J-j')/(1-\varepsilon)})2^{3j'/2}2^{3j/2} \\
&\lesssim 2^{2J}(j'-J)^2 2^{(j'-J)(\frac{1}{2}-\frac{2}{1-\varepsilon})}2^{3(j-J)/2}.
\end{aligned}$$

For $j \le j' \le J$ the same term is evidently bounded by $2^{(3j+j')/2}$.

Inserting these estimates and then proceeding similarly for the expectation we obtain

$$\begin{aligned}
&\mathbb{E}\Big[\|S_Y^{1/2}((\hat P_{\Delta,\Lambda}-P_{\Delta,\Lambda})v)'\|_{L^2}^2\Big]\\
&\lesssim N^{-1}\sum_{\substack{(j,k),(j',k')\in\Lambda\\ j'\ge\max(j,J+1)}} 2^{2J}(j'-J)^2 2^{(j'-J)(\frac12-\frac{2}{1-\varepsilon})}2^{3(j-J)/2}\,\mathbb{E}\Big[\psi_{jk}(Y_0)\psi_{j'k'}(Y_0)v^2(Y_\Delta)\Big]\\
&\quad+N^{-1}\sum_{\substack{(j,k),(j',k')\in\Lambda\\ j\le j'\le J}} 2^{(3j+j')/2}\,\mathbb{E}\Big[\psi_{jk}(Y_0)\psi_{j'k'}(Y_0)v^2(Y_\Delta)\Big]\\
&\lesssim N^{-1}2^{2J}\sum_{j'\ge\max(j,J+1)}(j'-J)^2 2^{(j'-J)(\frac12-\frac{2}{1-\varepsilon})}2^{3(j-J)/2}\|v\|_{L^2}^2 2^{(J-j')/(1-\varepsilon)}2^{j/2}2^{j'/2}\\
&\quad+N^{-1}\sum_{j\le j'\le J}2^{(3j+j')/2}\|v\|_{L^2}^2 2^{j/2}2^{j'/2}\\
&\lesssim N^{-1}2^{3J}\Big(\sum_{j\le J,\,j'>0}(j')^2 2^{j'(1-\frac{3}{1-\varepsilon})}2^{2(j-J)}+\sum_{j>0,\,j'\ge j}(j')^2 2^{j'(1-\frac{3}{1-\varepsilon})}2^{2j}+\sum_{j\le j'\le J}2^{2j+j'-3J}\Big)\\
&\lesssim N^{-1}2^{3J}.
\end{aligned}$$

□

The next result is essential for the spectral approximation to work. It is stated as Proposition 2.9 and Corollary 2.13 in (Gobet, Hoffmann, and Reiß 2002).

4.6 Proposition. *Suppose a selfadjoint bounded linear operator T on a Hilbert space has a simple eigenvalue κ such that κ has distance ρ from the remaining spectrum. Let T_ε be a second linear operator with $\|T_\varepsilon - T\| < \frac12\rho^{-1}$. Then the operator T_ε has a simple eigenvalue κ_ε and there are normalized eigenvectors u and u_ε with $Tu=\kappa u$, $T_\varepsilon u_\varepsilon = \kappa_\varepsilon u_\varepsilon$ satisfying*

$$|\kappa_\varepsilon-\kappa|+\|u_\varepsilon-u\|\lesssim\rho\|(T_\varepsilon-T)u\|.$$

Proof of Theorem 3.3. We apply the preceding proposition to the Hilbert space $H=\mathcal{D}((\mathrm{Id}-L_Y)^{1/2})$, the domain of the operator $(\mathrm{Id}-L_Y)^{1/2}$ on $L^2(0,1)$, and with the operators $P_{\Delta,Y}$ and $\hat P_{\Delta,\Lambda}$. The functional calculus shows that L_Y and $P_{\Delta,Y}$ are selfadjoint on H. For any normalized eigenfunction u_Y of $P_{\Delta,Y}$ we obtain from Proposition 4.3 using $P_{\Delta,Y}u_Y=\kappa u_Y$ and from Proposition 4.5 that

$$\mathbb{E}\big[\|(\mathrm{Id}-L)^{1/2}(\hat P_{\Delta,\Lambda}-P_{\Delta,Y})u_Y\|_{L^2}^2\big]\lesssim 2^{-2Js}+N^{-1}2^{3J}.$$

The spectral approximation result in Proposition 4.6 thus gives

$$\mathbb{E}\big[\big(|\hat\kappa_1-\kappa_1|^2+\|(\mathrm{Id}-L)^{1/2}(\hat u_{1,Y}-u_{1,Y})\|_{L^2}^2\big)\mathbf{1}_{\mathfrak{A}}\big]\lesssim 2^{-2Js}+N^{-1}2^{3J}$$

on the random set $\mathfrak{A} := \{\|\hat{P}_{\Delta,\Lambda} - P_{\Delta,Y}\| < \frac{1}{2}\rho^{-1}\}$. From Propositions 4.3 and 4.5 we infer further for the corresponding H-norms

$$\begin{aligned}
&\mathbb{E}[\|\hat{P}_{\Delta,\Lambda} - P_{\Delta,Y}\|^2] \\
&\leq 2\,\mathbb{E}[\|(\hat{P}_{\Delta,\Lambda} - P_{\Delta,\Lambda})|_{V_\Lambda}\|^2] + 2\|(\mathrm{Id}-\Pi_\Lambda)P_{\Delta,Y} + \Pi_\Lambda P_{\Delta,Y}(\mathrm{Id}-\Pi_\Lambda)\|^2 \\
&\lesssim |\Lambda| \sup_{v\in V_\Lambda,\,\|v\|=1} \mathbb{E}[\|(\hat{P}_{\Delta,\Lambda} - P_{\Delta,\Lambda})v\|^2] + \|(\mathrm{Id}-\Pi_\Lambda)P_{\Delta,Y}\|^2 \\
&\lesssim 2^J N^{-1} 2^{3J} + 2^{-2sJ} = N^{-1}2^{4J} + 2^{-2sJ}.
\end{aligned}$$

For $s \geq 1$ and $2^J \sim N^{1/(2s+3)}$ Chebyshev's inequality implies $\mathbb{P}(\Omega \setminus \mathfrak{A}) \lesssim N^{-s/(2s+3)}$. Since we assess the risk by convergence in probability, the loss of the estimator on $\Omega \setminus \mathfrak{A}$ is not larger than $O(N^{-s/(2s+3)})$.

Keeping $|\hat{\kappa}_1| + \|(\mathrm{Id}-L)^{1/2}\hat{u}_{1,Y}\|_{L^2}$ uniformly bounded, we obtain using $S_Y \sim \mu_M^2$

$$\mathbb{E}\big[|\hat{\nu}_1 - \nu_1|^2\big] + \mathbb{E}\big[\|\hat{u}_{1,Y} - u_{1,Y}\|_{L^2}^2\big] + \mathbb{E}\big[\|\mu_M(\hat{u}'_{1,Y} - u'_{1,Y})\|_{L^2}^2\big] \lesssim 2^{-2Js} + N^{-1}2^{3J}.$$

Note that we have bounded the estimation risk for $\hat{\nu}_1$ by that of $\hat{\kappa}_1$ due to the continuity of the log-transformation involved. From

$$\mathbb{E}\Big[\Big\|\int_0^\bullet (\hat{u}_{1,Y} - u_{1,Y})\Big\|_{L^2}^2\Big] \lesssim 2^{-2Js} + N^{-1}2^{3J}$$
$$\mathbb{E}\big[\|(\hat{u}'_{1,Y} - u'_{1,Y})\mu_M^2\|_{L^2}^2\big] \lesssim 2^{-2Js} + N^{-1}2^{3J}$$

and the fact that $u'_{1,Y}$ does not vanish inside $(0,1)$ we infer for any fixed $R, \delta > 0$ by the usual triangle inequality argument and the exclusion of explosions

$$(2^{-2Js} + N^{-1}2^{3J})^{-1}\,\mathbb{E}\Big[\int_\delta^{1-\delta} |\hat{\sigma}_M^2(y) - \sigma_M^2(y)|^2\,dy \wedge R\Big] \lesssim 1.$$

Hence, transforming back to the real line gives

$$(2^{-2Js} + N^{-1}2^{3J})^{-1}\,\mathbb{E}\Big[\int_{M^{-1}(\delta)}^{M^{-1}(1-\delta)} |\hat{\sigma}^2(x) - \sigma^2(x)|^2 \mu(x)\,dx \wedge R\Big] \lesssim 1.$$

The fact that $d_R(X,Y) := \mathbb{E}[|X - Y| \wedge R]$ is a metric for convergence in probability then gives the result. □

References

Aït-Sahalia, Y. (1996): "Nonparametric pricing of interest rate derivative securities.," *Econometrica*, 64(3), 527–560.

Bakry, D. (1994): "L'hypercontractivite et son utilisation en theorie des semigroupes. (Hypercontractivity and its usage in semigroup theory).," in *Bakry, Dominique (ed.) et al., Lectures on probability theory. Ecole d'Ete de Probabilites de Saint-Flour XXII-1992. Berlin: Springer. Lect. Notes Math. 1581, 1-114* .

BASS, R. F. (1998): *Diffusions and elliptic operators.* Probability and Its Applications. New York, NY: Springer.

CHEN, X., L. P. HANSEN, AND J. A. SCHEINKMAN (1997): "Shape preserving estimation of diffusions," Manuscript.

COHEN, A. (2000): "Wavelet methods in numerical analysis.," in *Handbook of numerical analysis. Vol. 7*, ed. by P. G. Ciarlet. North-Holland/ Elsevier, Amsterdam.

GOBET, E., M. HOFFMANN, AND M. REISS (2002): "Nonparametric estimation of scalar diffusions based on low-frequency data is ill-posed," Discussion paper 57, Sonderforschungsbreich 373, Humboldt University Berlin (to appear in Annals of Statistics).

HANSEN, L. P., J. A. SCHEINKMAN, AND N. TOUZI (1998): "Spectral methods for identifying scalar diffusions," *Journal of Econometrics*, 86, 1–32.

KARLIN, S., AND H. M. TAYLOR (1981): *A second course in stochastic processes.* New York etc.: Academic Press.

KERKYACHARIAN, G., AND D. PICARD (2003): "Regression in random design and warped wavelets.," Preprint, Laboratoire de probabilités et modèles aléatoires, Universités de Paris 6 et 7.

KESSLER, M., AND M. SØRENSEN (1999): "Estimating equations based on eigenfunctions for a discretely observed diffusion process.," *Bernoulli*, 5(2), 299–314.

KLEINOW, T. (2002): "Testing the diffusion coefficient," Discussion paper 38, Sonderforschungsbereich 373, Humboldt University Berlin.

LEDOUX, M. (1998): "The geometry of Markov diffusion generators," Lecture notes, ETH Zürich.