

# Asymptotic equivalence and sufficiency for volatility estimation under microstructure noise

Markus Reiß

*Institute of Mathematics*

*Humboldt-Universität zu Berlin*

mreiss@mathematik.hu-berlin.de

January 26, 2010

## **Abstract**

The basic model for high-frequency data in finance is considered, where an efficient price process is observed under microstructure noise. It is shown that this nonparametric model is in Le Cam's sense asymptotically equivalent to a Gaussian shift experiment in terms of the square root of the volatility function  $\sigma$ . As an application, simple rate-optimal estimators of the volatility and efficient estimators of the integrated volatility are constructed.

*Key words and Phrases:* High-frequency data, integrated volatility, spot volatility estimation, Le Cam deficiency, equivalence of experiments, Gaussian shift.

*AMS subject classification:* 62G20, 62B15, 62M10, 91B84

## 1 Introduction

In recent years volatility estimation from high-frequency data has attracted a lot of attention in financial econometrics and statistics. Due to empirical evidence that the observed transaction prices of assets cannot follow a semi-martingale model, a prominent approach is to model the observations as the superposition of the true (or efficient) price process with some measurement error, conceived as microstructure noise. The main features are already present in the basic model of observing

$$Y_i = X_{i/n} + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

with an efficient price process  $X_t = \int_0^t \sigma(s) dB_s$ ,  $B$  a standard Brownian motion, and  $\varepsilon_i \sim N(0, \delta^2)$  all independent. The aim is to perform statistical inference on the volatility function  $\sigma : [0, 1] \rightarrow \mathbb{R}^+$ , e.g. estimating the so-called integrated volatility  $\int_0^1 \sigma^2(t) dt$  over the trading day.

The mathematical foundation on the parametric formulation of this model has been laid by Gloter and Jacod (2001a) who prove the interesting result that the model is locally asymptotically normal (LAN) as  $n \rightarrow \infty$ , but with the unusual rate  $n^{-1/4}$ , while without microstructure noise the rate is  $n^{-1/2}$ . Starting with Zhang, Mykland, and Aït-Sahalia (2005), the nonparametric model has come into the focus of research. Mainly three different, but closely related approaches have been proposed afterwards to estimate the integrated volatility: multi-scale estimators (Zhang 2006), realized kernels or autocovariances (Barndorff-Nielsen, Hansen, Lunde, and Shephard 2008) and preaveraging (Jacod, Li, Mykland, Podolskij, and Vetter 2009). Under various degrees of generality, especially also for stochastic volatility, all authors provide central limit theorems with convergence rate  $n^{-1/4}$  and an asymptotic variance involving the so-called quarticity  $\int_0^1 \sigma^4(t) dt$ . Recently, also the problem of estimating the spot volatility  $\sigma^2(t)$  itself has found some interest (Munk and Schmidt-Hieber 2009).

The aim of the present paper is to provide a thorough mathematical understand-

ing of the basic model, to explain why statistical inference is not so canonical and to propose a simple estimator of the integrated volatility which is efficient. To this end we employ Le Cam's concept of asymptotic equivalence between experiments. In fact, our main theoretical result in Theorem 6.2 states under some regularity conditions that observing  $(Y_i)$  in (1.1) is for  $n \rightarrow \infty$  asymptotically equivalent to observing the Gaussian shift experiment

$$dY_t = \sqrt{2\sigma(t)} dt + \delta^{1/2} n^{-1/4} dW_t, \quad t \in [0, 1],$$

with Gaussian white noise  $dW$ . Not only the large noise level  $\delta^{1/2} n^{-1/4}$  is apparent, but also a non-linear  $\sqrt{\sigma(t)}$ -form of the signal, from which optimal asymptotic variance results can be derived. Note that a similar form of a Gaussian shift was found to be asymptotically equivalent to nonparametric density estimation (Nussbaum 1996). A key ingredient of our asymptotic equivalence proof are the results by Grama and Nussbaum (2002) on asymptotic equivalence for generalized nonparametric regression, but also ideas from Carter (2006) and Reiß (2008) play a role. Moreover, fine bounds on Hellinger distances for Gaussian measures with different covariance operators turn out to be essential.

Roughly speaking, asymptotic equivalence means that any statistical inference procedure can be transferred from one experiment to the other such that the asymptotic risk remains the same, at least for bounded loss functions. Technically, two sequences of experiments  $\mathcal{E}^n$  and  $\mathcal{G}^n$ , defined on possibly different sample spaces, but with the same parameter set, are asymptotically equivalent if the Le Cam distance  $\Delta(\mathcal{E}^n, \mathcal{G}^n)$  tends to zero. For  $\mathcal{E}_i = (\mathcal{X}_i, \mathcal{F}_i, (\mathbb{P}_\vartheta^i)_{\vartheta \in \Theta})$ ,  $i = 1, 2$ , by definition,  $\Delta(\mathcal{E}_1, \mathcal{E}_2) = \max(\delta(\mathcal{E}_1, \mathcal{E}_2), \delta(\mathcal{E}_2, \mathcal{E}_1))$  holds in terms of the deficiency  $\delta(\mathcal{E}_1, \mathcal{E}_2) = \inf_M \sup_{\vartheta \in \Theta} \|MP_\vartheta^1 - P_\vartheta^2\|_{TV}$ , where the infimum is taken over all randomisations or Markov kernels  $M$  from  $(\mathcal{X}_1, \mathcal{F}_1)$  to  $(\mathcal{X}_2, \mathcal{F}_2)$ , see e.g. Le Cam and Yang (2000) for details. In particular,  $\delta(\mathcal{E}_1, \mathcal{E}_2) = 0$  means that  $\mathcal{E}_1$  is more informative than  $\mathcal{E}_2$  in the sense that any observation in  $\mathcal{E}_2$  can be obtained from  $\mathcal{E}_1$ , possibly using additional randomisations. Here, we shall always explicitly con-

struct the transformations and randomisations and we shall then only use that  $\Delta(\mathcal{E}_1, \mathcal{E}_2) \leq \sup_{\vartheta \in \Theta} \|P_{\vartheta}^1 - P_{\vartheta}^2\|_{TV}$  holds when both experiments are defined on the same sample space.

The asymptotic equivalence is deduced stepwise. In Section 2 the regression-type model (1.1) is shown to be asymptotically equivalent to a corresponding white noise model with signal  $X$ . Then in Section 3, a very simple construction yields a Gaussian shift model with signal  $\log(\sigma^2(\bullet) + c)$ ,  $c > 0$  some constant, which is asymptotically less informative, but only by a constant factor in the Fisher information. Inspired by this construction, we present a generalisation in Section 4 where the information loss can be made arbitrarily small (but not zero), before applying nonparametric local asymptotic theory in Section 5 to derive asymptotic equivalence with our final Gaussian shift model for shrinking local neighbourhoods of the parameters. Section 6 yields the global result, which is based on an asymptotic sufficiency result for simple independent statistics.

Extensions and restrictions are discussed in Section 7 before we use the theoretical insight to construct in Section 8 a rate-optimal estimator of the spot volatility and an efficient estimator of the integrated volatility by a locally-constant approximation. Remarkably, the asymptotic variance is found to depend on the third moment  $\int_0^1 \sigma^3(t) dt$  and for non-constant  $\sigma^2(\bullet)$  our estimator outperforms previous approaches applied to the basic model. Constructions needed for the proof are presented and discussed alongside the mathematical results, deferring more technical parts to the Appendix, which in Section 9.1 also contains a summary of results on white noise models, the Hellinger distance and Hilbert-Schmidt norm estimates.

## 2 The regression and white noise model

In the main part we shall work in the white noise setting, which is more intuitive to handle than the regression setting, which in turn is the observation model in

practice. Let us define both models formally. For that we introduce the Hölder ball

$$C_\alpha(R) := \{f \in C^\alpha([0, 1]) \mid \|f\|_{C^\alpha} \leq R\} \text{ with } \|f\|_{C^\alpha} = \|f\|_\infty + \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x - y|^\alpha}.$$

**2.1 Definition.** Let  $\mathcal{E}_0 = \mathcal{E}_0(n, \delta, \alpha, R, \underline{\sigma}^2)$  with  $n \in \mathbb{N}$ ,  $\delta > 0$ ,  $\alpha \in (0, 1)$ ,  $R > 0$ ,  $\underline{\sigma}^2 \geq 0$  be the statistical experiment generated by observing (1.1). The volatility  $\sigma^2$  belongs to the class

$$\mathcal{S}(\alpha, R, \underline{\sigma}^2) := \left\{ \sigma^2 \in C^\alpha(R) \mid \min_{t \in [0, 1]} \sigma^2(t) \geq \underline{\sigma}^2 \right\}.$$

Let  $\mathcal{E}_1 = \mathcal{E}_1(\varepsilon, \alpha, R, \underline{\sigma}^2)$  with  $\varepsilon > 0$ ,  $\alpha \in (0, 1)$ ,  $R > 0$ ,  $\underline{\sigma}^2 \geq 0$  be the statistical experiment generated by observing

$$dY_t = X_t dt + \varepsilon dW_t, \quad t \in [0, 1],$$

with  $X_t = \int_0^t \sigma(s) dB_s$  as above, independent standard Brownian motions  $W$  and  $B$  and  $\sigma^2 \in \mathcal{S}(\alpha, R, \underline{\sigma}^2)$ .

From Brown and Low (1996) it is well known that the white noise and the Gaussian regression model are asymptotically equivalent for noise level  $\varepsilon = \delta/\sqrt{n} \rightarrow 0$  as  $n \rightarrow \infty$ , provided the signal is  $\beta$ -Hölder continuous for  $\beta > 1/2$ . Since Brownian motion and thus also our price process  $X$  is only Hölder continuous of order  $\beta < 1/2$  (whatever  $\alpha$  is), it is not clear whether asymptotic equivalence can hold for the experiments  $\mathcal{E}_0$  and  $\mathcal{E}_1$ . Yet, this is true. Subsequently, we employ the notation  $A_n \lesssim B_n$  if  $A_n = O(B_n)$  and  $A_n \sim B_n$  if  $A_n \lesssim B_n$  as well as  $B_n \lesssim A_n$  and obtain:

**2.2 Theorem.** For any  $\alpha > 0$ ,  $\underline{\sigma}^2 \geq 0$  and  $\delta, R > 0$  the experiments  $\mathcal{E}_0$  and  $\mathcal{E}_1$  with  $\varepsilon = \delta/\sqrt{n}$  are asymptotically equivalent ; more precisely:

$$\Delta(\mathcal{E}_0(n, \delta, \alpha, R, \underline{\sigma}^2), \mathcal{E}_1(\delta/\sqrt{n}, h, \alpha, R, \underline{\sigma}^2)) \lesssim R\delta^{-2}n^{-\alpha}.$$

Interestingly, the asymptotic equivalence holds for any positive Hölder regularity  $\alpha > 0$ . In particular, the volatility  $\sigma^2$  could be itself a continuous semimartingale, but such that  $X$  conditionally on  $\sigma^2$  remains Gaussian. As the proof in

Section 9.2 of the appendix reveals, we construct the equivalence by rate-optimal approximations of the anti-derivative of  $\sigma^2$  which lies in  $C^{1+\alpha}$ . Similar techniques have been used by Carter (2006) and Reiß (2008), but here we have to cope with the random signal for which we need to bound the Hilbert-Schmidt norm of the respective covariance operators. Note further that the asymptotic equivalence even holds when the level of the microstructure noise  $\delta$  tends to zero, provided  $\delta^2 n^\alpha \rightarrow \infty$  remains valid.

### 3 Less informative Gaussian shift experiments

From now on we shall work with the white noise observation experiment  $\mathcal{E}_1$ , where the main structures are more clearly visible. In this section we shall find easy Gaussian shift models which are asymptotically not more informative than  $\mathcal{E}_1$ , but already permit rate-optimal estimation results. The whole idea is easy to grasp once we can replace the volatility  $\sigma^2$  by a piecewise constant approximation on small blocks of size  $h$ . That this is no loss of generality, is shown by the subsequent asymptotic equivalence result, proved in Section 9.3 of the appendix.

**3.1 Definition.** Let  $\mathcal{E}_2 = \mathcal{E}_2(\varepsilon, h, \alpha, R, \underline{\sigma}^2)$  be the statistical experiment generated by observing

$$dY_t = X_t^h dt + \varepsilon dW_t, \quad t \in [0, 1],$$

with  $X_t^h = \int_0^t \sigma([\cdot]_h) dB_s$ ,  $[s]_h := [s/h]h$  for  $h > 0$  and  $h^{-1} \in \mathbb{N}$ , and independent standard Brownian motions  $W$  and  $B$ . The volatility  $\sigma^2$  belongs to the class  $\mathcal{S}(\alpha, R, \underline{\sigma}^2)$ .

**3.2 Proposition.** Assume  $\alpha > 1/2$  and  $\underline{\sigma}^2 > 0$ . Then for  $\varepsilon \rightarrow 0$ ,  $h^\alpha = o(\varepsilon^{1/2})$  the experiments  $\mathcal{E}_1$  and  $\mathcal{E}_2$  are asymptotically equivalent ; more precisely:

$$\Delta(\mathcal{E}_1(\varepsilon, \alpha, R, \underline{\sigma}^2), \mathcal{E}_2(\varepsilon, h, \alpha, R, \underline{\sigma}^2)) \lesssim R \underline{\sigma}^{-3/2} h^\alpha \varepsilon^{-1/2}.$$

In the sequel we always assume  $h^\alpha = o(\varepsilon^{1/2})$  to hold such that we can work equivalently with  $\mathcal{E}_2$ . Recall that observing  $Y$  in a white noise model is equivalent

to observing  $(\int e_m dY)_{m \geq 1}$  for an orthonormal basis  $(e_m)_{m \geq 1}$  of  $L^2([0, 1])$ , cf. also Subsection 9.1 below. Our first step is thus to find an orthonormal system (not a basis) which extracts as much *local* information on  $\sigma^2$  as possible. For any  $\varphi \in L^2([0, 1])$  with  $\|\varphi\|_{L^2} = 1$  we have by partial integration

$$\begin{aligned} \int_0^1 \varphi(t) dY_t &= \int_0^1 \varphi(t) X_t^h dt + \varepsilon \int_0^1 \varphi(t) dW_t \\ &= \Phi(1)X_1^h - \Phi(0)X_0^h - \int_0^1 \Phi(t)\sigma(\lfloor t \rfloor_h) dB_t + \varepsilon \int_0^1 \varphi(t) dW_t \\ &= \left( \int_0^1 \Phi^2(t)\sigma^2(\lfloor t \rfloor_h) dt + \varepsilon^2 \right)^{1/2} \zeta_\varphi \end{aligned} \quad (3.1)$$

where  $\Phi(t) = -\int_t^1 \varphi(s) ds$  is the antiderivative of  $\varphi$  with  $\Phi(1) = 0$  and  $\zeta_\varphi \sim N(0, 1)$  holds. To ensure that  $\Phi$  has only support in some interval  $[kh, (k+1)h]$ , we require  $\varphi$  to have support in  $[kh, (k+1)h]$  and to satisfy  $\int \varphi(t) dt = 0$ . The function  $\varphi_k$  with  $\text{supp}(\varphi_k) = [kh, (k+1)h]$ ,  $\|\varphi_k\|_{L^2} = 1$ ,  $\int \varphi_k(t) dt = 0$  that maximizes the information load  $\int \Phi_k^2(t) dt$  for  $\sigma^2(kh)$  is given by (use Lagrange theory)

$$\varphi_k(t) = \sqrt{2}h^{-1/2} \cos(\pi(t - kh)/h) \mathbf{1}_{[kh, (k+1)h]}(t), \quad t \in [0, 1]. \quad (3.2)$$

The  $L^2$ -orthonormal system  $(\varphi_k)$  for  $k = 0, 1, \dots, h^{-1} - 1$  is now used to construct Gaussian shift observations. In  $\mathcal{E}_2$  we obtain from (3.1) the observations

$$y_k := \int \varphi_k(t) dY_t = \left( h^2 \pi^{-2} \sigma^2(kh) + \varepsilon^2 \right)^{1/2} \zeta_k, \quad k = 0, \dots, h^{-1} - 1, \quad (3.3)$$

with independent standard normal random variables  $(\zeta_k)_{k=0, \dots, h^{-1}-1}$ . Observing  $(y_k)$  is clearly equivalent to observing

$$z_k := \log(y_k^2 h^{-2} \pi^2) - \mathbb{E}[\log(\zeta_k^2)] = \log\left(\sigma^2(kh) + \varepsilon^2 h^{-2} \pi^2\right) + \eta_k \quad (3.4)$$

for  $k = 0, \dots, h^{-1} - 1$  with  $\eta_k := \log(\zeta_k^2) - \mathbb{E}[\log(\zeta_k^2)]$ .

We have found a nonparametric regression model with regression function  $\log(\sigma^2(\bullet) + \varepsilon^2 h^{-2} \pi^2)$  and  $h^{-1}$  equidistant observations corrupted by non-Gaussian, but centered noise  $(\eta_k)$  of variance 2. To ensure that the regression function does not change under the asymptotics  $\varepsilon \rightarrow 0$ , we specify the block size  $h = h(\varepsilon) = h_0 \varepsilon$  with some fixed constant  $h_0 > 0$ .

It is not surprising that the nonparametric regression experiment in (3.4) is equivalent to a corresponding Gaussian shift experiment. Indeed, this follows readily from results by Grama and Nussbaum (2002) who in their Section 4.2 derive asymptotic equivalence already for our Gaussian scale model (3.3). Note, however, that their Fisher information should be  $I(\vartheta) = \frac{1}{2}\vartheta^{-2}$  and we thus have asymptotic equivalence of (3.3) with the Gaussian regression model

$$w_k = \frac{1}{\sqrt{2}} \log(\sigma^2(kh) + h_0^{-2}\pi^2) + \gamma_k, \quad k = 0, \dots, h^{-1} - 1,$$

where  $\gamma_k \sim N(0,1)$  i.i.d. Since by the classical result of Brown and Low (1996) the Gaussian regression is equivalent to the corresponding white noise experiment (note that  $\log(\sigma^2(\bullet) + h_0^{-2}\pi^2)$  is also  $\alpha$ -Hölder continuous), we have already derived an important and far-reaching result.

**3.3 Theorem.** *For  $\alpha > 1/2$  and  $\underline{\sigma}^2 > 0$  the high frequency experiment  $\mathcal{E}_1(\varepsilon, \alpha, R, \underline{\sigma}^2)$  is asymptotically more informative than the Gaussian shift experiment  $\mathcal{G}_1(\varepsilon, \alpha, R, \underline{\sigma}^2, h_0)$  of observing*

$$dZ_t = \frac{1}{\sqrt{2}} \log(\sigma^2(t) + h_0^{-2}\pi^2) dt + h_0^{1/2} \varepsilon^{1/2} dW_t, \quad t \in [0, 1].$$

Here  $h_0 > 0$  is an arbitrary constant and  $\sigma^2 \in \mathcal{S}(\alpha, R, \underline{\sigma}^2)$ .

**3.4 Remark.** *Moving the constants from the diffusion to the drift part, the experiment  $\mathcal{G}_1$  is equivalent to observing*

$$d\tilde{Z}_t = (2h_0)^{-1/2} \log(\sigma^2(t) + h_0^{-2}\pi^2) dt + \varepsilon^{1/2} dW_t, \quad t \in [0, 1]. \quad (3.5)$$

The Gaussian shift experiment is nonlinear in  $\sigma^2$  which is to be expected. Writing  $\varepsilon = \delta/\sqrt{n}$  gives us the noise level  $\delta^{1/2}n^{-1/4}$  which appears in all previous work on the model  $\mathcal{E}_0$ .

To quantify the amount of information we have lost, let us study the LAN-property of the constant parametric case  $\sigma^2(t) = \sigma^2 > 0$  in  $\mathcal{G}_1$ . We consider the local alternatives  $\sigma_\varepsilon^2 = \sigma_0^2 + \varepsilon^{1/2}$  for which we obtain the Fisher information  $I_{h_0} =$



$(2h_0)^{-1}h_0^4/(\pi^2 + h_0^2\sigma_0^2)^2$ . Maximizing over  $h_0$  yields  $h_0 = \sqrt{3}\pi\sigma_0^{-1}$  and the Fisher information is at most equal to

$$\sup_{h_0 > 0} I_{h_0} = \sigma_0^{-3} 3^{3/2}/(32\pi) \approx 0.0517\sigma_0^{-3}.$$

By the LAN-result of Gloter and Jacod (2001a) for  $\mathcal{E}_0$  the best value is  $I(\sigma_0) = \frac{1}{8}\sigma_0^{-3}$  which is clearly larger. Note, however, that the relative (normalized) efficiency is already  $\frac{\sqrt{3^{3/2}/(32\pi)}}{\sqrt{1/8}} \approx 0.64$ , which means that we attain about 64% of the precision when working with  $\mathcal{G}_1$  instead of  $\mathcal{E}_0$  or  $\mathcal{E}_1$ .

## 4 A close sequence of simple models

In order to decrease the information loss in  $\mathcal{G}_1$ , we now take into account higher frequencies in each block  $[kh, (k+1)h]$ . In a frequency-location notation  $(j, k)$  we consider for  $k = 0, 1, \dots, h^{-1} - 1, j \geq 1$

$$\varphi_{jk}(t) = \sqrt{2}h^{-1/2} \cos(j\pi(t - kh)/h) \mathbf{1}_{[kh, (k+1)h]}(t), \quad t \in [0, 1]. \quad (4.1)$$

This gives the corresponding antiderivatives

$$\Phi_{jk}(t) = \frac{\sqrt{2h}}{\pi j} \sin(j\pi(t - kh)/h) \mathbf{1}_{[kh, (k+1)h]}(t), \quad t \in [0, 1].$$

Not only the  $(\varphi_{jk})$  and  $(\Phi_{jk})$  are localized on each block, also each single family of functions is orthogonal in  $L^2([0, 1])$ . Working again on the piecewise constant experiment  $\mathcal{E}_2$ , we extract the observations

$$y_{jk} := \int_0^1 \varphi_{jk}(t) dY_t = \left( h^2 \pi^{-2} j^{-2} \sigma^2(kh) + \varepsilon^2 \right)^{1/2} \zeta_{jk}, \quad j \geq 1, k = 0, \dots, h^{-1} - 1, \quad (4.2)$$

with  $\zeta_{jk} \sim N(0, 1)$  independent over all  $(j, k)$ . The same transformation as before leads for each  $j \geq 1$  to the regression model for  $k = 0, \dots, h^{-1} - 1$

$$z_{jk} := \log(y_{jk}^2) - \log(h^2 \pi^{-2} j^{-2}) - \mathbb{E}[\log(\zeta_{jk}^2)] = \log(\sigma^2(t) + \varepsilon^2 h^{-2} \pi^2 j^2) + \eta_{jk}. \quad (4.3)$$

Applying the asymptotic equivalence result by Grama and Nussbaum (2002) for each independent level  $j$  separately, we immediately generalize Theorem 3.3.

**4.1 Theorem.** For  $\alpha > 1/2$  and  $\underline{\sigma}^2 > 0$  the high frequency experiment  $\mathcal{E}_1(\varepsilon, \alpha, R, \underline{\sigma}^2)$  is asymptotically more informative than the combined experiment  $\mathcal{G}_2(\varepsilon, \alpha, R, \underline{\sigma}^2, h_0, J)$  of independent Gaussian shifts

$$dZ_t^j = \frac{1}{\sqrt{2}} \log(\sigma^2(t) + h_0^{-2} \pi^2 j^2) dt + h_0^{1/2} \varepsilon^{1/2} dW_t^j, \quad t \in [0, 1], j = 1, \dots, J,$$

with independent Brownian motions  $(W^j)_{j=1, \dots, J}$  and  $\sigma^2 \in \mathcal{S}(\alpha, R, \underline{\sigma}^2)$ . The constants  $h_0 > 0$  and  $J \in \mathbb{N}$  are arbitrary, but fixed.

**4.2 Remark.** Let us again study the LAN-property of the constant parametric case  $\sigma^2(t) = \sigma^2 > 0$  for the local alternatives  $\sigma_\varepsilon^2 = \sigma_0^2 + \varepsilon^{1/2}$ . We obtain the Fisher information

$$I_{h_0, J} = \sum_{j=1}^J (2h_0)^{-1} h_0^4 (\pi^2 j^2 + h_0^2 \sigma_0^2)^{-2} = \sum_{j=1}^J \frac{h_0^{-1}}{2(\pi^2 (j h_0^{-1})^2 + \sigma_0^2)^2}.$$

In the limit  $J \rightarrow \infty$  and  $h_0 \rightarrow \infty$  we obtain by Riemann sum approximation

$$\lim_{h_0 \rightarrow \infty} \lim_{J \rightarrow \infty} I_{h_0, J} = \int_0^\infty \frac{dx}{2(\pi^2 x^2 + \sigma_0^2)^2} = \frac{1}{8\sigma_0^3}.$$

This is exactly the optimal Fisher information, obtained by Gloter and Jacod (2001a) in this case. Note, however, that it is not at all obvious that we may let  $J, h_0 \rightarrow \infty$ , in the asymptotic equivalence result. Moreover, in our theory the restriction  $h^\alpha = o(\varepsilon^{1/2})$  is necessary, which translates into  $h_0 = o(\varepsilon^{(1-2\alpha)/2\alpha})$ . Still, the positive aspect is that we can come as close as we wish to an asymptotically almost equivalent, but much simpler model.

## 5 Localisation

We know from standard regression theory (Stone 1982) that in the experiment  $\mathcal{G}_1$  we can estimate  $\sigma^2 \in C^\alpha$  in sup-norm with rate  $(\varepsilon \log(\varepsilon^{-1}))^{\alpha/(2\alpha+1)}$ , using that the log-function is a  $C^\infty$ -diffeomorphism for arguments bounded away from zero and infinity. Since  $\mathcal{E}_1$  is for  $\alpha > 1/2$  asymptotically more informative than  $\mathcal{G}_1$ , we can therefore localize  $\sigma^2$  in a neighbourhood of some  $\sigma_0^2$ . Using the local coordinate  $s^2$  in  $\sigma^2 = \sigma_0^2 + v_\varepsilon s^2$  for  $v_\varepsilon \rightarrow 0$  we define a localized experiment, cf. Nussbaum (1996).

**5.1 Definition.** Let  $\mathcal{E}_{i,loc} = \mathcal{E}_{i,loc}(\sigma_0, \varepsilon, \alpha, R, \underline{\sigma}^2)$  for  $\sigma_0 \in \mathcal{S}(\alpha, R, \underline{\sigma}^2)$  be the statistical subexperiment obtained from  $\mathcal{E}_i(\varepsilon, \alpha, R, \underline{\sigma}^2)$  by restricting to the parameters  $\sigma^2 = \sigma_0^2 + v_\varepsilon s^2$  with  $v_\varepsilon = \varepsilon^{\alpha/(2\alpha+1)} \log(\varepsilon^{-1})$  and unknown  $s^2 \in C^\alpha(R)$ .

We shall consider the observations  $(y_{jk})$  in (4.2) derived from  $\mathcal{E}_{2,loc}$  and multiplied by  $\pi j/h$ . The model is then a generalized nonparametric regression family in the sense of Grama and Nussbaum (2002). On the sequence space  $(\mathcal{X}, \mathcal{F}) = (\mathbb{R}^{\mathbb{N}}, \mathfrak{B}^{\otimes \mathbb{N}})$  we consider for  $\vartheta \in \Theta = [\underline{\sigma}^2, R]$  the Gaussian product measure

$$\mathbb{P}_\vartheta = \bigotimes_{j \geq 1} N(0, \vartheta + h_0^{-2} \pi^2 j^2). \quad (5.1)$$

The parameter  $\vartheta$  plays the role of  $\sigma^2(kh)$  for each  $k$ . By independence and the result for the one-dimensional Gaussian scale model, the Fisher information for  $\vartheta$  is given by

$$I(\vartheta) := \sum_{j \geq 1} \frac{1}{2(\vartheta + h_0^{-2} \pi^2 j^2)^2} = \frac{h_0}{8\vartheta^{3/2}} \left( \frac{1 + 4\vartheta^{1/2} h_0 e^{-2\vartheta^{1/2} h_0} - e^{-4\vartheta^{1/2} h_0}}{(1 - e^{-2\vartheta^{1/2} h_0})^2} - \frac{2}{\vartheta^{1/2} h_0} \right), \quad (5.2)$$

where the series is evaluated in Section 9.6 using Fourier analysis. Since we shall later let  $h_0$  tend to infinity, an essential point is the asymptotics  $I(\vartheta) \sim h_0$ .

We split our observation design  $\{kh \mid k = 0, \dots, h^{-1}\}$  into blocks  $A_m = \{kh \mid k = (m-1)\ell, \dots, m\ell - 1\}$ ,  $m = 1, \dots, (\ell h)^{-1}$ , of length  $\ell$  such that the radius  $v_\varepsilon$  of our nonparametric local neighbourhood has the order of the *parametric* noise level  $(I(\vartheta)\ell)^{-1/2}$  in each block:

$$v_\varepsilon \sim (I(\vartheta)\ell)^{-1/2} \Rightarrow \ell \sim h_0^{-1} v_\varepsilon^{-2}.$$

For later convenience we consider odd and even indices  $k$  separately, assuming that  $h^{-1}$  and  $\ell$  are even integers. This way, for each block  $m$  observing  $(y_{jk} \pi j/h)$  for  $j \geq 1$  and  $k \in A_m$ ,  $k$  odd respectively  $k$  even, can be modeled by the experiments

$$\mathcal{E}_{3,m}^{odd} = \left( \mathcal{X}^{\ell/2}, \mathcal{F}^{\otimes \ell/2}, \left( \bigotimes_{k \in A_m \text{ odd}} \mathbb{P}_{\sigma_0^2(k/n) + v_\varepsilon s^2(k/n)} \right)_{s^2 \in C_\alpha(R)} \right), \quad (5.3)$$

$$\mathcal{E}_{3,m}^{even} = \left( \mathcal{X}^{\ell/2}, \mathcal{F}^{\otimes \ell/2}, \left( \bigotimes_{k \in A_m \text{ even}} \mathbb{P}_{\sigma_0^2(k/n) + v_\varepsilon s^2(k/n)} \right)_{s^2 \in C_\alpha(R)} \right), \quad (5.4)$$

where all parameters are the same as for  $\mathcal{E}_{2,loc}$ . Using the nonparametric local asymptotic theory developed by Grama and Nussbaum (2002) and the independence of the experiments  $(\mathcal{E}_{3,m}^{odd})_m$  (resp.  $(\mathcal{E}_{3,m}^{even})_m$ ), we are able to prove in Section 9.4 the following asymptotic equivalence.

**5.2 Proposition.** *Assume  $\alpha > 1/2$ ,  $\underline{\sigma}^2 > 0$  and  $h_0 \sim \varepsilon^{-p}$  with  $p \in (0, 1 - (2\alpha)^{-1})$  such that  $(2h)^{-1} \in \mathbb{N}$ . Then observing  $\{y_{j,2k+1} \mid j \geq 1, k = 0, \dots, (2h)^{-1} - 1\}$  in experiment  $\mathcal{E}_{2,loc}$  is asymptotically equivalent to the local Gaussian shift experiment  $\mathcal{G}_{3,loc}$  of observing*

$$dY_t = \frac{1}{\sqrt{8}\sigma_0^{3/2}(t)} \left(1 - \frac{2}{\sigma_0(t)h_0}\right)^{1/2} v_\varepsilon s^2(t) dt + (2\varepsilon)^{1/2} dW_t, \quad t \in [0, 1], \quad (5.5)$$

where the unknown  $s^2$  and all parameters are the same as in  $\mathcal{E}_{2,loc}$ . The Le Cam distance tends to zero uniformly over the center of localisation  $\sigma_0^2 \in \mathcal{S}(\alpha, R, \underline{\sigma}^2)$ .

The same asymptotic equivalence result holds true for observing  $\{y_{j,2k} \mid j \geq 1, k = 0, \dots, (2h)^{-1} - 1\}$  in experiment  $\mathcal{E}_{2,loc}$ .

Note that in this model, combining even and odd indices  $k$ , we can already infer the LAN-result by Gloter and Jacod (2001a), but we still face a second order term of order  $h_0^{-1}v_\varepsilon$  in the drift. This term is asymptotically negligible only if it is of smaller order than the noise level  $\varepsilon^{1/2}$ . To be able to choose  $h_0$  sufficiently large, we have to require a larger Hölder smoothness of the volatility.

**5.3 Corollary.** *Assume  $\alpha > \frac{1+\sqrt{17}}{8} \approx 0.64$ ,  $\underline{\sigma}^2 > 0$  and  $(2h)^{-1} \in \mathbb{N}$ . Then observing  $\{y_{j,2k+1} \mid j \geq 1, k = 0, \dots, (2h)^{-1} - 1\}$  in experiment  $\mathcal{E}_{2,loc}$  is asymptotically equivalent to the local Gaussian shift experiment  $\mathcal{G}_{4,loc}$  of observing*

$$dY_t = \frac{1}{\sqrt{8}\sigma_0^{3/2}(t)} v_\varepsilon s^2(t) dt + (2\varepsilon)^{1/2} dW_t, \quad t \in [0, 1], \quad (5.6)$$

where the unknown  $s^2$  and all parameters are the same as in  $\mathcal{E}_{2,loc}$ . The Le Cam distance tends to zero uniformly over the center of localisation  $\sigma_0^2 \in \mathcal{S}(\alpha, R, \underline{\sigma}^2)$ .

The same asymptotic equivalence result holds true for observing  $\{y_{j,2k} \mid j \geq 1, k = 0, \dots, (2h)^{-1} - 1\}$  in experiment  $\mathcal{E}_{2,loc}$ .

*Proof.* For  $\alpha > \frac{1+\sqrt{17}}{8}$  the choice of  $h_0 = \varepsilon^{-p}$  for some  $p \in (\frac{1}{4\alpha+2}, \frac{2\alpha-1}{2\alpha})$  is possible and ensures that  $h^\alpha = o(\varepsilon^{1/2})$  holds as well as  $h_0^{-2} = o(v_\varepsilon^{-2}\varepsilon)$ . Therefore the Kullback-Leibler divergence between the observations in  $\mathcal{G}_3^{loc}$  and in  $\mathcal{G}_4^{loc}$  evaluates by the Cameron-Martin (or Girsanov) formula to

$$\varepsilon^{-1} \int_0^1 \frac{1}{8\sigma_0^3(t)} \left( \left(1 - \frac{2}{\sigma_0(t)h_0}\right)^{1/2} - 1 \right)^2 v_\varepsilon^2 s^4(t) dt \lesssim \varepsilon^{-1} h_0^2 v_\varepsilon^2.$$

Consequently, the Kullback-Leibler and thus also the total variation distance tends to zero.  $\square$

In a last step we find local experiments  $\mathcal{G}_{5,loc}$ , which are asymptotically equivalent to  $\mathcal{G}_{4,loc}$  and do not depend on the center of localisation  $\sigma_0^2$ . To this end we use a variance-stabilizing transform, based on the Taylor expansion

$$\sqrt{2x^{1/4}} = \sqrt{2x_0^{1/4}} + \frac{1}{\sqrt{8}}x_0^{-3/4}(x - x_0) + O((x - x_0)^2)$$

which holds uniformly over  $x, x_0$  on any compact subset of  $(0, \infty)$ . Inserting  $x = \sigma^2(t) = \sigma_0^2(t) + v_\varepsilon s^2(t)$  and  $x_0 = \sigma_0^2$  from our local model, we obtain

$$\sqrt{2\sigma(t)} = \sqrt{2\sigma_0(t)} + \frac{1}{\sqrt{8}}\sigma_0^{-3/2}(t)v_\varepsilon s^2(t) + O(v_\varepsilon^2). \quad (5.7)$$

Since  $v_\varepsilon^2 = o(\varepsilon^{1/2})$  holds for  $\alpha > 1/2$ , we can add the uninformative signal  $\sqrt{2}\sigma_0^{1/2}(t)$  to  $Y$  in  $\mathcal{G}_{4,loc}$ , replace the drift by  $\sqrt{2}\sigma^{1/2}(t)$  and still keep convergence of the total variation distance, compare the preceding proof. Consequently, from Corollary 5.3 we obtain the following result.

**5.4 Corollary.** *Assume  $\alpha > \frac{1+\sqrt{17}}{8} \approx 0.64$ ,  $\underline{\sigma}^2 > 0$  and  $(2h)^{-1} \in \mathbb{N}$ . Then observing  $\{y_{j,2k+1} \mid j \geq 1, k = 0, \dots, (2h)^{-1} - 1\}$  in the experiment  $\mathcal{E}_{2,loc}$  is asymptotically equivalent to the local Gaussian shift experiment  $\mathcal{G}_{5,loc}$  of observing*

$$dY_t = \sqrt{2\sigma(t)} dt + (2\varepsilon)^{1/2} dW_t, \quad t \in [0, 1], \quad (5.8)$$

where the unknown is  $\sigma^2 = \sigma_0^2 + v_\varepsilon s^2$  and all parameters are the same as in  $\mathcal{E}_{2,loc}$ . The Le Cam distance tends to zero uniformly over the center of localisation  $\sigma_0^2 \in \mathcal{S}(\alpha, R, \underline{\sigma}^2)$ .

The same asymptotic equivalence result holds true for observing  $\{y_{j,2k} \mid j \geq 1, k = 0, \dots, (2h)^{-1} - 1\}$  in experiment  $\mathcal{E}_{2,loc}$ .

## 6 Globalisation

The globalisation now basically follows the usual route, first established by Nussbaum (1996). Essential for us is to show that observing  $(y_{jk})$  for  $j \geq 1$  is asymptotically sufficient in  $\mathcal{E}_2$ . Then we can split the white noise observation experiment  $\mathcal{E}_2$  into two independent sub-experiments obtained from  $(y_{jk})$  for  $k$  odd and  $k$  even, respectively. Usually, a white noise experiment can be split into two independent subexperiments with the same drift and an increase by  $\sqrt{2}$  in the noise level. Here, however, this does not work since the two diffusions in the *random* drift remain the same and thus independence fails.

Let us introduce the  $L^2$ -normalized step functions

$$\begin{aligned}\varphi_{0,k}(t) &:= (2h)^{-1/2}(\mathbf{1}_{[(k-1)h, kh]}(t) - \mathbf{1}_{[kh, (k+1)h]}(t)), \quad k = 1, \dots, h^{-1} - 1, \\ \varphi_{0,0}(t) &:= h^{-1/2}\mathbf{1}_{[0,h]}(t).\end{aligned}$$

We obtain a normalized complete basis  $(\varphi_{jk})_{j \geq 0, 0 \leq k \leq h^{-1} - 1}$  of  $L^2([0, 1])$  such that observing  $Y$  in experiment  $\mathcal{E}_2$  is equivalent to observing

$$y_{jk} := \int_0^1 \varphi_{jk}(t) dY_t, \quad j \geq 0, k = 0, \dots, h^{-1} - 1.$$

Calculating the Fourier series, we can express the tent function  $\Phi_{0,k}$  with  $\Phi'_{0,k} = \varphi_{0,k}$  and  $\Phi_{0,k}(1) = 0$  as an  $L^2$ -convergent series over the dilated sine functions  $\Phi_{jk}$  and  $\Phi_{j,k-1}$ ,  $j \geq 1$ :

$$\Phi_{0,k}(t) = \sum_{j \geq 1} (-1)^{j+1} \Phi_{j,k-1}(t) + \sum_{j \geq 1} \Phi_{jk}(t), \quad k = 1, \dots, h^{-1} - 1. \quad (6.1)$$

We also have  $\Phi_{0,0}(t) = 2 \sum_{j \geq 1} \Phi_{j,0}(t)$ . By partial integration, this implies (with

$L^2$ -convergence)

$$\beta_{0,k} := \langle \varphi_{0,k}, X \rangle = - \int_0^1 \Phi_{0,k}(t) dX(t) = \sum_{j \geq 1} (-1)^{j+1} \beta_{j,k-1} + \sum_{j \geq 1} \beta_{jk},$$

where  $\beta_{jk} := \langle \varphi_{jk}, X \rangle$

for  $k \geq 1$  and similarly  $\beta_{0,0} = 2 \sum_{j \geq 1} \beta_{j,0}$ . This means that the signal  $\beta_{0,k}$  in  $y_{0,k}$  can be perfectly reconstructed from the signals in the  $y_{j,k-1}, y_{jk}$ . For jointly Gaussian random variables we obtain the conditional law in  $\mathcal{E}_2$

$$\mathcal{L}(\beta_{jk} | y_{jk}) = N\left(\frac{\text{Var}(\beta_{jk})}{\text{Var}(y_{jk})} y_{jk}, \frac{\varepsilon^2 \text{Var}(\beta_{jk})}{\text{Var}(y_{jk})}\right)$$

Given the results by Stone (1982) and our less informative Gaussian shift experiment  $\mathcal{G}_1$  for  $\alpha > 1/2$ ,  $\underline{\sigma}^2 > 0$ , there is an estimator  $\hat{\sigma}_\varepsilon^2$  based on  $(y_{1,k})_k$  in  $\mathcal{E}_2$  with

$$\lim_{\varepsilon \rightarrow 0} \inf_{\sigma^2 \in \mathcal{S}} \mathbb{P}_{\sigma^2, \varepsilon}(\|\hat{\sigma}_\varepsilon^2 - \sigma^2\|_\infty \leq Rv_\varepsilon) = 1, \quad (6.2)$$

where  $v_\varepsilon = \varepsilon^{\alpha/(2\alpha+1)} \log(\varepsilon^{-1})$  as in the definitions of the localized experiments.

We can thus generate independent  $N(0, 1)$ -distributed random variables  $\rho_{jk}$  to construct from  $(y_{jk})_{j \geq 1, k}$

$$\tilde{\beta}_{jk} := \frac{\text{Var}_\varepsilon(\beta_{jk})}{\text{Var}_\varepsilon(y_{jk})} y_{jk} + \frac{\varepsilon \text{Var}_\varepsilon(\beta_{jk})^{1/2}}{\text{Var}_\varepsilon(y_{jk})^{1/2}} \rho_{jk},$$

where the variance  $\text{Var}_\varepsilon$  is the expression for  $\text{Var}$  where all unknown values  $\sigma^2(kh)$  are replaced by the estimated values  $\hat{\sigma}_\varepsilon^2(kh)$ . From this we can generate artificial observations  $(\tilde{y}_{0,k})_k$  such that the conditional law  $\mathcal{L}((\tilde{y}_{0,k})_k | (\tilde{\beta}_{j,k})_k)$  coincides with  $\mathcal{L}((y_{0,k})_k | (\beta_{0,k})_k)$ , which is just a multivariate normal law with mean zero and tri-diagonal covariance matrix  $\varepsilon^2(\langle \varphi_{0,k}, \varphi_{0,k'} \rangle)_{k,k'}$ .

In Section 9.5 we shall prove that the Hellinger distance between the families of centered Gaussian random variables  $\mathcal{Y} := \{y_{jk} | j \geq 0, k = 0, \dots, h^{-1} - 1\}$  and  $\tilde{\mathcal{Y}} := \{\tilde{y}_{0,k} | k = 0, \dots, h^{-1} - 1\} \cup \{y_{jk} | j \geq 1, k = 0, \dots, h^{-1} - 1\}$  tends to zero, provided  $h_0^{-1} v_\varepsilon^2 = o(\varepsilon)$ , which is possible when  $\alpha > \frac{1+\sqrt{5}}{4}$  with the choice  $h_0 = \varepsilon^{-p}$  for some  $p \in (\frac{1}{2\alpha+1}, \frac{2\alpha-1}{2\alpha})$ .

**6.1 Proposition.** *Assume  $\alpha > \frac{1+\sqrt{5}}{4} \approx 0.81$ ,  $\underline{\sigma}^2 > 0$  and  $h^{-1}$  an even integer. Then the experiment  $\mathcal{E}_2$  is asymptotically equivalent to the product experiment  $\mathcal{E}_{2,odd} \otimes \mathcal{E}_{2,even}$  where  $\mathcal{E}_{2,odd}$  is obtained from the observations  $\{y_{j,2k+1} \mid j \geq 1, k = 0, \dots, (2h)^{-1} - 1\}$  and  $\mathcal{E}_{2,even}$  from the observations  $\{y_{j,2k} \mid j \geq 1, k = 0, \dots, (2h)^{-1} - 1\}$  in experiment  $\mathcal{E}_2$ .*

This key result permits to globalize the local result. In the sequel we always assume  $\alpha > \frac{1+\sqrt{5}}{4}$  and  $\underline{\sigma}^2 > 0$ . We start with the asymptotic equivalence between  $\mathcal{E}_2$  and  $\mathcal{E}_{2,odd} \otimes \mathcal{E}_{2,even}$ . Using again an estimator  $\hat{\sigma}_\varepsilon^2$  in  $\mathcal{E}_{2,odd}$  satisfying (6.2) we can localize the second factor  $\mathcal{E}_{2,even}$  around  $\hat{\sigma}_\varepsilon^2$  and therefore by Corollary 5.4 replace it by experiment  $\mathcal{G}_{5,loc}$ , see Theorem 3.2 in Nussbaum (1996) for a formal proof. Since  $\mathcal{G}_{5,loc}$  does not depend on the center  $\hat{\sigma}_\varepsilon^2$ , we conclude that  $\mathcal{E}_2$  is asymptotically equivalent to the product experiment  $\mathcal{E}_{2,odd} \otimes \mathcal{G}_5$  where  $\mathcal{G}_5$  has the same parameters as  $\mathcal{E}_2$  and is given by observing  $Y$  in (5.8). Now we use an estimator  $\hat{\sigma}_\varepsilon^2$  in  $\mathcal{G}_5$  satisfying (6.2), whose existence is ensured by Stone (1982), to localize  $\mathcal{E}_{2,odd}$ . Corollary 5.4 then allows again to replace the localized  $\mathcal{E}_{2,odd}$ -experiment by  $\mathcal{G}_5$  such that  $\mathcal{E}_2$  is asymptotically equivalent to the product experiment  $\mathcal{G}_5 \otimes \mathcal{G}_5$ . Finally, taking the mean of the independent observations (5.8) in both factors, which is a sufficient statistics, (or, abstractly, due to identical likelihood processes) we see that  $\mathcal{G}_5 \otimes \mathcal{G}_5$  is equivalent to the experiment  $\mathcal{G}_0$  of observing  $dY_t = \sqrt{2\sigma(t)} dt + \sqrt{\varepsilon} dW_t$ ,  $t \in [0, 1]$ . Our final result then follows from the asymptotic equivalence between  $\mathcal{E}_0$  and  $\mathcal{E}_1$  as well as between  $\mathcal{E}_1$  and  $\mathcal{E}_2$ .

**6.2 Theorem.** *Assume  $\alpha > \frac{1+\sqrt{5}}{4} \approx 0.81$  and  $\delta, \underline{\sigma}^2, R > 0$ . Then the regression experiment  $\mathcal{E}_0(n, \delta, \alpha, R, \underline{\sigma}^2)$  is for  $n \rightarrow \infty$  asymptotically equivalent to the Gaussian shift experiment  $\mathcal{G}_0(\delta n^{-1/2}, \alpha, R, \underline{\sigma}^2)$  of observing*

$$dY_t = \sqrt{2\sigma(t)} dt + \delta^{1/2} n^{-1/4} dW_t, \quad t \in [0, 1], \quad (6.3)$$

for  $\sigma^2 \in \mathcal{S}(\alpha, R, \underline{\sigma}^2)$ .



## 7 Discussion

Our results show that inference for the volatility in the high-frequency observation model under microstructure noise  $\mathcal{E}_0$  is asymptotically as difficult as in the well understood Gaussian shift model  $\mathcal{G}_0$ . Remark that the constructions in Gloter and Jacod (2001a), Gloter and Jacod (2001b) rely on preliminary estimators at the boundary of suitable blocks, while we require  $\text{supp } \Phi_{jk} = [kh, (k+1)h]$  to obtain independence among blocks. In this context Proposition 6.1 shows asymptotic sufficiency of observing only the pinned process  $X_t - \frac{(k+1)h-t}{h}X_{kh} - \frac{t-kh}{h}X_{(k+1)h}$ ,  $t \in [kh, (k+1)h]$ , on each block due to  $\int(\alpha t + \beta)\varphi_{jk}(t) dt = 0$  for  $j \geq 1$ ,  $\alpha, \beta \in \mathbb{R}$ . Naturally, the  $(\Phi_{jk})_{j \geq 1}$  form exactly the eigenfunctions of the covariance operator of the Brownian bridge.

It is interesting to note that both, model  $\mathcal{E}_0$  and model  $\mathcal{G}_0$ , are homogeneous in the sense that factors from the noise (i.e. the  $dW_t$ -term) can be moved to the drift term and vice versa such that for example high volatility can counterbalance a high noise level  $\delta$  or a large observation distance  $1/n$ . Another phenomenon is that observing  $\mathcal{E}_0$   $m$ -times independently, in particular with different realisations of the process  $X$ , is asymptotically as informative as observing  $\mathcal{E}_0$  with  $m^2$  as many observations: both experiments are asymptotically equivalent to  $dY_t = \sqrt{2\sigma(t)}dt + m^{1/2}\delta^{1/2}n^{-1/4}dW_t$ . Similarly, by rescaling we can treat observations on intervals  $[0, T]$  with  $T > 0$  fixed: Observing  $Y_i = X_{iT/n} + \varepsilon_i$ ,  $i = 1, \dots, n$ , in  $\mathcal{E}_0$  with  $X_t = \int_0^t \sigma(s) dB_s$ ,  $t \in [0, T]$ , is under the same conditions asymptotically equivalent to observing

$$dY_u = \sqrt{2\sigma(Tu)} du + \delta^{1/2}T^{-1/4}n^{-1/4}dW_u, \quad u \in [0, 1],$$

or equivalently,

$$d\tilde{Y}_v = \sqrt{2\sigma(v)} dv + \delta^{1/2}T^{1/4}n^{-1/4}dW_v, \quad v \in [0, T].$$

Concerning the various restrictions on the smoothness  $\alpha$  of the volatility  $\sigma^2$ , one might wonder whether the critical index is  $\alpha = 1/2$  in view of the classical asymp-

otic equivalence results (Brown and Low 1996, Nussbaum 1996). In our approach, we still face the second order term in (5.5) and using the localized results, a much easier globalisation yields for  $\alpha > 1/2$  only that  $\mathcal{E}_0$  is asymptotically not less informative than observing

$$dY_t = F(\sigma^2(t)) dt + \delta^{1/2} n^{-1/4} dW_t, \quad t \in [0, 1],$$

with  $F(x) = \int_1^x (y^{1/2} - 2h_0^{-1})^{1/2} y^{-1} dy / \sqrt{8}$ , which includes a small, but non-negligible second-order term since  $h_0$  cannot tend to infinity too quickly.

On the other hand, it is quite easy to see that for  $\alpha \leq 1/4$  asymptotic equivalence fails. In the regression model  $\mathcal{E}_0$  with  $n$  observations we cannot distinguish between  $X_n(t) = \int_0^t \sigma_n(t) dB_t$  with  $\sigma_n^2(t) = 1 + n^{-1/4} \cos(\pi nt)$ ,  $\|\sigma_n^2\|_{C^{1/4}} = 2 + n^{-1/4}$ , and standard Brownian motion ( $\sigma^2 = 1$ ) since  $X_n(i/n) - X_n((i-1)/n) \sim N(0, 1/n)$  i.i.d. holds. On the other hand, we have  $\int_0^1 (\sqrt{2\sigma_n(t)} - \sqrt{2})^2 dt \sim n^{-1/2}$ , which shows that the signal to noise ratio in the Gaussian shift  $\mathcal{G}_0$  is of order 1 and a Neyman-Pearson test between  $\sigma_n^2$  and 1 can distinguish both signals with a positive probability. This different behaviour for testing in  $\mathcal{E}_0$  and  $\mathcal{G}_0$  implies that both models cannot be asymptotically equivalent for  $\alpha = 1/4$ . Note that Gloter and Jacod (2001a) merely require  $\alpha \geq 1/4$  for their LAN-result, but our counterexample is excluded by their parametric setting. In conclusion, the behaviour in the zone  $\alpha \in (1/4, (1 + \sqrt{5})/4]$  remains unexplored.

## 8 Applications

Let us first consider the nonparametric problem of estimating the spot volatility  $\sigma^2(t)$ . From our asymptotic equivalence result in Theorem 6.2 we can deduce, at least for bounded loss functions, the usual nonparametric minimax rates, but with the number  $n$  of observations replaced by  $\sqrt{n}$  provided  $\sigma^2 \in C^\alpha$  for  $\alpha > (1 + \sqrt{5})/4$  as the mapping  $\sqrt{\sigma(t)} \mapsto \sigma^2(t)$  is a  $C^\infty$ -diffeomorphism for volatilities  $\sigma^2$  bounded away from zero. Since the results so far obtained only deal with rate results, it is even

simpler to use our less informative model  $\mathcal{G}_1$  or more concretely the observations  $(y_k)$  in (3.3) which are independent in  $\mathcal{E}_2$ , centered and of variance  $h^2\pi^{-2}\sigma^2(kh)+\varepsilon^2$ . With  $h = \varepsilon$  a local (kernel or wavelet) averaging over  $\varepsilon^{-2}\pi^2 y_k^2 - \pi^2$  therefore yields rate-optimal estimators for classical pointwise or  $L^p$ -type loss functions.

For later use we choose  $h = \varepsilon$  in  $\mathcal{E}_2$  and propose the simple estimator

$$\hat{\sigma}_b^2(t) := \frac{\varepsilon}{2b} \sum_{k:|k\varepsilon-t|\leq b} (\varepsilon^{-2}\pi^2 y_k^2 - \pi^2)$$

for some bandwidth  $b > 0$ . Since  $\zeta_k^2$  is  $\chi^2(1)$ -distributed, it is standard (Stone 1982) to show that with the choice  $b \sim (\varepsilon \log(\varepsilon^{-1}))^{1/(2\alpha+1)}$  we have the sup-norm risk bound

$$\mathbb{E}[\|\hat{\sigma}_b^2 - \sigma^2\|_\infty^2] \lesssim (\varepsilon \log(\varepsilon^{-1}))^{2\alpha/(2\alpha+1)},$$

especially we shall need that  $\hat{\sigma}_b^2$  is consistent in sup-norm loss.

In terms of the regression experiment  $\mathcal{E}_0$  we work (in an asymptotically equivalent way) with the linear interpolation  $\hat{Y}'$  of the observations  $(Y_i)$ , see the proof of Theorem 2.2. By partial integration we can thus take for any  $j, k$

$$y_{jk}^0 := - \int_0^1 \Phi_{jk}(t) \hat{Y}''(t) dt = \sum_{i=1}^n \left( - \int_{(i-1)/n}^{i/n} \Phi_{jk}(t) dt \right) (Y_i - Y_{i-1}), \quad (8.1)$$

setting  $Y_0 := 0$ . Note that we have the uniform approximation  $y_{jk}^0 = \frac{-1}{n} \sum_{i=1}^n \Phi_{jk}(i/n)(Y_i - Y_{i-1}) + O(h^{-1/2}n^{-1})$  due to  $\|\varphi_{jk}\|_\infty \leq (2h)^{-1/2}$ . We see the relationship with the pre-averaging approach. The idea of using disjoint averages is present in Podolskij and Vetter (2009), where in our terminology Haar functions are used as  $\Phi_k$ . They were aware of the fact that discretized sine functions would slightly increase the Fisher information (personal communication, see also their discussion after Corollary 2), but they have not used higher frequencies.

Since we use the concrete coupling by linear interpolation to define  $y_{jk}^0$  in  $\mathcal{E}_0$  and since convergence in total variation is stronger than weak convergence, all asymptotics for probabilities and weak convergence results for functionals  $F((y_{jk})_{jk})$  in  $\mathcal{E}_2$  remain true for  $F((y_{jk}^0)_{jk})$  in  $\mathcal{E}_0$ , uniformly over the parameter class. The formal

argument for the latter is that whenever  $\|\mathbb{P}_n - \mathbb{Q}_n\|_{TV} \rightarrow 0$  and  $\mathbb{P}_n^{X_n} \rightarrow \mathbb{P}$  weakly for some random variables  $X_n$  we have for all bounded and continuous  $g$

$$\mathbb{E}_{\mathbb{Q}_n}[g(X_n)] = \mathbb{E}_{\mathbb{P}_n}[g(X_n)] + O(\|g\|_\infty \|\mathbb{P}_n - \mathbb{Q}_n\|_{TV}) \xrightarrow{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}}[g(X)].$$

Thus, for  $\alpha > 1/2$ ,  $\underline{\sigma}^2 > 0$  and  $b \sim (n^{-1/2} \log n)^{-1/(2\alpha+1)}$  the estimator

$$\tilde{\sigma}_n^2(t) := \frac{\delta}{2b\sqrt{n}} \sum_{k: |kn^{-1/2}-t| \leq b} (n\delta^{-2}\pi^2(y_k^0)^2 - \pi^2) \quad (8.2)$$

satisfies in the regression experiment  $\mathcal{E}_0$

$$\lim_{n \rightarrow \infty} \inf_{\sigma^2 \in \mathcal{S}(\alpha, R, \underline{\sigma}^2)} \mathbb{P}_{\sigma^2, n}(n^{\alpha/(4\alpha+2)}(\log n)^{-1} \|\tilde{\sigma}_n^2 - \sigma^2\|_\infty \leq R) = 1. \quad (8.3)$$

The asymptotic equivalence can be applied to construct estimators for the integrated volatility  $\int_0^1 \sigma^2(t) dt$  or more generally  $p$ -th order integrals  $\int_0^1 \sigma^p(t) dt$  using the approach developed by Ibragimov and Khas'minskii (1991) for white noise models like  $\mathcal{G}_0$ . In our notation their Theorem 7.1 yields an estimator  $\hat{\vartheta}_{p,n}$  of  $\int_0^1 \sigma^p(t) dt$  in  $\mathcal{G}_0$  such that

$$\mathbb{E}_{\sigma^2} \left[ \left( \hat{\vartheta}_{p,n} - \int_0^1 \sigma^p(t) dt - \delta^{1/2} n^{-1/4} \sqrt{2} p \int_0^1 \sigma^{p-1/2}(t) dW_t \right)^2 \right] = o(n^{-1/2})$$

holds uniformly over  $\sigma^2 \in \mathcal{S}(\alpha, R, \underline{\sigma}^2)$  for any  $\alpha, R, \underline{\sigma}^2 > 0$  since the functional  $\sqrt{\sigma(\bullet)} \mapsto \int_0^1 \sigma^p(t) dt$  is smooth on  $L^2$ . A LAN-result shows that asymptotic normality with rate  $n^{-1/4}$  and variance  $\delta 2p^2 \int_0^1 \sigma^{2p-1}(t) dt$  is minimax optimal. Specializing to the case  $p = 2$  for integrated volatility, the asymptotic variance is  $8\delta \int_0^1 \sigma^3(t) dt$ . It should be stressed here that the existing estimation procedures for integrated volatility are globally sub-optimal for our idealized model in the sense that their asymptotic variances involve the integrated quarticity  $\int_0^1 \sigma^4(t) dt$  which can at most yield optimal variance for constant values of  $\sigma^2$ , because otherwise  $\int_0^1 \sigma^4(t) dt > (\int_0^1 \sigma^3(t) dt)^{4/3}$  follows from Jensen's inequality. The fundamental reason is that all these estimators are based on quadratic forms of the increments depending on global tuning parameters, whereas optimizing weights locally permits to attain the above efficiency bound as we shall see.

Instead of following these more abstract approaches, we use our analysis to construct a simple estimator of the integrated volatility with optimal asymptotic variance. First we use the statistics  $(y_{jk})$  in  $\mathcal{E}_2$  and then transfer the results to  $\mathcal{E}_0$  using  $(y_{jk}^0)$  from (8.1).

On each block  $k$  we dispose in  $\mathcal{E}_2$  of independent  $N(0, h^2 j^{-2} \pi^{-2} \sigma^2(kh) + \varepsilon^2)$ -observations  $y_{jk}$  for  $j \geq 1$ . A maximum-likelihood estimator  $\hat{\sigma}^2(kh)$  in this exponential family satisfies the estimating equation

$$\hat{\sigma}^2(kh) = \sum_{j \geq 1} w_{jk}(\hat{\sigma}^2) h^{-2} j^2 \pi^2 (y_{jk}^2 - \varepsilon^2), \quad (8.4)$$

$$\text{where } w_{jk}(\sigma^2) := \frac{(\sigma^2(kh) + h_0^{-2} \pi^2 j^2)^{-2}}{\sum_{l \geq 1} (\sigma^2(kh) + h_0^{-2} \pi^2 l^2)^{-2}}. \quad (8.5)$$

This can be solved numerically, yet it is a non-convex problem (personal communication by J. Schmidt-Hieber). Classical MLE-theory, however, asserts for fixed  $h, k$  and consistent initial estimator  $\tilde{\sigma}_n^2(kh)$  that only one Newton step suffices to ensure asymptotic efficiency. Because of  $h \rightarrow 0$  this immediate argument does not apply here, but still gives rise to the estimator

$$\widehat{IV}_\varepsilon := \sum_{k=0}^{h^{-1}-1} h \sum_{j \geq 1} w_{jk}(\tilde{\sigma}_n^2) h^{-2} j^2 \pi^2 (y_{jk}^2 - \varepsilon^2)$$

of the integrated volatility  $IV := \int_0^1 \sigma^2(t) dt$ . Assuming the  $L^\infty$ -consistency  $\|\tilde{\sigma}_n^2 - \sigma^2\|_\infty \rightarrow 0$  in probability for the initial estimator, we assert in  $\mathcal{E}_2$  the efficiency result

$$\varepsilon^{-1/2} (\widehat{IV}_\varepsilon - IV) \xrightarrow{\mathcal{L}} N\left(0, 8 \int_0^1 \sigma^3(t) dt\right).$$

To prove this, it suffices by Slutsky's lemma to show

$$\varepsilon^{-1/2} \sum_{k=0}^{h^{-1}-1} h \sum_{j \geq 1} w_{jk}(\sigma^2) h^{-2} j^2 \pi^2 (y_{jk}^2 - \varepsilon^2) \xrightarrow{\mathcal{L}} N\left(0, 8 \int_0^1 \sigma^3(t) dt\right), \quad (8.6)$$

$$\sup_{jk} |w_{jk}(\tilde{\sigma}_n^2) - w_{jk}(\sigma^2)| \lesssim w_{jk}(\sigma^2) \|\tilde{\sigma}_n^2 - \sigma^2\|_\infty. \quad (8.7)$$

The second assertion (8.7) follows from inserting the Lipschitz property that  $W(x) := (x + h_0^{-2} \pi^2 j^2)^{-2}$  satisfies  $|W'(x)| \lesssim W(x)$  and thus  $|W(x) - W(y)| \lesssim W(x)|x - y|$  uniformly over  $x, y \geq \underline{\sigma}^2 > 0$ .

For the first assertion (8.6) note that in  $\mathcal{E}_2$  the estimator  $\widehat{IV}_\varepsilon$  is unbiased and

$$\text{Var} \left( \sum_{j \geq 1} w_{jk}(\sigma^2) h^{-2} j^2 \pi^2 (y_{jk}^2 - \varepsilon^2) \right) = \frac{2}{\sum_{j \geq 1} (\sigma^2(kh) + h_0^{-2} \pi^2 j^2)^{-2}}$$

such that by formula (9.14) and Riemann sum approximation as  $h_0 \rightarrow \infty$  (with arbitrary speed)

$$\varepsilon^{-1} \text{Var}(\widehat{IV}_\varepsilon) = \sum_{k=0}^{h^{-1}-1} \frac{2hh_0}{\sum_{j \geq 1} (\sigma^2(kh) + h_0^{-2} \pi^2 j^2)^{-2}} \rightarrow 8 \int_0^1 \sigma^3(t) dt.$$

Due to the independence and Gaussianity of the  $(y_{jk})$  we deduce also

$$\mathbb{E} \left[ \left( \sum_{j \geq 1} w_{jk}(\sigma^2) h^{-2} j^2 \pi^2 (y_{jk}^2 - \mathbb{E}[y_{jk}^2]) \right)^4 \right] \lesssim \text{Var} \left( \sum_{j \geq 1} w_{jk}(\sigma^2) h^{-2} j^2 \pi^2 (y_{jk}^2 - \varepsilon^2) \right)^2$$

such that the central limit theorem under a Lyapounov condition with power  $p = 4$  (e.g. Shiryaev (1995)) proves assertion (8.6), assuming  $h \rightarrow 0$  and  $h_0 \rightarrow \infty$ . A feasible estimator is obtained by neglecting frequencies larger than some  $J = J(\varepsilon)$ :

$$\widehat{IV}_{\varepsilon, J} := \sum_{k=0}^{h^{-1}-1} h \sum_{j=1}^J w_{jk}^J(\tilde{\sigma}_n^2) h^{-2} j^2 \pi^2 (y_{jk}^2 - \varepsilon^2) \quad (8.8)$$

$$\text{where } w_{jk}^J(\sigma^2) := \frac{(\sigma^2(kh) + h_0^{-2} \pi^2 j^2)^{-2}}{\sum_{l=1}^J (\sigma^2(kh) + h_0^{-2} \pi^2 l^2)^{-2}}. \quad (8.9)$$

A simple calculation yields  $\mathbb{E}[|\widehat{IV}_{\varepsilon, J} - \widehat{IV}_\varepsilon|^2] \lesssim \varepsilon(h_0/J)^3$  such that for  $h_0/J \rightarrow 0$  convergence in probability implies again by Slutsky's lemma

$$\varepsilon^{-1/2} (\widehat{IV}_{\varepsilon, J} - IV) \xrightarrow{\mathcal{L}} N \left( 0, 8 \int_0^1 \sigma^3(t) dt \right).$$

By the above argument, weak convergence results transfer from  $\mathcal{E}_2$  to  $\mathcal{E}_0$  and we obtain the following result where we give a concrete choice of the initial estimator, the block size  $h$  and the spectral cut-off  $J$  (we just need some consistent estimator  $\tilde{\sigma}_n^2$ ,  $h^{2\alpha} n^{1/2} \rightarrow 0$  as well as  $hn^{1/2} \rightarrow \infty$  and  $J^{-1} = o(h^{-1}n^{-1/2})$ ).

**8.1 Theorem.** *Let  $y_{jk}^0$  for  $j \geq 1$ ,  $k = 0, h^{-1} - 1$  be the statistics (8.1) from model  $\mathcal{E}_0$ . For  $h \sim n^{-1/2} \log(n)$  and  $J/\log(n) \rightarrow \infty$  consider the estimator of integrated volatility*

$$\widehat{IV}_n := \sum_{k=0}^{h^{-1}-1} h \sum_{j=1}^J w_{jk}^J(\tilde{\sigma}_n^2) h^{-2} j^2 \pi^2 ((y_{jk}^0)^2 - \delta n^{-1})$$

with weights  $w_{jk}^J$  from (8.9) and the initial estimator  $\tilde{\sigma}_n^2$  from (8.2). Then  $\widehat{IV}_n$  is asymptotically efficient in the sense that

$$n^{1/4}(\widehat{IV}_n - IV) \xrightarrow{\mathcal{L}} N\left(0, 8\delta \int_0^1 \sigma^3(t) dt\right) \text{ as } n \rightarrow \infty,$$

provided  $\sigma^2$  is strictly positive and  $\alpha$ -Hölder continuous with  $\alpha > 1/2$ .

This might serve as a benchmark for more general models, whereas we, in the spirit of Mykland (2009), focus on elucidating the underlying fundamental structures. In particular, we should dispense with the Gaussianity of the microstructure noise ( $\varepsilon_i$ ) as well as with the deterministic nature of the volatility  $\sigma^2$ . The analysis in both cases, however, cannot simply rely on model  $\mathcal{E}_2$ , since  $\mathcal{E}_0$  is non-Gaussian. Different tools are required.

## 9 Appendix

### 9.1 Gaussian measures, Hellinger distance and Hilbert-Schmidt norm

We gather basic facts about cylindrical Gaussian measures, the Hellinger distance and their interplay.

Formally, we realize the white noise experiments, as  $L^2$ -indexed Gaussian variables, e.g. in experiment  $\mathcal{E}_1$  we observe for any  $f \in L^2([0, 1])$

$$Y_f := \langle f, dY \rangle := \int_0^1 f(t) \left( \int_0^t \sigma(s) dB(s) \right) dt + \varepsilon \int_0^1 f(t) dW_t.$$

Canonically, we thus define  $\mathbb{P}^{\sigma, \varepsilon}$  on the set  $\Omega = \mathbb{R}^{L^2([0, 1])}$  with product Borel  $\sigma$ -algebra  $\mathcal{F} = \mathfrak{B}^{\otimes L^2([0, 1])}$  (realizing a cylindrical centered Gaussian measure). Its covariance structure is given by

$$\mathbb{E}[Y_f Y_g] = \langle Cf, g \rangle, \quad f, g \in L^2([0, 1]),$$

with the covariance operator  $C : L^2([0, 1]) \rightarrow L^2([0, 1])$  given by

$$Cf(t) = \int_0^1 \left( \int_0^{t \wedge u} \sigma^2(s) ds \right) f(u) du + \varepsilon^2 f(t), \quad f \in L^2([0, 1]).$$

Note that  $C$  is not trace class and thus does not define a Gaussian measure on  $L^2([0, 1])$  itself.

In the construction, it suffices to prescribe  $(Y_{e_m})_{m \geq 1}$  for an orthonormal basis  $(e_m)_{m \geq 1}$  and to set

$$Y_f := \sum_{m=1}^{\infty} \langle f, e_m \rangle Y_{e_m}.$$

This way, we can define  $\mathbb{P}^{\sigma, \varepsilon}$  equivalently on the sequence space  $\Omega = \mathbb{R}^{\mathbb{N}}$  with product  $\sigma$ -algebra  $\mathcal{F} = \mathfrak{B}^{\otimes \mathbb{N}}$ . This is useful when extending results from finite dimensions.

The Hellinger distance between two probability measures  $\mathbb{P}$  and  $\mathbb{Q}$  on  $(\Omega, \mathcal{F})$  is defined as

$$H(\mathbb{P}, \mathbb{Q}) = \left( \int_{\Omega} (\sqrt{p(\omega)} - \sqrt{q(\omega)})^2 \mu(d\omega) \right)^{1/2},$$

where  $\mu$  denotes a dominating measure, e.g.  $\mu = \mathbb{P} + \mathbb{Q}$ , and  $p$  and  $q$  denote the respective densities. The total variation distance is smaller than the Hellinger distance:

$$\|\mathbb{P} - \mathbb{Q}\|_{TV} \leq H(\mathbb{P}, \mathbb{Q}). \quad (9.1)$$

The identity  $H^2(\mathbb{P}, \mathbb{Q}) = 2 - 2 \int \sqrt{p} \sqrt{q} d\mu$  implies the bound for finite or countably infinite product measures

$$H^2\left(\bigotimes_n \mathbb{P}_n, \bigotimes_n \mathbb{Q}_n\right) \leq \sum_n H^2(\mathbb{P}_n, \mathbb{Q}_n). \quad (9.2)$$

Moreover, the Hellinger distance is invariant under bi-measurable bijections  $T : \Omega \rightarrow \Omega'$  since with the densities  $p \circ T^{-1}, q \circ T^{-1}$  of the image measures  $\mathbb{P}^T$  and  $\mathbb{Q}^T$  with respect to  $\mu^T$  we have

$$H^2(\mathbb{P}^T, \mathbb{Q}^T) = \int_{\Omega'} (\sqrt{p \circ T^{-1}} - \sqrt{q \circ T^{-1}})^2 d\mu^T = \int_{\Omega} (\sqrt{p} - \sqrt{q})^2 d\mu = H^2(\mathbb{P}, \mathbb{Q}). \quad (9.3)$$

For the one-dimensional Gaussian laws  $N(0, 1)$  and  $N(0, \sigma^2)$  we derive

$$H^2(N(0, 1), N(0, \sigma^2)) = 2 - \sqrt{8\sigma/(\sigma^2 + 1)} \leq 2(\sigma^2 - 1)^2.$$



For the multi-dimensional Gaussian laws  $N(0, \Sigma_1)$  and  $N(0, \Sigma_2)$  with invertible covariance matrices  $\Sigma_1, \Sigma_2 \in \mathbb{R}^{d \times d}$  we obtain by linear transformation and independence, denoting by  $\lambda_1, \dots, \lambda_d$  the eigenvalues of  $\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}$ :

$$H^2(N(0, \Sigma_1), N(0, \Sigma_2)) = H^2(N(0, \text{Id}), N(0, \Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2})) \leq \sum_{k=1}^d 2(\lambda_k - 1)^2.$$

The last sum is nothing, but the squared Hilbert-Schmidt (or Frobenius norm) of  $\Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} - \text{Id}$  such that

$$H^2(N(0, \Sigma_1), N(0, \Sigma_2)) \leq 2 \|\Sigma_1^{-1/2} (\Sigma_2 - \Sigma_1) \Sigma_1^{-1/2}\|_{HS}^2. \quad (9.4)$$

Observing that (9.2) and (9.3) also apply to Gaussian measures on the sequence space  $\mathbb{R}^{\mathbb{N}}$ , the bound (9.4) is also valid for (cylindrical) Gaussian measures  $N(0, \Sigma_i)$  with self-adjoint positive definite covariance operators  $\Sigma_i : L^2([0, 1]) \rightarrow L^2([0, 1])$ .

The Hilbert-Schmidt norm of a linear operator  $A : H \rightarrow H$  on any separable real Hilbert space  $H$  can be expressed by its action on an orthonormal basis  $(e_m)$  via

$$\|A\|_{HS}^2 = \sum_{m,n} \langle Ae_m, e_n \rangle^2,$$

which for a matrix is just the usual Frobenius norm. For self-adjoint operators  $A, B$  with  $|\langle Av, v \rangle| \leq |\langle Bv, v \rangle|$  for all  $v \in H$  we use the eigenbasis  $(e_m)$  of  $A$  and obtain

$$\|A\|_{HS}^2 = \sum_m \langle Ae_m, e_m \rangle^2 \leq \sum_{m,n} \langle Be_m, e_n \rangle^2 = \|B\|_{HS}^2. \quad (9.5)$$

Furthermore, it is straight-forward to see for any bounded operator  $T$

$$\|TA\|_{HS} \leq \|T\| \|A\|_{HS}, \quad \|AT\|_{HS} \leq \|T\| \|A\|_{HS} \quad (9.6)$$

with the usual operator norm  $\|T\|$  of  $T$ . Finally, for integral operators  $Kf(x) = \int_0^1 k(x, y) f(y) dy$  on  $L^2([0, 1])$  it is well known that

$$\|K\|_{HS} = \|k\|_{L^2([0, 1]^2)}. \quad (9.7)$$

For two Gaussian laws with different mean vectors  $\mu_1, \mu_2$  and with the same invertible covariance matrix  $\Sigma$  we can similarly use the transformation  $\Sigma^{-1/2}$  and

the scalar case  $H^2(N(\mu_1, 1), N(\mu_2, 1)) = 2(1 - e^{-(\mu_1 - \mu_2)^2/8}) \leq (\mu_1 - \mu_2)^2/4$  to conclude by independence

$$H^2(N(\mu_1, \Sigma), N(\mu_2, \Sigma)) \leq \frac{1}{4} \|\Sigma^{-1/2}(\mu_1 - \mu_2)\|^2. \quad (9.8)$$

Combining (9.4) and (9.8) we obtain by the triangle inequality the bound

$$H^2(N(\mu_1, \Sigma_1), N(\mu_2, \Sigma_2)) \lesssim \|\Sigma_1^{-1/2}(\mu_1 - \mu_2)\|^2 + \|\Sigma_1^{-1/2}(\Sigma_2 - \Sigma_1)\Sigma_1^{-1/2}\|_{HS}^2. \quad (9.9)$$

## 9.2 Proof of Theorem 2.2

We first show that  $\mathcal{E}_1$  is asymptotically at least as informative as  $\mathcal{E}_0$  for  $\varepsilon = \delta/\sqrt{n}$  and  $\alpha > 0$ . From  $\mathcal{E}_1$  with  $\varepsilon = \delta/\sqrt{n}$  we can generate the observations (statistics)

$$\begin{aligned} \tilde{Y}_i &:= n \int_{(2i-1)/2n}^{(2i+1)/2n} dY_t = n \int_{(2i-1)/2n}^{(2i+1)/2n} X_t dt + \tilde{\varepsilon}_i, \quad i = 1, \dots, n-1, \\ \tilde{Y}_n &:= 2n \int_{(2n-1)/2n}^1 dY_t = 2n \int_{(2n-1)/2n}^1 X_t dt + \tilde{\varepsilon}_n, \end{aligned}$$

with  $\tilde{\varepsilon}_i = n\varepsilon(W_{(2i+1)/2n} - W_{(2i-1)/2n}) \sim N(0, \delta^2)$  and similarly  $\tilde{\varepsilon}_n \sim N(0, \delta^2)$ , all independent. In contrast to standard equivalence proofs, it turns out to be essential here to take  $\tilde{Y}_i$  as a mean symmetric around the point  $i/n$ . Since  $(Y_i)$  and  $(\tilde{Y}_i)$  are defined on the same sample space, using inequality (9.1) it suffices to prove that the Hellinger distance between the law of  $(Y_i)$  and the law of  $(\tilde{Y}_i)$  tends to zero as  $n$  tends to infinity.

For the integrated volatility function we introduce the notation

$$a(t) := \int_0^t \sigma^2(s) ds, \quad 0 \leq t \leq 1.$$

For notational convenience we also set  $a(1+s) := a(1-s)$  for  $s > 0$ .

The covariance matrix  $\Sigma^Y$  of the centered Gaussian vector  $(Y_i)$  is given by

$$\Sigma_{kl}^Y := \mathbb{E}[Y_k Y_l] = a(k/n) + \delta^2 \mathbf{1}(k = l), \quad 1 \leq k \leq l \leq n.$$

Similarly, the covariance matrix  $\Sigma^{\tilde{Y}}$  of the centered Gaussian vector  $(\tilde{Y}_i)$  is given by

$$\Sigma_{kl}^{\tilde{Y}} := \mathbb{E}[\tilde{Y}_k \tilde{Y}_l] = n \int_{(2k-1)/2n}^{(2k+1)/2n} a(t) dt + \delta^2 \mathbf{1}(k = l), \quad 1 \leq k \leq l \leq n,$$

where for  $k = l = n$  we used the convention for  $a(1 + s)$  above. We bound the Hellinger distance using consecutively (9.4),  $\Sigma^Y \geq \delta^2 \text{Id}$  in (9.5) and (9.2), a Taylor expansion for  $a$  and treating the case  $k = l = n$  by a Lipschitz bound separately:

$$\begin{aligned}
& H^2(\mathcal{L}(Y_i, i = 1, \dots, n), \mathcal{L}(\tilde{Y}_i, i = 1, \dots, n)) \\
& \leq 2\|(\Sigma^Y)^{-1/2}(\Sigma^Y - \Sigma^{\tilde{Y}})(\Sigma^Y)^{-1/2}\|_{HS}^2 \\
& \leq 2\delta^{-4}\|\Sigma^{\tilde{Y}} - \Sigma^Y\|_{HS}^2 \\
& \leq 4\delta^{-4} \sum_{1 \leq k \leq l \leq n} \left( n \int_{(2k-1)/2n}^{(2k+1)/2n} (a(t) - a(k/n)) dt \right)^2 \\
& \leq 4\delta^{-4} \left( O(R^2 n^{-2}) + n \sum_{k=1}^n \left( n \int_{(2k-1)/2n}^{(2k+1)/2n} (a'(k/n)(t - k/n) + O(Rn^{-1-\alpha})) dt \right)^2 \right) \\
& = 4\delta^{-4} \left( O(R^2 n^{-2}) + O(R^2 n^{2-2-2\alpha}) \right) \\
& = O(\delta^{-4} R^2 n^{-2\alpha}).
\end{aligned}$$

Consequently, by (9.1) the total-variation and thus also the Le Cam distance between the experiments of observing  $(Y_i)$  and of observing  $(\tilde{Y}_i)$  tends to zero for  $n \rightarrow \infty$ , which proves that the white noise experiment  $\mathcal{E}_1$  is asymptotically at least as informative as the regression experiment  $\mathcal{E}_0$ .

To show the converse, we build from the regression experiment  $\mathcal{E}_0$  a continuous time observation by linear interpolation. To this end we introduce the linear  $B$ -splines (or hat functions)  $b_i(t) = b(t - i/n)$  with  $b(t) = \min(1 + nt, 1 - tn)\mathbf{1}_{[-1/n, 1/n]}(t)$  and set

$$\hat{Y}'_t := \sum_{i=1}^n Y_i b_i(t) = \sum_{i=1}^n X_{i/n} b_i(t) + \sum_{i=1}^n \varepsilon_i b_i(t), \quad t \in [0, 1].$$

Note that  $(\hat{Y}'_t)$  is a centered Gaussian process with covariance function

$$\hat{c}(t, s) := \mathbb{E}[\hat{Y}'_t \hat{Y}'_s] = \sum_{i,j=1}^n a((i \wedge j)/n) b_i(t) b_j(s) + \delta^2 \sum_{i=1}^n b_i(t) b_i(s), \quad 0 \leq t, s \leq 1.$$

For any  $f \in L^2([0, 1])$  we thus obtain

$$\begin{aligned} \mathbb{E}[\langle f, \hat{Y}' \rangle^2] &= \sum_{i,j=1}^n a((i \wedge j)/n) \langle f, b_i \rangle \langle f, b_j \rangle + \delta^2 \sum_{i=1}^n \langle f, b_i \rangle^2 \\ &\leq \sum_{i,j=1}^n a((i \wedge j)/n) \langle f, b_i \rangle \langle f, b_j \rangle + \delta^2 n^{-1} \|f\|^2, \end{aligned}$$

because  $\int n b_i = 1$  yields by Jensen's inequality  $\langle f, n b_i \rangle^2 \leq \langle f^2, n b_i \rangle$  and we have  $\sum_i b_i \leq 1$ . This means that the covariance operator  $\hat{C}$  induced by the kernel  $\hat{c}$  is smaller than

$$\bar{C}f(t) := \sum_{i,j=1}^n a((i \wedge j)/n) \langle f, b_j \rangle b_i(t) + \delta^2 n^{-1} f(t), \quad f \in L^2([0, 1])$$

in the sense that  $\hat{C} - \bar{C}$  is positive (semi-)definite. Now observe that  $\bar{C}$  is the covariance operator of the white noise observations

$$d\bar{Y}_t = \sum_{i=1}^n X_{i/n} b_i(t) + \frac{\delta}{\sqrt{n}} dW_t, \quad t \in [0, 1]. \quad (9.10)$$

Hence, we can generate these observations from  $(\hat{Y}'_t)$  by randomisation, i.e. by adding uninformative  $N(0, \bar{C} - \hat{C})$ -noise to  $\hat{Y}'$ . Now it is easy to see that observing  $\bar{Y}$  in (9.10) and  $Y$  from  $\mathcal{E}_1$  is asymptotically equivalent, since in terms of the respective covariance operators, using again (9.4), (9.5) and (9.2), the squared Hellinger distance satisfies

$$\begin{aligned} H^2(\mathcal{L}(\bar{Y}), \mathcal{L}(Y)) &\leq 2\|(C^Y)^{-1/2}(\bar{C} - C^Y)(C^Y)^{-1/2}\|_{HS}^2 \\ &\leq 2\delta^{-4}n^2 \int_0^1 \int_0^1 \left( a(t \wedge s) - \sum_{i,j=1}^n a((i \wedge j)/n) b_i(t) b_j(s) \right)^2 dt ds \\ &= 2\delta^{-4}n^2 \int_0^1 \int_0^1 \left( \sum_{i,j=0}^n (a(t \wedge s) - a((i \wedge j)/n)) b_i(t) b_j(s) \right)^2 dt ds, \end{aligned}$$

where for the last line we have used  $\sum_{i=0}^n b_i(t) = 1$  and  $a(0) = 0$ . Since  $b_i(t) \neq 0$  can only hold when  $i - \lfloor nt \rfloor \in \{0, 1\}$ , the  $\alpha$ -Hölder regularity of  $\sigma^2$  implies for

$t \leq s - 1/n$ :

$$\begin{aligned}
& \left( \sum_{i,j=0}^n (a(t \wedge s) - a((i \wedge j)/n)) b_i(t) b_j(s) \right)^2 \\
&= \left( \sum_{k,l=0}^1 (a'(\lfloor nt \rfloor/n)(t - (k + \lfloor nt \rfloor)/n) + O(Rn^{-1-\alpha})) b_{k+\lfloor nt \rfloor}(t) b_{l+\lfloor ns \rfloor}(s) \right)^2 \\
&= O(R^2 n^{-2-2\alpha}) + \left( a'(\lfloor nt \rfloor/n) \sum_{k=0}^1 (t - (k + \lfloor nt \rfloor)/n) b_{k+\lfloor nt \rfloor}(t) \right)^2 \\
&= O(R^2 n^{-2-2\alpha}).
\end{aligned}$$

A symmetric argument gives the same bound for  $s \leq t - 1/n$ . For  $|t - s| < 1/n$  we use only the Lipschitz continuity of  $a$  to obtain the bound  $O(R^2 n^{-2})$ . Altogether we have found

$$H^2(\mathcal{L}(\bar{Y}), \mathcal{L}(Y)) \leq 2\delta^{-4} n^2 \left( O(R^2 n^{-2-2\alpha}) + n^{-1} O(R^2 n^{-2}) \right) = O(\delta^{-4} R^2 n^{-2\alpha}),$$

which together with the transformation in the other direction shows that the Le Cam distance between  $\mathcal{E}_0$  and  $\mathcal{E}_1$  is of order  $O(\delta^{-2} R n^{-\alpha})$ .

### 9.3 Proof of Proposition 3.2

The main tool is Proposition 9.1 below. Together with the Hölder bound

$$|\sigma^2(\lfloor s \rfloor_h) - \sigma^2(s)| \leq R h^\alpha, \quad s \in [0, 1],$$

it implies that for fixed  $\sigma$  the observation laws in  $\mathcal{E}_1$  and  $\mathcal{E}_2$  have a Hellinger distance of order  $R h^\alpha \underline{c}^{-3/2} \varepsilon^{-1/2}$ . By inequality (9.1) this translates to the total variation and thus to the Le Cam distance.

**9.1 Proposition.** *For  $\varepsilon > 0$  and continuous  $\sigma : [0, 1] \rightarrow (0, \infty)$  consider the law  $\mathbb{P}^{\sigma, \varepsilon}$  generated by*

$$dY_t = \left( \int_0^t \sigma(s) dB(s) \right) dt + \varepsilon dW_t, \quad t \in [0, 1],$$

with independent Brownian motions  $B$  and  $W$ . Then the Hellinger distance between two laws  $\mathbb{P}^{\sigma_1, \varepsilon}$  and  $\mathbb{P}^{\sigma_2, \varepsilon}$  satisfies

$$H(\mathbb{P}^{\sigma_1, \varepsilon}, \mathbb{P}^{\sigma_2, \varepsilon}) \lesssim \|\sigma_1^2 - \sigma_2^2\|_\infty \left( \max_{t \in [0, 1]} \sigma_1^{-3}(t) \right) \varepsilon^{-1/2}.$$

*Proof.* The covariance operator  $C_\sigma$  of  $\mathbb{P}^{\sigma, \varepsilon}$  is for  $f, g \in L^2([0, 1])$  given by

$$\langle C_\sigma f, g \rangle = \mathbb{E}[\langle f, dY \rangle \langle g, dY \rangle] = \mathbb{E}[\langle f, X \rangle \langle g, X \rangle] + \varepsilon^2 \langle f, g \rangle = \int FG\sigma^2 + \varepsilon^2 \int fg.$$

For covariance operators corresponding to  $\sigma_1, \sigma_2$  we have with  $F(t) = -\int_t^1 f(s) ds$

by twofold partial integration

$$\begin{aligned} |\langle (C_{\sigma_1} - C_{\sigma_2})f, f \rangle| &= \left| \int_0^1 \int_0^1 \int_0^{t \wedge s} (\sigma_1^2 - \sigma_2^2)(u) du f(t) f(s) ds dt \right| \\ &= \left| \int_0^1 F(u)^2 (\sigma_1^2 - \sigma_2^2)(u) du \right| \\ &\leq \|\sigma_1^2 - \sigma_2^2\|_\infty \int_0^1 F(u)^2 du = \|\sigma_1^2 - \sigma_2^2\|_\infty \langle C_{BM} f, f \rangle \end{aligned}$$

with  $C_{BM}g(t) := \int_0^1 (t \wedge s)g(s) ds$ , the covariance operator of standard Brownian motion. Using further the ordering  $C_{\sigma_1} \geq \min_t \sigma_1^2(t)C_{BM} + \varepsilon^2 \text{Id}$  and (9.5), (9.2) we obtain

$$\begin{aligned} &\|C_{\sigma_1}^{-1/2}(C_{\sigma_2} - C_{\sigma_1})C_{\sigma_1}^{-1/2}\|_{HS} \\ &\leq \|\sigma_1^2 - \sigma_2^2\|_\infty \|C_{\sigma_1}^{-1/2}C_{BM}C_{\sigma_1}^{-1/2}\|_{HS} \\ &\leq \|\sigma_1^2 - \sigma_2^2\|_\infty \|(\min_t \sigma_1^2(t)C_{BM} + \varepsilon^2 \text{Id})^{-1/2}C_{BM}(\min_t \sigma_1^2(t)C_{BM} + \varepsilon^2 \text{Id})^{-1/2}\|_{HS} \\ &= \|\sigma_1^2 - \sigma_2^2\|_\infty \|F(C_{BM})\|_{HS}, \end{aligned}$$

employing functional calculus with  $F(x) = (\min_t \sigma_1^2(t)x + \varepsilon^2)^{-1}x$ . The spectral properties of  $C_{BM}$  imply that  $F(C_{BM})$  has eigenfunctions  $e_k(t) = \sqrt{2} \sin(\pi(k - 1/2)t)$ ,  $k \geq 1$ , with eigenvalues  $\lambda_k = \frac{4}{4 \min_t \sigma_1^2(t) + (2k-1)^2 \pi^2 \varepsilon^2}$ , whence its Hilbert-Schmidt norm is of order  $\max_t \sigma_1^{-3}(t) \varepsilon^{-1/2}$ . This yields the result.  $\square$

## 9.4 Proof of Proposition 5.2

We only consider the case of odd indices  $k$ , both cases are treated analogously. The result of Theorem 6.1 in conjunction with Theorem 5.2 of Grama and Nuss-

baum (2002) establishes that  $\mathcal{E}_{3,m}^{odd}$  and the Gaussian regression experiment  $\mathcal{G}_{3,m}$  of observing

$$Y_k = v_\varepsilon s^2(kh) + I(\sigma_0^2(kh))^{-1/2} \gamma_k, \quad k \in A_m \text{ odd}, \quad \gamma_k \sim N(0, 1) \text{ i.i.d.} \quad (9.11)$$

are equivalent to experiments  $\tilde{\mathcal{E}}_{3,m} = (\mathcal{Y}, \mathcal{G}, (\tilde{\mathbb{P}}_{s^2}^m)_{s^2 \in C_\alpha(R)})$  and  $\tilde{\mathcal{G}}_{3,m} = (\mathcal{Y}, \mathcal{G}, (\tilde{\mathbb{Q}}_{s^2}^m)_{s^2 \in C_\alpha(R)})$ , respectively, on the same space  $(\mathcal{Y}, \mathcal{G})$  such that

$$\sup_{s^2 \in C_\alpha(R)} H^2(\tilde{\mathbb{P}}_{s^2}^m, \tilde{\mathbb{Q}}_{s^2}^m) \lesssim \ell^{-2\rho} \quad (9.12)$$

holds for all  $\rho < 1$ .

To be precise, it must be checked that the regularity conditions  $R1 - R3$  of Grama and Nussbaum (2002) are satisfied for all values  $\delta$ . One complication is that in our parametric model the probabilities  $\mathbb{P}_\vartheta$  and the Fisher information  $I(\vartheta)$  depend on  $h_0$  which tends to infinity. Yet, inspecting the proofs it becomes clear that the results remain valid if (a) the conditions  $R1 - R3$  hold for varying models, but with uniform constants and (b) the Fisher information is renormalized by the localisation such that the parametric rate  $\ell^{-1/2}$  (in our block length notation) is attained. From the fact that  $\mathbb{P}_\vartheta$  is the product of one-dimensional exponential family models we easily check condition  $R1$  for  $\delta = 1$  and condition  $R2$  for any  $\delta > 0$ . Both conditions hold uniformly over  $h_0$  once the score  $\dot{l}$  has been renormalized through multiplication by  $h_0^{-1/2}$ . In (5.2) we have already calculated the Fisher information and we infer directly condition  $R3$  that  $h_0^{-1}I(\vartheta)$  is uniformly bounded away from zero and infinity. We thus infer (9.12).

In view of the independence among the experiments  $(\mathcal{E}_{3,m}^{odd})_m$  and equally among the experiments  $(\mathcal{G}_{3,m})_m$  we infer from (9.12) and (9.2)

$$\sup_{s^2 \in C_\alpha(R)} H^2(\otimes_{m=1}^{(\ell h)^{-1}} \tilde{\mathbb{P}}_{s^2}^m, \otimes_{m=1}^{(\ell h)^{-1}} \tilde{\mathbb{Q}}_{s^2}^m) \lesssim (\ell h)^{-1} \ell^{-2\rho} \lesssim \varepsilon^{-1} v_\varepsilon^2 h_0^{2\rho} v_\varepsilon^{4\rho}.$$

Since we assume  $h_0 = o(\varepsilon^{(1-2\alpha)/2\alpha})$ , the right-hand side tends to zero provided

$$-1 + 2\frac{\alpha}{2\alpha + 1} + \frac{\rho(1 - 2\alpha)}{\alpha} + \frac{4\rho\alpha}{2\alpha + 1} = \frac{\rho - \alpha}{\alpha(2\alpha + 1)} > 0$$

holds. Since  $\rho < 1$  is arbitrary, this is always satisfied for  $\alpha < 1$ . In the case  $\alpha = 1$  we use  $h_0 \lesssim \varepsilon^{-p}$  for some  $p < 1/2$ . We have derived asymptotic equivalence between the product experiments  $\otimes_m \tilde{\mathcal{E}}_{3,m}^{loc}$  and  $\otimes_m \tilde{\mathcal{G}}_{3,m}$ . A fortiori, applying the Brown and Low (1996) result, this leads to asymptotic equivalence between observing  $(y_{jk})$  in experiments  $\mathcal{E}_{2,loc}$  and the corresponding Gaussian shift models of observing

$$dY_t = I(\sigma_0^2(t))^{1/2} v_\varepsilon s^2(t) dt + (2h)^{1/2} dW_t, \quad t \in [0, 1]. \quad (9.13)$$

From the explicit form (5.2) of the Fisher information we infer for  $h_0 \rightarrow \infty$

$$\left| \frac{2\vartheta^{3/2}}{h_0} I(\vartheta) - \frac{1}{4} + \frac{1}{2\vartheta^{1/2}h_0} \right| \lesssim e^{-\sigma h_0}.$$

Consequently, by the polynomial growth of  $h_0$  in  $\varepsilon^{-1}$ , the Kullback-Leibler divergence between the observation laws from (9.13) and the model  $\mathcal{G}_{3,loc}$  converges to zero. This gives the result.

## 9.5 Proof of Proposition 6.1

Since the observations  $y_{jk}$  for  $j \geq 1$  are the same in  $\mathcal{Y}$  and  $\tilde{\mathcal{Y}}$ , we can work conditionally on those. Moreover, it suffices to consider only the event  $\Omega_\varepsilon := \{\|\hat{\sigma}_\varepsilon^2 - \sigma^2\|_\infty \leq Rv_\varepsilon\}$  because the squared Hellinger distance satisfies (with obvious notation)

$$\begin{aligned} H^2(\mathcal{L}(\mathcal{Y}), \mathcal{L}(\tilde{\mathcal{Y}})) &= \mathbb{E}[H^2(\mathcal{L}((y_{0k})_k | (y_{jk})_{j \geq 1, k}), \mathcal{L}((\tilde{y}_{0k})_k | (y_{jk})_{j \geq 1, k}))] \\ &\leq \mathbb{E}[H^2(\mathcal{L}((y_{0k})_k | (y_{jk})_{j \geq 1, k}), \mathcal{L}((\tilde{y}_{0k})_k | (y_{jk})_{j \geq 1, k})) \mathbf{1}_{\Omega_\varepsilon}] + 2\mathbb{P}(\Omega_\varepsilon^c) \end{aligned}$$

with  $\mathbb{P}(\Omega_\varepsilon^c) \rightarrow 0$ . Conditional on  $(y_{jk})_{j \geq 1, k}$ , both laws are Gaussian,  $(y_{0,k})_k$  has mean  $\mu$  with

$$\mu_0 = 2 \sum_{j \geq 1} \frac{\text{Var}(\beta_{jk})}{\text{Var}(y_{jk})} y_{jk}, \quad \mu_k = \sum_{j \geq 1} \frac{\text{Var}(\beta_{jk})}{\text{Var}(y_{jk})} ((-1)^{j+1} y_{j,k-1} + y_{jk}), \quad k \geq 1,$$



and covariance matrix  $\Sigma$  with

$$\Sigma_{k,k'} = \begin{cases} 2c_k \sum_{j \geq 1} \frac{\varepsilon^2 \text{Var}(\beta_{jk})}{\text{Var}(y_{jk})} + \varepsilon^2, & \text{if } k' = k, \\ c_{k \wedge k'} \sum_{j \geq 1} \frac{\varepsilon^2 \text{Var}(\beta_{jk})}{\text{Var}(y_{jk})} - \frac{\varepsilon^2}{2}, & \text{if } |k' - k| = 1, \\ 0, & \text{otherwise,} \end{cases}$$

where  $c_k := 1 \vee (2 - k) \in \{1, 2\}$ . Conditional mean  $\tilde{\mu}$  and covariance matrix  $\tilde{\Sigma}$  of  $(\tilde{y}_{0k})_k$  have the same representation, but replacing  $\text{Var}$  each time by  $\text{Var}_\varepsilon$ .

From the tri-diagonal structure of  $\Sigma$  and from

$$\sum_{j \geq 1} \frac{\text{Var}(\beta_{jk})}{\text{Var}(y_{jk})} \sim \sum_{j \geq 1} \frac{(h_0/j)^2}{(h_0/j)^2 + 1} \sim h_0, \quad h_0 \rightarrow \infty,$$

we infer  $\Sigma \gtrsim (\varepsilon^2 h_0 + \varepsilon^2) \text{Id} \geq \varepsilon h \text{Id}$  in matrix order. Combining this with the Hellinger bound (9.9) we arrive at the estimate

$$\begin{aligned} & \mathbb{E}[H^2(\mathcal{L}((y_{0k})_k | (y_{jk})_{j \geq 1, k}), \mathcal{L}((\tilde{y}_{0k})_k | (y_{jk})_{j \geq 1, k}))] \\ & \lesssim \mathbb{E} \left[ \frac{\|\mu - \tilde{\mu}\|^2}{\varepsilon h} \right] + \frac{\|\Sigma - \tilde{\Sigma}\|_{HS}^2}{\varepsilon^2 h^2} \\ & \lesssim \sum_{j \geq 1, k} \left( \frac{\text{Var}(\beta_{jk})}{\text{Var}(y_{jk})} - \frac{\text{Var}_\varepsilon(\beta_{jk})}{\text{Var}_\varepsilon(y_{jk})} \right)^2 \frac{\text{Var}(y_{jk})}{\varepsilon h} + \sum_{j \geq 1, k} \left( \frac{\varepsilon^2 \text{Var}(\beta_{jk})}{\text{Var}(y_{jk})} - \frac{\varepsilon^2 \text{Var}_\varepsilon(\beta_{jk})}{\text{Var}_\varepsilon(y_{jk})} \right)^2 \varepsilon^{-2} h^{-2}. \end{aligned}$$

The function  $G(z) := \frac{\|\Phi_{jk}\|^2 z}{\|\Phi_{jk}\|^2 z + \varepsilon^2}$  has derivative  $G'(z) = \frac{\|\Phi_{jk}\|^2 \varepsilon^2}{(\|\Phi_{jk}\|^2 z + \varepsilon^2)^2}$  and thus satisfies uniformly over all  $z$  bounded away from zero  $|G(w) - G(z)| \lesssim \frac{\|\Phi_{jk}\|^2 \varepsilon^2 |w - z|}{(\|\Phi_{jk}\|^2 z + \varepsilon^2)^2}$ .

Inserting  $|\sigma^2 - \sigma_0^2| \lesssim v_\varepsilon$  and  $\|\Phi_{jk}\| \sim h/j$ , we thus find the uniform bound on  $\Omega_\varepsilon$

$$\left( \frac{\text{Var}(\beta_{jk})}{\text{Var}(y_{jk})} - \frac{\text{Var}_\varepsilon(\beta_{jk})}{\text{Var}_\varepsilon(y_{jk})} \right)^2 \lesssim \frac{v_\varepsilon^2 \varepsilon^4 h^4 / j^4}{(\varepsilon^2 + h^2 / j^2)^4} \sim v_\varepsilon^2 \min(h_0/j, j/h_0)^4.$$

Putting the estimates together, we arrive at

$$\begin{aligned} H^2(\mathcal{L}(\mathcal{Y}), \mathcal{L}(\tilde{\mathcal{Y}})) & \lesssim v_\varepsilon^2 \sum_{j \geq 1, k} \min(h_0/j, j/h_0)^4 \left( \frac{1 + h_0^2/j^2}{h_0} + \frac{1}{h_0^2} \right) + \mathbb{P}(\Omega_\varepsilon^c) \\ & \leq 2v_\varepsilon^2 h^{-1} \sum_{j \geq 1} \min(h_0/j, j/h_0)^2 h_0^{-1} + \mathbb{P}(\Omega_\varepsilon^c) \\ & \sim v_\varepsilon^2 h_0^{-1} \varepsilon^{-1} + \mathbb{P}(\Omega_\varepsilon^c) \end{aligned}$$

such that the Hellinger distance tends to zero uniformly if  $h_0^{-1} v_\varepsilon^2 = o(\varepsilon)$ , which is ensured by our choice of  $h_0$ . This implies asymptotic equivalence of observing

$\mathscr{Y}$  and  $\tilde{\mathscr{Y}}$  and thus of experiment  $\mathcal{E}_2$  and of just observing  $(y_{jk})_{j \geq 1, k}$  in  $\mathcal{E}_2$ . By independence the latter is equivalent to  $\mathcal{E}_{2, \text{odd}} \otimes \mathcal{E}_{2, \text{even}}$ .

## 9.6 An explicit series representation

We aim at deriving the formula

$$\sum_{j \geq 1} \frac{\lambda^3}{(\lambda^2 + \pi^2 j^2)^2} = \frac{1 + 4\lambda e^{-2\lambda} - e^{-4\lambda}}{4(1 - e^{-2\lambda})^2} - \frac{1}{2\lambda} \quad (9.14)$$

for any  $\lambda > 0$ . We employ Fourier techniques and consider the Fourier coefficients of  $g(x) = e^{-\lambda x/\pi}$ :

$$\hat{g}(j) := (2\pi)^{-1/2} \int_0^{2\pi} g(x) e^{ijx} dx = \frac{\pi(1 - e^{-2\lambda})}{\sqrt{2\pi}(\lambda - i\pi j)}, \quad j \in \mathbb{Z}.$$

For the  $2\pi$ -periodic convolution  $g * g(x) = x e^{-\lambda x/\pi} + (2\pi - x) e^{-\lambda(2+x/\pi)}$  we obtain the Fourier coefficient as a product:

$$\widehat{g * g}(j) = \frac{\pi^2(1 - e^{-2\lambda})^2}{\sqrt{2\pi}(\lambda - i\pi j)^2}.$$

The Parseval formula therefore yields

$$\sum_{j \in \mathbb{Z}} \frac{1}{(\lambda^2 + \pi^2 j^2)^2} = \frac{2}{\pi^3(1 - e^{-2\lambda})^4} \sum_{j \in \mathbb{Z}} |\widehat{g * g}(j)|^2 = \frac{2}{\pi^3(1 - e^{-2\lambda})^4} \|g * g\|_{L^2}^2.$$

We infer that  $(\frac{\lambda}{\pi})^3 \|g * g\|_{L^2}^2$  equals

$$\begin{aligned} & \left(\frac{\lambda}{\pi}\right)^3 \int_0^{2\pi} \left( (1 - e^{-2\lambda})^2 x^2 + 4\pi^2 e^{-4\lambda} + 4\pi(e^{-2\lambda} - e^{-4\lambda})x \right) e^{-2\lambda x/\pi} dx \\ &= \left( \lambda e^{-2\lambda} + (1 - e^{-4\lambda})/4 \right) (1 - e^{-2\lambda})^2 \end{aligned}$$

and thus obtain

$$\sum_{j \in \mathbb{Z}} \frac{\lambda^3}{(\lambda^2 + \pi^2 j^2)^2} = \frac{2}{(1 - e^{-2\lambda})^2} \left( \lambda e^{-2\lambda} + (1 - e^{-4\lambda})/4 \right).$$

Using the symmetry in  $j$ , we establish (9.14).

## Acknowledgement

I am grateful to Marc Hoffmann, Mark Podolskij and Johannes Schmidt-Hieber for very useful discussions.

## References

- BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): “Designing Realized Kernels to Measure the ex post Variation of Equity Prices in the Presence of Noise,” *Econometrica*, 76(6), 1481–1536.
- BROWN, L. D., AND M. G. LOW (1996): “Asymptotic equivalence of nonparametric regression and white noise.,” *Ann. Stat.*, 24(6), 2384–2398.
- CARTER, A. (2006): “A continuous Gaussian process approximation to a nonparametric regression in two dimensions.,” *Bernoulli*, 12(1), 143–156.
- GLOTER, A., AND J. JACOD (2001a): “Diffusions with measurement errors. I: Local asymptotic normality.,” *ESAIM, Probab. Stat.*, 5, 225–242.
- (2001b): “Diffusions with measurement errors. II: Optimal estimators.,” *ESAIM, Probab. Stat.*, 5, 243–260.
- GRAMA, I., AND M. NUSSBAUM (2002): “Asymptotic equivalence for nonparametric regression.,” *Math. Methods Stat.*, 11(1), 1–36.
- IBRAGIMOV, I., AND R. KHAS’MINSKII (1991): “Asymptotically normal families of distributions and efficient estimation.,” *Ann. Stat.*, 19(4), 1681–1724.
- JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2009): “Microstructure noise in the continuous case: the pre-averaging approach.,” *Stochastic Processes Appl.*, 119(7), 2249–2276.
- LE CAM, L., AND G. L. YANG (2000): *Asymptotics in statistics. Some basic concepts. 2nd ed.* Springer Series in Statistics. New York, Springer.
- MUNK, A., AND J. SCHMIDT-HIEBER (2009): “Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise.,” arxiv:0908.3163v2, Math arXiv Preprint.
- MYKLAND, P. (2009): “A Gaussian Calculus for Inference from High Frequency Data,” *Annals of Finance*, to appear.

- NUSSBAUM, M. (1996): “Asymptotic equivalence of density estimation and Gaussian white noise,” *Ann. Stat.*, 24(6), 2399–2430.
- PODOLSKIJ, M., AND M. VETTER (2009): “Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps,” *Bernoulli*, 15(3), 634–658.
- REISS, M. (2008): “Asymptotic equivalence for nonparametric regression with multivariate and random design,” *Ann. Stat.*, 36(4), 1957–1982.
- SHIRYAEV, A. (1995): *Probability. 2nd ed.* Graduate Texts in Mathematics. 95. New York, Springer.
- STONE, C. J. (1982): “Optimal global rates of convergence for nonparametric regression,” *Ann. Stat.*, 10, 1040–1053.
- ZHANG, L. (2006): “Efficient estimation of stochastic volatility using noisy observations: a multi-scale approach,” *Bernoulli*, 12(6), 1019–1043.
- ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): “A tale of two time scales: Determining integrated volatility with noisy high-frequency data,” *J. Am. Stat. Assoc.*, 100(472), 1394–1411.