ESTIMATING THE QUADRATIC COVARIATION MATRIX FROM NOISY OBSERVATIONS: LOCAL METHOD OF MOMENTS AND EFFICIENCY

By Markus Bibinger[†], Nikolaus Hautsch[†], Peter Malec[†] and Markus Reiss[*,†]

An efficient estimator is constructed for the quadratic covariation or integrated covolatility matrix of a multivariate continuous martingale based on noisy and non-synchronous observations under high-frequency asymptotics. Our approach relies on an asymptotically equivalent continuous-time observation model where a local generalised method of moments in the spectral domain turns out to be optimal. Asymptotic semiparametric efficiency is established in the Cramér-Rao sense. Main findings are that non-synchronicity of observation times has no impact on the asymptotics and that major efficiency gains are possible under correlation. Simulations illustrate the finite-sample behaviour.

1. Introduction. The estimation of the quadratic covariation (or integrated covolatility) matrix of a multi-dimensional semi-martingale is studied. Martingales are central objects in stochastics and the estimation of their quadratic covariation from noisy observations is certainly a fundamental topic on its own. Because of its key importance in finance this question moreover attracts high attention from high-frequency financial statistics with implications for portfolio allocation, risk quantification, hedging or asset pricing. While the univariate case has been studied extensively from both angles (see, e.g., the survey of Andersen *et al.* [3] or recent work by Reiß [22] and Jacod and Rosenbaum [15]), statistical inference for the quadratic covariation matrix is not yet well understood. This is, on the one hand, due to a richer geometry, e.g. induced by non-commuting matrices, generating new effects and calling for a deeper mathematical understanding. On the

[*]Corresponding author.

[†]Financial support from the Deutsche Forschungsgemeinschaft via SFB 649 *Ökonomisches Risiko* and FOR 1735 *Structural Inference in Statistics: Adaptation and Efficiency* is gratefully acknowledged. Hautsch also acknowledges financial support from the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF).

AMS 2000 subject classifications: Primary 62M10; secondary 62G05

Keywords and phrases: adaptive estimation, asymptotic equivalence, asynchronous observations, integrated covolatility matrix, quadratic covariation, semiparametric efficiency, microstructure noise, spectral estimation

other hand, statistical challenges arise by the use of underlying multivariate high-frequency data which are typically polluted by noise. Though they open up new ways for statistical inference, their noise properties, significantly different sample sizes (induced by different trading frequencies) as well as irregular and asynchronous spacing in time make estimation in these models far from obvious. Different approaches exist, partly furnish unexpected results, but are rather linked to the method than to the statistical problem. In this paper, we strive for a general understanding of the statistical problem itself, in particular the question of efficiency, while at the same time we develop a local method of moments approach which yields a simple and efficient estimator.

To remain concise, we consider the basic statistical model where the d-dimensional discrete-time process

$$(\mathcal{E}_0) \qquad Y_i^{(l)} = X^{(l)}_{t_i^{(l)}} + \varepsilon_i^{(l)}, 0 \le i \le n_l, 1 \le l \le d,$$

is observed with the d-dimensional continuous martingale

$$X_t = X_0 + \int_0^t \Sigma^{1/2}(s)\, dB_s,\ t \in [0,1],$$

in terms of a d-dimensional standard Brownian motion B and the squared (instantaneous or spot) covolatility matrix

$$\Sigma(t) = \left(\Sigma_{lr}(t)\right)_{1\le l,r\le d} \in \mathbb{R}^{d\times d}.$$

In financial applications, X_t corresponds to the multi-dimensional process of fundamental asset prices whose martingale property complies with market efficiency and exclusion of arbitrage. The major quantity of interest is the quadratic covariation matrix $\int_0^1 \Sigma(t)\, dt$, computed over a normalised interval such as, e.g., a trading day.

The signal part X is assumed to be independent of the observation errors $(\varepsilon_i^{(l)}), 1 \le l \le d, 1 \le i \le n_l$, which are mutually independent and centered normal with variances η_l^2. In the literature on financial high-frequency data these errors capture microstructure frictions in the market (*microstructure noise*). The observation times are given via quantile transformations as $t_i^{(l)} = F_l^{-1}(i/n_l)$ for some distribution functions F_l. While the model ($\mathcal{E}_0$) is certainly an idealisation of many real data situations, its precise analysis delivers a profound understanding and thus serves as a basis for developing procedures in more complex models.

Estimation of the quadratic covariation of a price process is a core research topic in current financial econometrics and various approaches have been put forward in the literature. The realised covariance estimator was studied by Barndorff-Nielsen and Shephard [5] for a setting that neglects both microstructure noise and effects due to the non-syncronicity of observations. Hayashi and Yoshida [14] propose an estimator which is efficient under the presence of asynchronicity, but without noise. Methods accounting for both types of frictions are the quasi-maximum-likelihood approach by Aït-Sahalia *et al.* [1], realised kernels by Barndorff-Nielsen *et al.* [4], pre-averaging by Christensen *et al.* [7], the two-scale estimator by Zhang [24] and the local spectral estimator by Bibinger and Reiß [6]. In contrast to the univariate case, the asymptotic properties of these estimators are involved and the structure of the terms in the asymptotic variance deviate significantly. None of the methods outperforms the others for all settings, calling for a lower efficiency bound as a benchmark.

In this paper, we propose a local method of moments (LMM) estimator, which is optimal in a semiparametric Cramér-Rao sense under the presence of noise and the non-synchronicity of observations. The idea rests on the (strong) asymptotic equivalence in Le Cam's sense of model ($\mathcal{E}_0$) with the continuous time signal-in-white-noise model

$$(\mathcal{E}_1) \qquad dY_t = X_t\,dt + \operatorname{diag}(H_{n,l}(t))_{1\le l\le d}\,dW_t, \quad t\in[0,1]\,,$$

where W is a standard d-dimensional Brownian motion independent of B and the component-wise local noise level is

$$H_{n,l}(t) := \eta_l (n_l F_l'(t))^{-1/2}. \tag{1.1}$$

Here, $F_l'(t)$ represents the local frequency of occurrences ("observation density") and thus $n_l F_l'(t)$ corresponds to the local sample size, which is the continuous-time analogue of the so called *quadratic variation of time*, discussed in the literature. The advantage of the continuous-time model ($\mathcal{E}_1$) is particularly distinctive in the multivariate setting where asynchronicity and different sample sizes in the discrete data ($\mathcal{E}_0$) blur the fundamental statistical structure. If two sequences of statistical experiments are asymptotically equivalent, then any statistical procedure in one experiment has a counterpart in the other experiment with the same asymptotic properties, see Le Cam and Yang [19] for details. Our equivalence proof is constructive such that the procedure we shall develop for ($\mathcal{E}_1$) has a concrete equivalent in ($\mathcal{E}_0$) with the same asymptotic properties.

A remarkable theoretical consequence of the equivalence between ($\mathcal{E}_0$) and ($\mathcal{E}_1$) is that under noise, the asynchronicity of the data does not affect the asymptotically efficient procedures. In fact, in model ($\mathcal{E}_1$), the distribution functions F_l only generate time-varying local noise levels $H_{n,l}(t)$, but the shift between observation times of the different processes does not matter. Hence, locally varying observation frequencies have the same effect as locally varying variances of observation errors and may be pooled. This is in sharp contrast to the noiseless setting where the variance of the Hayashi-Yoshida estimator [14] suffers from errors due to asynchronicity, which carries over to the pre-averaged version by Christensen *et al.* [7] designed for the noisy case. Only if the noise level is assumed to tend to zero so fast that the noiseless case is asymptotically dominant, then the non-synchronicity may induce additional errors.

Our proposed estimator builds on a locally constant approximation of the continuous-time model ($\mathcal{E}_1$) with equi-distant blocks across all dimensions. We show that the errors induced by this approximation vanish asymptotically. Empirical local Fourier coefficients allow for a simple moment estimator for the block-wise spot covolatility matrix. The final estimator then corresponds to a generalized method of moments estimator of $\int_0^1 \Sigma(t)\,dt$, computed as a weighted sum of all individual local estimators (across spectral frequencies and time). Asymptotic efficiency of the resulting LMM estimator is shown to be achieved by an optimal weighting scheme based on the Fisher information matrices of the underlying local moment estimators.

As a result of the non-commutativity of the Fisher information matrices, the LMM estimator for one element of the covariation matrix generally depends on *all* entries of the underlying local covariances. Consequently, the volatility estimator in one dimension substantially gains in efficiency when using data of all other potentially correlated processes. These efficiency gains in the multi-dimensional setup constitute a fundamental difference to the case of i.i.d. observations of a Gaussian vector where the empirical variance of one component is an efficient estimator. Here, using the other entries cannot improve the variance estimator unless the correlation is known, cf. the classical Example 6.6.4 in Lehmann and Casella [20]. This finding is natural for covariance estimation under non-homogeneous noise and because of its general interest we shall discuss a related i.i.d. example in Section 2. The possibility of these efficiency gains in specific cases has been announced by the authors several years ago and is also discussed in Shephard and Xiu [23] and Liu and Tang [21], but a general view and a lower bound were missing.

The next Section 2 gives an overview of the estimation methodology and explains the major implications in a compact and intuitive way with the subsequent sections establishing the general results in full rigour. Emphasis is put on the concrete form of the efficient asymptotic variance-covariance structure which provides a rich geometry and has surprising consequences in practice.

In Section 3 we establish the asymptotic equivalence in Le Cam's sense of models ($\mathcal{E}_0$) and ($\mathcal{E}_1$) in Theorem 3.4. The regularity assumptions required for Σ are less restrictive than in Reiß [22] and particularly allow Σ to jump.

Section 4 introduces the LMM estimator in the spectral domain. Theorem 4.2 provides a multivariate central limit theorem (CLT) for an oracle LMM estimator, using the unknown optimal weights and an information-type matrix for normalisation, which allows for asymptotically diverging sample sizes in the coordinates. Specifying to sample sizes of the same order n, Corollary (4.3) yields a CLT with rate $n^{1/4}$ and a covariance structure between matrix entries, which is explicitly given by concise matrix algebra. Then pre-estimated weight matrices generate a fully adaptive version of the LMM-estimator, which by Theorem 4.4 shares the same asymptotic properties as the oracle estimator. This allows intrinsically feasible confidence sets without pre-estimating asymptotic quantities.

In Section 5 we show that the asymptotic covariance matrix of the LMM estimator attains a lower bound in the Cramér-Rao sense. This lower bound is achieved by a combination of space-time transformations and advanced calculus for covariance operators.

Finally, the discretisation and implementation of the estimator for model ($\mathcal{E}_0$) is briefly described in Section 6 and presented together with some numerical results. We apply the method for a complex and realistic simulation scenario, obtained by a superposition of time-varying seasonality functions, calibrated to real data, and a semi-martingale process with stochastic volatilities. We conclude that the finite sample behaviour of the LMM estimators is well predicted by the asymptotic theory (even in cases where it does not apply formally). Some comparison with competing procedures is provided.

2. Principles and major implications.

2.1. *Spectral LMM methodology.* The time interval $[0,1]$ is partitioned into small blocks $[kh,(k+1)h)$, $k=0,\ldots,h^{-1}-1$, such that on each block a constant parametric covolatility matrix estimate can be sought for (cf. the *local-likelihood* approach). The main estimation idea is then to use block-wise spectral statistics (S_{jk}), which represent localised Fourier coefficients as in Reiß [22]. Specifying to the original discrete data ($\mathcal{E}_0$), they are calculated as

$$S_{jk}=\pi j h^{-1}\Big(\sum_{\nu=1}^{n_l}\Big(Y_\nu-Y_{\nu-1}\Big)\Phi_{jk}\Big(\frac{t_{\nu-1}^{(l)}+t_\nu^{(l)}}{2}\Big)\Big)_{1\le l\le d}\in\mathbb{R}^d\,, \quad (2.1)$$

with sine functions Φ_{jk} of frequency index j on each block $[kh,(k+1)h]$ given by

$$\Phi_{jk}(t) = \tfrac{\sqrt{2h}}{j\pi} \sin\left(j\pi h^{-1}(t-kh)\right) \mathbb{1}_{[kh,(k+1)h]}(t), \quad j \geq 1\,. \tag{2.2}$$

The same blocks are used across all dimensions d with their size h being determined by the least frequently observed process.

The statistics (S_{jk}) are Riemann-Stieltjes sum approximations to Fourier integrals based on a possibly non-equidistant grid. The discrete-time processes $(Y_i^{(l)})$ can be transformed into a continuous-time process via linear interpolation in each dimension, which yields piecewise constant (weak) derivatives, with the S_{jk} being interpreted as integrals over these derivatives. Mathematically, the asymptotic equivalence of ($\mathcal{E}_0$) and ($\mathcal{E}_1$) based on this linear interpolation is made rigorous in Theorem 3.4. The required regularity condition is that $\Sigma(t)$ is the sum of an L^2-Sobolev function of regularity β and an L^2-martingale and the size of β accommodates for asymptotically separating sample sizes $(n_l)_{1\leq l\leq d}$. In model ($\mathcal{E}_1$) by partial integration, the statistics S_{jk} then correspond to

$$S_{jk}^{(l)} = \pi j h^{-1} \int_{kh}^{(k+1)h} \varphi_{jk}(t)\, dY^{(l)}(t) \tag{2.3}$$

with block-wise cosine functions $\varphi_{jk} = \Phi'_{jk}$ which form an orthonormal system in $L^2([0,1])$. As they serve also as the eigenfunctions of the Karhunen-Loève decomposition of a Brownian motion, they carry maximal information for Σ. What is more, the spectral statistics S_{jk} de-correlate the observations and thus form their (block-wise) principal components, assuming that Σ and the noise levels are block-wise constant. Then the entire family $(S_{jk})_{jk}$ is independent and

$$S_{jk} \sim \mathbf{N}(0, C_{jk}), \quad C_{jk} = \Sigma^{kh} + \pi^2 j^2 h^{-2}\,\mathrm{diag}(H_{n,l}^{kh})_l^2, \tag{2.4}$$

with the k-th block average Σ^{kh} of Σ and $H_{n,l}^{kh}$ encoding the local noise level, cf. (4.2) below.

This relationship suggests to estimate Σ^{kh} in each frequency j by bias-corrected spectral covariance matrices $S_{jk}S_{jk}^{\top} - \pi^2 j^2 h^{-2}\,\mathrm{diag}\left((H_{n,l}^{kh})^2\right)_l$. The resulting *local method of moment (LMM) estimator* then takes weighted sums across all frequencies and blocks

$$\mathrm{LMM}^{(n)} := \sum_{k=0}^{h^{-1}-1} h \sum_{j=1}^{\infty} W_{jk} vec\left(S_{jk}S_{jk}^{\top} - \pi^2 j^2 h^{-2}\,\mathrm{diag}\left((H_{n,l}^{kh})^2\right)_l\right),$$

where $W_{jk} \in \mathbb{R}^{d^2 \times d^2}$ are weight matrices and matrices $A \in \mathbb{R}^{d \times d}$ are transformed into vectors via

$$vec(A) := \big(A_{11}, A_{21}, \dots, A_{d1}, A_{12}, A_{22}, \dots, A_{d2}, \dots, A_{d(d-1)}, A_{dd}\big)^\top \in \mathbb{R}^{d^2}.$$

To ensure efficiency, the oracle and adaptive choice of the weight matrices W_{jk} are based on Fisher information calculus, see Section 4 below. Let us mention that scalar weights for each matrix estimator entry as in Bibinger and Reiß [6] will not be sufficient to achieve (asymptotic) efficiency and the W_{jk} will be densely populated.

The matrix estimator *per se* is not ensured to be positive semi-definite, but it is symmetric and can be projected onto the cone of positive semi-definite matrices by putting negative eigenvalues to zero. This projection only improves the estimator, while the adjustment is asymptotically negligible in the CLT. For the relevant question of confidence sets, the estimated non-asymptotic Fisher information matrices are positive-semidefinite (basically, estimating C_{jk} from above) and finite sample inference is always feasible.

2.2. *The efficiency bound.* Deriving the covariance structure of a matrix estimator requires tensor notation, see e.g. Fackler [11] or textbooks on multivariate analysis. Kronecker products $A \otimes B \in \mathbb{R}^{d^2 \times d^2}$ for $A, B \in \mathbb{R}^{d \times d}$ are defined as

$$(A \otimes B)_{d(p-1)+q, d(p'-1)+q'} = A_{pp'} B_{qq'}, \quad p, q, p', q' = 1, \dots, d.$$

The covariance structure for the empirical covariance matrix of a standard Gaussian vector is defined as

$$\mathcal{Z} = \mathbb{C}\mathrm{OV}(vec(ZZ^\top)) \in \mathbb{R}^{d^2 \times d^2} \text{ for } Z \sim \mathbf{N}(0, E_d). \tag{2.5}$$

We can calculate $\mathcal{Z}$ explicitly as

$$\mathcal{Z}_{d(p-1)+q, d(p'-1)+q'} = (1 + \delta_{p,q}) \delta_{\{p,q\},\{p',q'\}}, p, q, p', q' = 1, \dots, d,$$

exploiting the property $\mathcal{Z} vec(A) = vec(A + A^\top)$ for all $A \in \mathbb{R}^{d \times d}$. It is classical, cf. Lehmann and Casella [20], that for n i.i.d. Gaussian observations $Z_i \sim \mathbf{N}(0, \Sigma)$, the empirical covariance matrix $\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n Z_i Z_i^\top$ is an asymptotically efficient estimator of Σ satisfying

$$\sqrt{n}\, vec(\hat{\Sigma}_n - \Sigma) \xrightarrow{\mathcal{L}} \mathbf{N}(0, (\Sigma \otimes \Sigma)\mathcal{Z}).$$

The asymptotic variance can be easily checked by the rule $vec(ABC) = (C^\top \otimes A) vec(B)$ and the fact that $\mathcal{Z}$ commutes with $(\Sigma \otimes \Sigma)^{1/2} = \Sigma^{1/2} \otimes \Sigma^{1/2}$ such that $\mathbb{C}\mathrm{OV}(vec(\hat{\Sigma}_n))$ equals

$$\mathbb{C}\mathrm{OV}(vec(\Sigma^{1/2} ZZ^\top \Sigma^{1/2})) = (\Sigma^{1/2} \otimes \Sigma^{1/2}) \mathcal{Z} (\Sigma^{1/2} \otimes \Sigma^{1/2}) = (\Sigma \otimes \Sigma)\mathcal{Z}.$$

Before proceeding, let us provide an intuitive understanding of the efficiency gains from other dimensions by looking at another easy case with independent observations. Suppose an i.i.d. sample $Z_1, \ldots, Z_n \sim \mathbf{N}(0, \Sigma)$, $\Sigma \in \mathbb{R}^{d\times d}$ unknown, is observed indirectly via $Y_j = Z_j + \varepsilon_j$, blurred by independent non-homogeneous noise $\varepsilon_j \sim \mathbf{N}(0, \eta_j^2 E_d)$, $j = 1, \ldots, n$, with identity matrix E_d and $\eta_1, \ldots, \eta_n > 0$ known. Then the sample covariance matrix $\hat{C}_Y = \sum_{j=1}^n Y_j Y_j^\top$ and a bias correction yields a first natural estimator $\hat{\Sigma}^{(1)} = \hat{C}_Y - \eta^2 E_d$, $\eta^2 = \sum_j \eta_j^2/n$. Yet, we can weight each observation differently by some $w_j \in \mathbb{R}$ with $\sum_j w_j = 1$ and obtain a second estimator $\hat{\Sigma}^{(2)} = \sum_{j=1}^n w_j(Y_jY_j^\top - \eta_j^2 E_d)$. For optimal estimation of the first variance Σ_{11}, we should choose (as in a weighted least squares approach) $w_j = (\Sigma_{11} + \eta_j^2)^{-2}/(\sum_i (\Sigma_{11} + \eta_i^2)^{-2})$ to obtain

$$\mathbb{V}\mathrm{ar}(\hat{\Sigma}_{11}^{(2)}) = 2\Big(\sum_{j=1}^n (\Sigma_{11} + \eta_j^2)^{-2}\Big)^{-1} \le \frac{2}{n^2}\sum_{j=1}^n (\Sigma_{11} + \eta_j^2)^2 = \mathbb{V}\mathrm{ar}(\hat{\Sigma}_{11}^{(1)}),$$

where the bound is due to Jensen's inequality. More generally, we can use weight matrices $W_j \in \mathbb{R}^{d^2\times d^2}$ and introduce $\hat{\Sigma}^{(3)} = \sum_{j=1}^n W_j vec(Y_jY_j^\top - \eta_j^2 E_d)$. Since the matrices $C_j = \Sigma + \eta_j E_d$ commute, its covariance structure is given by $\mathbb{COV}(\hat{\Sigma}^{(3)}) = \sum_{j=1}^n W_j (C_j \otimes C_j)\mathcal{Z} W_j^\top$. This is minimal for $W_j = (\sum_i C_i^{-1} \otimes C_i^{-1})^{-1}(C_j^{-1} \otimes C_j^{-1})$, which gives $\mathbb{COV}(\hat{\Sigma}^{(3)}) = (\sum_j C_j^{-1} \otimes C_j^{-1})^{-1}\mathcal{Z}$. The matrices W_j are diagonal if all η_j coincide or if Σ is diagonal. Otherwise, the estimator for one matrix entry involves in general all other entries in $Y_jY_j^\top$ and in particular $\mathbb{V}\mathrm{ar}(\hat{\Sigma}_{11}^{(3)}) < \mathbb{V}\mathrm{ar}(\hat{\Sigma}_{11}^{(2)})$ holds. Considering as $(Y_j)_{j\ge 1}$ the spectral statistics $(S_{jk}))_{j\ge 1}$ on a fixed block k, this example reveals the heart of our analysis for the LMM estimator.

Similar to the i.i.d. case, for equidistant observations $(X_{i/n})_{1\le i\le n}$ of $X_t = \int_0^t \Sigma(s)dB_s$ without noise, the realised covariation matrix

$$\widehat{RCV}_n = \sum_{i=1}^n (X_{i/n} - X_{(i-1)/n})(X_{i/n} - X_{(i-1)/n})^\top$$

satisfies the d^2-dimensional central limit theorem

$$\sqrt{n}\, vec\Big(\widehat{RCV}_n - \int_0^1 \Sigma(t)\,dt\Big) \xrightarrow{\mathcal{L}} \mathbf{N}\Big(0, \Big(\int_0^1 \Sigma(t)\otimes\Sigma(t)\,dt\Big)\mathcal{Z}\Big),$$

provided $t \mapsto \Sigma(t)$ is Riemann-integrable. In the one-dimensional case, it is known that in the presence of noise the optimal rate of convergence not only changes from $n^{-1/2}$ to $n^{-1/4}$, but also the optimal variance changes from $2\sigma^4$ to $8\sigma^3$. The corresponding analogue of $(\Sigma\otimes\Sigma)\mathcal{Z}$ in the noisy case is not obvious at all. So far, only the result by Barndorff-Nielsen et al. [4], establishing $(\Sigma\otimes\Sigma)\mathcal{Z}$ as limiting variance under the suboptimal rate $n^{-1/5}$, was available and even a conjecture concerning the efficiency bound was lacking.

To illustrate our multivariate efficiency results under noise let us for simplicity illustrate a special case of Corollary 4.3 for equidistant observations, i.e. $t_i^{(l)}=i/n$, and homogeneous noise level $\eta_l=\eta$. Then, the oracle (and also the adaptive) estimator $\mathrm{LMM}^{(n)}$ satisfies under mild regularity conditions (omitting the integration variable t)

$$n^{1/4}\Big(\mathrm{LMM}^{(n)}-\int_0^1 vec(\Sigma)\Big)\xrightarrow{\mathcal{L}}\mathbf{N}\left(0,2\eta\int_0^1\left(\Sigma\otimes\Sigma^{1/2}+\Sigma^{1/2}\otimes\Sigma\right)\mathcal{Z}\right).$$

In Theorem 5.3, it will be shown that this asymptotic covariance structure is optimal in a semiparametric Cramér-Rao sense. Consequently, the efficient asymptotic variance **AVAR** for estimating $\int_0^1\Sigma_{pp}(t)\,dt$ is

$$\mathbf{AVAR}\Big(\int_0^1\Sigma_{pp}(t)\,dt\Big)=8\eta\int_0^1\Sigma_{pp}(t)(\Sigma^{1/2}(t))_{pp}\,dt.$$

For the asymptotic variance of the estimator of $\int_0^1\Sigma_{pq}(t)\,dt$ we obtain

$$2\eta\int_0^1\Big((\Sigma^{1/2})_{pp}\Sigma_{qq}+(\Sigma^{1/2})_{qq}\Sigma_{pp}+2(\Sigma^{1/2})_{pq}\Sigma_{pq}\Big)(t)\,dt.$$

Let us illustrate specific examples. First, in the case $d=1$ and $\Sigma=\sigma^2$, the asymptotic variance simplifies to

$$\mathbf{AVAR}\Big(\int_0^1\sigma^2(t)\,dt\Big)=8\eta\int_0^1\sigma^3(t)\,dt,$$

coinciding with the efficiency bound in Reiß [22]. For $d>1$, $p\neq q$ in the independent case $\Sigma=\mathrm{diag}(\sigma_p^2)_{1\le p\le d}$, we find

$$\mathbf{AVAR}\Big(\int_0^1\Sigma_{pq}(t)\,dt\Big)=2\eta\int_0^1(\sigma_p^2\sigma_q+\sigma_p\sigma_q^2)(t)\,dt.$$

An interesting example is the case $d=2$ with spot volatilities $\sigma_1^2(t)=\sigma_2^2(t)=\sigma^2(t)$ and general correlation $\rho(t)$, i.e. $\sigma_{12}(t)=(\rho\sigma_1\sigma_2)(t)$. In this case we obtain

$$\mathbf{AVAR}\Big(\int_0^1\sigma_1^2(t)\,dt\Big)=4\eta\int_0^1\sigma^3(t)\Big(\sqrt{1+\rho(t)}+\sqrt{1-\rho(t)}\Big)\,dt,$$
$$\mathbf{AVAR}\Big(\int_0^1\sigma_{12}(t)\,dt\Big)=2\eta\int_0^1\sigma^3(t)((1+\rho(t))^{3/2}+(1-\rho(t))^{3/2})\,dt.$$

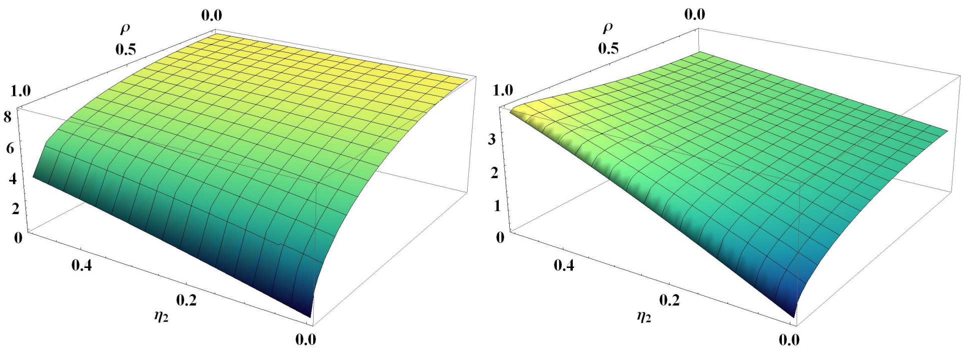

FIG 1. *Asymptotic variances of LMM for volatility σ_1^2 (left) and covolatility σ_{12} (right) plotted against correlation ρ and noise level η_2 (constant in time).*

With time-constant parameters these bounds decay for σ_1^2 (resp. grow for σ_{12}) in $|\rho|$ from $8\eta\sigma^3$ (resp. $4\eta\sigma^3$) at $\rho = 0$ to $4\sqrt{2}\eta\sigma^3$ at $|\rho| = 1$ for both cases.

Figure 1 illustrates the asymptotic variance in the case of volatilities $\sigma_1^2 = \sigma_2^2 = 1$ and covolatility $\sigma_{12} = \rho$ (constant in time) and the first noise level given by $\eta_1 = 1$. The left plot shows the asymptotic variance of the estimator of σ_1^2 as a function of ρ and η_2. It is shown that using observations from the other (correlated) process induces clear efficiency gains rising in ρ. If the noise level η_2 for the second process is small, the asymptotic variance can even approach zero. The plot on the right shows the same dependence for estimating the covolatility σ_{12}. For comparable size of η_2 and η_1 the asymptotic variance increases in ρ, which is explained by the fact that also the value to be estimated increases. For small values of η_2, however, the efficiency gain by exploiting the correlation prevails.

For larger dimensions d, the variance can even be of order $\mathcal{O}(1/\sqrt{d})$: in the concrete case where all volatilities and noise levels equal 1, the asymptotic variance for estimating σ_1^2 can be reduced from 8 (using only observations from the first component or if Σ is diagonal) down to $8/\sqrt{d}$ (in case of perfect correlation).

All the preceding examples can be worked out for different noise levels η_p. For a fixed entry (p, q) generally all noise levels enter and can be only decoupled in case of a diagonal covariation matrix $\Sigma = \operatorname{diag}(\sigma_p^2)_{1\le p\le d}$. Then, the covariance simplifies to

$$p \neq q:\ 2\int_0^1 \left(\eta_p\sigma_p\sigma_q^2 + \eta_q\sigma_q\sigma_p^2\right)(t)\,dt; \quad p = q:\ 8\int_0^1 \left(\eta_p\sigma_p^3\right)(t)\,dt.$$

Finally, we can also investigate the estimation of the entire quadratic covariation matrix $\int_0^1 \Sigma(t)\,dt$ under homogeneous noise level and measure its loss by the squared $(d\times d)$-Hilbert-Schmidt norm. Summing up the variances for each entry, we obtain the asymptotic risk

$$\frac{4\eta}{\sqrt{n}}\int_0^1 \Big(\operatorname{trace}(\Sigma^{1/2})\operatorname{trace}(\Sigma)+\operatorname{trace}(\Sigma^{3/2})\Big)(t)\,dt.$$

This can be compared with the corresponding Hilbert-Schmidt norm error $\frac{1}{n}(\operatorname{trace}(\Sigma)^2+\operatorname{trace}(\Sigma^2))$ for the empirical covariance matrix in the i.i.d. Gaussian $\mathbf{N}(0,\Sigma)$-setting.

3. From discrete to continuous-time observations.

3.1. *Setting.* First, let us specify different regularity assumptions. For functions $f:[0,1]\to\mathbb{R}^m$, $m\geq 1$ or also $m=d\times d$ for matrix values, we introduce the L^2-Sobolev ball of order $\alpha\in(0,1]$ and radius $R>0$ given by

$$H^\alpha(R)=\{f\in H^\alpha([0,1],\mathbb{R}^m)|\|f\|_{H^\alpha}\leq R\}\text{ where }\|f\|_{H^\alpha}:=\max_{1\leq i\leq m}\|f_i\|_{H^\alpha},$$

which for matrices means $\|f\|_{H^\alpha}:=\max_{1\leq i,j\leq d}\|f_{ij}\|_{H^\alpha}$. We also consider Hölder spaces $C^\alpha([0,1])$ and Besov spaces $B^\alpha_{p,q}([0,1])$ of such functions. Canonically, for matrices we use the spectral norm $\|\cdot\|$ and we set $\|f\|_\infty:=\sup_{t\in[0,1]}\|f(t)\|$.

In order to pursue asymptotic theory, we impose that the deterministic samplings in each component can be transferred to an equidistant scheme by respective quantile transformations independent of $n_l, 1\leq l\leq d$.

ASSUMPTION 3.1.-*(α) Suppose that there exist differentiable distribution functions F_l with $F_l'\in C^\alpha([0,1])$, $F_l(0)=0$, $F_l(1)=1$ and $F_l'>0$ such that the observation times in ($\mathcal{E}_0$) are generated by $t_i^{(l)}=F_l^{-1}(i/n_l)$, $0\leq i\leq n_l$, $1\leq l\leq d$.*

We gather all assertions on the instantaneous covolatility matrix function $\Sigma(t)$, $t\in[0,1]$, which we shall require at some point.

ASSUMPTION 3.2. *Let $\Sigma:[0,1]\to\mathbb{R}^{d\times d}$ be a possibly random function with values in the class of symmetric, positive semi-definite matrices, independent of X and the observational noise, satisfying:*

(i-β) $\Sigma\in H^\beta([0,1])$ *for* $\beta>0$.
(ii-α) $\Sigma=\Sigma^B+\Sigma^M$ *with* $\Sigma^B\in B^\alpha_{1,\infty}([0,1])$ *for* $\alpha>0$ *and* Σ^M *a matrix-valued* L^2*-martingale.*

(iii-$\underline{\Sigma}$) $\Sigma(t) \geq \underline{\Sigma}$ *for a strictly positive definite matrix* $\underline{\Sigma}$ *and all* $t \in [0,1]$.

We briefly discuss the different function spaces, see e.g. Cohen [9, Section 3.2] for a survey. First, any α-Hölder-continuous function lies in the L^2-Sobolev space H^α and any H^α-function lies in the Besov space $B^\alpha_{1,\infty}$, where differentiability is measured in an L^1-sense. The important class of bounded variation functions (e.g., modeling jumps in the volatility) lies in $B^1_{1,\infty}$, but only in H^α for $\alpha < 1/2$. In particular, part (ii-α), $\alpha \leq 1$, covers L^2-semi-martingales by separate bounds on the drift (bounded variation) and martingale part. Beyond classical theory in this area is the fact that also non-semi-martingales like fractional Brownian motion B^H with hurst parameter $H > 1/2$ give rise to feasible volatility functions in the results below, using $B^H \in C^{H-\varepsilon} \cap B^H_{1,\infty}$ for any $\varepsilon > 0$ as in Ciesielski *et al.* [8].

In the sequel, the potential randomness of Σ is often not discussed additionally because by independence we can always work conditionally on Σ. Finally, let us point out that we could weaken the Hölder-assumptions on $F_1, \ldots, F_d$ towards Sobolev or Besov regularity at the cost of tightening the assumptions on Σ. For the sake of clarity this is not pursued here.

Throughout the article we write $Z_n = \mathcal{O}_P(\delta_n)$ and $Z_n = \scriptstyle\mathcal{O}_P(\delta_n)$ for a sequence of random variables Z_n and a sequence δ_n, to express that $\delta_n^{-1} Z_n$ is tight and tends to zero in probability, respectively. Analogously $\mathcal{O}$ (or equivalently $\lesssim$) and $\scriptstyle\mathcal{O}$ refer to deterministic sequences. We write $Z_n \asymp Y_n$ if $Z_n = \mathcal{O}_P(Y_n)$ and $Y_n = \mathcal{O}_P(Z_n)$ and the same for deterministic quantities.

3.2. *Continuous-time experiment.*

DEFINITION 3.3. *Let* $\mathcal{E}_0((n_l)_{1\leq l\leq d}, \beta, R)$ *with* $n_l \in \mathbb{N}, \beta \in (0,1], R > 0$, *be the statistical experiment generated by observations from* ($\mathcal{E}_0$) *with* $\Sigma \in H^\beta(R)$. *Analogously, let* $\mathcal{E}_1((n_l)_{1\leq l\leq d}, \beta, R)$ *be the statistical experiment generated by observing* ($\mathcal{E}_1$) *with the same parameter class.*

As we shall establish next, experiments ($\mathcal{E}_0$) and ($\mathcal{E}_1$) will be asymptotically equivalent as $n_l \to \infty, 1 \leq l \leq d$, at a comparable speed, denoting

$$n_{min} = \min_{1\leq l\leq d} n_l \text{ and } n_{max} = \max_{1\leq l\leq d} n_l \,.$$

THEOREM 3.4. *Grant Assumption 3.1-(β) on the design. The statistical experiments* $\mathcal{E}_0((n_l)_{1\leq l\leq d}, \beta, R)$ *and* $\mathcal{E}_1((n_l)_{1\leq l\leq d}, \beta, R)$ *are asymptotically equivalent for any* $\beta \in (0, 1/2]$ *and* $R > 0$, *provided*

$$n_{min} \to \infty, \quad n_{max} = \scriptstyle\mathcal{O}((n_{min})^{1+\beta}).$$

More precisely, the Le Cam distance Δ is of order

$$\Delta\left(\mathcal{E}_0((n_l)_{1\le l\le d},\beta,R),\mathcal{E}_1((n_l)_{1\le l\le d},\beta,R)\right)=\mathcal{O}\left(R^2\left(\sum_{l=1}^d n_l/\eta_l^2\right)n_{min}^{-1-\beta}\right).$$

By inclusion, the result also applies for $\beta > 1/2$ when in the remaining expressions β is replaced by $\min(\beta, 1/2)$. A standard Sobolev smoothness of Σ is β almost $1/2$ for diffusions with finitely many or absolutely summable jumps. In that case, the asymptotic equivalence result holds if n_{max} grows more slowly than $n_{min}^{3/2}$. Theorem 3.4 is proved in the appendix in a constructive way by warped linear interpolation, which yields a readily implementable procedure, cf. Section 6 below.

4. Localisation and method of moments.

4.1. *Construction.* We partition the interval $[0,1]$ in blocks $[kh,(k+1)h)$ of length h. On each block a parametric MLE for a constant model could be sought for. Its numerical determination, however, is difficult and unstable due to the non-concavity of the ML objective function and its analysis is quite involved. Yet, the likelihood equation leads to spectral statistics whose empirical covariances estimate the quadratic covariation matrix. We therefore prefer a localised method of moments (LMM) for these spectral statistics where for an adaptive version the theoretically optimal weights are determined in a pre-estimation step, in analogy with the classical (multi-step) GMM (generalised method of moments) approach by Hansen [13].

As motivated in Section 2, let us consider the local spectral statistics S_{jk} in (2.3) from the continuous-time experiment ($\mathcal{E}_1$). In order to specify our estimator, we consider a locally constant approximation of the general non-parametric model.

DEFINITION 4.1. *Set $\bar f_h(t) := h^{-1}\int_{kh}^{(k+1)h} f(s)ds$ for $t\in[kh,(k+1)h)$, $k\in\mathbb{N}_0$, a function f on $[0,1]$ and $h\in(0,1)$. Assume $h^{-1}\in\mathbb{N}$ and let $X_t^h = X_0+\int_0^t \bar\Sigma_h^{1/2}(s)\,dB_s$ with a d-dimensional standard Brownian motion B. Define the process*

$$(\mathcal{E}_2)\qquad d\tilde Y_t = X_t^h\,dt+\operatorname{diag}\left(\sqrt{\overline{H^2}_{n,l,h}(t)}\right)_{1\le l\le d} dW_t\,,t\in[0,1]\,,$$

where W is a standard Brownian motion independent of B and with noise level (1.1). The statistical model generated by the observations from ($\mathcal{E}_2$) for $\Sigma\in H^\beta(R)$ is denoted by $\mathcal{E}_2((n_l)_{1\le l\le d},h,\beta,R)$.

In experiment ($\mathcal{E}_2$) we thus observe a process with a covolatility matrix which is constant on each block $[kh,(k+1)h)$ and corrupted by noise of block-wise constant magnitude. Our approach is founded on the idea that for small block sizes h and sufficient regularity this piecewise constant approximation is close to ($\mathcal{E}_1$).

The LMM estimator is built from the data in experiment $\mathcal{E}_1$, but designed for the block-wise parametric model ($\mathcal{E}_2$). In ($\mathcal{E}_2$), the L^2-orthogonality of (φ_{jk}) as well as that of (Φ_{jk}) imply (cf. Reiß [22])

$$S_{jk} \sim \mathbf{N}\left(0, C_{jk}\right) \text{ independent for all } (j,k) \tag{4.1}$$

with covariance matrix

$$C_{jk} = \Sigma^{kh} + \pi^2 j^2 h^{-2} \operatorname{diag}(H^{kh}_{n,l})^2_l, \Sigma^{kh} = \bar{\Sigma}_h(kh),\ H^{kh}_{n,l} = (\overline{H^2}_{n,l,h}(kh))^{1/2}. \tag{4.2}$$

Let us further introduce the Fisher information-type matrices

$$I_{jk} = C_{jk}^{-1} \otimes C_{jk}^{-1}, \quad I_k = \sum_{j=1}^{\infty} I_{jk}, \quad j \geq 1,\ k = 0,\dots,h^{-1}-1.$$

Our local method of moments estimator with oracle weights $\mathrm{LMM}^{(n)}_{or}$ exploits that on each block a natural second moment estimator of Σ^{kh} is given as a convex combination of the bias-corrected empirical covariances:

$$\mathrm{LMM}^{(n)}_{or} := \sum_{k=0}^{h^{-1}-1} h \sum_{j=1}^{\infty} W_{jk} vec\left(S_{jk}S_{jk}^{\top} - \tfrac{\pi^2 j^2}{h^2} \operatorname{diag}\left((H^{kh}_{n,l})^2\right)_{1\leq l\leq d}\right). \tag{4.3}$$

The optimal weight matrices W_{jk} in the oracle case are obtained as

$$W_{jk} := I_k^{-1} I_{jk} \quad \in \mathbb{R}^{d^2\times d^2}. \tag{4.4}$$

Note that C_{jk}, I_{jk}, I_k and W_{jk} all depend on $(n_l)_{1\leq l\leq d}$ and h, which is omitted in the notation. Finally, observe that (4.2) and $\sum_j W_{jk} = E_{d^2}$ imply that $\mathrm{LMM}^{(n)}_{or}$ is unbiased under model ($\mathcal{E}_2$).

4.2. *Asymptotic properties of the estimators.* We formulate the main result of this section that the oracle estimator (4.3) and also a fully adaptive version for the quadratic covariation matrix satisfy central limit theorems.

THEOREM 4.2. *Let Assumptions 3.1-(α), 3.2(ii-α) and 3.2(iii-$\underline{\Sigma}$) with $\alpha > 1/2$ hold true for observations from model ($\mathcal{E}_1$). The oracle estimator (4.3) yields a consistent estimator for $vec(\int_0^1 \Sigma(s)\,ds)$ as $n_{min} \to \infty$ and*

$h = h_0 n_{min}^{-1/2}$ with $h_0 \to \infty$. Moreover, if $n_{max} = \mathcal{O}(n_{min}^{2\alpha})$ and $h = \mathcal{O}(n_{max}^{-1/4})$, then a multivariate central limit theorem holds:

$$\mathbf{I}_n^{1/2}\left(\mathrm{LMM}_{or}^{(n)} - vec\left(\int_0^1 \Sigma(s)\,ds\right)\right) \xrightarrow{\mathcal{L}} \mathbf{N}\left(0, \mathcal{Z}\right) \;\; in\ \mathcal{E}_1 \tag{4.5}$$

with $\mathcal{Z}$ from (2.5) and $\mathbf{I}_n^{-1} = \sum_{k=0}^{h^{-1}-1} h^2 I_k^{-1}$.

While the preceding result is most useful in applications, it is, of course, important to understand the asymptotic covariance structure of the estimator as well, cf. the discussion of efficiency above. Therefore, we consider comparable sample sizes and normalise with $n_{min}^{1/4}$ in the following result.

COROLLARY 4.3. *Under the assumptions of Theorem 4.2 suppose $n_{min}/n_p \to \nu_p \in (0,1]$ for $p = 1,\dots,d$ and introduce $\mathcal{H}(t) = \mathrm{diag}(\eta_p \nu_p^{1/2} F_p'(t)^{-1/2})_p \in \mathbb{R}^{d\times d}$ and $\Sigma_{\mathcal{H}}^{1/2} := \mathcal{H}(\mathcal{H}^{-1}\Sigma\mathcal{H}^{-1})^{1/2}\mathcal{H}$. Then*

$$n_{min}^{1/4}\left(\mathrm{LMM}_{or}^{(n)} - vec\left(\int_0^1 \Sigma(s)\,ds\right)\right) \xrightarrow{\mathcal{L}} \mathbf{N}\left(0, \mathbf{I}^{-1}\mathcal{Z}\right) \;\; in\ \mathcal{E}_1 \tag{4.6}$$

with

$$\mathbf{I}^{-1} = 2\int_0^1 (\Sigma \otimes \Sigma_{\mathcal{H}}^{1/2} + \Sigma_{\mathcal{H}}^{1/2} \otimes \Sigma)(t)\,dt.$$

In particular, the entries satisfy for $p, q = 1,\dots,d$

$$n_{min}^{1/4}\left((\mathrm{LMM}_{or}^{(n)})_{p(d-1)+q} - \int_0^1 \Sigma_{pq}(s)\,ds\right) \xrightarrow{\mathcal{L}} \tag{4.7}$$
$$\mathbf{N}\left(0, 2(1+\delta_{p,q})\int_0^1 (\Sigma_{pp}(\Sigma_{\mathcal{H}}^{1/2})_{qq} + \Sigma_{qq}(\Sigma_{\mathcal{H}}^{1/2})_{pp} + 2\Sigma_{pq}(\Sigma_{\mathcal{H}}^{1/2})_{pq})(t)\,dt\right).$$

The variance (4.7) will coincide with the lower bound obtained in Section 5 below. The local noise level in $\mathcal{H}(t)$ depends on the observational noise level η_p and the local sample size $\nu_p^{-1}F_p'(t)$, $p = 1,\dots,d$, after normalisation by n_{min}. It is easy to see that in the case $n_{min}/n_p \to 0$ the asymptotic variance vanishes for all entries (p,q), $q = 1,\dots,d$. We infer the structure of the asymptotic covariance matrix using block-wise diagonalisation in Appendix B.

To obtain a feasible estimator, the optimal weight matrices $W_{jk} = W_j(\Sigma^{kh})$ and the information-type matrices $I_{jk} = I_j(\Sigma^{kh})$ are estimated in a preliminary step from the same data. To reduce variability in the estimate, a coarser grid of r^{-1} equidistant intervals, $r/h \in \mathbb{N}$ is employed for $\hat{W}_{jk}$. As derived in Bibinger and Reiß [6] for supremum norm loss and extended to L^1-loss and Besov regularity using the L^1-modulus of continuity as in the case of wavelet estimators (Cor. 3.3.1 in Cohen [9]), a preliminary estimator $\hat{\Sigma}(t)$ of the instantaneous covolatility matrix $\Sigma(t)$ exists with

$$\|\hat{\Sigma}-\Sigma\|_{L^1}=\mathcal{O}_P\Big(n_{min}^{-\alpha/(4\alpha+2)}\Big) \tag{4.8}$$

for $\Sigma \in B^{\alpha}_{1,\infty}([0,1])$. For block k with $kh \in [mr,(m+1)r)$ we set

$$\hat{W}_{jk}=W_j(\hat{\Sigma}^{mr}),\ \hat{I}_{jk}=I_j(\hat{\Sigma}^{kh}) \text{ with } \hat{\Sigma}^{mr}=\overline{\hat{\Sigma}}_r(mr),\ \hat{\Sigma}^{kh}=\overline{\hat{\Sigma}}_h(kh).$$

The LMM estimator with adaptive weights is then given by

$$\mathrm{LMM}_{ad}^{(n)}=\sum_{k=0}^{h^{-1}-1}h\sum_{j=1}^{\infty}\hat{W}_{jk}\,vec\left(S_{jk}S_{jk}^{\top}-\tfrac{\pi^2 j^2}{h^2}\operatorname{diag}\left((H^{kh}_{n,l})^2\right)_{1\le l\le d}\right). \tag{4.9}$$

We estimate the total covariance matrix via

$$\hat{\mathbf{I}}_n^{-1}=\sum_{k=0}^{h^{-1}-1}h^2\Big(\sum_{j=1}^{\infty}\hat{I}_{jk}\Big)^{-1}. \tag{4.10}$$

As $j\to\infty$, the weights $W_j(\Sigma)$ and the matrices $I_j(\Sigma)$ decay like j^{-4} in norm, compare Lemma C.1 below, such that in practice a finite sum over frequencies j suffices. By a tight bound on the derivatives of $\Sigma \mapsto W_j(\Sigma)$ we show in Appendix C.4 the following general result.

THEOREM 4.4. *Suppose* $\Sigma \in B^{\alpha}_{1,\infty}([0,1])$ *for* $\alpha \in (1/2,1]$ *satisfying* $\alpha/(2\alpha+1) > \log(n_{max})/\log(n_{min})-1$. *Choose* $h,r\to 0$ *such that* $h_0 = hn_{min}^{1/2} \asymp \log(n_{min})$ *and* $n_{min}^{-\alpha/(2\alpha+1)} \lesssim r \lesssim (n_{min}/n_{max})^{1/2}$, $h^{-1},r^{-1},r/h \in \mathbb{N}$. *If the pilot estimator* $\hat{\Sigma}$ *satisfies* (4.8), *then under the conditions of Theorem 4.2 the adaptive estimator* (4.9) *satisfies*

$$\hat{\mathbf{I}}_n^{1/2}\left(\mathrm{LMM}_{ad}^{(n)}-vec\left(\int_0^1\Sigma(s)\,ds\right)\right)\xrightarrow{\mathcal{L}}\mathbf{N}\left(0,\mathcal{Z}\right), \tag{4.11}$$

with $\hat{\mathbf{I}}_n$ *from* (4.10).

Moreover, Corollary 4.3 applies equally to the adaptive estimator (4.9).

Since the estimated $\hat{\mathbf{I}}_n$ appears in the CLT, we have obtained a feasible limit theorem and (asymptotic) inference statements are immediate.

Some assumptions of Theorem 4.4 are tighter than for the oracle estimator. To some extent this is for the sake of clarity. Here, we have restricted Assumption 3.2(*ii*-α) to the Besov-regular part. A generalisation of the pilot estimator to martingales seems feasible, but is non-standard and might require additional conditions. We have also proposed a concrete order of h and r, less restrictive bounds are used in the proof, see e.g. (C.3) below.

The lower bound for α in terms of the sample-size ratio n_{max}/n_{min} is due to rough norm bounds for (estimated) information-type matrices. For $\alpha = 1$ (bounded variation case) the restriction imposes n_{max} to be slightly smaller than $n_{min}^{4/3}$. By the Sobolev embedding $B^1_{1,\infty} \subseteq H^\beta$ for all $\beta < 1/2$, the restriction $n_{max} = \mathcal{O}(n_{min}^{1+\beta})$ from Theorem 3.4 is clearly also satisfied in this case. It is not clear whether a more elaborate analysis can avoid these restrictions. Still, to the best of our knowledge, a feasible CLT for asymptotically separating sample sizes has not been obtained before.

5. Semiparametric efficiency.

5.1. *Semiparametric Cramér-Rao bound.* We shall derive an efficiency bound for the following basic case of observation model ($\mathcal{E}_1$):

$$dY_t = X_t\,dt + \frac{1}{\sqrt{n}}\,dW_t, \quad X_t = \int_0^t \Sigma(s)^{1/2} dB_s, \quad t \in [0,1], \tag{5.1}$$

where

$$\Sigma(t) = \Sigma_0(t) + \varepsilon \mathbb{H}(t), \quad \Sigma_0(t)^{1/2} = O(t)^\top \Lambda(t) O(t). \tag{5.2}$$

We assume $\Sigma_0(t)$ and $\mathbb{H}(t)$ to be known symmetric matrices, $O(t)$ orthogonal matrices, $\Lambda(t) = \operatorname{diag}(\lambda_1(t), \ldots, \lambda_d(t))$ diagonal and consider $\varepsilon \in [-1,1]$ as unknown parameter. Furthermore, we require Assumption 3.2(iii-$\underline{\Sigma}$) for all Σ. Finally, we impose throughout this section the regularity assumption that the matrix functions $O(t), \mathbb{H}(t), \Lambda(t)$ are continuously differentiable.

The key idea is to transform the observation of dY_t in such a manner that the white noise part remains invariant in law while for the central parameter $\Sigma(t) = \Sigma_0(t)$ the process X is transformed to a process with independent coordinates and constant volatility. It turns out that this can only be achieved at the cost of an additional drift in the signal. The construction first rotates the observations via $O(t)$, which diagonalises $\Sigma_0(t)$, and then applies a coordinate-wise time-transformation, corrected by a multiplication term to ensure L^2-isometry such that the white noise remains law-invariant.

We introduce the coordinate-wise time changes by

$$r_i(t) = \frac{\int_0^t \lambda_i(s)ds}{\int_0^1 \lambda_i(s)ds} \text{ and } (T_r g)(t) := (g_1(r_1(t)), \ldots, g_d(r_d(t)))^\top$$

for $g = (g_1, \ldots, g_d) : \mathbb{R} \to \mathbb{R}^d$. Moreover, we set

$$\bar{\Lambda} := \int_0^1 \Lambda(s)ds, \quad R'(t) := \bar{\Lambda}^{-1}\Lambda(t) = \operatorname{diag}(r_1'(t), \ldots, r_d'(t)).$$

LEMMA 5.1. *By transforming $d\bar{Y} = T_r^{-1}\mathcal{M}_{(R')^{-1/2}O}dY$, the observation model* (5.1), (5.2) *is equivalent to observing*

$$d\bar{Y}(t) = S(t)\,dt + \frac{1}{\sqrt{n}}d\bar{W}(t) \qquad \textit{with} \tag{5.3}$$

$$S(t) = T_r^{-1}\Big((R')^{-1}\Big(\int_0^{\cdot}((R')^{-\frac{1}{2}}O)'(s)X(s)\,ds + \int_0^{\cdot}(R'(s))^{-\frac{1}{2}}O(s)\,dX(s)\Big)\Big)(t)$$

for $t \in [0,1]$. At $\varepsilon = 0$ the observation $d\bar{Y}(t)$ reduces to

$$\Big(\int_0^t T_r^{-1}((R')^{-1}((R')^{-1/2}O)'X)(s)\,ds + \bar{\Lambda}\bar{B}(t)\Big)\,dt + \frac{1}{\sqrt{n}}d\bar{W}(t). \tag{5.4}$$

Here $\bar{W}$ and $\bar{B}$ are Brownian motions obtained from W and B, respectively, via rotation and time shift, as defined in (D.1) *below.*

If we may forget in (5.4) the first term, which is a drift term with respect to the martingale part $\bar{\Lambda}\bar{B}(t)$, then the central observation is indeed a constant volatility model in white noise. The lemma is proved in Appendix D.1.

Let us introduce the multiplication operator $\mathcal{M}_A g := Ag$ and the integration operator

$$Ig(t) = -\int_t^1 g(s)\,ds \text{ and its adjoint } I^*g(t) = -\int_0^t g(s)\,ds.$$

The covariance operator $C_{n,\varepsilon}$ on $L^2([0,1],\mathbb{R}^d)$ obtained from observing the differential in (5.3) is then given by

$$C_{n,\varepsilon} = T_r^*\mathcal{M}_{(R')^{1/2}O}I^*\mathcal{M}_{\Sigma_0+\varepsilon\mathbb{H}}I\mathcal{M}_{O^\top(R')^{1/2}}T_r + n^{-1}\,\mathrm{Id}\,.$$

The covariance operator $Q_{n,\varepsilon}$ when omitting the drift part is given by

$$Q_{n,\varepsilon} = Q_{n,0} + \varepsilon I^*T_r^*\mathcal{M}_M T_r I \text{ with } M(t) := ((R')^{-1/2}O\mathbb{H}O^\top(R')^{-1/2})(t),$$

where for $\varepsilon = 0$ the one-dimensional Brownian motion covariance operator C_{BM} appears in

$$Q_{n,0} = \mathrm{diag}(\bar{\lambda}_{ii}C_{BM} + n^{-1}\,\mathrm{Id})_{1\le i\le d}, \quad C_{BM} = I^*I.$$

We set $\dot{C}_0 = (C_{n,\varepsilon} - C_{n,0})/\varepsilon$ and $\dot{Q}_0 = (Q_{n,\varepsilon} - Q_{n,0})/\varepsilon$.

Standard Fisher information calculations for the finite-dimensional Gaussian scale model, e.g. [20, Chapter 6.6], transfer one-to-one to the infinite-dimensional case of observing $\mathbf{N}(0, Q_{n,\varepsilon})$ and yield as Fisher information for the parameter ε at $\varepsilon = 0$ the value

$$I_n^Q = \frac{1}{2}\|Q_{n,0}^{-1/2}\dot{Q}_0 Q_{n,0}^{-1/2}\|_{HS}^2\ ,$$

because $Q_{n,0}^{-1/2}Q_{n,\varepsilon}Q_{n,0}^{-1/2}$ is differentiable at $\varepsilon = 0$ in Hilbert-Schmidt norm. In Appendix D.2, we show by Hilbert-Schmidt calculus, the Feldman-Hajek Theorem and the Girsanov Theorem that the models with and without drift do not separate:

LEMMA 5.2. *We have*

$$\limsup_{n\to\infty} \|Q_{n,0}^{-1/2}\dot{Q}_0 Q_{n,0}^{-1/2} - C_{n,0}^{-1/2}\dot{C}_0 C_{n,0}^{-1/2}\|_{HS} < \infty.$$

Lemma 5.2 implies that the drift only contributes the negligible order $\mathcal{O}(1) = \mathcal{o}(\sqrt{n})$ to the Fisher information. By identifying the hardest parametric subproblem for observations $\mathbf{N}(0, Q_{n,\varepsilon})$ we thus establish in Appendix D.3 a semiparametric Cramér-Rao bound for estimating any linear functional of the covolatility matrix. Further classical asymptotic statements like the local asymptotic minimax theorem would require the LAN-property of the parametric subproblem.

THEOREM 5.3. *For a continuous matrix-valued function $A : [0,1] \to \mathbb{R}^{d\times d}$ consider the estimation of*

$$\vartheta := \int_0^1 \langle A(t), \Sigma(t)\rangle_{HS}\, dt = \int_0^1 \sum_{i,j=1}^d A_{ij}(t)\Sigma_{ij}(t)\, dt \in \mathbb{R}. \tag{5.5}$$

Then a hardest parametric subproblem in model (5.1), (5.2) *is obtained for the perturbation of* Σ_0 *by*

$$\mathbb{H}^*(t) = (\Sigma_0(A + A^\top)\Sigma_0^{1/2} + \Sigma_0^{1/2}(A + A^\top)\Sigma_0)(t).$$

There any estimator $\hat{\vartheta}_n$ of ϑ, which is asymptotically unbiased in the sense $\frac{d}{d\vartheta}(\mathbb{E}_\vartheta[\hat{\vartheta}_n] - \vartheta) \to 0$, satisfies as $n \to \infty$

$$\mathbb{V}\mathrm{ar}_{\varepsilon=0}(\hat{\vartheta}_n) \geq \frac{(2 + \mathcal{o}(1))}{\sqrt{n}} \int_0^1 \langle (\Sigma_0 \otimes \Sigma_0^{1/2} + \Sigma_0^{1/2} \otimes \Sigma_0)\mathcal{Z}vec(A), \mathcal{Z}vec(A)\rangle(t)dt.$$

6. Implementation and numerical results.

6.1. *Discrete-time estimator.* The construction to transfer discrete-time to continuous-time observations in the proof of Theorem 3.4 paves the way to the discrete approximation of the local spectral statistics (2.3). Using the interpolated process and integration by parts yields

$$\int \varphi_{jk}(t)dY^{(l)}(t) \asymp -\sum_{\nu=1}^{n_l} \int_{t_{\nu-1}^{(l)}}^{t_\nu^{(l)}} \Phi_{jk}(t) \frac{Y_\nu^{(l)} - Y_{\nu-1}^{(l)}}{t_\nu^{(l)} - t_{\nu-1}^{(l)}}\, dt.$$

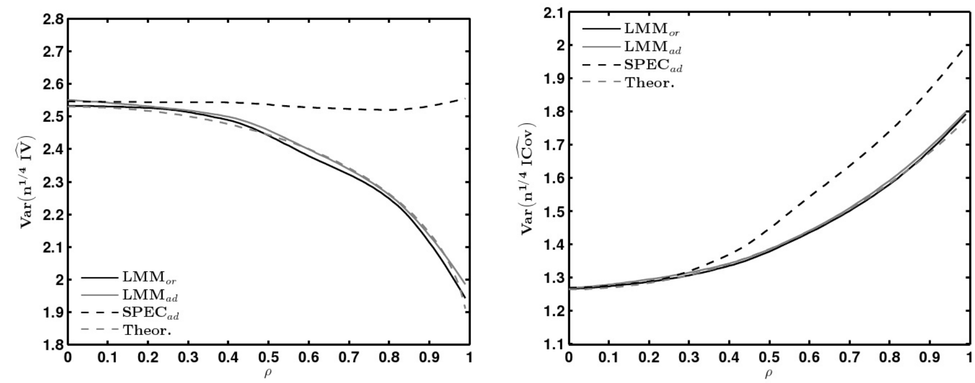

FIG 2. *Variances of estimators of σ_1^2 (left) and σ_{12} (right) in time-constant scenario ($n = 30,000$).*

Hence, for discrete-time observations from ($\mathcal{E}_0$) we use the local spectral statistics S_{jk} in (2.1). The noise terms in (4.2) translate from $\mathcal{E}_1$ to $\mathcal{E}_0$ via substituting $n_l^{-1}\int_{kh}^{(k+1)h}(F_l'(s))^{-1}\,ds$ by $\sum_{\nu:\,kh\le t_\nu^{(l)}\le(k+1)h}(t_\nu^{(l)}-t_{\nu-1}^{(l)})^2$. The discrete sum times h^{-1} can be understood as a block-wise quadratic variation of time in the spirit of Zhang *et al.* [25]. The bias is discretised analogously.

For the adaptive estimator we are in need of local estimates of n_lF_l', Σ and estimators for $\eta_l^2, 1\le l\le d$. It is well known how to estimate noise variances with faster $\sqrt{n_l}$-rates, see e.g. Zhang *et al.* [25]. Local observation densities can be estimated with block-wise quadratic variation of time as above, which then yield estimates $\hat H_{n,l}^{kh}$ of $H_{n,l}$ around time kh. Uniformly consistent estimators for $\Sigma(t), t\in[0,1]$, are feasible, e.g., averaging spectral statistics for $j=1,\dots,J$ over a set $\mathcal{K}_t$ of K adjacent blocks containing t:

$$\widehat{\Sigma}(t) = K^{-1}\sum_{k\in\mathcal{K}_t} J^{-1}\sum_{j=1}^{J}\left(S_{jk}S_{jk}^{\top} - \pi^2 j^2 h^{-2}\,\mathrm{diag}((\hat H_{n,l}^{kh})_l^2)\right). \tag{6.1}$$

We refer to Bibinger and Reiß [6] for details on the non-parametric pilot estimator with $J=1$.

6.2. *Simulations.* We examine the finite-sample properties of the LMM for the case $d=2$ in two scenarios. First, we compare the finite-sample variance with the asymptotic variances from Sections 3 and 4, for a parametric setup with $\eta_1^2=\eta_2^2=0.1$, $\sigma_1=\sigma_2=1$ and constant correlation ρ. We simulate $n_1=n_2=30,000$ synchronous observations on $[0,1]$. For estimating σ_1^2 and $\sigma_{12}=\rho$, Figure 2 displays the rescaled Monte-Carlo variance based on $20,000$ replications of the oracle and adaptive LMM (LMM$_{or}$ and LMM$_{ad}$), as well as the adaptive spectral estimator (SPEC_{ad}) by Bibinger and Reiß [6]. The latter relies on the same spectral approach, but uses only scalar weighting instead of the full information matrix approach.

In practice the pilot estimator from (6.1) for J not too large performed well. As configuration we use $h^{-1} = 10$, $J = 30$ and $K = 8$, which turned out to be an accurate choice, but the estimators are reasonably robust to alternative input choices. For the LMM of σ_1^2, we observe the variance reduction effect associated with a growing signal correlation ρ, while the simulation-based variances of both LMM_{or} and LMM_{ad} are close to their theoretical asymptotic counterpart (Theor). The results for σ_{12} underline the precision gains compared to SPEC_{ad} with univariate weights when ρ increases.

Next, we consider a complex and realistic stochastic volatility setting that relies on an extension of the widely-used Heston model, as e. g. employed by Aït-Sahalia *et al.* [1], accounting for both leverage effects and an intraday seasonality of volatility. The signal process for $l = 1, 2$ evolves as

$$dX_t^{(l)} = \varphi_l(t)\ \sigma_l(t)\ dZ_t^{(l)}, \qquad d\sigma_l^2(t) = \alpha_l\left(\mu_l - \sigma_l^2(t)\right)dt + \psi_l\,\sigma_l(t)\ dV_t^{(l)},$$

where $Z_t^{(l)}$ and $V_t^{(l)}$ are standard Brownian motions with $dZ_t^{(1)}dZ_t^{(2)} = \rho\,dt$ and $dZ_t^{(l)}dV_t^{(m)} = \delta_{l,m}\,\gamma_l\,dt$. $\varphi_l(t)$ is a non-stochastic seasonal factor with $\int_0^1 \varphi_l^2(t)\ dt = 1$. The unit time interval can represent one trading day, e.g. 6.5 hours or 23,400 seconds at NYSE.

We initialise the variance process $\sigma_l^2(t)$ by sampling from its stationary distribution $\Gamma\big(2\,\alpha_l\,\mu_l/\psi_l^2, \psi_l^2/(2\alpha_l)\big)$ and vary the value of the instantaneous signal correlation ρ, while setting $(\mu_l, \alpha_l, \psi_l, \gamma_l) = (1, 6, 0.3, -0.3)$, $l = 1, 2$, which under the stationary distribution, implies $\mathbb{E}\big[\int_0^1 \varphi_l^2(t)\,\sigma_l^2(t)\,dt\big] = 1$. The seasonal factor $\varphi_l(t)$ is specified in terms of intraday volatility functions estimated for S&P 500 equity data by the procedure in Andersen and Bollerslev [2]. $\varphi_1(t)$ and $\varphi_2(t)$ are based on cross-sectional averages of the 50 most and 50 least liquid stocks, respectively, which yields a pronounced L-shape in both cases (see Figure 3). We add noise processes that are i.i.d. $\mathbf{N}\big(0, \eta_l^2\big)$ and mutually independent with $\eta_l = 0.1(\mathbb{E}\big[\int_0^1 \varphi_l^4(t)\,\sigma_l^4(t)\,dt\big])^{1/4}$, computed under the stationary distribution of $\sigma_l^2(t)$. Finally, asynchronicity effects are introduced by drawing observation times $t_i^{(l)}$, $1 \le i \le n_l$, $l = 1, 2$, from two independent Poisson processes with intensities $\lambda_1 = 1$ and $\lambda_2 = 2/3$ such that, on average, $n_1 = 23,400$ and $n_2 = 15,600$.

As a representative example, Figure 3 depicts the root mean-squared errors (RMSEs) based on 40,000 replications of the following estimators of $\int_0^1 \varphi_1^2(t)\,\sigma_1^2(t)\,dt$: the oracle and adaptive LMM using $h^{-1} = 20$, $J = 15$ and $K = 8$, the quasi-maximum likelihood (QML) estimator by Aït-Sahalia *et.*

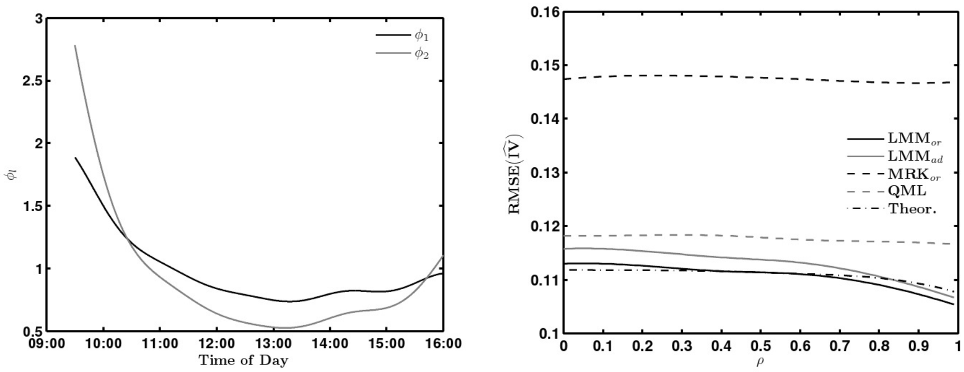

FIG 3. *Non-stochastic volatility seasonality factors (left) and RMSE for estimators of* $\int_0^1 \varphi_1^2(t)\,\sigma_1^2(t)\,dt$ *(right) in stochastic volatility scenario.*

al. [1] as well as an oracle version of the widely-used multivariate realised kernel (MRK_{or}) by Barndorff-Nielsen *et al.* [4]. For the latter, we employ the average univariate mean-squared error optimal bandwidth based on the true value of $\int_0^1 \varphi_l^4(t)\,\sigma_l^4(t)\,dt,\ l = 1, 2$. Finally, we include the theoretical variance from the asymptotic theory (Theor), which is computed as the variance (4.7) averaged across all replications.

Three major results emerge. First, the LMM offers considerable precision gains when compared to both benchmarks. Second, a rising instantaneous signal correlation ρ is associated with a declining RMSE of the LMM, which is due to the decreasing variance, and thus confirms the findings from Section 3 in a realistic setting. Finally, the adaptive LMM closely tracks its oracle counterpart.

In summary, the simulation results show that the estimator has promising properties even in settings which are more general than those assumed in ($\mathcal{E}_1$), allowing, for instance, for random observation times, stochastic intraday volatility as well as leverage effects. Even if the latter effects are not yet covered by our theory, the proposed estimator seems to be quite robust to deviations from the idealised setting.

APPENDIX A: FROM DISCRETE TO CONTINUOUS EXPERIMENTS

PROOF OF THEOREM 3.4. To establish Le Cam equivalence, we give a constructive proof to transfer observations in $\mathcal{E}_0$ to the continuous-time model $\mathcal{E}_1$ and the other way round. We bound the Le Cam distance by estimates for the squared Hellinger distance between Gaussian measures and refer to Section **A.1** in [22] for information on Hellinger distances between Gaussian measures and bounds with the Hilbert-Schmidt norm. The crucial difference here is that linear interpolation is carried out for non-synchronous irregular observation schemes. Consider the linear B-splines or hat functions

$$b_{i,n}(t) = \mathbb{1}_{[\frac{i-1}{n},\frac{i+1}{n}]}(t)\min\left(1+n\left(t-\frac{i}{n}\right)\,,\,1-n\left(t-\frac{i}{n}\right)\right).$$

Define $b_i^l(t) := b_{i,n_l}(F_l(t)), 1 \le i \le n_l, 1 \le l \le d$, which are warped spline functions satisfying $b_{i_1}^l(t_{i_2}^{(l)}) = \delta_{i_1,i_2}$. A centered Gaussian process $\hat{Y}$ is derived from linearly interpolating each component of Y:

$$\hat{Y}_t^{(l)} = \sum_{i=1}^{n_l} Y_i^{(l)} b_i^l(t) = \sum_{i=1}^{n_l} X_{t_i^{(l)}}^{(l)} b_i^l(t) + \sum_{i=1}^{n_l} \varepsilon_i^{(l)} b_i^l(t)\,. \tag{A.1}$$

The covariance matrix function $\mathbb{E}\left[\hat{Y}_t\hat{Y}_s^\top\right]$ of the interpolated process $\hat{Y}$ is determined by

$$\mathbb{E}\left[\hat{Y}_t^{(l)}\hat{Y}_s^{(r)}\right] = \sum_{i=1}^{n_l}\sum_{\nu=1}^{n_r} a_{lr}(t_i^{(l)}\wedge t_\nu^{(r)}) b_i^l(t) b_\nu^r(s) + \delta_{l,r}\eta_l^2\sum_{i=1}^{n_l} b_i^l(t) b_i^l(s)$$
$$\text{with } A(t) = (a_{lr}(t))_{l,r=1,\dots,d} = \int_0^t \Sigma(s)\,ds\,.$$

For any $g = (g^{(1)},\dots,g^{(d)})^\top \in L^2([0,1],\mathbb{R}^d)$ we have

$$\mathbb{E}\left[\langle \mathbf{g},\hat{Y}\rangle^2\right] = \mathbb{E}\Big[\Big(\sum_{\nu=1}^d \langle g^{(v)},\hat{Y}^{(v)}\rangle\Big)^2\Big]$$
$$= \sum_{l,r=1}^d\sum_{i=1}^{n_l}\sum_{\nu=1}^{n_r} a_{lr}(t_i^{(l)}\wedge t_\nu^{(r)})\langle g^{(l)},b_i^l\rangle\langle g^{(r)},b_\nu^r\rangle + \sum_{l=1}^d\sum_{i=1}^{n_l}\langle g^{(l)},b_i^l\rangle^2\eta_l^2.$$

The sum of the addends induced by the observation noise in diagonal terms is bounded from above by

$$\sum_{l=1}^d \frac{\eta_l^2}{n_l}\left\|g^{(l)}/\sqrt{F_l'}\right\|_{L^2}^2 = \sum_{l=1}^d \left\|g^{(l)}H_{n,l}\right\|_{L^2}^2\,,$$

since by virtue of $0 \le \sum_i b_{i,n} \le 1$, $\int b_{i,n} = 1/n$ and Jensen's inequality:

$$\sum_{i=1}^{n_l}\langle g^{(l)},b_i^l\rangle^2 \le \frac{1}{n_l}\sum_{i=1}^{n_l}\int_0^1((g^{(l)}\circ F_l^{-1})\cdot(F_l^{-1})')^2 b_{i,n_l}$$
$$\le \frac{1}{n_l}\int_0^1\left((g^{(l)}\circ F_l^{-1})\cdot(F_l^{-1})'\right)^2 = \frac{1}{n_l}\int_0^1\frac{(g^{(l)})^2}{F_l'}\,.$$

On the other hand, we have $\mathbb{E}[\langle g, \operatorname{diag}(H_{n,l})_l\, dW\rangle] = \sum_{l=1}^d \|g^{(l)}H_{n,l}\|_{L^2}^2$ for a d-dimensional standard Brownian motion W. Consequently, a process $\bar{Y}$ with continuous-time white noise and the same signal part as $\hat{Y}$ can be obtained by adding uninformative noise. Introduce the process

$$d\bar{Y} = \Big(\sum_{i=1}^{n_l} X_{t_i^{(l)}} b_i^l(t)\Big)_{1\le l\le d} dt + \operatorname{diag}(H_{n,l}(t))_{1\le l\le d}\, dW_t\,, \quad \text{(A.2)}$$

and its associated covariance operator $\bar{C}: L^2 \to L^2$, given by

$$\bar{C}g(t) = \Big(\sum_{r=1}^{d}\sum_{i=1}^{n_l}\sum_{\nu=1}^{n_r} a_{lr}(t_i^{(l)} \wedge t_\nu^{(r)})\langle g^{(r)}, b_\nu^r\rangle\Big)_{1\le l\le d} + \Big(H_{n,l}(t)^2 g^{(l)}(t)\Big)_{1\le l\le d}\,.$$

In fact, it is possible to transfer observations from our original experiment $\mathcal{E}_0$ to observations of (A.2) by adding $\mathbf{N}(0,\bar{C}-\hat{C})$-noise, where $\hat{C}: L^2 \to L^2$ is the covariance operator of $\hat{Y}$. Now, consider the covariance operator

$$Cg(t) = \int_0^1 \left(\int_0^{t\wedge u} A(s)\, ds\right) g(u)\, du + \left(\frac{\eta_l^2}{n_l F_l'(t)} g^{(l)}(t)\right)_{1\le l\le d},$$

associated with the continuous-time experiment $\mathcal{E}_1$.
We can bound $C^{-1/2}$ on $L^2([0,1],\mathbb{R}^d)$ from below (by partial ordering of operators) by a simple matrix multiplication operator:

$$C^{-1/2} \le \mathcal{M}_{\operatorname{diag}(H_{n,l}(t))_l}.$$

Denote the Hilbert-Schmidt norm by $\|\cdot\|_{\mathrm{HS}}$. The asymptotic equivalence of observing $\bar{Y}$ and Y in $\mathcal{E}_1$ is ensured by the Hellinger distance bound

$$\begin{aligned}
&\mathrm{H}^2\left(\mathcal{L}\left(\bar{Y}\right), \mathcal{L}\left(Y\right)\right) \le 2\, \|C^{-1/2}\left(\bar{C}-C\right)C^{-1/2}\|_{\mathrm{HS}}^2 \\
&\le 2\int_0^1\int_0^1 \Bigg(\sum_{l=1}^d\sum_{r=1}^d H_{n,l}(t)^{-2}H_{n,r}(t)^{-2} \\
&\qquad \Big(\sum_{i=1}^{n_l}\sum_{\nu=1}^{n_r} a_{lr}(t_i^{(l)}\wedge t_\nu^{(r)}) b_i^l(t) b_\nu^r(s) - a_{lr}(t\wedge s)\Big)^2\Bigg)\, dt\, ds \\
&= 2\int_0^1\int_0^1 \Bigg(\sum_{l=1}^d\sum_{r=1}^d \frac{n_l n_r}{\eta_l^2\eta_r^2} \\
&\qquad \Big(\sum_{i=1}^{n_l}\sum_{\nu=1}^{n_r} a_{lr}(t_i^{(l)}\wedge t_\nu^{(r)}) b_{i,n_l}(u) b_{\nu,n_r}(z) - a_{lr}(F_l^{-1}(u)\wedge F_r^{-1}(z))\Big)^2\Bigg)\, du\, dz \\
&= \mathcal{O}\Big(R^4\sum_{l=1}^d\sum_{r=1}^d \eta_l^{-2}\eta_r^{-2} n_l n_r n_{min}^{-2-2\beta}\Big).
\end{aligned}$$

The estimate for the L^2-distance between the function $(t,s)\mapsto A(F_l^{-1}(t)\wedge F_r^{-1}(s)), (l,r)\in\{1,\dots,d\}^2$, and its coordinate-wise linear interpolation by $\mathcal{O}(n_{min}^{-1-\beta}\vee n_{min}^{-3/2})$ relies on a standard approximation result on a rectangular grid of maximal width $(n_{min})^{-1}$ based on the fact that this function lies in the Sobolev class $H^{1+\beta}([0,1]^2)$ with corresponding norm bounded by $2R^4$. This follows immediately by the product rule from $A'=\Sigma\in H^\beta$ and $(F_l^{-1})'\in C^\beta$, together with an L^2-error bound at the skewed diagonal $\{(t,s):F_l(t)=F_r(s)\}$.

Next, we explicitly show that $\mathcal{E}_1$ is at least as informative as $\mathcal{E}_0$. To this end we discretise in each component on the intervals $I_{i,l}=[\frac{i}{n_l}-\frac{1}{2n_l},\frac{i}{n_l}+\frac{1}{2n_l}]\cap[0,1]$ for $i=0,\dots,n_l$. Define

$$\left(Y_i'\right)^{(l)}=\frac{1}{|I_{i,l}|}\int_{F_l^{-1}(I_{i,l})}F_l'(t)\,dY_t^{(l)}=\frac{1}{|I_{i,l}|}\int_{F_l^{-1}(I_{i,l})}X_t^{(l)}F_l'(t)\,dt+\varepsilon_i^{(l)}$$
$$=\frac{1}{|I_{i,l}|}\int_{I_{i,l}}X^{(l)}_{F^{-1}(u)}\,du+\varepsilon_i^{(l)}\,, \quad \text{(A.3)}$$

for $0\le i\le n_l$ with i. i. d. random variables:

$$\varepsilon_i^{(l)}=\frac{1}{|I_{i,l}|}\int_{F_l^{-1}(I_{i,l})}\eta_l(F_l'/n_l)^{1/2}\,dW_t^{(l)}\overset{i.\,i.\,d.}{\sim}\mathbf{N}\left(0,\eta_l^2\right). \quad \text{(A.4)}$$

The covariances are calculated as

$$\mathbb{E}\left[(Y_i')^{(l)}(Y_\nu')^{(r)}\right]=\frac{1}{|I_{i,l}||I_{\nu,r}|}\int_{I_{i,l}}\int_{I_{\nu,r}}a_{lr}(F_l^{-1}(u)\wedge F_r^{-1}(u'))\,du du'+\delta_{l,r}\delta_{i,\nu}\eta_l^2.$$

We obtain for the squared Hellinger distance between the laws of observation

$$\begin{aligned}&\mathrm{H}^2\left(\mathcal{L}\left((Y_i^{(l)})_{l=1,\dots,d;i=0,\dots,n_l}\right),\,\mathcal{L}\left(((Y_i')^{(l)})_{l=1,\dots,d;i=0,\dots,n_l}\right)\right)\\&\le\sum_{l,r=1}^d\eta_l^{-2}\eta_r^{-2}\sum_{i=0}^{n_l}\sum_{\nu=0}^{n_r}\Big(\frac{1}{|I_{i,l}||I_{\nu,r}|}\int_{I_{i,l}}\int_{I_{\nu,r}}a_{lr}(F_l^{-1}(u)\wedge F_r^{-1}(u'))\\&\qquad\qquad-a_{lr}(F_l^{-1}(i/n_l\wedge\nu/n_r))\,du du'\Big)^2.\end{aligned}$$

Write $A_{lr}^F(u,u')=a_{lr}(F_l^{-1}(u)\wedge F_r^{-1}(u'))$ and note $A_{lr}^F\in H^{1+\beta}([0,1]^2)$ due to $A'=\Sigma\in H^\beta$ and $F_l^{-1},F_r^{-1}\in C^\beta$. For $(i,\nu)\notin\mathcal{C}:=\{(0,0),(0,n_r),(n_l,0),(n_l,n_r)\}$ the rectangle $I_{i,l}\times I_{\nu,r}$ is symmetric around $(i/n_l,\nu/n_r)$ such that the integral in the preceding display equals (∇ denotes the gradient)

$$\int_{I_{i,l}\times I_{\nu,r}}\int_0^1\Big(\Big\langle\nabla A_{lr}^F\left(\tfrac{i}{n_l}+\vartheta\left(u-\tfrac{i}{n_l}\right),\tfrac{\nu}{n_r}+\vartheta\left(u'-\tfrac{\nu}{n_r}\right)\right),\left(u-\tfrac{i}{n_l},u'-\tfrac{\nu}{n_r}\right)\Big\rangle$$

$$- \left\langle \nabla A_{lr}^F\left(\tfrac{i}{n_l},\tfrac{\nu}{n_r}\right),\left(u-\tfrac{i}{n_l},u'-\tfrac{\nu}{n_r}\right)\right\rangle\Big)\, d\vartheta du du'.$$

Using Jensen's inequality we thus obtain further the bound for the squared Hellinger distance:

$$\sum_{l,r=1}^d \eta_l^{-2}\eta_r^{-2}\sum_{i=0}^{n_l}\sum_{\nu=0}^{n_r}\frac{(n_l\vee n_r)^{-2}}{|I_{i,l}||I_{\nu,r}|}\int_{I_{i,l}\times I_{\nu,r}}\int_0^1 \|\nabla A_{lr}^F(i/n_l+\vartheta(u-i/n_l),$$
$$\nu/n_r+\vartheta(u'-\nu/n_r))-\nabla A_{lr}^F(i/n_l,\nu/n_r)\mathbb{1}((i,\nu)\notin\mathcal{C})\|^2\,d\vartheta\,dudu'$$
$$=\sum_{l,r=1}^d \eta_l^{-2}\eta_r^{-2}\frac{n_l n_r}{(n_l\vee n_r)^2}\mathcal{O}\left(R^4(n_l\wedge n_r)^{-2\beta}\right)=\mathcal{O}\left(R^4\left(\sum_{l=1}^d n_l/\eta_l^2\right)^2 n_{min}^{-2-2\beta}\right)$$

where the order estimate is due to $\|\nabla A_{lr}^F\|_{H^\beta}\le R^2$ and a standard L^2-approximation result for Sobolev spaces, observing that for the four corner rectangles in $\mathcal{C}$ the boundedness of the respective integrals only adds the total order $4n_{min}^{-2}<n_l n_r n_{min}^{-2-2\beta}$. □

APPENDIX B: ASYMPTOTICS IN THE BLOCK-WISE CONSTANT EXPERIMENT

PROOF OF THEOREM 4.2. As we have seen, the estimator is unbiased in $\mathcal{E}_2$. For the covariance structure we use the independence between blocks and frequencies and the commutativity with $\mathcal{Z}$ to infer

$$\mathbb{C}\mathrm{OV}_{\mathcal{E}_2}\big(\mathbf{I}_n^{1/2}\,\mathrm{LMM}_{or}^{(n)}\big)=\mathbf{I}_n^{1/2}\sum_{k=0}^{h^{-1}-1}h^2\sum_{j=1}^{\infty}W_{jk}\mathbb{C}\mathrm{OV}_{\mathcal{E}_2}\left(vec\left(S_{jk}S_{jk}^\top\right)\right)W_{jk}^\top\mathbf{I}_n^{1/2}$$
$$=\mathbf{I}_n^{1/2}\sum_{k=0}^{h^{-1}-1}h^2 I_k^{-1}\mathbf{I}_n^{1/2}\mathcal{Z}=\mathcal{Z}\,. \tag{B.1}$$

Since the local Fisher-type informations are strictly positive definite and thus invertible by Assumption 3.2(iii), the multivariate CLT (4.5) for the oracle estimator follows by applying a standard CLT for triangular schemes as Theorem 4. 12 from [16]. The Lindeberg condition is implied by the stronger Lyapunov condition which is easily verified here by bounding moments of order 4.

In Appendix C below we prove that in experiment $\mathcal{E}_1$ the estimator $\mathrm{LMM}_{or}^{(n)}$ has an additional bias of order $\mathcal{O}(n_{min}^{-\alpha/2})+\mathcal{O}_P(h)$ and a difference in the covariance of order $\mathcal{O}(hn_{min}^{-\alpha/2})+\mathcal{O}_P(h^2)$ under our Assumption 3.2(ii-α),(iii-$\underline{\Sigma}$), which by Slutsky's lemma yields an asymptotically negligible term compared to the best attainable rate (in any entry) $n_{max}^{-1/4}$, cf. Theorem 5.3. □

Proof of Corollary 4.3. An important property of our oracle estimator is its equivariance with respect to invertible linear transformations A_k on each block k in the sense that for observed statistics $\tilde{S}_{jk} := A_k S_{jk} \sim \mathbf{N}(0, \tilde{C}_{jk})$ under $\mathcal{E}_2$ we obtain ($A^{-\top} := (A^\top)^{-1}$ for short)

$$C_{jk} = A_k^{-1}\tilde{C}_{jk}A_k^{-\top},\ I_{jk} = (A_k \otimes A_k)^\top \tilde{I}_{jk}(A_k \otimes A_k),\ I_k = (A_k \otimes A_k)^\top \tilde{I}_k(A_k \otimes A_k)$$

and hence with some (deterministic) bias correction terms $B_{jk}, \tilde{B}_{jk}$

$$\begin{aligned}\mathrm{LMM}_{or}^{(n)} &= \sum_{k=0}^{h^{-1}-1} h(A_k \otimes A_k)^{-1}\tilde{I}_k^{-1}\sum_{j\geq 0}\tilde{I}_{jk}(A_k \otimes A_k)vec(S_{jk}S_{jk}^\top - B_{jk}) \\ &= \sum_{k=0}^{h^{-1}-1}(A_k \otimes A_k)^{-1}\Big(h\tilde{I}_k^{-1}\sum_{j\geq 0}\tilde{I}_{jk}vec(\tilde{S}_{jk}\tilde{S}_{jk}^\top - \tilde{B}_{jk})\Big).\end{aligned}$$

For the covariance we use commutativity with $\mathcal{Z}$ and obtain likewise

$$\mathbb{C}\mathrm{OV}_{\mathcal{E}_2}(\mathrm{LMM}_{or}^{(n)}) = \sum_{k=0}^{h^{-1}-1} h^2(A_k \otimes A_k)^{-1}\tilde{I}_k^{-1}(A_k \otimes A_k)^{-\top}\mathcal{Z}. \tag{B.2}$$

We use this property to diagonalise the problem on each block. In terms of the noise level matrix $\mathcal{H}_k := \mathrm{diag}(H_{l,n}^k)_{l=1,\dots,d}$, let O_k be an orthogonal matrix such that

$$\Lambda^{kh} = O_k\mathcal{H}_k^{-1}\Sigma^{kh}\mathcal{H}_k^{-1}O_k^\top \tag{B.3}$$

is diagonal. Note that Λ^{kh} grows with n, but we drop the dependence on n in the notation for all matrices Λ^{kh}, O_k and $\mathcal{H}_k$. Use $A_k = O_k\mathcal{H}_k^{-1}$ to obtain the spectral statistics (2.3) transformed:

$$\tilde{S}_{jk} = O_k\mathcal{H}_k^{-1}S_{jk} \sim \mathbf{N}\left(\mathbf{0}, \tilde{C}_{jk}\right) \text{ independent for all } (j,k)$$

which yields a simple-structured diagonal covariance matrix:

$$\tilde{C}_{jk} = O_k\mathcal{H}_k^{-1}C_{jk}\mathcal{H}_k^{-1}O_k^\top = \Lambda^{kh} + \tfrac{\pi^2 j^2}{h^2}E_d.$$

A key point is that the covariance structure (B.2) in $\mathbb{R}^{d^2\times d^2}$ is for independent components $\tilde{S}_{jk}$ also diagonal, up to symmetry in the covolatility matrix entries. Summing $\tilde{I}_{jk}$ over j is explicitly solvable and gives for

$p, q = 1, \dots, d$

$$
\begin{aligned}
(h\tilde{I}_k^{-1})_{p,q} &= \Big(h^{-1}\sum_{j=1}^{\infty}(\tilde{C}_{jk}^{-1}\otimes\tilde{C}_{jk}^{-1})_{p,q}\Big)^{-1}\\
&= \Big(h^{-1}\sum_{j=1}^{\infty}(\Lambda_{pp}^{kh}+\pi^2j^2h^{-2})^{-1}(\Lambda_{qq}^{kh}+\pi^2j^2h^{-2})^{-1}\Big)^{-1}\\
&= \Big(\frac{\sqrt{\Lambda_{qq}^{kh}}\coth(h\sqrt{\Lambda_{pp}^{kh}})-\sqrt{\Lambda_{pp}^{kh}}\coth(h\sqrt{\Lambda_{qq}^{kh}})}{2\sqrt{\Lambda_{pp}^{kh}\Lambda_{qq}^{kh}}(\Lambda_{qq}^{kh}-\Lambda_{pp}^{kh})}-\frac{1}{2h\Lambda_{pp}^{kh}\Lambda_{qq}^{kh}}\Big)^{-1}\\
&= 2(\Lambda_{pp}^{kh}\sqrt{\Lambda_{qq}^{kh}}+\Lambda_{qq}^{kh}\sqrt{\Lambda_{pp}^{kh}})\Big(1+\mathcal{O}\Big(e^{-2h\sqrt{\Lambda_{pp}^{kh}\wedge\Lambda_{qq}^{kh}}}+h^{-1}(\Lambda_{pp}^{kh}\wedge\Lambda_{qq}^{kh})^{-1/2}\Big)\Big),
\end{aligned}
$$

using $\Lambda^{kh}\geq(\min_{l,t}n_lF_l'(t)\eta_l^{-2})\underline{\Sigma}\gtrsim n_{min}E_d$, $h^2n_{min}\to\infty$ and $\coth(x)=1+\mathcal{O}(e^{-2x})$ for $x\to\infty$. We thus obtain uniformly over k

$$h\tilde{I}_k^{-1}=(2+\scriptstyle\mathcal{O}(1))(\Lambda^{kh}\otimes\sqrt{\Lambda^{kh}}+\sqrt{\Lambda^{kh}}\otimes\Lambda^{kh}).$$

By formula (B.2) we infer in terms of $(\Sigma_{\mathcal{H}}^{kh})^{1/2}:=\mathcal{H}_k(\mathcal{H}_k^{-1}\Sigma^{kh}\mathcal{H}_k^{-1})^{1/2}\mathcal{H}_k$

$$\mathbb{C}\mathrm{OV}_{\mathcal{E}_2}(\mathrm{LMM}_{or}^{(n)})=(2+o(1))\sum_{k=0}^{h^{-1}-1}h(\Sigma^{kh}\otimes(\Sigma_{\mathcal{H}}^{kh})^{1/2}+(\Sigma_{\mathcal{H}}^{kh})^{1/2}\otimes\Sigma^{kh})\mathcal{Z}.$$

The final step consists in combining $n_{min}^{1/2}H_{n,l}(t)\to H_l(t)$ uniformly in t together with a Riemann sum approximation to conclude

$$
\begin{aligned}
&\lim_{n_{min}\to\infty}n_{min}^{1/2}\mathbb{C}\mathrm{OV}_{\mathcal{E}_2}(\mathrm{LMM}_{or}^{(n)})\\
&\quad=2\Big(\int_0^1\Big(\Sigma\otimes(\mathcal{H}(\mathcal{H}^{-1}\Sigma\mathcal{H}^{-1})^{1/2}\mathcal{H})+(\mathcal{H}(\mathcal{H}^{-1}\Sigma\mathcal{H}^{-1})^{1/2}\mathcal{H})\otimes\Sigma\Big)(t)\,dt\Big)\mathcal{Z}.
\end{aligned}
$$

□

APPENDIX C: PROOFS FOR CONTINUOUS MODELS

C.1. Weight matrix estimates. We shall often need general norm bounds on the weight matrices W_{jk}.

LEMMA C.1. *The oracle weight matrices satisfy uniformly over* (j,k) *and the matrices* Σ^{kh} *with* $\|\Sigma^{kh}\|_\infty+\|(\Sigma^{kh})^{-1}\|_\infty\lesssim 1$

$$\|W_{jk}\|\lesssim h_0^{-1}(1+j^4/h_0^4)^{-1}.$$

PROOF. From the proof of Corollary 4.3 we infer

$$W_{jk} = (H_k O_k^\top \otimes H_k O_k^\top)\tilde{W}_{jk}(O_k H_k^{-1} \otimes (O_k H_k^{-1}) \text{ with}$$
$$\tilde{W}_{jk} = (2+\mathcal{O}(1))h^{-1}\Big((\Lambda^{kh}\tilde{C}_{jk}^{-1}) \otimes (\sqrt{\Lambda^{kh}}\tilde{C}_{jk}^{-1}) + (\sqrt{\Lambda^{kh}}\tilde{C}_{jk}^{-1}) \otimes (\Lambda^{kh}\tilde{C}_{jk}^{-1})\Big).$$

We evaluate one factor in W_{jk} using

$$\|H_k O_k^\top \Lambda^{kh}\tilde{C}_{jk}^{-1} O_k H_k^{-1}\| = \|\Sigma^{kh}(\Sigma^{kh}+\pi^2 j^2 h^{-2} H_k^2)^{-1}\| \lesssim (1+j^2 h^{-2} n_{min}^{-2})^{-1}.$$

By $\|A \otimes B\| \le \|A\|\|B\|$ and $\sqrt{\Lambda^{kh}}\tilde{C}_{jk}^{-1} = (\Lambda^{kh}\tilde{C}_{jk}^{-1})(\Lambda^{kh})^{-1/2}$ (the matrices are diagonal), we infer

$$\|W_{jk}\| \lesssim h^{-1}(1+j^2 h_0^{-2})^{-2}\|H_k O_k^\top (\Lambda^{kh})^{-1/2} O_k H_k^{-1}\|.$$

To evaluate the last norm, despite matrix multiplication is non-commutative, we note

$$\Big(O_k^\top (\Lambda^{kh})^{-\frac{1}{2}} O_k H_k^{-1}\Big)^\top O_k^\top (\Lambda^{kh})^{-\frac{1}{2}} O_k H_k^{-1} = H_k^{-1} O_k^\top (\Lambda^{kh})^{-1} O_k H_k^{-1} = (\Sigma^{kh})^{-1},$$

whence by polar decomposition $|O_k^\top (\Lambda^{kh})^{-1/2} O_k H_k^{-1}| = (\Sigma^{kh})^{-1/2}$ implies

$$\|O_k^\top (\Lambda^{kh})^{-\frac{1}{2}} O_k H_k^{-1}\| = \|(\Sigma^{kh})^{-\frac{1}{2}}\| \lesssim 1.$$

Together with $\|H_k\| \lesssim n_{min}^{-1/2}$ this yields

$$\|W_{jk}\| \lesssim h^{-1}(1+j^2 h_0^{-2})^{-2} n_{min}^{-1/2},$$

which gives the result. □

Moreover, for the adaptive estimator we have to control the dependence of the weight matrices $W_{jk} = W_j(\Sigma^{kh})$ on Σ^{kh}. We use the notion of matrix differentiation as introduced in [11]: define the derivative dA/dB of a matrix-valued function $A(B) \in \mathbb{R}^{o\times p}$ with respect to $B \in \mathbb{R}^{q\times r}$ as the $\mathbb{R}^{op\times qr}$ matrix with row vectors $(d/dB_{ab})vec(A), 1 \le a \le q, 1 \le b \le r$.

LEMMA C.2. *For the derivatives of the oracle weight matrices $W_j(\Sigma^{kh})$, assuming $\|\Sigma^{kh}\|_\infty + \|(\Sigma^{kh})^{-1}\|_\infty \lesssim 1$, we have uniformly over (j,k):*

$$\left\|\frac{d}{d\Sigma^{kh}} W_j(\Sigma^{kh})\right\| \lesssim h_0^{-1}(1+j^4 h_0^{-4})^{-1}. \tag{C.1}$$

PROOF. Since the notion of matrix derivatives relies on vectorisation, the identities $vec(I_k^{-1}I_{jk}) = (E_{d^2} \otimes I_k^{-1})vec(I_{jk}) = (I_{jk}^{\top} \otimes E_{d^2})vec(I_k^{-1})$ give rise to the matrix differentiation product rule

$$\frac{d}{d\Sigma^{kh}} W_{jk} = (I_{jk} \otimes E_{d^2})\,\frac{dI_k^{-1}}{d\Sigma^{kh}} + \left(E_{d^2} \otimes I_k^{-1}\right)\frac{dI_{jk}}{d\Sigma^{kh}}\,. \tag{C.2}$$

Applying the mixed product rule $(A\otimes B)(C\otimes D) = (AC\otimes BD)$ repeatedly, and the differentiation product rule and chain rule to $I_{jk} = C_{jk}^{-1} \otimes C_{jk}^{-1}$, we obtain

$$\frac{d}{dC_{jk}}\left(C_{jk}^{-1} \otimes C_{jk}^{-1}\right) = -\left(\left(C_{jk}^{-1} \otimes C_{jk}^{-1}\right) \otimes \left(C_{jk}^{-1} \otimes C_{jk}^{-1}\right)\right)\Big(\big((C_{jk} \otimes E_d \otimes E_{d^2}) \\ +(E_{d^2} \otimes E_d \otimes C_{jk})\big)(E_d \otimes C_{d,d} \otimes E_d)\big((vec(E_d) \otimes E_{d^2}) + (E_{d^2} \otimes vec(E_d))\big)\Big),$$

with the so-called commutation matrix $C_{d,d} = \mathcal{Z} - E_{d^2}$. By orthogonality of the last factors in both addends, $\|A \otimes B\| = \|A\|\|B\|$, and the mixed product rule, we infer for the norm of the second addend in (C.2)

$$\left\|(E_{d^2} \otimes I_k^{-1})\frac{dI_{jk}}{d\Sigma^{kh}}\right\| \le 2\,\left\|\left(E_d \otimes C_{jk}^{-1}\right) \otimes \left(I_k^{-1}(C_{jk}^{-1} \otimes C_{jk}^{-1})\right)\right\| \\ = 2\,\|W_{jk}\|\left\|C_{jk}^{-1}\right\| \lesssim \|W_{jk}\|\;.$$

By virtue of

$$(I_k^{-1} \otimes E_{d^2})\frac{dI_k}{d\Sigma^{kh}} = -(E_{d^2} \otimes I_k)\frac{dI_k^{-1}}{d\Sigma^{kh}}\,,$$

it follows with the mixed product rule that

$$dI_k^{-1}/d\Sigma^{kh} = -(I_k^{-1} \otimes I_k^{-1})(dI_k/d\Sigma^{kh}).$$

This yields for the norm of the first addend in (C.2)

$$\left\|(I_{jk} \otimes E_{d^2})\frac{dI_k^{-1}}{d\Sigma^{kh}}\right\| = \left\|(W_{jk}^{\top} \otimes I_k^{-1})\frac{dI_k}{d\Sigma^{kh}}\right\| \lesssim \|W_{jk}\|\left\|(E_{d^2} \otimes I_k^{-1})\sum_{j'}\frac{dI_{j'k}}{d\Sigma^{kh}}\right\| \\ \lesssim \|W_{jk}\|\Big(\sum_{j'}\|W_{j'k}\|\Big) \lesssim \|W_{jk}\|$$

since we can differentiate inside the sum by the absolute convergence of $\sum_{j'}\|W_{j'k}\|$. This proves our claim by Lemma C.1. □

C.2. Bias bound. Using the formula $1-2\sin^2(x)=\cos(2x)$ and Itô isometry, the $(d\times d)$-matrix of (negative) biases (in the signal) of the addends in (4.3) as an estimator of Σ^{kh} in experiment $\mathcal{E}_1$ is given by

$$B_{j,k}:=2h^{-1}\int_{kh}^{(k+1)h}\Sigma(t)\cos(2j\pi h^{-1}(t-kh))\,dt,$$

which has the structure of a j-th Fourier cosine coefficient. We introduce the corresponding weighting function in the time domain:

$$G_k(u)=2\sum_{j=1}^{\infty}W_{jk}\cos(2j\pi u)\in\mathbb{R}^{d^2\times d^2},\quad u\in[0,1].$$

Parseval's identity then shows for the d^2-dimensional block-wise bias vector of (4.3):

$$\sum_{j=1}^{\infty}W_{jk}vec(B_{j,k})=\int_{kh}^{(k+1)h}h^{-1}G_k(h^{-1}(t-kh))vec(\Sigma(t))\,dt.$$

The vector of total biases of (4.3) is then the linear functional of Σ:

$$\sum_{k=0}^{h^{-1}-1}h\sum_{j=1}^{\infty}W_{jk}vec(B_{jk})=\int_0^1 G^h(t)vec(\Sigma(t))\,dt$$

where for $t\in[kh,(k+1)h))$

$$G^h(t)=G_k(h^{-1}(t-kh))=2\sum_{j=1}^{\infty}W_{jk}\cos(2\pi jh^{-1}t)\,.$$

For Σ in the Besov space $B^{\alpha}_{1,\infty}([0,1])$, $0<\alpha\le 1$, the L^1-modulus of continuity satisfies $\omega_{L^1([0,1])}(\Sigma,\delta)\le\|\Sigma\|_{B^{\alpha}_{1,\infty}}\delta^{\alpha}$, see e.g. [9, Section 3.2]. We have for $\delta\in(0,1)$ and $s\in[0,1-\delta]$

$$\Big|\int_0^{\delta}vec(\Sigma(t+s))\cos(\tfrac{2\pi t}{\delta})\,dt\Big|=\frac{1}{\delta}\Big|\int_0^{\delta}\int_0^{\delta}vec(\Sigma(t+s)-\Sigma(u+s))\,du$$
$$\cos(\tfrac{2\pi t}{\delta})\,dt\Big|\le\sup_{0\le v\le\delta}\int_0^{\delta}|vec(\Sigma(t+s)-\Sigma(t+v+s))|\,dt\le\omega_{L^1([s,s+\delta])}(\Sigma,\delta).$$

This shows for the total bias in estimation of the volatility in X by the bound on $\|W_{jk}\|$ in Lemma C.1

$$\Big|\int_0^1 G^h(t)vec(\Sigma(t))\,dt\Big|\le 2\sum_{k=0}^{h^{-1}-1}\sum_{j=1}^{\infty}\|W_{jk}\|\omega_{L^1([kh,(k+1)h])}(\Sigma,h/j)$$
$$\lesssim\sum_{j=1}^{\infty}h_0^{-1}(1+(h_0/j)^4)^{-1}(h/j)^{\alpha}\asymp(h/h_0)^{\alpha}=n_{min}^{-\alpha/2}.$$

We thus have a bias of order $\mathcal{O}(n_{min}^{-\alpha/2})$. Remark that it is quite surprising that this bias bound is independent of h, which is also at the heart of the quasi-maximum likelihood method [1].

If $vec(\Sigma)$ is a (vector-valued) square-integrable martingale, then we use that martingale differences are uncorrelated and write for the total bias

$$\int_0^1 G^h(t) vec(\Sigma(t))\, dt = \int_0^1 G^h(t) vec(\Sigma(t) - \Sigma(\lfloor h^{-1}t \rfloor h))\, dt,$$

using $\int G_k = 0$. This expression is centred with covariance matrix

$$\sum_{k=0}^{h^{-1}-1} \int_{[kh,(k+1)h]^2} G_k(h^{-1}(t-kh)) \mathbb{E}[vec(\Sigma(t)-\Sigma(kh)) vec(\Sigma(s)-\Sigma(kh))^\top] \\ G_k(h^{-1}(s-kh))\, dtds.$$

The expected value in the display is smaller than (in matrix ordering) $\mathbb{E}[vec(\Sigma((k+1)h)-\Sigma(kh)) vec(\Sigma((k+1)h)-\Sigma(kh))^\top]$. Because of $\|G_k\|_\infty \lesssim 1$ the covariance matrix (in any norm) is of order $\mathcal{O}(h^2 \mathbb{E}[\|\Sigma(1)-\Sigma(0)\|^2]) = \mathcal{O}(h^2)$.

If $\Sigma = \Sigma^B + \Sigma^M$ is the sum of a function Σ^B in $B^\alpha_{1,\infty}([0,1])$ and a square-integrable martingale Σ^M, then the preceding estimations apply for each summand and the total bias has maximal order $\mathcal{O}(n_{min}^{-\alpha/2}) + \mathcal{O}_P(h)$.

C.3. Variance for general continuous-time model. The covariance for the estimator under model $\mathcal{E}_1$ can be calculated as under model $\mathcal{E}_2$, but we lose independence between different frequencies j, j' on the same block. For that we use the formula for Gaussian random vectors A, B

$$\mathbb{C}\mathrm{OV}(vec(AA^\top), vec(BB^\top)) = \Big(\mathbb{C}\mathrm{OV}(B,B) \otimes \mathbb{C}\mathrm{OV}(A,B) + \mathbb{C}\mathrm{OV}(A,A) \otimes \\ \mathbb{C}\mathrm{OV}(A,B) + \mathbb{C}\mathrm{OV}(A,B) \otimes \mathbb{C}\mathrm{OV}(A,A) + \mathbb{C}\mathrm{OV}(A,B) \otimes \mathbb{C}\mathrm{OV}(B,B)\Big)\mathcal{Z}/4,$$

obtained by polarisation. This implies

$$\|\mathbb{C}\mathrm{OV}_{\mathcal{E}_1}(\mathrm{LMM}_{or}^{(n)}) - \mathbb{C}\mathrm{OV}_{\mathcal{E}_2}(\mathrm{LMM}_{or}^{(n)})\| \\ \lesssim \sum_{k=0}^{h^{-1}-1} h^2 \sum_{j,j'=1}^{\infty} \|W_{j'k}\| \|W_{jk}(\mathbb{C}\mathrm{OV}_{\mathcal{E}_1}(S_{jk}, S_{jk}) \otimes \mathbb{C}\mathrm{OV}_{\mathcal{E}_1}(S_{jk}, S_{j'k}))\|.$$

From Lemma C.1 and $\|A \otimes B\| \leq \|A\|\|B\|$ for matrices A, B we infer that the series over j, j' is bounded in order by

$$\sum_{j,j'=1}^{\infty} h_0^{-2}(1+j'/h_0)^{-4}(1+j/h_0)^{-2}\Big(\Big\|\int_0^1 (\Sigma - \bar{\Sigma}_h)(t) \frac{\Phi_{jk}(t)\Phi_{j'k}(t)}{\|\Phi_{jk}\|_{L^2}\|\Phi_{j'k}\|_{L^2}}\, dt\Big\| \\ + \Big\|\int_0^1 \mathrm{diag}(H_{n,l}^2 - \overline{H^2}_{n,l,h})(t)\, \varphi_{jk}(t)\varphi_{j'k}(t)\, dt\Big\|\Big).$$

The identities $2\cos(a)\cos(b) = \cos(a+b) + \cos(a-b)$, $2\sin(a)\sin(b) = \cos(a-b) - \cos(a+b)$ and the same bound as in Section C.2 imply for $\Sigma, (F_1')^{-1}, \dots, (F_d')^{-1} \in B^{\alpha}_{1,\infty}([0,1])$ (note that even $(F_l')^{-1} \in C^{\alpha}([0,1])$)

$$\Big\|\int_0^1 (\Sigma - \bar\Sigma_h)(t) \frac{\Phi_{jk}(t)\Phi_{j'k}(t)}{\|\Phi_{jk}\|_{L^2}\|\Phi_{j'k}\|_{L^2}}\,dt\Big\| \lesssim h^{-1}\Big(\frac{h}{j+j'} + \frac{h(1-\delta_{j,j'})}{|j-j'|}\Big)^{\alpha} \times \|\Sigma\|_{B^{\alpha}_{1,\infty}([kh,(k+1)h])}$$

and similarly the bound

$$h^{-1}\Big(\frac{h}{j+j'} + \frac{h(1-\delta_{j,j'})}{|j-j'|}\Big)^{\alpha} jj' h_0^{-2} \max_l \|(F_l')^{-1}\|_{B^{\alpha}_{1,\infty}([kh,(k+1)h])}$$

for the norm over $H^2_{n,l}$. Putting all estimates together gives

$$\begin{aligned}&\|\mathbb{COV}_{\mathcal{E}_1}(\mathrm{LMM}^{(n)}_{or}) - \mathbb{COV}_{\mathcal{E}_2}(\mathrm{LMM}^{(n)}_{or})\|\\ &\lesssim h \sum_{j,j'=1}^{\infty} h_0^{-2}(1+j'/h_0)^{-4}(1+j/h_0)^{-2} h^{\alpha}(1+|j-j'|)^{-\alpha}(1+jj'h_0^{-2}).\end{aligned}$$

By comparison with the double integral (in terms of $x \approx j/h_0$, $y \approx j'/h_0$)

$$\int_0^{\infty}\int_0^{\infty} (1+y)^{-4}(1+x)^{-2}|x-y|^{-\alpha}(1+xy)\,dxdy \lesssim 1$$

we conclude

$$\|\mathbb{COV}_{\mathcal{E}_1}(\mathrm{LMM}^{(n)}_{or}) - \mathbb{COV}_{\mathcal{E}_2}(\mathrm{LMM}^{(n)}_{or})\| \lesssim h n_{min}^{-\alpha/2}.$$

Arguing exactly as in Section C.2 for the case of Σ being a sum of a $B^{\alpha}_{1,\infty}$-function and an L^2-martingale, the difference of covariances is in general of order $\mathcal{O}(hn_{min}^{-\alpha/2}) + \mathcal{O}_P(h^2)$.

C.4. Proof of Theorem 4.4. Let us denote the rate of convergence of $\hat\Sigma$ by $\delta_n = n_{min}^{-\alpha/(4\alpha+2)}$. For later use we note the order bounds

$$\delta_n = \scriptstyle{\mathcal{O}}\Big(r^{1/2}h_0^{-1/2}(n_{min}/n_{max})^{1/4}\Big), \quad \delta_n = \scriptstyle{\mathcal{O}}\Big(h_0^{-1}(n_{min}/n_{max})^{1/2}\Big). \tag{C.3}$$

First, we show that

$$\|\,\mathrm{LMM}^{(n)}_{or} - \mathrm{LMM}^{(n)}_{ad}\,\| = \scriptstyle{\mathcal{O}}_P(n_{max}^{-1/4}) \tag{C.4}$$

which by Slutsky's Lemma implies the CLT with normalisation matrix $\mathbf{I_n}$. This in turn is already sufficient for obtaining the result of Corollary 4.3 for $\text{LMM}_{ad}^{(n)}$. Let us start with proving that

$$T_n^m := \Big\| \sum_{m=0}^{r^{-1}-1} h \sum_{k=mr/h}^{(m+1)r/h-1} \sum_{j=1}^{\infty} \left(W_j(\hat{\Sigma}^{mr}) - W_j(\Sigma^{mr})\right) Z_{jk} \Big\| = \scriptstyle\mathcal{O}_P(n_{max}^{-1/4})$$

where the random variables

$$Z_{jk} = vec\left(S_{jk}S_{jk}^{\top} - \pi^2 j^2 h^{-2} \operatorname{diag}\left((H_{n,l}^{kh})^2\right)_{1\le l\le d} - \Sigma^{kh}\right)$$

are independent, $\mathbb{E}_{\mathcal{E}_2}[Z_{jk}] = 0$, $\mathbb{COV}_{\mathcal{E}_2}(Z_{jk}) = I_{jk}^{-1}\mathcal{Z}$. We have

$$T_n^m \le \sum_{m=0}^{r^{-1}-1} h \sum_{j=1}^{\infty} \Big\| W_j(\hat{\Sigma}^{mr}) - W_j(\Sigma^{mr}) \Big\| \Big\| \sum_{k=mr/h}^{(m+1)r/h-1} Z_{jk} \Big\|, \tag{C.5}$$

since the weight matrices do not depend on k on the same block of the coarse grid. Using Lemma C.2 and that $\|\hat{\Sigma} - \Sigma\|_{L^1} = \mathcal{O}_P(\delta_n)$, we obtain

$$\begin{aligned}\Big\| W_j(\hat{\Sigma}^{mr}) - W_j(\Sigma^{mr}) \Big\| &\le \max_k \Big\| \frac{dW_j(\Sigma^{kh})}{d\Sigma^{kh}} \Big\| \|\hat{\Sigma}^{mr} - \Sigma^{mr}\| \\ &= \mathcal{O}_P\left(\left(h_0^{-1} \wedge h_0^3 j^{-4}\right) r^{-1} \|\hat{\Sigma} - \Sigma\|_{L^1([mr,(m+1)r])}\right) .\end{aligned}$$

For the second factor in (C.5) we employ $\|\mathbb{COV}_{\mathcal{E}_2}(Z_{jk})\| = 2\|C_{jk}\|^2$. Consequently, (C.3) implies for T_n^m the bound

$$\begin{aligned}&\sum_{m=0}^{r^{-1}-1} h\|\hat{\Sigma}^{mr} - \Sigma^{mr}\| \sum_{j=1}^{\infty} \mathcal{O}\left((h_0^{-1} \wedge h_0^3 j^{-4})\,(1 \vee j^2 h_0^{-2})\right) \\ &= \|\hat{\Sigma} - \Sigma\|_{L^1([0,1])}\, \|\mathcal{O}\left(r^{-1/2} h^{1/2}\right) = \mathcal{O}_P\left(r^{-1/2} h^{1/2} \delta_n\right) = \scriptstyle\mathcal{O}_P(n_{max}^{-1/4})\, .\end{aligned}$$

The asymptotics (C.4) follow if we can ensure that the coarse grid approximations of the weights induce a negligible error, i.e. if also

$$\sum_{m=0}^{r^{-1}-1} \sum_{k=mr/h}^{(m+1)r/h-1} h^2 \sum_{j=1}^{\infty} \left(W_j(\Sigma^{kh}) - W_j(\Sigma^{mr})\right) Z_{jk} = \scriptstyle\mathcal{O}_P(n_{max}^{-1/4})$$

holds. The term is centred and its covariance matrix is bounded in norm by

$$\sum_{m=0}^{r^{-1}-1} \sum_{k=mr/h}^{(m+1)r/h-1} h^2 \sum_{j=1}^{\infty} \Big\| W_j(\Sigma^{kh}) - W_j(\Sigma^{mr}) \Big\|^2 \|I_{jk}^{-1}\|.$$

From Lemma C.2, $\|I_{jk}^{-1}\| = 2\|C_{jk}\|^2 \lesssim 1 + j^4 h_0^{-4}$ and $\Sigma \in B^{\alpha}_{1,\infty}([0,1])$ we derive the upper bound

$$\mathcal{O}\Big(\sum_{k=0}^{h^{-1}-1} h^2 \sum_{j=1}^{\infty} r^2 h_0^{-2}(1+j^4h_0^{-4})^{-1}\Big) = \mathcal{O}\Big(n_{min}^{-1/2} r^{2\alpha}\Big) = \scriptstyle\mathcal{O}\displaystyle(n_{max}^{-1/2})$$

by the choice of r and $\alpha > 1/2$.

Another application of Slutsky's Lemma yields the CLT with normalisation matrix $\hat{\mathbf{I}}_n$ provided $\mathbf{I}_n^{1/2}\hat{\mathbf{I}}_n^{-1/2} \to E_{d^2}$ in probability. The proof of Lemma C.2, more specifically the bound on the last term in (C.2), yields also

$$\Big\|\frac{d}{d\Sigma^{kh}} I_j(\Sigma^{kh})\Big\| \lesssim h_0^{-1}(1+j^4h_0^{-4})^{-1}.$$

This implies $\sum_{k,j}\|\hat{I}_{jk} - I_{jk}\| = \mathcal{O}_P(h^{-1}\delta_n)$. Using $\hat{A}^{-1} - A^{-1} = A^{-1}(\hat{A} - A)\hat{A}^{-1}$ and $\|I_k^{-1}\| \lesssim h_0^{-1}$, we infer

$$\|\hat{\mathbf{I}}_n^{-1} - \mathbf{I}_n^{-1}\| \le \sum_{k=0}^{h^{-1}-1} h^2 \Big\| \Big(\sum_{j=1}^{\infty} \hat{I}_{jk}\Big)^{-1} - \Big(\sum_{j=1}^{\infty} I_{jk}\Big)^{-1}\Big\| = \mathcal{O}_P\big(h\delta_n h_0^{-2}\big).$$

The smallest eigenvalue of $\mathbf{I}_n^{-1}$ equals $\|\mathbf{I}_n\|^{-1}$ which has order at least $n_{max}^{-1/2}$. The global Lipschitz constant L_n of $f(x) = x^{1/2}$ for $x \ge \|\mathbf{I}_n\|^{-1}$ is therefore of order $n_{max}^{1/4}$. The perturbation result from [18] for functional calculus therefore implies

$$\|\mathbf{I}_n^{1/2}\hat{\mathbf{I}}_n^{-1/2} - E_d\| \le L_n \|\mathbf{I}_n^{1/2}\|\|\mathbf{I}_n^{-1} - \hat{\mathbf{I}}_n^{-1}\| = \mathcal{O}_P\Big(n_{max}^{1/2} h\delta_n h_0^{-2}\Big).$$

The order is $(n_{max}/n_{min})^{1/2} h_0^{-1}\delta_n$ and tends to zero by (C.3).

APPENDIX D: PROOF OF THE LOWER BOUND

D.1. Proof of Lemma 5.1. Since $\mathcal{M}_{(R')^{1/2}} T_r$ is an isometry on $L^2([0,1];\mathbb{R}^d)$, we obtain directly for the adjoint $T_r^* = T_r^{-1}\mathcal{M}_{(R')^{-1}}$. We observe in a formal differential notation:

$$\begin{aligned} T_r^*\mathcal{M}_{(R')^{1/2}O} dY &= T_r^{-1}\mathcal{M}_{(R')^{-1/2}O}(X dt + \tfrac{1}{\sqrt{n}} dW) \\ &= -T_r^{-1} I^*(\mathcal{M}_{((R')^{-1/2}O)'} X dt + \mathcal{M}_{(R')^{-1/2}O} dX) + \tfrac{1}{\sqrt{n}} d\bar{W} \\ &= -I^* T_r^*(\mathcal{M}_{((R')^{-1/2}O)'} X dt + \mathcal{M}_{(R')^{-1/2}O} dX) + \tfrac{1}{\sqrt{n}} d\bar{W}. \end{aligned}$$

Here, we use that $T_r^*\mathcal{M}_{(R')^{1/2}O}$ is an L^2-isometry and we introduce the independent Brownian motions $\bar{W}$, $\bar{B}$ via the differentials

$$d\bar{W} = T_r^*\mathcal{M}_{(R')^{1/2}O}\, dW, \quad d\bar{B} = T_r^*\mathcal{M}_{(R')^{1/2}O}\, dB$$

or alternatively (apply $-I^*$) via their coordinates $i = 1, \dots, d$ as

$$\bar{W}_i(u) = \sum_{j=1}^{d} \int_0^{r_i^{-1}(u)} R'_{ii}(s)^{1/2} O_{ij}(s)\, dW_j(s), \tag{D.1}$$

and $\bar{B}_i(u)$ analogously.

The formal derivations are made rigorous by duality, that is testing stochastic differentials with deterministic L^2-functions. We infer from the coordinate-wise definition of $\bar{W}$ for $f \in L^2([0,1];\mathbb{R}^d)$ (e.g., check via indicator functions f)

$$\int_0^1 \langle O(t)^\top R'(t)^{1/2}(T_r f)(t), dW_t\rangle = \int_0^1 \langle f(u), d\bar{W}(u)\rangle$$

and equally for $\bar{B}$. Now consider for functions $g \in L^2([0,1];\mathbb{R}^d)$ the real observations

$$\begin{aligned}
&\int_0^1 \langle O(t)^\top R'(t)^{1/2}(T_r g)(t), dY_t\rangle = \int_0^1 \langle O(t)^\top R'(t)^{-1/2}(T_r Ig)'(t), dY_t\rangle \\
&= \int_0^1 \langle (O(t)^\top R'(t)^{-1/2}(T_r Ig))' - (O(t)^\top R'(t)^{-1/2})'(T_r Ig)(t), X_t\rangle\, dt \\
&\quad + \frac{1}{\sqrt{n}} \int_0^1 \langle O(t)^\top R'(t)^{1/2}(T_r g)(t), dW_t\rangle \\
&= \int_0^1 \langle -(O(t)^\top R'(t)^{-1/2})'(T_r Ig)(t), X_t\rangle\, dt \\
&\quad - \int_0^1 \langle O(t)^\top R'(t)^{-1/2}(T_r Ig)(t), dX_t\rangle + \frac{1}{\sqrt{n}} \int_0^1 \langle g(u), d\bar{W}_u\rangle \\
&= \int_0^1 \langle -(O(t)^\top R'(t)^{-1/2})'(T_r Ig)(t), X_t\rangle\, dt \\
&\quad - \int_0^1 \langle \Sigma(t)^{1/2} O(t)^\top R'(t)^{-1/2}(T_r Ig)(t), dB_t\rangle + \frac{1}{\sqrt{n}} \int_0^1 \langle g(u), d\bar{W}_u\rangle.
\end{aligned}$$

For $\varepsilon = 0$ we use $(R')^{-1/2}\Lambda(R')^{-1/2} = \bar{\Lambda}$ and evaluate the first two terms of the last display as

$$\int_0^1 \langle -(O(t)^\top R'(t)^{-1/2})'(T_r Ig)(t), X_t\rangle\, dt - \int_0^1 \langle (T_r^{-1}\bar{\Lambda}(T_r Ig))(u), d\bar{B}_u\rangle.$$

As $\bar{\Lambda}$ is constant in time, the second term is equal to $-\int_0^1 \langle Ig, \bar{\Lambda} d\bar{B}\rangle$ and the formal derivations above are confirmed.

D.2. Proof of Lemma 5.2. In a first step note that for general operators A, B we have

$$\begin{aligned}\|AA^* - BB^*\|_{HS}^2 &= \tfrac{1}{2}\|(2A + B - A)(A - B)^* + (2A + B - A)^*(A - B)\|_{HS} \\ &\le 2\|A\|\|A - B\|_{HS} + \|A - B\|_{HS}^2.\end{aligned}$$

Hence, it suffices to show

$$\|Q_{n,0}^{-1/2} Q_{n,1}^{1/2}\| \lesssim 1 \text{ and } \|Q_{n,0}^{-1/2} Q_{n,1}^{1/2} - C_{n,0}^{-1/2} C_{n,1}^{1/2}\|_{HS} \lesssim 1.$$

A further reduction is achieved by splitting terms to obtain

$$\begin{aligned}\|Q_{n,0}^{-1/2} Q_{n,1}^{1/2} - C_{n,0}^{-1/2} C_{n,1}^{1/2}\|_{HS} &\le \|\mathrm{Id} - C_{n,0}^{-1/2} Q_{n,0}^{1/2}\|_{HS}\|Q_{n,0}^{-1/2} Q_{n,1}^{1/2}\| \\ &\quad + \|C_{n,0}^{-1/2} Q_{n,0}^{1/2}\|\|Q_{n,0}^{-1/2} Q_{n,1}^{1/2}\|\|\mathrm{Id} - Q_{n,1}^{-1/2} C_{n,1}^{1/2}\|_{HS}.\end{aligned}$$

Owing to $\|C_{n,0}^{-1/2} Q_{n,0}^{1/2}\| \le 1 + \|\mathrm{Id} - C_{n,0}^{-1/2} Q_{n,0}^{1/2}\|_{HS}$ it remains to show

$$\|Q_{n,0}^{-1/2} Q_{n,1}^{1/2}\| \lesssim 1, \quad \|\mathrm{Id} - C_{n,0}^{-1/2} Q_{n,0}^{1/2}\|_{HS} \lesssim 1 \text{ and } \|\mathrm{Id} - Q_{n,0}^{-1/2} C_{n,0}^{1/2}\|_{HS} \lesssim 1.$$

Finally, we can use $Q_{n,1} - Q_{n,0} = Q_{\infty,1} - Q_{\infty,0}$, $Q_{n,1} \ge Q_{\infty,1}$ in operator order (and similarly for $C_{n,\varepsilon}$) as well as $|a - 1| \le |a^2 - 1|$ for $a \ge 0$ implying $\|A - \mathrm{Id}\|_{HS} \le \|AA^* - \mathrm{Id}\|_{HS}$ for positive operators A. We are thus left with proving that the following three quantities are uniformly bounded

$$\|Q_{n,0}^{-1/2} Q_{n,1}^{1/2}\|, \|C_{\infty,0}^{-1/2}(Q_{\infty,0} - C_{\infty,0}) C_{\infty,0}^{-1/2}\|_{HS}, \|Q_{\infty,1}^{-1/2}(C_{\infty,1}^{1/2} - Q_{\infty,1}^{1/2}) Q_{\infty,1}^{-1/2}\|_{HS}.$$

By the Feldman-Hajek Theorem for Gaussian measures, see e.g. [10], the latter two quantities are finite iff the Gaussian laws $\mathbf{N}(0, C_{\infty,\varepsilon})$ and $\mathbf{N}(0, Q_{\infty,\varepsilon})$, are equivalent for $\varepsilon \in \{0, 1\}$. Using again differential notation, these are the laws of

$$Z^C := T_r^* \mathcal{M}_{(R')^{1/2} O} X, \quad Z^Q := -I^* T_r^* \mathcal{M}_{(R')^{-1/2} O} dX$$

where $dX = \Sigma^{1/2} dB$ for the ε at hand. Both processes are images in $C([0,1], \mathbb{R}^d)$ under the linear (and thus measurable) map $T_r^{-1} = T_r^* \mathcal{M}_{R'}$ of the respective processes

$$\tilde{Z}^C := \mathcal{M}_{(R')^{-1/2} O} X, \quad \tilde{Z}^Q := -I^* \mathcal{M}_{(R')^{-1/2} O} dX.$$

By the product rule we see

$$\begin{aligned}\tilde{Z}^C(t) &= -I^*\{\mathcal{M}_{(R')^{-1/2} O} dX + \mathcal{M}_{((R')^{-1/2} O)'} X\}(t) \\ &= \tilde{Z}^Q(t) + \int_0^t ((R')^{-1/2} O)'(s) X(s)\, ds.\end{aligned}$$

Hence, $\tilde{Z}^C$ equals the Brownian martingale $\tilde{Z}^Q$ plus an adapted linear drift in X. By Girsanov's theorem, noting that all deterministic quantities are continuous and bounded away from zero, the laws of $\tilde{Z}^C$ and $\tilde{Z}^Q$ are equivalent, e.g. use Thm. 3.5.1. together with Cor. 3.5.16 in [17]. Hence, so are the laws of their images Z^C and Z^Q, as required.

Let us finally consider $Q_{n,0}^{-1/2}Q_{n,1}^{1/2}$. Its squared norm equals

$$\begin{aligned}\sup_{f\in L^2}\frac{\langle Q_{n,1}f,f\rangle}{\langle Q_{n,0}f,f\rangle} &= \sup_{f\in L^2}\frac{\langle \mathcal{M}_{(R')^{-1/2}O\Sigma_1 O^\top (R')^{-1/2}}T_r If, T_r If\rangle + \frac{1}{n}\|f\|^2}{\|If\|^2+\frac{1}{n}\|f\|^2}\\ &\le (1+\|M\|_\infty)\max_{i=1,\dots,d}\|(r_i^{-1})'\|_\infty.\end{aligned}$$

This uniform bound is finite under our regularity assumptions.

D.3. Proof of Theorem 5.3. Without loss of generality we may assume that $A(t)$ is symmetric for all t because $\Sigma(t)$ is symmetric. Owing to $\mathcal{Z}vec(A) = 2vec(A)$ and $(\Sigma_0\otimes\Sigma_0^{1/2})vec(A) = vec(\Sigma_0^{1/2}A\Sigma_0)$, we thus have to show in terms of the Hilbert-Schmidt scalar product

$$\mathbb{V}\mathrm{ar}_{\varepsilon=0}(\hat{\vartheta}_n) \ge \frac{(8+\mathcal{O}(1))}{\sqrt{n}}\int_0^1 \langle(\Sigma_0\otimes\Sigma_0^{1/2}+\Sigma_0^{1/2}\otimes\Sigma_0)A)(t),A(t)\rangle_{HS}dt.$$

Since C_{BM} is a positive operator on $L^2([0,1];\mathbb{R})$, we can define the bounded self-adjoint operator

$$\Delta_n^\sigma = I(\sigma^2 C_{BM}+\frac{1}{n}\operatorname{Id})^{-1}I^* = (\sigma^2\operatorname{Id}+\frac{1}{n}C_{BM}^{-1})^{-1}.$$

C_{BM} is Hilbert-Schmidt and so is Δ_n^σ. We identify its kernel $\delta_n^\sigma:[0,1]^2\to\mathbb{R}$ (or Green function) as

$$\delta_n^\sigma(t,s) = \frac{\sqrt{n}}{2\sigma\cosh(\sigma\sqrt{n})}\Big(\sinh(\sigma\sqrt{n}(1-|s-t|))+\sinh(\sigma\sqrt{n}(t+s-1))\Big).$$

This can be formally derived from the properties $C_{BM}^{-1}=-D^2$ on its domain, δ_n^σ in the domain (i.e. $\delta_n^\sigma(0,s)=0$, $(\delta_n^\sigma)'(1,s)=0$) and $\delta_n^\sigma(t,s)=\Delta_n^\sigma\delta_s(t)$. Alternatively use the eigenvalue-eigenfunction decomposition of C_{BM} and apply functional calculus. The main observation is that δ_n^σ has all the properties of a smoothing kernel, which for $n\to\infty$ concentrates on the diagonal $\{t=s\}$, where it approximates the uniform law. This is best seen by the approximation for large n

$$\delta_n^\sigma(t,s)\asymp\frac{\sqrt{n}}{2\sigma}\Big(\exp(-\sigma\sqrt{n}|s-t|)+\operatorname{sgn}(t+s-1)\exp(\sigma\sqrt{n}(|t+s-1|-1))\Big),$$

observing $|t+s-1|-1 \le |s-t|+|2t-1|-1$ such that the second exponential asymptotically only contributes at the corners $(0,0)$ and $(1,1)$ of the diagonal.

We shall see, however, that for the Hilbert-Schmidt norm evaluation we face $(\delta_n^\sigma)^2$ as the operator kernel, which also behaves like a smoothing kernel on the diagonal, but needs to be rescaled by $\|\delta_n^\sigma\|_{L^2}^2 = (1+\mathcal{O}(1))\sqrt{n}/(4\sigma^3)$.

Consequently, in terms of $\Delta_n = \operatorname{diag}(\Delta_n^{\bar\Lambda_{ii}})_{i=1,\dots,d}$ and its kernel $\delta_n(t,s) = \operatorname{diag}(\delta_n^{\bar\Lambda_{ii}}(t,s))_{i=1,\dots,d}$, the Fisher information evaluates as

$$\begin{aligned} I_n^Q &= \frac{1}{2}\|Q_{n,0}^{-1/2}\dot{Q}_0 Q_{n,0}^{-1/2}\|_{HS}^2 \\ &= \frac{1}{2}\operatorname{trace}(T_r I Q_{n,0}^{-1} I^* T_r^* \mathcal{M}_M T_r I Q_{n,0}^{-1} I^* T_r^* \mathcal{M}_M) \\ &= \frac{1}{2}\operatorname{trace}((T_r\Delta_n T_r^*)\mathcal{M}_M (T_r\Delta_n T_r^*)\mathcal{M}_M) \\ &= \frac{1}{2}\int_0^1\int_0^1 \operatorname{trace}_{\mathbb{R}^{d\times d}}(\delta_n(r(t),r(s))M(s)\delta_n(r(s),r(t))M(t))\,dtds. \end{aligned}$$

We now use $\int_0^1 M(s)a_n e^{-|t-s|a_n}ds = 2M(t)(1+\mathcal{O}(1))$ uniformly over $t\in[b_n,1-b_n]$ whenever $a_n\to\infty$, $a_nb_n\to\infty$ and $M(t)$ is continuously differentiable. Together with the asymptotic behaviour of δ_n^σ we obtain

$$\begin{aligned} &\int_0^1 \delta_n^{\bar\Lambda_{ii}}(r_i(t),r_i(s))M_{ij}(s)\delta_n^{\bar\Lambda_{jj}}(r_j(s),r_j(t))\,ds \\ &= (1+\mathcal{O}(1))\frac{n}{4\bar\Lambda_{ii}\bar\Lambda_{jj}}\int_0^1 \exp(-\sqrt{n}(\bar\Lambda_{ii}r_i'(t)+\bar\Lambda_{jj}r_j'(t))|t-s|)M_{ij}(s)\,ds \\ &= (1+\mathcal{O}(1))\frac{\sqrt{n}}{2\bar\Lambda_{ii}\bar\Lambda_{jj}(\bar\Lambda_{ii}r_i'(t)+\bar\Lambda_{jj}r_j'(t))}M_{ij}(t) \\ &= \frac{\sqrt{n}M_{ij}(t)(1+\mathcal{O}(1))}{2\bar\Lambda_{ii}\bar\Lambda_{jj}(\lambda_i(t)+\lambda_j(t))} \end{aligned}$$

with $\mathcal{O}(1)$ uniformly in n and $t\in[n^{-p},1-n^{-p}]$ for any $p\in(0,1/2)$ to infer

$$\begin{aligned} I_n^Q &= \frac{\sqrt{n}}{4}\sum_{i,j=1}^d \bar\Lambda_{ii}^{-1}\bar\Lambda_{jj}^{-1}\int_0^1 (\lambda_i(t)+\lambda_j(t))^{-1}M_{ij}(t)^2(1+\mathcal{O}(1))dt \\ &= \frac{\sqrt{n}(1+\mathcal{O}(1))}{4}\int_0^1\sum_{i,j=1}^d \frac{(O\mathbb{H}O^\top)_{ij}^2}{\lambda_i(\lambda_i+\lambda_j)\lambda_j}(t)\,dt. \end{aligned}$$

Asymptotically for $n\to\infty$ neglecting terms of smaller order, this bound is obtained by the worst parametric perturbation $\mathbb{H}^*(t) = \Sigma_0 A\Sigma_0^{1/2} +$

$\Sigma_0^{1/2} A\Sigma_0$, which we evaluate using duality with respect to the scalar product $\int_0^1 \sum_{i,j} A_{ij}(t)B_{ij}(t)dt$ as

$$\sup_{\mathbb{H}:\mathbb{H}(t)=\mathbb{H}(t)^\top} \frac{(\int_0^1 \sum_{i,j=1}^d A_{ij}(t)\mathbb{H}_{ij}(t)\,dt)^2}{\frac{\sqrt{n}}{4}\int_0^1 \sum_{k,l=1}^d \frac{(O\mathbb{H}O^\top)_{kl}^2}{\lambda_k(\lambda_k+\lambda_l)\lambda_l}(t)\,dt} = \frac{(\int_0^1 \sum_{i,j=1}^d A_{ij}(t)\mathbb{H}^*_{ij}(t)\,dt)^2}{\frac{\sqrt{n}}{4}\int_0^1 \sum_{k,l=1}^d \frac{(O\mathbb{H}^*O^\top)_{kl}^2}{\lambda_k(\lambda_k+\lambda_l)\lambda_l}(t)\,dt}$$
$$= \frac{4}{\sqrt{n}}\int_0^1 \langle A(t), (\Sigma A\Sigma^{1/2} + \Sigma^{1/2}A\Sigma)(t)\rangle_{HS}\,dt.$$

Finally, remark that the Cramér-Rao inequality, e.g. [20, Thm. 2.5.10], is applicable since $(\mathbf{N}(0, Q_{n,\varepsilon}))_\varepsilon$ forms an exponential family in $(Q_{n,\varepsilon}^{-1})_\varepsilon$, which is differentiable at $\varepsilon = 0$, and thus the models $(\mathbf{N}(0, Q_{n,\varepsilon}))_\varepsilon$ as well as $(\mathbf{N}(0, C_{n,\varepsilon}))_\varepsilon$ are regular.

REFERENCES

[1] Aït-Sahalia, Y. and Fan, J. and Xiu, D. (2010). High frequency covariance estimates with noisy and asynchronous financial data. *J. Amer. Statist. Assoc.* **105** 1504-1517. MR2796567

[2] Andersen, T. and Bollerslev, T. (1997). Intraday perdiodicity and volatility persistence in financial markets. *J. Empir. Financ.* **4** 115-158.

[3] Andersen, T.G. and Bollerslev, T. and Diebold, F.X. (2010). Parametric and Nonparametric Volatility Measurement *Handbook of Financial Econometrics, Vol. 1, Ed. Aït-Sahalia, Y. and Hansen, L.P., Elsevier* **105** 67-137.

[4] Barndorff-Nielsen, O. E. and Hansen, P. R. and Lunde, A. and Shephard, N. (2011). Multivariate realised kernels: consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J. Econometrics* **162(2)** 149–169.

[5] Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: high frequency based covariance, regression, and correlation in financial economics. *Econometrica* **72(3)** 885–925.

[6] Bibinger, M. and Reiss, M. (2013). Spectral estimation of covolatility from noisy observations using local weights. *Scand. J. Stat.*, DOI: 10.1111/sjos.12019.

[7] Christensen, K. and Podolskij, M. and Vetter, M. (2013). On covariation estimation for multivariate continuous Itô semimartingales with noise in non-synchronous observation schemes, *J. Multivar. Anal.* **120**, 59–84.

[8] Ciesielski, Z., Kerkyacharian, G. and Roynette, B. (1993). Quelques espaces fonctionnels associés à des processus gaussiens, *Studia Math.* **107**(2), 171–203. MR1244574

[9] Cohen, A. (2003). *Numerical Analysis of Wavelet Methods*, Studies in Math. Appl. **32**, Elsevier. MR1990555

[10] Da Prato, G. and Zabczyk, J. (2008). *Stochastic Equations in Infinite Dimensions*, corr. ed., Cambridge University Press. MR1207136

[11] Fackler, P. L. (2005). Notes on matrix calculus, *Lecture Notes, North Carolina State University.* http://www4.ncsu.edu/~pfackler/MatCalc.pdf

[12] Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors 1 and 2. *ESAIM, Probab. Stat.* **5** 225–242. MR1875672 & 1875673

[13] Hansen, L. P. (1982). Large sample properties of generalized methods of moment estimation. *Econometrica* **50(4)** 1029–1054.

[14] Hayashi, T. and Yoshida, N. (2011). Nonsynchronous covariation process and limit theorems. *Stochastic Processes Appl.* **121(10)** 2416-2454. MR2822782

[15] Jacod, J. and Rosenbaum, M. (2013) Quarticity and other functionals of volatility: Efficient estimation *Ann. Statist.* **41(3)** 1462–1484.

[16] Kallenberg, O. (2002), *Foundations of Modern Probability*, 2nd ed., Springer. MR1876169

[17] Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed., Springer Graduate Texts **113**. MR1121940

[18] Kittaneh, F. (1985). On Lipschitz functions of normal operators. *Proc. Amer. Math. Soc.* **94**, 416–418. MR0787884

[19] Le Cam, L. and Yang, L. G. (2000). *Asymptotics in Statistics. Some basic concepts*, 2nd ed., Springer. MR1784901

[20] Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation*, 2nd ed., Springer. MR1639875

[21] Liu, C. and Tang, C. Y. (2012). A Quasi-Maximum Likelihood Approach for Integrated Covariance Matrix Estimation with High Frequency Data, *Preprint.*

[22] Reiss, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *Ann. Statist.* **2** 772–802. MR2816338

[23] Shephard, N. and Xiu, D. (2012). Econometric analysis of multivariate realised QML: efficient positive semi-definite estimators of the covariation of equity prices, *Preprint.*

[24] Zhang, L. (2011). Estimating covariation: epps effect and microstructure noise. *J. Econometrics* **160(1)** 33–47.

[25] Zhang, L. and Mykland, P. A. and Aït-Sahalia, Y. (2005). A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data. *J. Amer. Statist. Assoc.* **100** 1394–1411. MR2236450

Institut für Mathematik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin, Germany
E-mail: bibinger@math.hu-berlin.de

Department of Statistics
and Operations Research
University of Vienna
Oskar-Morgenstern-Platz 1
1090 Vienna, Austria
E-mail: nikolaus.hautsch@univie.ac.at

School of Business and Economics
Humboldt-Universität zu Berlin
Spandauer Str. 1
10178 Berlin, Germany
E-mail: malecpet@hu-berlin.de

Institut für Mathematik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin, Germany
E-mail: mreiss@math.hu-berlin.de