



FiNANCE FOR ENERGY MARKET RESEARCH CENTRE



EFFICIENT VOLATILITY ESTIMATION IN A TWO-FACTOR MOREL

Olivier FERON, Pierre GRUET, and Marc HOFFMANN

Working Paper
RR-FiME-19-06

June 2019

EFFICIENT VOLATILITY ESTIMATION IN A TWO-FACTOR MODEL

OLIVIER FÉRON, PIERRE GRUET, AND MARC HOFFMANN

ABSTRACT. We statistically analyse a multivariate HJM diffusion model with stochastic volatility. The volatility process of the first factor is left totally unspecified while the volatility of the second factor is the product of an unknown process and an exponential function of time to maturity. This exponential term includes some real parameter measuring the rate of increase of the second factor as time goes to maturity. From historical data, we efficiently estimate the time to maturity parameter in the sense of constructing an estimator that achieves an optimal information bound in a semiparametric setting. We also identify nonparametrically the paths of the volatility processes and achieve minimax bounds. We address the problem of degeneracy that occurs when the dimension of the process is greater than two, and give in particular optimal limit theorems under suitable regularity assumptions on the drift process. We consistently analyse the numerical behaviour of our estimators on simulated and real datasets of prices of forward contracts on electricity markets.

Mathematics Subject Classification (2010): 62M86, 60J75, 60G35, 60F05.

Keywords: HJM models, time-to-maturity factor, Financial statistics, Discrete observations, semiparametric efficient bounds, nonparametric estimation, Electricity market modelling.

1. INTRODUCTION

1.1. Motivation and setting. We address statistical estimation for multidimensional diffusion processes from historical data, with a volatility structure including both a parametric and a nonparametric components. We aim at achieving efficient estimation of a scalar parameter in the volatility, in presence of nonparametric nuisance, while providing point estimates of nonparametric components simultaneously. The processes of interest follow the multiple Brownian factor representation, as in the Heath-Jarrow-Morton (HJM) framework for forward rates, for instance in Heath *et al.* [6], or for electricity forward contracts in Benth and Koekebakker [1].

Our setting is motivated by the context of prices of specific forward contracts, which are available on the electricity market. Interest rate models have been applied to the pricing of such contracts: see for instance Hinz *et al.* [7], in which an analogy between interest rate models and forward contracts prices models is performed, the maturity in the former framework being a date of delivery in the latter. The factorial representation of the HJM framework has been precisely studied in Benth and Koekebakker [1] to model the electricity forward curve, giving constraints in the volatility terms to ensure no arbitrage. Koekebakker and Ollmar [15] perform a Principal Component Analysis to point out that two factors can explain 75% of the electricity forward contracts in the Norwegian market, and more than 10 factors are needed to explain 95%. They argue that, due to the non-storability of electricity, there is a weak correlation between short-term and long-term events. In Keppo *et al.* [13], a one-factor model is designed for each maturity date, having correlations between the Brownian motions for distinct dates. In Kiesel *et al.* [14], a two-factor model is described, with a specification of the volatility terms allowing to reproduce the classical behaviour of prices, especially the empirical evidence of the Samuelson effect (the volatility of prices increases as time to maturity decreases) and to ensure non-zero volatility for long-term forward prices.

On some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$, we consider a d -dimensional Itô semimartingale $X = (X_t)_{t \geq 0}$ with components X^j , for $j = 1, \dots, d$, of the form

$$(1) \quad X_t^j = X_0^j + \int_0^t b_s^j ds + \int_0^t e^{-\vartheta(T_j-s)} \sigma_s dB_s + \int_0^t \bar{\sigma}_s d\bar{B}_s,$$

where $X_0^j \in \mathbb{R}$ is an initial condition, $B = (B_t)_{t \geq 0}$ and $\overline{B} = (\overline{B}_t)_{t \geq 0}$ are two independent Brownian motions, ϑ and T_j are positive numbers and $\sigma = (\sigma_t)_{t \geq 0}$, $\overline{\sigma} = (\overline{\sigma}_t)_{t \geq 0}$, $b^j = (b_t^j)_{t \geq 0}$ are càdlàg adapted processes. To avoid trivial situations, we assume that for some $T > 0$, we have

$$T \leq T_1 < \dots < T_d$$

and that the T_j are known. Moreover, we observe X at times

$$0, \Delta_n, 2\Delta_n, \dots, n\Delta_n = T.$$

Asymptotics are taken as $n \rightarrow \infty$ with fixed T once for all and throughout the paper. In this setting, it is impossible to identify the components b^i , so we are left with trying to estimate the parameter ϑ and the random components $t \rightsquigarrow \sigma_t$ (or rather σ_t^2) and $t \rightsquigarrow \overline{\sigma}_t$ (or $\overline{\sigma}_t^2$) over the time interval $[0, T]$ with the best possible rate of convergence. This is not always possible and will require regularity assumptions.

The statistical estimation of the volatility of a diffusion process observed over some period $[0, T]$ has long been studied for asymptotic regimes in which observation times asymptotically recover the whole observation period. This carries over to the setting, considered here, where the unknown volatility – as a parameter – is random w.r.t. the filtration generated by the observation itself, see for instance [4, 5, 8, 9, 10] and the references therein for a comprehensive study in both parametric and nonparametric settings. Concerning estimating a functional of the trajectory of the diffusion process the chapters of Mykland and Zhang [16] and Jacod [11] present the most advanced problems related to the estimation of diffusion processes, together with important estimation results, stated in a general way. As integrated volatility can be estimated with the usual $\Delta_n^{-1/2}$ -rate of convergence, the quality of its estimators may be assessed by looking at the limit law that one can get when writing a central limit theorem, and by looking at a minimal variance in some sense (usually, the limiting distribution is a mixture of a centred Gaussian variable, with random variance). Clément *et al.* [2] estimated some functionals of the volatility; in the diffusion model that they introduced, they prove an extension of Hájek convolution theorem, and are able to define some notion of efficiency, which is somehow related to our setting. The present paper is in line with these results from a methodological point of view. However, we face several new structural aspects and questions that require a novel treatment and that lead to new results in some cases.

- (1) The dimension d dictates in some sense the underlying regularity of the model. The (componentwise) quadratic variation identifies the parameters of the model asymptotically via its integrated volatility by the (componentwise) convergence

$$\sum_{i=1}^{\lfloor \Delta_n^{-1} t \rfloor} (\Delta_i^n X^j)^2 \rightarrow \int_0^t (e^{-2\vartheta(T_j-t)} \sigma_t^2 + \overline{\sigma}_t^2) dt, \quad j = 1, \dots, d$$

in probability as $\Delta_n \rightarrow 0$, valid for every $t \in [0, T]$ with the notation $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$ (componentwise). We thus identify the (random) function

$$t \rightsquigarrow e^{-2\vartheta(T_j-t)} \sigma_t^2 + \overline{\sigma}_t^2, \quad j = 1, \dots, d$$

for $t \in [0, T]$. Since ϑ , σ and $\overline{\sigma}$ are unknown, we readily see that the model cannot be identified for $d = 1$. For $d = 2$, we asymptotically recover the function

$$t \rightsquigarrow \begin{cases} e^{-2\vartheta(T_1-t)} \sigma_t^2 + \overline{\sigma}_t^2 \\ e^{-2\vartheta(T_2-t)} \sigma_t^2 + \overline{\sigma}_t^2 \end{cases}$$

and it is not obvious that the three parameters can be identified. This is possible by considering specific linear and quadratic combinations of the components of the vectors $\Delta_i^n X$, $i = 1, \dots, n$ and this yields to estimators of ϑ and $\sigma, \overline{\sigma}$ that achieve usual rates of convergence. In particular, the case $d = 3$ (and beyond) becomes somehow degenerate, since we (discretely) observe a 3-dimensional diffusion model driven by a two dimensional Brownian motion $(B, \overline{B})$. For applications, this had been reported by Jeffrey *et al.* [12] in a similar context. In that case, the same kind of techniques lead to estimator of ϑ that achieve the fast rate Δ_n^{-1} .

- (2) It is interesting to note that it is possible in dimension $d = 2$ to separate ϑ from the nuisance nonparametric components σ and $\bar{\sigma}$. However, the substitution estimators are only rate-optimal. We need to modify our estimator of ϑ and introduce a completely different technique in order to obtain an optimal asymptotic variance. This is done by using semiparametric efficiency theory see *e.g.* the classical textbook of van der Vaart [17] in a context where the nonparametric nuisance parameters is a bivariate random volatility process $t \rightsquigarrow (\sigma_t^2, \bar{\sigma}_t^2)$, which is new to the best of our knowledge.
- (3) In particular, we need to exhibit nonparametric estimators of σ and $\bar{\sigma}$ to be used in the correction efficient estimator of ϑ . While nonparametric estimation of the time-dependent diffusion coefficient is classical, we do not have a completely usual nonparametric problem, since
 - (a) σ and $\bar{\sigma}$ are random themselves, so that we do not estimate them pointwise, instead we estimate the trajectories $t \rightsquigarrow (\sigma^2(\omega)_t, \bar{\sigma}^2(\omega)_t)$ pointwise, as realisations of the volatility process;
 - (b) an increment $\Delta_i^n X$ is the sum of two stochastic integrals, in which the volatility processes have different regularities.

1.2. Main results and organisation of the paper. In Section 2.1, we provide an estimator of ϑ , based on quadratic variation, in the above observation scheme. While we cannot perform estimation when the number of observed processes d is equal to 1, the case $d = 2$ is statistically regular, and by approaching the quadratic variation of X^1 , X^2 and $X^2 - X^1$, we derive an estimator $\hat{\vartheta}_{2,n}$ of ϑ , which is $\Delta_n^{-1/2}$ -consistent. Using the theory of statistics for diffusion processes and relying on the tools of stable convergence in law, which are for instance summarized in [11, 16], we show that

$$\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta) \rightarrow \mathcal{N}(0, V_\vartheta(\sigma, \bar{\sigma})),$$

stably in law, where, conditional on $\mathcal{F}_T$, the random variable $\mathcal{N}(0, V_\vartheta(\sigma, \bar{\sigma}))$ is centred Gaussian, with conditional variance $V_\vartheta(\sigma, \bar{\sigma})$, and possibly defined on an extension of the original probability space. When $d = 3$ (and beyond) the model is somehow degenerate because the 3 marginal components of the process are driven by 2 Brownian motions. The remaining source of randomness is the drift process, and we are able to construct a Δ_n^{-1} -consistent estimator $\hat{\vartheta}_{3,n}$ for ϑ . This is possible as soon as b has some integrated regularity in expectation, reminiscent of the so-called Besov regularity, as will be made precise by Assumption 1. This enables us to obtain a satisfying limit theorem for $\Delta_n^{-1}(\hat{\vartheta}_{3,n} - \vartheta)$, namely the convergence in probability to some $\mathcal{F}$ -measurable random variable. All our results in dimension $d = 2, 3$ are stated in Theorem 1. The estimation of ϑ in $d = 2$ however misses the optimal variance and we next turn to modifying $\hat{\vartheta}_{3,n}$ in order to achieve the best possible variance.

In Section 2.2, we turn to the problem of estimating ϑ optimally when $d = 2$. As a preliminary technical result, We first perform a relatively classical nonparametric estimation procedure to get point estimates of σ_t^2 and $\bar{\sigma}_t^2$ when $d = 2$ that will serve our later purposes of estimating ϑ in an optimal way. We have to separate, in some way, the parts of the random increments that are linked to each of the Brownian integrals, to be able to get estimates of each process. We derive estimators $(\hat{\sigma}_n^2, \hat{\bar{\sigma}}_n^2)$ of $(\sigma^2, \bar{\sigma}^2)$. If the volatility processes are Hölder in expectation (Assumption 2) we show in Proposition 1 the convergence of the estimators with rate $\Delta_n^{-\alpha/(2\alpha+1)}$ -consistent, where α is the lowest of two values of the Hölder regularities of σ^2 and $\bar{\sigma}^2$. In Section 2.3, relying to the theory of asymptotically efficient semiparametric estimation (for instance [17]) we compute a lower bound $V_\vartheta^{\text{opt}}(\sigma, \bar{\sigma})$ for the limit variance while estimating ϑ with $d = 2$ observed processes, for deterministic volatility functions, in Theorem 2. As soon as $\bar{\sigma}$ is not constant, this bound is lower than $V_\vartheta(\sigma, \bar{\sigma})$. We subsequently derive an estimator $\tilde{\vartheta}_{2,n}$ such that

$$\Delta_n^{-1/2}(\tilde{\vartheta}_{2,n} - \vartheta) \rightarrow \mathcal{N}(0, V_\vartheta^{\text{opt}}(\sigma, \bar{\sigma}))$$

stably in law, where conditional on $\mathcal{F}$, the random variable $\mathcal{N}(0, V_\vartheta^{\text{opt}}(\sigma, \bar{\sigma}))$ is centred Gaussian with conditional variance $V_\vartheta^{\text{opt}}(\sigma, \bar{\sigma})$. This is the main result of the paper. The estimator $\tilde{\vartheta}_{2,n}$ is built upon the preliminary estimators $\hat{\vartheta}_{2,n}$ and $(\hat{\sigma}_n^2, \hat{\bar{\sigma}}_n^2)$ of $(\sigma^2, \bar{\sigma}^2)$. Moreover, it is asymptotically efficient in the sense that it achieves the minimal conditional variance $V_\vartheta^{\text{opt}}(\sigma, \bar{\sigma})$ among all possible $\Delta_n^{-1/2}$ -consistent estimators that are asymptotically centred mixed normal. We perform numerical experiments in Section 3, using both

simulated and real data from the electricity forward markets in order to compare the behaviours of the estimators in various configurations. The proofs are delayed until Section 4.

2. CONSTRUCTION OF THE ESTIMATORS AND CONVERGENCE RESULTS

2.1. Rate-optimal estimation of ϑ . In this section, we build a (preliminary) estimator of ϑ , denoted by $\widehat{\vartheta}_{d,n}$ in each of the situation $d = 2$ or $d = 3$.

The case $d = 1$. In that setting, it is impossible to identify ϑ from data $X_{i\Delta_n}, i = 1, \dots, n$ asymptotically when $t \rightsquigarrow \sigma_t$ and $t \rightsquigarrow \bar{\sigma}_t$ are unknown. Indeed X has the same law under the choice of $(\vartheta, \sigma, \bar{\sigma})$ and $(\vartheta + 1, e^{T_1 - \cdot} \sigma, \bar{\sigma})$.

The case $d = 2$. This is the statistically most regular case. Set, as usual $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$ (componentwise). From the convergences

$$\sum_{i=1}^n (\Delta_i^n X^j)^2 \rightarrow \int_0^T (e^{-2\vartheta(T_j-t)} \sigma_t^2 + \bar{\sigma}_t^2) dt, \quad j = 1, 2$$

and

$$\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2 \rightarrow \int_0^T (e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 e^{2\vartheta t} \sigma_t^2 dt$$

in probability, we also obtain the convergence of the ratio

$$\Psi_{T_1, T_2}^n = \frac{\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2}{\sum_{i=1}^n ((\Delta_i^n X^2)^2 - (\Delta_i^n X^1)^2)} \rightarrow \frac{(e^{-\vartheta T_2} - e^{-\vartheta T_1})^2}{e^{-2\vartheta T_2} - e^{-2\vartheta T_1}} = \psi_{T_1, T_2}(\vartheta),$$

in probability. The function $\vartheta \rightsquigarrow \psi_{T_1, T_2}(\vartheta)$ maps $(0, \infty)$ onto $(-1, 0)$ and this leads to a first estimation strategy by setting

$$\widehat{\vartheta}_{2,n} = \psi_{T_1, T_2}^{-1}(\Psi_{T_1, T_2}^n)$$

whenever $\Psi_{T_1, T_2}^n \in (-1, 0)$ and 0 otherwise.

The case $d = 3$. Since X is driven by two Brownian motions, the underlying statistical model becomes degenerate. Indeed, assume first that $b^1 = b^2 = b^3$. Then, we readily obtain

$$\frac{\Delta_i^n X^2 - \Delta_i^n X^1}{\Delta_i^n X^3 - \Delta_i^n X^2} = \frac{e^{-\vartheta T_2} - e^{-\vartheta T_1}}{e^{-\vartheta T_3} - e^{-\vartheta T_2}}$$

which is invertible as a function of ϑ . It is thus possible to identify ϑ exactly from the observation of a single increment of X itself! When the b^j are not all equal, the situation is still somehow degenerate, as we can eliminate all volatility components by taking linear combinations of the observed increments. The lowest-order remaining term is the drift process, with increments that are of smaller order than that of the diffusion part. Therefore, we may hope to exhibit estimators of ϑ that converge at a faster rate than $\Delta_n^{1/2}$. More specifically, we have the convergence

$$\Psi_{T_1, T_2, T_3}^n = \frac{\sum_{i=1}^n (\Delta_i^n X^3 - \Delta_i^n X^2)^2}{\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2} \rightarrow \left(\frac{e^{-\vartheta T_3} - e^{-\vartheta T_2}}{e^{-\vartheta T_2} - e^{-\vartheta T_1}} \right)^2 = \psi_{T_1, T_2, T_3}(\vartheta),$$

say, in probability. The function $\vartheta \rightsquigarrow \psi_{T_1, T_2, T_3}(\vartheta)$ maps $(0, \infty)$ onto $(0, (\frac{T_3 - T_2}{T_2 - T_1})^2)$ and it can be checked with elementary calculus arguments that it is also invertible, leading to the estimator

$$\widehat{\vartheta}_{3,n} = \psi_{T_1, T_2, T_3}^{-1}(\Psi_{T_1, T_2, T_3}^n)$$

whenever $\Psi_{T_1, T_2, T_3}^n \in (0, (\frac{T_3 - T_2}{T_2 - T_1})^2)$ and 0 otherwise.

Convergence results. Remember that T is fixed and thus asymptotics are taken as $n \rightarrow \infty$ or equivalently $\Delta_n \rightarrow 0$. We need some assumption about the regularity of the processes b , σ and $\bar{\sigma}$. For a random process $X = (X_t)_{0 \leq t \leq T}$, introduce the following modulus of continuity:

$$\omega(X)_t = \sup_{|h| \leq t} \left(\int_0^T \mathbb{E}[(X_{s+h} - X_s)^2] ds \right)^{1/2}.$$

Assumption 1. For some constant $c_{\min} > 0$, we have $\inf_{t \in [0, T]} \min(\sigma_t, \bar{\sigma}_t) \geq c_{\min}$. Moreover, for some $s > 1/2$, we have $\sup_{t \in [0, T]} t^{-s} \omega(b^j)_t < \infty$ for every $j = 1, \dots, d$.

Remark 1. Assumption 1 is technical. The first condition on the ellipticity of the volatility processes enables one to apply Girsanov theorem to get rid of the drift terms b^j in some cases. The second assumption is related to the pathwise smoothness property of the drift b^j in L^2 and is reminiscent of a Besov regularity. For instance, a sufficient condition is that $t \rightsquigarrow b_t^j$ is almost-surely s -Hölder continuous. In particular, we need a slightly stronger smoothness than that of typical Brownian paths.

To state the convergence results, we need some notation. Set

$$\bar{b}_t = 2(e^{-\vartheta T_2} - e^{-\vartheta T_1})(e^{-\vartheta T_3} - e^{-\vartheta T_2})((e^{-\vartheta T_2} - e^{-\vartheta T_1})(b_t^3 - b_t^2) - (e^{-\vartheta T_3} - e^{-\vartheta T_2})(b_t^2 - b_t^1))$$

and

$$\tilde{b}_T = (e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 \int_0^T (b_t^3 - b_t^2)^2 dt - (e^{-\vartheta T_3} - e^{-\vartheta T_2})^2 \int_0^T (b_t^2 - b_t^1)^2 dt.$$

Define also

$$D_3 = (e^{-\vartheta T_3} - e^{-\vartheta T_2})((e^{-\vartheta T_3} - e^{-\vartheta T_2})(T_2 e^{-\vartheta T_2} - T_1 e^{-\vartheta T_1}) - (e^{-\vartheta T_2} - e^{-\vartheta T_1})(T_3 e^{-\vartheta T_3} - T_2 e^{-\vartheta T_2})).$$

Theorem 1. Work under Assumption 1.

(1) For $d = 2$, we have

$$\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta) \rightarrow \mathcal{N}(0, V_\vartheta(\sigma, \bar{\sigma}))$$

in distribution as $n \rightarrow \infty$, where $\mathcal{N}(0, V_\vartheta(\sigma, \bar{\sigma}))$ is a random variable which, conditionally on $\mathcal{F}_T$, is centred Gaussian with variance

$$V_\vartheta(\sigma, \bar{\sigma}) = \frac{1}{(T_2 - T_1)^2} (e^{\vartheta T_2} - e^{\vartheta T_1})^2 \frac{\int_0^T e^{2\vartheta t} \sigma_t^2 \bar{\sigma}_t^2 dt}{\left(\int_0^T e^{2\vartheta t} \sigma_t^2 dt \right)^2}.$$

(2) Assume moreover that $b^j = 0$. For $d = 3$ we have

$$\Delta_n^{-1}(\hat{\vartheta}_{3,n} - \vartheta) \rightarrow \frac{\tilde{b}_T + \int_0^T \bar{b}_t e^{\vartheta t} \sigma_t dB_t}{2(e^{-\vartheta T_2} - e^{-\vartheta T_1}) D_3 \int_0^T e^{2\vartheta t} \sigma_t^2 dt}$$

in probability as $n \rightarrow \infty$.

Remark 2. In the proof in Section 4.2 below, we obtain (1) when $d = 2$ for $b^j = 0$, $j = 1, 2$. The general case is obtained via a change of probability argument thanks to Girsanov theorem that can be applied under Assumption 1. This can no longer be applied when $d = 3$ since the 3-dimensional semimartingale X^j relies on the 2-dimensional Brownian motion $(B, \bar{B})$ solely. We therefore obtain (2) under the restriction $b^j = 0$. The general case can however be obtained by specifically evaluating additional superoptimal error terms involving b^j at a significant cost of technical length.

2.2. Rate-optimal estimation of the volatility processes.

Construction of an estimator. In order to improve $\hat{\vartheta}_{2,n}$ and building an asymptotically efficient estimator of ϑ , we shall build pointwise rate-optimal nonparametric estimators of σ^2 and $\bar{\sigma}^2$. We start with the observation that for any sufficiently regular test function $g : [0, T] \rightarrow \mathbb{R}$, we have, for any $j = 1, \dots, d$,

$$(2) \quad \sum_{i=1}^n g((i-1)\Delta_n) (\Delta_i^n X^j)^2 \rightarrow \int_0^T g(s) d\langle X^j \rangle_s = \int_0^T g(s) (e^{-2\vartheta(T_j-s)} \sigma_s^2 + \bar{\sigma}_s^2) ds$$

in probability as $n \rightarrow \infty$. Therefore, picking a function g that mimics a Dirac mass at t , we can asymptotically identify

$$e^{-2\vartheta(T_1-t)}\sigma_t^2 + \bar{\sigma}_t^2 \quad \text{and} \quad e^{-2\vartheta(T_2-t)}\sigma_t^2 + \bar{\sigma}_t^2$$

by applying (2) for $j = 1, 2$ for a sequence g_n that converges to δ_t weakly. We thus identify σ_t^2 and $\bar{\sigma}_t^2$ as well by inverting a 2×2 linear system, namely

$$\begin{pmatrix} \sigma_t^2 \\ \bar{\sigma}_t^2 \end{pmatrix} = \mathcal{M}(\vartheta)_t \begin{pmatrix} e^{-2\vartheta(T_1-t)}\sigma_t^2 + \bar{\sigma}_t^2 \\ e^{-2\vartheta(T_2-t)}\sigma_t^2 + \bar{\sigma}_t^2 \end{pmatrix}$$

where

$$\mathcal{M}(\vartheta)_t = \frac{1}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}} \begin{pmatrix} 1 & -1 \\ -e^{-2\vartheta(T_2-t)} & e^{-2\vartheta(T_1-t)} \end{pmatrix}.$$

For a threshold $\varpi_n > 0$ and a bandwidth $h_n > 0$, define the estimators

$$(3) \quad \begin{pmatrix} \hat{\sigma}_{n,t}^2 \\ \hat{\bar{\sigma}}_{n,t}^2 \end{pmatrix} = h_n^{-1} \mathcal{M}(\max(\hat{\vartheta}_{2,n}, \varpi_n))_t \sum_{t-h_n \leq (i-1)/n < t} \begin{pmatrix} (\Delta_i^n X^1)^2 \\ (\Delta_i^n X^2)^2 \end{pmatrix}.$$

The bandwidth h_n is set below to balance both bias and variance, while $\varpi_n > 0$ guarantees the well-posedness of the estimator. Moreover, the exponential functions appearing in the denominator of $\hat{\sigma}_{n,t}^2$ and not in its numerator may imply an ill-posedness of the estimation problem. In particular, when $\hat{\vartheta}_{2,n}$ is high, one can expect very high values for $\hat{\sigma}_{n,t}^2$. This is what happens in the numerical results of Section 3.2.

Convergence result. We need an additional regularity assumption on the volatility processes σ and $\bar{\sigma}$.

Assumption 2. *There exists $\alpha \geq 1/2$ and a constant $c_\alpha > 0$ such that for every $t, s \in [0, T]$, we have*

$$(4) \quad \mathbb{E}[(\sigma_t^2 - \sigma_s^2)^2] + \mathbb{E}[(\bar{\sigma}_t^2 - \bar{\sigma}_s^2)^2] \leq c_\alpha |t - s|^{2\alpha}.$$

Remark 3. *Assumption 2 is technical and related to the pathwise smoothness property of the volatility processes σ and $\bar{\sigma}$ in L^∞ . Like Assumption 1, it is reminiscent of a Besov regularity. In particular, in Theorem 3, we need $\alpha > 1/2$, a slightly stronger smoothness than that of typical Brownian paths for which $\alpha = 1/2$ exactly.*

Proposition 1. *Work under Assumptions 1 and 2. Specify $h_n = \Delta_n^{1/(2\alpha+1)}$ and let $\varpi_n \rightarrow 0$. Then the sequences*

$$\Delta_n^{-\alpha/(2\alpha+1)}(\hat{\sigma}_{n,t}^2 - \sigma_t^2) \quad \text{and} \quad \Delta_n^{-\alpha/(2\alpha+1)}(\hat{\bar{\sigma}}_{n,t}^2 - \bar{\sigma}_t^2)$$

are tight, uniformly in t over compact sets included in $(0, T]$.

2.3. Efficient estimation of ϑ when $d = 2$. We may now look for the best attainable variance among rate-optimal estimators of ϑ that are asymptotically Gaussian. However, we do not have a statistical model in the classical sense, with parameters $(\vartheta, \sigma, \bar{\sigma})$ since $t \rightsquigarrow \sigma_t$ and $t \rightsquigarrow \bar{\sigma}_t$ are random processes themselves. In order to bypass this difficulty, we first restrict our attention to the case where σ and $\bar{\sigma}$ are deterministic functions, which enables us to identify our data within a semiparametric regular statistical model. Thanks to classical bounds on semiparametric estimation, we can explicitly compute the optimal (best achievable) variance $V_\vartheta^{\text{opt}}(\sigma, \bar{\sigma})$. In a second step, allowing σ and $\bar{\sigma}$ to be random again, we build a one-step correction of our preliminary estimator $\hat{\vartheta}_{2,n}$ which has the property of being asymptotically mixed Gaussian, with (conditional) variance equal to $V_\vartheta^{\text{opt}}(\sigma, \bar{\sigma})$, *i.e.* thus achieving the optimal variance along deterministic paths.

Lower bounds. Consider the statistical experiment $\mathcal{E}^n$ generated by data $(\Delta_i^n X^1, \Delta_i^n X^2, i = 1, \dots, n)$ with

$$(5) \quad X_t^j = X_0^j + \int_0^t e^{-\vartheta(T_j-s)} \sigma_s dB_s + \int_0^t \bar{\sigma}_s d\bar{B}_s, \quad j = 1, 2,$$

with parameter $(\vartheta, \sigma, \bar{\sigma}) \in \Theta \times \Sigma(c, \tilde{c})$, with $\Theta = (0, \infty)$ and $\Sigma(c, \tilde{c})$ being the space of positive (deterministic) functions $(\sigma, \bar{\sigma})$ defined on $[0, T]$, satisfying (4) of Assumption 2 with constant c and satisfying moreover $\tilde{c} \leq \inf_t \sigma_t \leq \sup_t \sigma_t \leq c$ for some $\tilde{c} > 0$. Note that (5) is a Gaussian diffusion obtained from (1) when restricting $(\sigma, \bar{\sigma})$ to deterministic functions and setting $b^j = 0$.

Theorem 2. Let $\widehat{\vartheta}_n$ be an estimator of ϑ in the experiment $\mathcal{E}^n$ such that $\Delta_n^{-1/2}(\widehat{\vartheta}_n - \vartheta)$ converges to $\mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma}))$ in distribution as $n \rightarrow \infty$. Then

$$V_\vartheta(\sigma, \overline{\sigma}) \geq V_\vartheta^{\text{opt}}(\sigma, \overline{\sigma}) = \frac{1}{(T_2 - T_1)^2} (e^{\vartheta T_2} - e^{\vartheta T_1})^2 \left(\int_0^T \frac{e^{2\vartheta t} \sigma_t^2}{\overline{\sigma}_t^2} dt \right)^{-1}.$$

Using Cauchy-Schwarz inequality, it is easy to prove that the expression of the limit variance is equal to the one we got in Theorem 1 for $\Delta_n^{-1/2}(\widehat{\vartheta}_{2,n} - \vartheta)$ if and only if $\overline{\sigma}$ is constant over the interval $[0, T]$. In the general case of an arbitrary volatility process, we have a suboptimal result. This is the motivation of our next step, toward asymptotic efficiency.

Construction of an efficient procedure. This is the most delicate part of the paper. By representation (5), we see that the $(\Delta_i^n X^1, \Delta_i^n X^2)$ are independent for $i = 1, \dots, n$. Moreover, $(\Delta_i^n X^1, \Delta_i^n X^2)$ is a centred Gaussian, with explicit covariance structure

$$\begin{aligned} \mathbb{E}[(\Delta_i^n X^1)^2] &= \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 dt + \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt, \\ \mathbb{E}[(\Delta_i^n X^2)^2] &= \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_2-t)} \sigma_t^2 dt + \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt, \\ \mathbb{E}[\Delta_i^n X^1 \Delta_i^n X^2] &= \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1+T_2-2t)} \sigma_t^2 dt + \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt. \end{aligned}$$

Let us further denote by $f_{\vartheta, \sigma, \overline{\sigma}}^{i,n}$ its density function w.r.t. the Lebesgue measure on $\mathbb{R}^2$. If the nuisance parameters $(\sigma, \overline{\sigma})$ were known, then an optimal (efficient) procedure could be obtained by a one-step correction of the type

$$\widehat{\vartheta}_n = \widehat{\vartheta}_{2,n} + \frac{\sum_{i=1}^n \ell_{\vartheta=\widehat{\vartheta}_{2,n}, \sigma, \overline{\sigma}}^i(\Delta_i^n X^1, \Delta_i^n X^2)}{\sum_{i=1}^n (\ell_{\vartheta=\widehat{\vartheta}_{2,n}, \sigma, \overline{\sigma}}^i(\Delta_i^n X^1, \Delta_i^n X^2))^2}$$

where

$$(6) \quad \ell_{\vartheta, \sigma, \overline{\sigma}}^{i,n}(\Delta_i^n X^1, \Delta_i^n X^2) = \partial_\vartheta \log f_{\vartheta, \sigma, \overline{\sigma}}^{i,n}(\Delta_i^n X^1, \Delta_i^n X^2)$$

is the score function associated to $(\Delta_i^n X^1, \Delta_i^n X^2)$, see for instance Section 8.9 in [17]. However, this oracle procedure is not achievable and we need to invoke the theory of semiparametric efficiency (see for instance Chapter 25 of [17]). In the presence of an extra nuisance parameter $(\sigma, \overline{\sigma})$, we consider instead the so-called efficient score

$$\widetilde{\ell}_{\vartheta, \sigma, \overline{\sigma}}^{i,n} = \ell_{\vartheta, \sigma, \overline{\sigma}}^{i,n} - \Pi \ell_{\vartheta, \sigma, \overline{\sigma}}^{i,n},$$

where Π is the projection operator onto the tangent space associated to a one-dimensional perturbation around the true (unknown) value $(\sigma, \overline{\sigma})$. It turns out that we indeed have a simple and explicit formula for $\widetilde{\ell}_{\vartheta, \sigma, \overline{\sigma}}^{i,n}$ which enables us to derive a one-step correction formula using $\widetilde{\ell}_{\vartheta, \sigma, \overline{\sigma}}^{i,n}$ and plug-in estimators in order to achieve the optimal bound.

For technical reason, we replace $\widehat{\vartheta}_{2,n}$ by $\Delta_n^{1/2} \lfloor \Delta_n^{-1/2} \widehat{\vartheta}_{2,n} \rfloor$ and we still write $\widehat{\vartheta}_{2,n}$ for simplicity. Likewise, we implicitly replace the estimators $\widehat{\sigma}_{n,t}^2$ defined in (3) by $\max(\min(\widehat{\sigma}_{n,t}^2, \tilde{c}^2), c)$, where $\tilde{c}$ is the lower bound associated to $\Sigma(c, \tilde{c})$ in the definition of the experiment $\mathcal{E}^n$. Set

$$\widetilde{\ell}_{\vartheta, \mu}^{i,n}(\Delta_i^n X^1, \Delta_i^n X^2) = \frac{(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)} \Delta_i^n X^1) e^{-\vartheta(T_2-T_1)} (T_2 - T_1)}{(1 - e^{-\vartheta(T_2-T_1)})^3 \Delta_n \mu}$$

for $i = 1, \dots, n$ and $(\vartheta, \mu) \in (0, \infty) \times [\tilde{c}, c]$. We are ready to state the main result of the paper. We may assume here that the data $(\Delta_i^n X^j)$ are obtained under (1) and therefore $b^j \neq 0$ here while the definition of $\widehat{\vartheta}_{n,3}$ remains unchanged.

Theorem 3. Work under Assumptions 1 and 2 with $\alpha > 1/2$.

- (1) In the Gaussian experiment $\mathcal{E}^n$ generated by (5) the efficient score for the parameter ϑ associated to $(\Delta_i^n X^1, \Delta_i^n X^2)$ is given by

$$\tilde{\ell}_{\vartheta, \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 dt}^{i,n}(\Delta_i^n X^1, \Delta_i^n X^2).$$

- (2) In the general experiment generated by (1) allowing for $b^j \neq 0$, if moreover $\mathbb{P}((\sigma, \bar{\sigma}) \in \Sigma(c, \tilde{c})) = 1$ for some $c, \tilde{c} > 0$, the estimator $\tilde{\vartheta}_{2,n}$ defined by

$$\tilde{\vartheta}_{2,n} = \hat{\vartheta}_{2,n} + \frac{\sum_{i \in \mathcal{I}_n} \tilde{\ell}_{\hat{\vartheta}_{2,n}, \hat{\sigma}_{n,(i-1)\Delta_n}^2}^{i,n}(\Delta_i^n X^1, \Delta_i^n X^2)}{\sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\hat{\vartheta}_{2,n}, \hat{\sigma}_{n,(i-1)\Delta_n}^2}^{i,n}(\Delta_i^n X^1, \Delta_i^n X^2))^2}$$

with $\mathcal{I}_n = \{i = 1, \dots, n, h_n \leq (i-1)\Delta_n < T\}$ satisfies

$$\Delta_n^{-1/2}(\tilde{\vartheta}_{2,n} - \vartheta) \rightarrow \mathcal{N}(0, V_{\vartheta}^{\text{opt}}(\sigma, \bar{\sigma}))$$

in distribution as $n \rightarrow \infty$, where $\mathcal{N}(0, V_{\vartheta}^{\text{opt}}(\sigma, \bar{\sigma}))$ denotes a random variable that conditionally on $\mathcal{F}_T$, centred Gaussian with (conditional) variance $V_{\vartheta}^{\text{opt}}(\sigma, \bar{\sigma})$.

This result shows that the lower bound $V_{\vartheta}^{\text{opt}}(\sigma, \bar{\sigma})$ can be attained, and therefore that efficient estimation can be performed beyond deterministic volatility functions. This is reminiscent of the LAMN property when considering random volatilities in parametric estimation and can be interpreted as a semiparametric analog, although we do not have a complete theory at this stage. A recommended reference in that direction is Genon-Catalot and Jacod [4] where the same methodology is applied in a parametric context.

3. NUMERICAL IMPLEMENTATION AND REAL DATA ANALYSIS

3.1. Electricity forward contracts. The prices of existing forward contracts in the electricity markets are characterised by three time components: the quotation date t and the dates T_s and T_e of respectively starting and ending power delivery. Therefore, a forward contract $F(t, T_s, T_e)$ will deliver to the holder 1 MWh of electricity continuously between dates T_s and T_e . Such a contract may be bought during a quotation period $[t_0, T]$ with $T < T_s$ and it is no more available once $t > T$. Typical observed contracts are of various delivery periods: one week, one month, one quarter (three months), one season (6 months) or one year. In this study we only consider the 6 observable monthly contracts (from 1 to 6 month-ahead) to estimate ϑ and the volatility processes σ and $\bar{\sigma}$. Also, for simplicity, we will drop T_e from the notation. In the context of simulated data, we will simulate prices of $F(t, T_s) = F(t, T_s, T_e)$, the forward delivering continuously 1 MWh during the period $[T_s, T_e]$. In the context of real data, the price $F(t, T_s)$ is observable.

3.2. Results on simulated data. The objective of this section is to study the estimators' behaviour on a simulated data set, where the log-prices of the forward contracts are simulated according to the two-factor model described in (1). The parameter values are chosen to be close to values estimated on real data: in [14], the volatility processes are constant, and the estimated values are $\sigma = 0.37 \text{ yr}^{-1/2}$ and $\bar{\sigma} = 0.15 \text{ yr}^{-1/2}$. Here we use a CIR-like model (the Cox-Ingersoll-Ross model for interest rates has been introduced in [3], in 1985), to emphasize the fact that our model may also be used in the context of interest rates modeling (this is indeed where it comes from, see [7]). Our parameters are

$$b_t^j = 3.65 \cdot 10^{-1}(\log(30) - X_t^j), \sigma_t = 0.37 \Sigma_t^d \text{ and } \bar{\sigma}_t = 0.15 \Sigma_t^d,$$

with $\Sigma_t^d = \sqrt{\frac{1}{d} \sum_{j=1}^d X_t^j}$, which is the square root of the average of the d quoted log-prices. We adopt various values of ϑ (values in yr^{-1}): 1.4, 10, 40. The first value is the estimated parameter shown in [14] and the others are chosen to cover a wide range of possible values to observe different behaviours of our estimators. Finally, the initial value of each simulated log-price series is the logarithm of a random variable taken uniformly over the interval $[20, 40]$, which is an usual range for prices in the market of forward contracts on electricity (see also the constant 30 in the drift, in the center of that interval). We consider different simulation configurations, all related to the situations we are facing on real data.

- 2 processes (1 month-ahead and 2 month-ahead) observed on $n = 100$ dates, with $T = T_1 = 150$ and $T_2 = 181$ days.

- 3 processes (1 month-ahead to 3 month-ahead) observed on $n = 80$ dates, with $T = T_1 = 120$, $T_2 = 150$ and $T_3 = 181$ days.
- 4 processes (1 month-ahead to 4 month-ahead) observed on $n = 60$ dates, with $T = T_1 = 90$, $T_2 = 120$, $T_3 = 151$ and $T_4 = 181$ days.
- 5 processes (1 month-ahead to 5 month-ahead) observed on $n = 40$ dates, with $T = T_1 = 59$, $T_2 = 90$, $T_3 = 120$, $T_4 = 151$ and $T_5 = 181$ days.
- 6 processes (1 month-ahead to 6 month-ahead) observed on $n = 20$ dates, with $T = T_1 = 31$, $T_2 = 59$, $T_3 = 90$, $T_4 = 120$, $T_5 = 151$ and $T_6 = 181$ days.

The decreasing number of observations corresponds to the configuration observed with real data: 2 monthly contracts with fixed delivery dates are jointly observed on working days during 5 months (around 100 quotation dates) whereas 6 monthly contracts can be jointly observed only during 1 month (around 20 quotation dates). The number of observations is a bit low, as we are relying on asymptotic results.

For each configuration, we perform 100,000 simulations. Recall that we denote by $\widehat{\vartheta}_{j,n}$ the estimator of ϑ from the configuration where j processes are observed, and also by $\widetilde{\vartheta}_{2,n}$ the efficient estimator as described in Section 2.3, available in the configuration of 2 observed processes. Although we have not proved that the estimator $\widetilde{\vartheta}_{2,n}$ is $\Delta_n^{-1/2}$ -consistent and that it reaches the lower bound for the limit variance when $\alpha = 1/2$, we have not got any numerical evidence against that possibility. Tables 1, 2 and 3 give the estimation results for $\vartheta = 1.4$, 10 and 40 yr^{-1} , respectively. In each configuration, these tables give the number of converging instances¹ of the estimator and their average, and the empirical quantile interval at 95% (issued from taking the quantiles of the sample of estimated values). We observe that the estimators perform quite well: except on three lines in Table 3, the true value of ϑ is always in the quantile interval. Finally, we empirically observe that adding new maturities does not improve the quality of estimation in all configurations. For instance, increasing the number of maturities may increase or decrease the length of the quantile interval, and it may shift it away from the true value of ϑ . Notice also that the one-step correction from $\widehat{\vartheta}_{2,n}$ to $\widetilde{\vartheta}_{2,n}$ never led to very different values.

Processes	Estimator	Instances that converged	Average	Quantile interval
2	$\widehat{\vartheta}_{2,n}$	100,000	1.4216	[1.2697,1.6048]
2	$\widetilde{\vartheta}_{2,n}$	100,000	1.4217	[1.2697,1.6048]
3	$\widehat{\vartheta}_{3,n}$	99,962	1.3799	[0.77864,1.9250]
4	$\widehat{\vartheta}_{4,n}$	100,000	1.3840	[1.0752,1.7646]
5	$\widehat{\vartheta}_{5,n}$	100,000	1.3807	[1.1274,1.6864]
6	$\widehat{\vartheta}_{6,n}$	100,000	1.3849	[1.0989,1.7644]

TABLE 1. Results of the estimation on simulated data with $\vartheta = 1.4 \text{ yr}^{-1}$.

Concerning the estimation results of the volatility processes σ_t^2 and $\overline{\sigma}_t^2$, we use the causal kernel $K(x) = \mathbf{1}_{(0,1](x)}$, and the bandwidth h_n for the two volatility functions is selected by cross validation. We also set $\varpi_n = 3.65 \cdot 10^{-2} \text{ yr}^{-1}$. In the following we show the estimators $\widehat{\sigma}_n^2$ and $\widehat{\overline{\sigma}}_n^2$ for the configuration where 2 processes are simulated on a period of 5 months (approximately 150 days), which means $T = T_1 = 150$ and $T_2 = 181$ days, with $n = 100$ dates and $\vartheta = 10 \text{ yr}^{-1}$. First we keep the specification $b_t^j = 3.65 \cdot 10^{-1}(\log(30) - X_t^j)$ for the drift process, but we use the constant volatility processes of [14], that is $\sigma = 0.37 \text{ yr}^{-1/2}$ and $\overline{\sigma} = 0.15 \text{ yr}^{-1/2}$. A deterministic specification allows us to compare the curve of point estimates with the deterministic function that was used to simulate the processes. For this exercise with constant volatility functions, the cross validation criterion does not have a minimum, as taking the longest possible bandwidth is the best way to estimate a constant volatility. Therefore we use a bandwidth of $h_n = 14 \text{ d}$, which is the value that is selected when we implement the cross validation on real data, see

¹We must notice that some occurrences may not lead to a solution in the estimation procedure because Ψ_{T_1, T_2}^n and Ψ_{T_1, T_2, T_3}^n , defined in Section 2.1, can sometimes take values outside the supports of ψ_{T_1, T_2}^{-1} and $\psi_{T_1, T_2, T_3}^{-1}$.

Processes	Estimator	Instances that converged	Average	Quantile interval
2	$\hat{\vartheta}_{2,n}$	99,953	10.507	[7.2997,16.500]
2	$\tilde{\vartheta}_{2,n}$	99,953	10.507	[7.2992,16.502]
3	$\hat{\vartheta}_{3,n}$	100,000	9.9498	[9.4916,10.258]
4	$\hat{\vartheta}_{4,n}$	100,000	9.9424	[9.6307,10.195]
5	$\hat{\vartheta}_{5,n}$	100,000	9.9388	[9.6538,10.180]
6	$\hat{\vartheta}_{6,n}$	100,000	9.9511	[9.6331,10.242]

TABLE 2. Results of the estimation on simulated data with $\vartheta = 10 \text{ yr}^{-1}$

Processes	Estimator	Instances that converged	Average	Quantile interval
2	$\hat{\vartheta}_{2,n}$	55,248	24.747	[10.215,56.650]
2	$\tilde{\vartheta}_{2,n}$	55,248	24.716	[10.210,56.663]
3	$\hat{\vartheta}_{3,n}$	100,000	33.904	[22.598,40.060]
4	$\hat{\vartheta}_{4,n}$	100,000	32.162	[22.204,39.689]
5	$\hat{\vartheta}_{5,n}$	100,000	31.075	[22.046,38.832]
6	$\hat{\vartheta}_{6,n}$	100,000	33.901	[26.134,39.320]

TABLE 3. Results of the estimation on simulated data with $\vartheta = 40 \text{ yr}^{-1}$.

below. Remember that the nonparametric estimation result, Proposition 1, gives convergence uniformly on $[h_n, T]$. Therefore we expect that the fit is not good for values of t being less than h_n , and this is why the lowest value on the x-axis is 14 days. We perform simulation and estimation 100,000 times, and then take the average and the quantiles of the 100,000 curves (that is, at each point t of the discretisation grid, we take the average and the quantiles at 2.5% and 97.5% of the 100,000 occurrences of $\hat{\sigma}_{n,t}^2$ and $\hat{\bar{\sigma}}_{n,t}^2$). Figures 1 and 2a give the averages of $\hat{\sigma}_{n,t}^2$ and $\hat{\bar{\sigma}}_{n,t}^2$, respectively, together with the true (constant) functions σ_t^2 and $\bar{\sigma}_t^2$. The average of $\hat{\sigma}_{n,t}^2$ is truncated, 0.4 % of the values at each date being removed because of extreme values. The quantile intervals are plotted as well. They show a good estimation of the volatility function $\bar{\sigma}_t^2$. However, we can observe in Figure 2a a bad performance of estimation of σ_t^2 , especially for large values of $T - t$, even when $t > h_n$. This reveals the ill-posedness of the problem, already pointed out in Section 2.2, due to the presence of the exponential terms in the denominator of the estimator of σ_t^2 : the latter can take very high values as soon as ϑ appears to be largely overestimated, which happens in a few simulations. Therefore, when high values appear, the estimation of σ_t should reasonably be taken into account only for small times to maturity $T - t$, where the estimation procedure seems to work well. Indeed, Figure 2b shows that the quality of the fit of σ_t is reasonably good at the very right of the time range. We put emphasis on the quantile intervals, that may be observed on all the plots. They show that the plot of the average exhibits a strange behaviour only because of some extreme values.

Now, we are back to the specification $\sigma_t = 0.37\Sigma_t^d, \bar{\sigma}_t = 0.15\Sigma_t^d$. As the volatility processes depend on the path of X , we cannot compare visually the real volatility and its point estimators. This is why we plot the differences $\hat{\sigma}_{n,t}^2 - \sigma_t^2$ and $\hat{\bar{\sigma}}_{n,t}^2 - \bar{\sigma}_t^2$ instead. Moreover, we let $\hat{\sigma}_{n,t}^2$ and $\hat{\bar{\sigma}}_{n,t}^2$ take negative values to emphasize symmetry in the differences. We plot the average and the quantile curves of the 100,000 estimators for the two volatility processes in Figures 3 and 4. The behaviours of the series of point estimators are very similar to the ones we described while considering deterministic volatility functions. The bandwidths for nonparametric estimation are chosen using cross validation on each simulation. The average bandwidths are 18.6 days for σ and 26.2 days for $\bar{\sigma}$. Because the estimator $\hat{\sigma}_{n,t}^2$ can take very high values, as we have noticed above, the average that is plotted in Figure 4 is computed using the 100,000 point estimates available at each plotting date, except the 100 lowest and the 100 highest ones. The quantiles are still plotted using all the available point estimates at each plotting date.

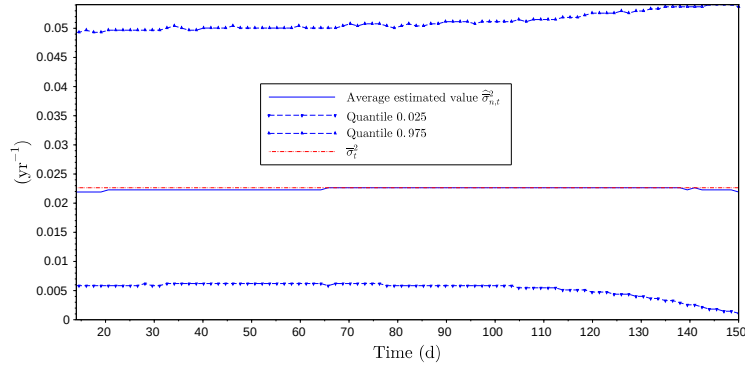


FIGURE 1. Quantiles for the square of the long-term volatility, with 2 processes, $\vartheta = 10 \text{ yr}^{-1}$ and deterministic constant volatilities

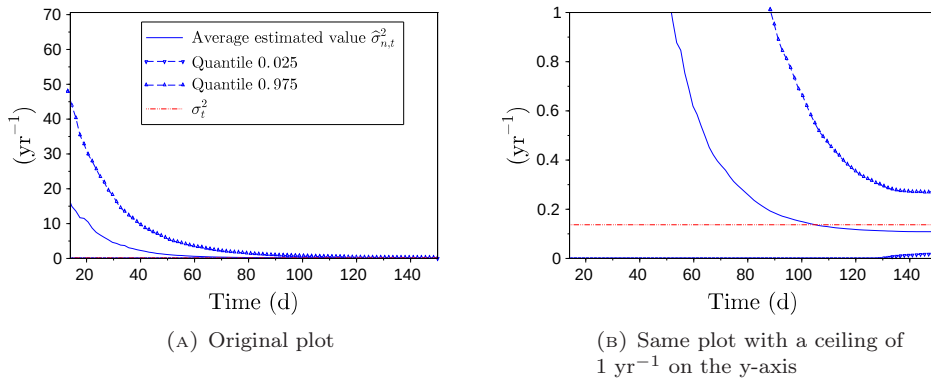


FIGURE 2. Quantiles for the square of the short-term volatility, with 2 processes, $\vartheta = 10 \text{ yr}^{-1}$ and deterministic constant volatilities

3.3. Results on real data from the French electricity market. The data used for estimation are the 6 available month-ahead forward contracts on the French market (www.eex.com) from December 6th, 2001 to November 30th, 2018. On this history, we get 204 periods of 1 month where 6 processes (the 6 month-ahead contracts) are jointly observed, whereas we get 200 periods of 5 months where 2 processes (the 1 month-ahead and the 2 month-ahead contracts) are jointly observed. These numbers of periods are given in Table 4 for all the configurations described in Section 3.2. In the same column, Table 4 also precises the number of periods on which the estimator converges. And the same table gives the estimation results of ϑ for all the possible configurations, with the average value and the standard deviation of the estimators. Contrary to the results on simulated data, the values of the estimators are different from one configuration to another. More precisely, the estimators from 2 processes are higher (of a factor between 5 and 8) than the ones from 3 to 6 processes. This can be explained by two different causes. First, we have proved the convergence of the estimators from 3 to 6 processes only if the drift is zero: this was stated in Theorem 1. Second, these differences may be due to the presence of errors linked to measurement or to the model. When we compare the estimators from more than 3 processes, we can observe that, like in the case of simulated data, adding processes does not significantly improve the estimation results.

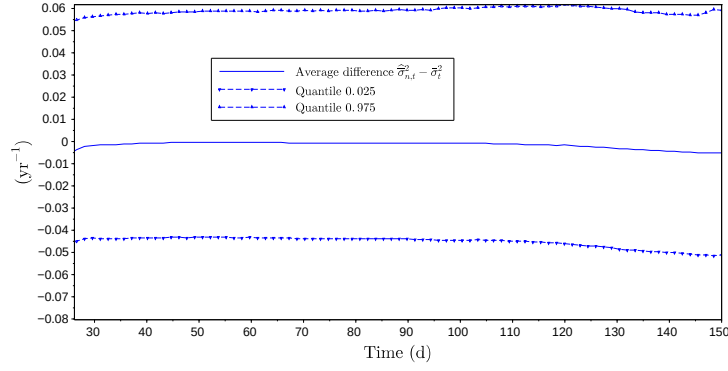


FIGURE 3. Estimation results of the long-term volatility process, in terms of $(\hat{\sigma}_{n,t}^2 - \sigma_t^2)$ with 2 processes, $\vartheta = 10 \text{ yr}^{-1}$ and the CIR-like specification for volatility processes

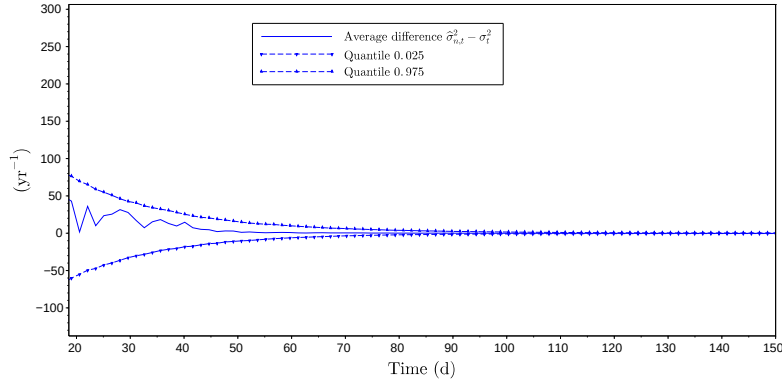


FIGURE 4. Estimation results of the short-term volatility process, in terms of $(\hat{\sigma}_{n,t}^2 - \sigma_t^2)$ with 2 processes, $\vartheta = 10 \text{ yr}^{-1}$ and the CIR-like specification for volatility processes

Estimator	Per. with convergence/	Number of per.	Average	Standard deviation
$\hat{\vartheta}_{2,n}$	71/200		24.966	10.734
$\tilde{\vartheta}_{2,n}$	71/200		25.049	10.743
$\hat{\vartheta}_{3,n}$	142/201		4.488	3.462
$\hat{\vartheta}_{4,n}$	149/202		3.387	2.732
$\hat{\vartheta}_{5,n}$	149/203		3.072	2.800
$\hat{\vartheta}_{6,n}$	132/204		3.514	2.946

TABLE 4. Estimators of ϑ on real data in France (unit: yr^{-1}).

We end this session by performing estimation on the last part of the dataset, which runs from July 1st, 2018 to November 30th, 2018. The value taken by $\hat{\vartheta}_{2,n}$ is 32.781 yr^{-1} , while $\tilde{\vartheta}_{2,n}$ has value 32.779 yr^{-1} . We use Theorems 1 and 3 to provide feasible central limit theorems, and thus we get confidence intervals using

the nonparametric estimates of the volatility processes. The confidence intervals are $[28.814, 36.748]$ for $\widehat{\vartheta}_{2,n}$ and $[29.918, 35.640]$ for $\widetilde{\vartheta}_{2,n}$ (values in yr^{-1}). As expected because of efficiency, the latter is narrower than the former.

We plot the estimates of the two volatility processes in Figure 5. The bandwidths, chosen using cross validation, are equal to 14 days. These results clearly show non constant volatility processes. In both cases, the estimates of the volatility process $\widehat{\sigma}_{n,t}$ evolve in a same order of magnitude, although we can observe higher values of the estimate $\widehat{\sigma}_{n,t}$ from $\widehat{\vartheta}_{2,n}$, in the last 70 days. The difference between $\widehat{\vartheta}_{2,n}$ and $\widehat{\vartheta}_{3,n}$ has a higher impact on the estimation of $\widehat{\sigma}_{n,t}^2$. And, the high value of $\widehat{\vartheta}_{2,n}$ leads to huge values of $\widehat{\sigma}_{n,t}^2$ especially when $T_1 - t$ is large, due to the exponential term in the short term volatility, as explained in Section 3.2. However, the estimates of the global volatility function $e^{-2\vartheta(T_1-t)}\sigma^2$ are close in both cases, although taking lower values in the case of $\widehat{\vartheta}_{2,n}$. This experiment reveals, like on simulated data, the ill-posedness of the problem of estimation of the volatility paths, especially for σ when $\widehat{\vartheta}_{2,n}$ is large.

4. PROOFS

4.1. Preliminaries: localisation. With no loss of generality, we may (and will) assume that the processes b^j , σ and $\bar{\sigma}$ are bounded, relying on a so-called localisation argument. For an integer $p \geq 1$, introduce the stopping time $\tau_p = \inf\{t \in [0, T], \min(b_t^j, \sigma_t, \bar{\sigma}_t) > p\}$. Replacing X_t^j by $X_{t \wedge \tau_p}^j$, we have bounded processes b^j , σ and $\bar{\sigma}$. Moreover since these processes are at least locally bounded, we have $\mathbb{P}(\tau_p > T) \rightarrow 1$ as $p \rightarrow \infty$. We refer to Section 3.6.3 in [11] for details.

4.2. Proof of Theorem 1.

Proof of Theorem 1 (1). **Step 1.** We first consider the case $b^j = 0$ for $j = 1, 2$. For notational simplicity, we set $\mathfrak{e}_{\ell,k}(\vartheta) = e^{-\vartheta T_k} - e^{-\vartheta T_\ell}$. Let us define

$$\zeta_i^n = (\Delta_i^n X^2)^2 - (\Delta_i^n X^1)^2$$

and

$$\xi_i^n = (\Delta_i^n X^2 - \Delta_i^n X^1)^2.$$

Clearly

$$\begin{aligned} & \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_2-t)} \sigma_t dB_t + \int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t d\bar{B}_t \right)^2 - \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)} \sigma_t dB_t + \int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t d\bar{B}_t \right)^2 \\ &= (e^{-2\vartheta T_2} - e^{-2\vartheta T_1}) \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t} \sigma_t dB_t \right)^2 + 2\mathfrak{e}_{1,2}(\vartheta) \int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t} \sigma_t dB_t \int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t d\bar{B}_t, \end{aligned}$$

therefore, setting $\chi_i^n = 2\mathfrak{e}_{1,2}(\vartheta) \int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t} \sigma_t dB_t \int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t d\bar{B}_t$, we obtain the following representation

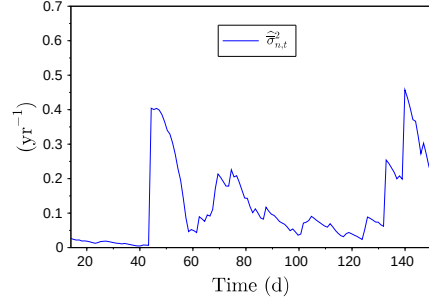
$$(7) \quad \zeta_i^n = \frac{1}{\psi_{T_1, T_2}(\vartheta)} \xi_i^n + \chi_i^n.$$

By standard convergence of the quadratic variation (see for instance Section 2.1.5 in [16]),

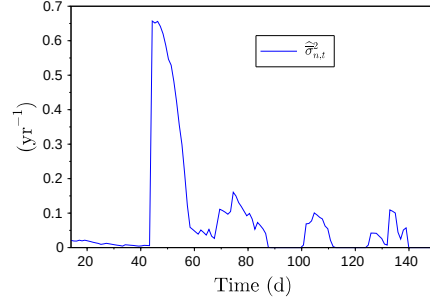
$$\sum_{i=1}^n \xi_i^n \rightarrow \mathfrak{e}_{1,2}(\vartheta)^2 \int_0^T e^{2\vartheta t} \sigma_t^2 dt$$

in probability. Note that the limit is almost surely positive by Assumption 1. Also, since B and $\bar{B}$ are independent, and since $\sigma_t^2 \leq M$ and $\bar{\sigma}_t^2 \leq M$ for some constant $M > 0$ by localization, we have that

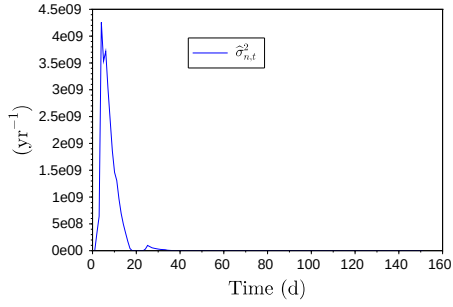
$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t} \sigma_t dB_t \int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t d\bar{B}_t \right)^2 \right] \\ &= \sum_{i=1}^n \mathbb{E} \left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t} \sigma_t dB_t \right)^2 \right] \mathbb{E} \left[\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t d\bar{B}_t \right)^2 \right] \leq \Delta_n e^{2\vartheta T} M^2 \rightarrow 0. \end{aligned}$$



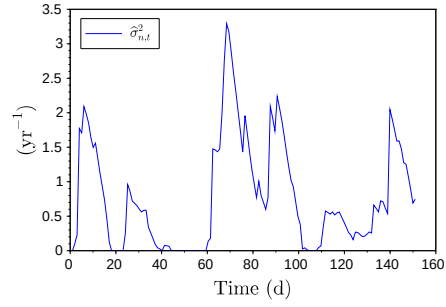
(A) Plot of $\hat{\sigma}_{n,t}^2$ obtained with the estimator $\hat{\vartheta}_{2,n}$



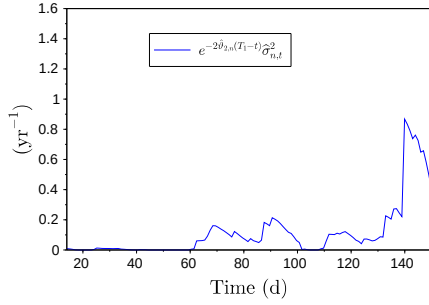
(B) Plot of $\hat{\sigma}_{n,t}^2$ obtained with the estimator $\hat{\vartheta}_{3,n}$



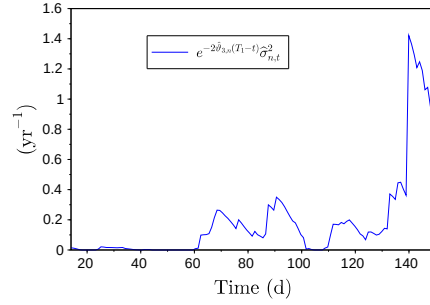
(C) Plot of $\hat{\sigma}_{n,t}^2$ obtained with the estimator $\hat{\vartheta}_{2,n}$



(D) Plot of $\hat{\sigma}_{n,t}^2$ obtained with the estimator $\hat{\vartheta}_{3,n}$



(E) Plot of $e^{-2\hat{\vartheta}_{2,n}(T_1-t)}\hat{\sigma}_{n,t}^2$ obtained with the estimator $\hat{\vartheta}_{2,n}$



(F) Plot of $e^{-2\hat{\vartheta}_{3,n}(T_1-t)}\hat{\sigma}_{n,t}^2$ obtained with the estimator $\hat{\vartheta}_{3,n}$

FIGURE 5. Plots of estimated paths of volatility processes. Left: using $\hat{\vartheta}_{2,n}$ as a preliminary estimator of ϑ . Right: using $\hat{\vartheta}_{3,n}$.

Therefore $\sum_{i=1}^n \zeta_i^n$ converges in probability as well, with the same limit as $\frac{1}{\psi_{T_1, T_2}(\vartheta)} \sum_{i=1}^n \xi_i^n$. It follows that

$$\Psi_{T_1, T_2}^n = \frac{\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2}{\sum_{i=1}^n (\Delta_i^n X^2)^2 - (\Delta_i^n X^1)^2} = \frac{\sum_{i=1}^n \xi_i^n}{\sum_{i=1}^n \zeta_i^n} \rightarrow \psi_{T_1, T_2}(\vartheta)$$

in probability. We derive the convergence

$$\psi_{T_1, T_2}(\widehat{\vartheta}_{2,n}) \rightarrow \psi_{T_1, T_2}(\vartheta)$$

in probability on the event $\{\Psi_{T_1, T_2}^n \in (-1, 0)\}$, hence the convergence $\widehat{\vartheta}_{2,n} \rightarrow \vartheta$ in probability as well since $\{\Psi_{T_1, T_2}^n \in (-1, 0)\}$ has asymptotically probability 1 and that $\vartheta \rightsquigarrow \psi_{T_1, T_2}(\vartheta)$ is invertible with continuous inverse.

Step 2. Using (7), we readily obtain

$$\Delta_n^{-1/2}(\Psi_{T_1, T_2}^n - \psi_{T_1, T_2}(\vartheta)) = \Delta_n^{-1/2} \left(\frac{\sum_{i=1}^n \xi_i^n}{\sum_{i=1}^n \zeta_i^n} - \psi_{T_1, T_2}(\vartheta) \right) = -\psi_{T_1, T_2}(\vartheta) \frac{\Delta_n^{-1/2} \sum_{i=1}^n \chi_i^n}{\sum_{i=1}^n \zeta_i^n}.$$

Consider next the sequence of 1-dimensional processes

$$\chi_n(t) = \Delta_n^{1/2} \sum_{i=1}^{\lfloor t \Delta_n^{-1} \rfloor} f(\Delta_n^{-1/2} \Delta_i^n Y^1, \Delta_n^{-1/2} \Delta_i^n Y^2),$$

where $Y_t = (Y_t^1, Y_t^2) = (\int_0^t e^{\vartheta s} \sigma_s dB_s, \int_0^t \bar{\sigma}_s d\bar{B}_s)$. By Theorem 3.21, p. 231 in [11] applied to the martingale Y with $f(x, y) = xy$ which has vanishing integral under the standard 2-dimensional-Gaussian measure, we have that the process $\chi_n(t)$ converges stably in law to a continuous process $\chi(t)$ defined on an extension of the original probability space and given by

$$\chi(t) = \int_0^t e^{\vartheta s} \sigma_s \bar{\sigma}_s dW_s,$$

where W is a Brownian motion independent of $\mathcal{F}$. Using successively $\Delta_n^{-1/2} \sum_{i=1}^n \chi_i^n = 2\mathfrak{e}_{1,2}(\vartheta) \chi_n(T)$, the fact that the convergence $\chi_n \rightarrow \chi$ holds stably in law and the convergence

$$\sum_{i=1}^n \zeta_i^n \rightarrow (e^{-2\vartheta T_2} - e^{-2\vartheta T_1}) \int_0^T e^{2\vartheta t} \sigma_t^2 dt,$$

in probability, we derive

$$\begin{aligned} -\Delta_n^{-1/2} \psi_{T_1, T_2}(\vartheta) \frac{\sum_{i=1}^n \chi_i^n}{\sum_{i=1}^n \zeta_i^n} &\rightarrow -\psi_{T_1, T_2}(\vartheta) \frac{2(e^{-\vartheta T_2} - e^{-\vartheta T_1}) \chi(T)}{(e^{-2\vartheta T_2} - e^{-2\vartheta T_1}) \int_0^T e^{2\vartheta t} \sigma_t^2 dt} \\ &= -\frac{2(e^{-\vartheta T_2} - e^{-\vartheta T_1})^3}{(e^{-2\vartheta T_2} - e^{-2\vartheta T_1})^2 \int_0^T e^{2\vartheta t} \sigma_t^2 dt} \chi(T) \end{aligned}$$

in distribution. Conditional on $\mathcal{F}$, the limiting variable is centred Gaussian, with conditional variance

$$v_\vartheta(\sigma, \bar{\sigma}) = 4 \frac{(e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 \int_0^T e^{2\vartheta t} \sigma_t^2 \bar{\sigma}_t^2 dt}{(e^{-\vartheta T_2} + e^{-\vartheta T_1})^4 (\int_0^T e^{2\vartheta t} \sigma_t^2 dt)^2}.$$

Step 3. On the event $\{\Psi_{T_1, T_2}^n \in (-1, 0)\}$, we have

$$\Delta_n^{-1/2}(\widehat{\vartheta}_{2,n} - \vartheta) = \Delta_n^{-1/2}(\Psi_{T_1, T_2}^n - \psi_{T_1, T_2}(\vartheta)) \partial_\vartheta \psi_{T_1, T_2}^{-1}(Z_n)$$

for some Z_n that converges to $\psi_{T_1, T_2}(\vartheta)$ in probability by Step 1. The conclusion follows from

$$(\partial_\vartheta \psi_{T_1, T_2}^{-1}(\psi_{T_1, T_2}(\vartheta)))^2 v_\vartheta(\sigma, \bar{\sigma}) = V_\vartheta(\sigma, \bar{\sigma})$$

together with the fact that $\{\Psi_{T_1, T_2}^n \in (-1, 0)\}$ has asymptotically probability 1.

Step 4. It remains to relax the restriction $b^j = 0$. When b^j is non-zero, by localization again, we may assume it is bounded. Moreover, the volatility processes $(\sigma, \bar{\sigma})$ are bounded below by Assumption 1. Then, by Girsanov theorem, we apply a change of measure which is $\mathcal{F}_T$ -measurable. Since the convergence in distribution in Step 2 holds stably in law, we may work under this change of measure (see Section 2.4.4 in [16] for a simple explanation). Finally, relaxing the boundedness assumption on $\sigma, \bar{\sigma}$ and b^j is standard, see Section 4.1 above.

Proof of Theorem 1 (2). Step 1. We have

$$\Psi_{T_1, T_2, T_3}^n = \frac{\sum_{i=1}^n (\Delta_i^n (X^3 - X^2))^2}{\sum_{i=1}^n (\Delta_i^n (X^2 - X^1))^2}$$

By standard convergence of the quadratic variation

$$(8) \quad \begin{aligned} \sum_{i=1}^n (\Delta_i^n (X^2 - X^1))^2 &\rightarrow \mathfrak{e}_{1,2}(\vartheta)^2 \int_0^T e^{2\vartheta t} \sigma_t^2 dt, \\ \sum_{i=1}^n (\Delta_i^n (X^3 - X^2))^2 &\rightarrow \mathfrak{e}_{2,3}(\vartheta)^2 \int_0^T e^{2\vartheta t} \sigma_t^2 dt \end{aligned}$$

in probability. Since $\psi_{T_1, T_2, T_3}(\vartheta) = \frac{\mathfrak{e}_{2,3}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}$, we derive $\psi_{T_1, T_2, T_3}(\widehat{\vartheta}_{n,3}) \rightarrow \psi_{T_1, T_2, T_3}(\vartheta)$ in probability on the event $\{\Psi_{T_1, T_2, T_3} \in (0, (\frac{T_3 - T_2}{T_2 - T_1})^2)\}$ which has asymptotically probability 1, hence the convergence $\widehat{\vartheta}_{n,3} \rightarrow \vartheta$ in probability.

Step 2. We further have

$$\Psi_{T_1, T_2, T_3}^n - \psi_{T_1, T_2, T_3}(\vartheta) = \frac{\sum_{i=1}^n (\Delta_i^n (X^3 - X^2))^2}{\sum_{i=1}^n (\Delta_i^n (X^2 - X^1))^2} - \frac{\mathfrak{e}_{2,3}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2} = \frac{\sum_{i=1}^n \eta_i^n}{\sum_{i=1}^n (\Delta_i^n (X^2 - X^1))^2},$$

with

$$\eta_i^n = (\Delta_i^n (X^3 - X^2))^2 - \frac{\mathfrak{e}_{2,3}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2} (\Delta_i^n (X^2 - X^1))^2.$$

Write $\overline{\Delta}_i^n f = \int_{(i-1)\Delta_n}^{i\Delta_n} f(t) dt$. One readily checks that the following decomposition holds: $\eta_i^n = (\eta')_i^n + (\eta'')_i^n$, with

$$(\eta')_i^n = (\overline{\Delta}_i^n (b^3 - b^2))^2 - \frac{\mathfrak{e}_{2,3}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2} (\overline{\Delta}_i^n (b^2 - b^1))^2$$

and

$$(\eta'')_i^n = 2\mathfrak{e}_{2,3}(\vartheta) \left(\overline{\Delta}_i^n ((b^3 - b^2) - \frac{\mathfrak{e}_{2,3}(\vartheta)}{\mathfrak{e}_{1,2}(\vartheta)} (b^2 - b^1)) \right) \int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t} \sigma_t dB_t.$$

We will need the following lemma, proof of which is relatively straightforward yet technical and given in Section 5.1.

Lemma 1. *Let $(Y_t)_{t \geq 0}$ and $(Z_t)_{t \geq 0}$ be two càdlàg and progressively measurable processes. Assume that for some $s > 1/2$, we have $\sup_{t \in [0, T]} t^{-s} \omega(Y)_t < \infty$. Then*

$$\Delta_n^{-1} \sum_{i=1}^n (\overline{\Delta}_i^n Y)^2 \rightarrow \int_0^T Y_t^2 dt$$

and

$$\Delta_n^{-1} \sum_{i=1}^n \overline{\Delta}_i^n(Y) \int_{(i-1)\Delta_n}^{i\Delta_n} Z_t dB_t \rightarrow \int_0^T Y_t Z_t dB_t$$

in probability.

We successively have

$$\Delta_n^{-1} \sum_{i=1}^n (\eta')_i^n \rightarrow \int_0^T \mu_\vartheta(b_t) dt$$

with $\mu_\vartheta(b_t) = (b_t^3 - b_t^2)^2 - \frac{\mathfrak{e}_{2,3}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2} (b_t^2 - b_t^1)^2$ and

$$\Delta_n^{-1} \sum_{i=1}^n (\eta'')_i^n \rightarrow 2 \int_0^T \lambda_\vartheta(b_t) e^{\vartheta t} \sigma_t dB_t$$

in probability, by Lemma 1 applied to $Y_t = (b_t^3 - b_t^2) - \frac{\mathfrak{e}_{2,3}(\vartheta)}{\mathfrak{e}_{1,2}(\vartheta)}(b_t^2 - b_t^1)$ and $Z_t = e^{\vartheta t}\sigma_t$, and Assumption 1, where $\lambda_\vartheta(b_t) = \mathfrak{e}_{2,3}(\vartheta)Y_t$. This, together with (8), implies the convergence

$$\Delta_n^{-1}(\Psi_{T_1, T_2, T_3}^n - \psi_{T_1, T_2, T_3}(\vartheta)) \rightarrow \frac{\int_0^T \mu_\vartheta(b_t)dt + 2 \int_0^T \lambda_\vartheta(b_t)e^{\vartheta t}\sigma_t dB_t}{\mathfrak{e}_{1,2}(\vartheta)^2 \int_0^T e^{2\vartheta t}\sigma_t^2 dt}$$

in probability.

Step 3. Finally, we have

$$\Delta_n^{-1}(\widehat{\vartheta}_{3,n} - \vartheta) = \Delta_n^{-1}(\Psi_{T_1, T_2, T_3}^n - \psi_{T_1, T_2, T_3}(\vartheta))\partial_\vartheta \psi_{T_1, T_2, T_3}^{-1}(Z_n),$$

for some Z_n that converges to $\psi_{T_1, T_2, T_3}(\vartheta)$ by Step 1. Hence

$$\Delta_n^{-1}(\widehat{\vartheta}_{3,n} - \vartheta) \rightarrow \frac{\int_0^T \mu_\vartheta(b_t)dt + 2 \int_0^T \lambda_\vartheta(b_t)e^{\vartheta t}\sigma_t dB_t}{\partial_\vartheta \psi_{T_1, T_2, T_3}(\vartheta)\mathfrak{e}_{1,2}(\vartheta)^2 \int_0^T e^{2\vartheta t}\sigma_t^2 dt}$$

and we conclude by noting that $\partial_\vartheta \psi_{T_1, T_2, T_3}(\vartheta) = \frac{2D_3}{\mathfrak{e}_{1,2}(\vartheta)^3}$.

4.3. Proof of Proposition 1. We may assume that $b^j = 0$, the case $b^j \neq 0$ being obtained exactly in the same line as in Step 4 of the proof of Theorem 1 (1). For ease of notation, we write $\widehat{\vartheta}_{2,n}$ for $\max(\widehat{\vartheta}_{2,n}, \varpi_n)$ and set $t_i = i\Delta_n$ for $i = 1, \dots, n$. We also define $K(t) = \mathbf{1}_{(0,1]}(t)$ and $K_h(t) = h^{-1}K(th^{-1})$ for $h > 0$. We have

$$\widehat{\sigma}_{n,t}^2 - \sigma_t^2 = \frac{\sum_{i=1}^n K_{h_n}(t - t_{i-1})((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2)}{e^{-2\widehat{\vartheta}_{2,n}(T_1-t)} - e^{-2\widehat{\vartheta}_{2,n}(T_2-t)}} - \sigma_t^2 = I + II,$$

with

$$I = \left(\frac{1}{e^{-2\widehat{\vartheta}_{2,n}(T_1-t)} - e^{-2\widehat{\vartheta}_{2,n}(T_2-t)}} - \frac{1}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}} \right) \times \sum_{i=1}^n K_{h_n}(t - t_{i-1})((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2)$$

and

$$II = \frac{\sum_{i=1}^n K_{h_n}(t - t_{i-1})((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2)}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}} - \sigma_t^2.$$

Step 1. Since $\mathbb{E}[(\Delta_i^n X^j)^2]$ is of order Δ_n by Burkholder-Davis-Gundy inequality, we have that $\mathbb{E}[|(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2|]$ is of order Δ_n as well and therefore

$$\mathbb{E}\left[\left|\sum_{i=1}^n K_{h_n}(t - t_{i-1})((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2)\right|\right] \lesssim \sum_{i=1}^n K_{h_n}(t - t_{i-1})\Delta_n \lesssim 1$$

since $K_{h_n}(t - t_{i-1})$ is of order h_n^{-1} for a number of terms that are at most of order $\Delta_n^{-1}h_n$. Therefore $\sum_{i=1}^n K_{h_n}(t - t_{i-1})((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2)$ is tight, and we conclude that I is of order $\Delta_n^{1/2}$ in probability by applying Theorem 1 (1).

Step 2. The term II further splits into $II = (e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)})^{-1}(B_n(t) + V_n(t))$, having

$$V_n(t) = \sum_{i=1}^n K_{h_n}(t - t_{i-1})((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - \mathbb{E}[(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 | \mathcal{F}_{i-1}])$$

and

$$B_n(t) = \sum_{i=1}^n \mathbb{E}[K_{h_n}(t - t_{i-1})((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2) | \mathcal{F}_{i-1}] - (e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)})\sigma_t^2.$$

Hereafter, we abbreviate $\mathcal{F}_{i\Delta_n}$ by $\mathcal{F}_i$.

Step 3. We first prove an upper bound for $\mathbb{E}[V_n(t)^2]$. We have

$$\begin{aligned} & \sup_{t \in [h_n, T]} \mathbb{E} \left[\left(\sum_{i=1}^n K_{h_n}(t - t_{i-1}) ((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - \mathbb{E}[(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 | \mathcal{F}_{i-1}]) \right)^2 \right] \\ &= \sup_{t \in [h_n, T]} h_n^{-2} \sum_{i=1}^n K \left(\frac{t - t_{i-1}}{h_n} \right)^2 \mathbb{E} \left[((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - \mathbb{E}[(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 | \mathcal{F}_{i-1}])^2 \right] \end{aligned}$$

because cross-terms in the development are zero due to conditioning. By compactness of the support of K , there are at most of order $\Delta_n^{-1} h_n$ nonvanishing terms in the sum and the estimate is uniform in $t \in [h_n, T]$. Finally, since

$$\mathbb{E} \left[((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - \mathbb{E}[(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 | \mathcal{F}_{i-1}])^2 \right] \lesssim \Delta_n^2,$$

we obtain $\sup_{t \in [h_n, T]} \mathbb{E}[V_n(t)^2] \lesssim \Delta_n h_n^{-1}$.

Step 4. In order to bound the bias we use the decomposition

$$B_n(t) = (e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)})(III + IV),$$

where

$$III = \int_0^T h_n^{-1} K \left(\frac{t-u}{h_n} \right) e^{-2\vartheta(t-u)} \sigma_u^2 du - \sigma_t^2$$

and

$$IV = \frac{\sum_{i=1}^n \mathbb{E} \left[h_n^{-1} K \left(\frac{t-t_{i-1}}{h_n} \right) ((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2) \middle| \mathcal{F}_{i-1} \right]}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}} - \int_0^T h_n^{-1} K \left(\frac{t-u}{h_n} \right) e^{-2\vartheta(t-u)} \sigma_u^2 du.$$

For every $t \in [h_n, T]$ we have $\int_{\frac{t-T}{h_n}}^{\frac{t}{h_n}} K(x) dx = 1$ hence

$$\begin{aligned} \mathbb{E}[III^2] &= \mathbb{E} \left[\left(\int_{\frac{t-T}{h_n}}^{\frac{t}{h_n}} K(x) e^{-2\vartheta h_n x} \sigma_{t-h_n x}^2 dx - \sigma_t^2 \right)^2 \right] \\ &= \mathbb{E} \left[\left(\int_{\text{supp}(K)} K(x) (e^{-2\vartheta h_n x} \sigma_{t-h_n x}^2 - \sigma_t^2) dx \right)^2 \right] \\ &\leq \int_{\text{supp}(K)} K(x)^2 \mathbb{E}[(e^{-2\vartheta h_n x} \sigma_{t-h_n x}^2 - \sigma_t^2)^2] dx \end{aligned}$$

by Jensen inequality since $\text{supp}(K) \subset [\frac{t-T}{h}, \frac{t}{h}]$ and since $\int_0^T K(s) ds = 1$. By convexity,

$$(e^{-2\vartheta h_n x} \sigma_{t-h_n x}^2 - \sigma_t^2)^2 \leq 2(e^{-2\vartheta h_n x} \sigma_{t-h_n x}^2 - \sigma_{t-h_n x}^2)^2 + 2(\sigma_{t-h_n x}^2 - \sigma_t^2)^2$$

follows. Bounding further the remainder in the expansion of $x \rightsquigarrow e^{-2\vartheta h_n x}$ at the point 0, we obtain $|e^{-2\vartheta h_n x} - 1| \leq M|2\vartheta h_n x|$ for some $M > 0$. By localization, we find some $M_\sigma > 0$ such that $\sigma_t < M_\sigma$. It follows that

$$\begin{aligned} \mathbb{E}[III^2] &\leq \int_{\text{supp}(K)} K^2(x) (2M_\sigma^4 (2\vartheta h_n x M)^2 + 2\mathbb{E}((\sigma_{t-h_n x}^2 - \sigma_t^2)^2)) dx \\ &\leq \int_{\text{supp}(K)} K^2(x) (2M_\sigma^4 (2\vartheta h_n x M)^2 + 2c|h_n x|^{2\alpha}) dx, \end{aligned}$$

using Assumption 2. This estimate is uniform in $t \in [h_n, T]$, therefore $\sup_{t \in [h_n, T]} \mathbb{E}[III^2] \lesssim h_n^{2\alpha}$. Next, we write $IV = \sum_{i=1}^n h_n^{-1} \delta_i(t)$, with

$$\delta_i(t) = \mathbb{E} \left(K \left(\frac{t - t_{i-1}}{h_n} \right) \int_{t_{i-1}}^{t_i} e^{-2\vartheta(t-u)} \sigma_u^2 du \middle| \mathcal{F}_{i-1} \right) - \int_{t_{i-1}}^{t_i} K \left(\frac{t-u}{h_n} \right) e^{-2\vartheta(t-u)} \sigma_u^2 du,$$

since the term

$$\begin{aligned} & \sum_{i=1}^n h_n^{-1} K\left(\frac{t-t_{i-1}}{h_n}\right) \mathbb{E}\left[\frac{((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2)}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}} - \int_{t_{i-1}}^{t_i} e^{-2\vartheta(t-u)} \sigma_u^2 du \middle| \mathcal{F}_{i-1}\right] \\ &= \sum_{i=1}^n h_n^{-1} K\left(\frac{t-t_{i-1}}{h_n}\right) e^{-2\vartheta t} \mathbb{E}\left[\left(\int_{t_{i-1}}^{t_i} e^{\vartheta u} \sigma_u dB_u\right)^2 - \int_{t_{i-1}}^{t_i} e^{2\vartheta u} \sigma_u^2 du \middle| \mathcal{F}_{i-1}\right] = 0 \end{aligned}$$

vanishes. Therefore, writing $\mathbb{E}[IV^2] = \sum_{i=1}^n \mathbb{E}(h_n^{-2}(\delta_i(t))^2) + 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[h_n^{-2} \delta_i(t) \delta_j(t)]$, we successively obtain

$$\mathbb{E}[h_n^{-2}(\delta_i^S(t))^2] \leq \frac{\Delta_n}{h_n^2} \int_{t_{i-1}}^{t_i} e^{-4\vartheta(t-u)} \mathbb{E}\left[\left(K\left(\frac{t-t_{i-1}}{h_n}\right) \mathbb{E}[\sigma_u^2 | \mathcal{F}_{i-1}] - K\left(\frac{t-u}{h_n}\right) \sigma_u^2\right)^2\right] du$$

by Jensen inequality, so that $\sum_{i=1}^n \mathbb{E}[h_n^{-2}(\delta_i^S(t))^2] \lesssim \Delta_n h_n^{-1}$ uniformly in t , since there are at most of order $\Delta_n h_n$ nonvanishing terms in the sum. Finally, conditioning on $\mathcal{F}_{j-1}$,

$$\mathbb{E}[h_n^{-2} \delta_i(t) \delta_j(t)] = h_n^{-2} \mathbb{E}\left[\delta_i(t) \mathbb{E}\left(\int_{t_{j-1}}^{t_j} \left(K\left(\frac{t-t_{j-1}}{h_n}\right) - K\left(\frac{t-u}{h_n}\right)\right) e^{-2\vartheta(t-u)} \sigma_u^2 du \middle| \mathcal{F}_{j-1}\right)\right],$$

and the difference $K\left(\frac{t-t_{j-1}}{h_n}\right) - K\left(\frac{t-u}{h_n}\right)$ is non-zero only if $t \in (t_{j-1}, u]$ or $t \in (t_{j-1} + h_n, u + h_n]$, which can be the case for j in some set $\mathcal{J}_t$, which contains at most three indexes. Therefore,

$$\left| \sum_{1 \leq i < j \leq n} \mathbb{E}[h_n^{-2} \delta_i(t) \delta_j(t)] \right| = \left| \sum_{i=1}^{n-1} \sum_{j \in \mathcal{J}_t} \mathbb{E}[h_n^{-2} \delta_i(t) \delta_j(t)] \right| \leq 3h_n^{-2} \sum_{i=1}^{n-1} \mathbb{E}[|\delta_i(t)|] M_\sigma^2 e^{2\vartheta T} \Delta_n,$$

which is of order $\Delta_n h_n^{-1}$. We infer $\sup_{t \in [h_n, T]} \mathbb{E}[IV^2] \lesssim \Delta_n h_n^{-1}$.

Step 5. From the estimates established in Steps 3. and 4. we derive

$$\sup_{t \in [h_n, T]} \mathbb{E}[(V_n(t))^2] \lesssim \Delta_n h_n^{-1}, \quad \sup_{t \in [h_n, T]} \mathbb{E}[B_n(t)^2] \lesssim h_n^{2\alpha} + \Delta_n h_n^{-1}.$$

The choice $h_n = \Delta_n^{1/(2\alpha+1)}$ implies that the two error terms $h_n^{2\alpha}$ and $\Delta_n h_n^{-1}$ are of the same order, namely $\Delta_n^{2\alpha/(2\alpha+1)}$, which ends the proof concerning σ .

Step 6. The proof is the same for $\hat{\sigma}^2$. We split $\hat{\sigma}_{n,t}^2 - \bar{\sigma}_t^2$ as follows

$$\begin{aligned} \hat{\sigma}_{n,t}^2 - \bar{\sigma}_t^2 &= \left(\frac{e^{-2\hat{\vartheta}_{2,n}T_2}}{e^{-2\hat{\vartheta}_{2,n}T_1} - e^{-2\hat{\vartheta}_{2,n}T_2}} - \frac{e^{-2\vartheta T_2}}{e^{-2\vartheta T_1} - e^{-2\vartheta T_2}} \right) \sum_{i=1}^n h_n^{-1} K\left(\frac{t-t_{i-1}}{h_n}\right) (\Delta_i^n X^1)^2 \\ &\quad + \left(\frac{e^{-2\hat{\vartheta}_{2,n}T_1}}{e^{-2\hat{\vartheta}_{2,n}T_1} - e^{-2\hat{\vartheta}_{2,n}T_2}} - \frac{e^{-2\vartheta T_1}}{e^{-2\vartheta T_1} - e^{-2\vartheta T_2}} \right) \sum_{i=1}^n h_n^{-1} K\left(\frac{t-t_{i-1}}{h_n}\right) (\Delta_i^n X^2)^2 \\ &\quad + \frac{\sum_{i=1}^n h_n^{-1} K\left(\frac{t-t_{i-1}}{h_n}\right) (e^{-2\vartheta(T_1-t)} (\Delta_i^n X^2)^2 - e^{-2\vartheta(T_2-t)} (\Delta_i^n X^1)^2)}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}} - \bar{\sigma}_t^2 \end{aligned}$$

and proceed analogously. The proof of Theorem 1 is complete.

4.4. Proof of Theorem 2.

Preliminaries on efficient semiparametric estimation. We refer to Sections 25.3–25.4 of [17] for a comprehensive presentation of efficient semiparametric estimation, that we need to adapt to our framework. Assuming $(\sigma, \bar{\sigma})$ to be deterministic, the data $(\Delta_i^n X^1, \Delta_i^n X^2), i = 1, \dots, n$ extracted from (5) generate a product experiment $\mathcal{E}^n = \otimes_{i=1}^n \mathcal{P}^{i,n}$, where

$$\mathcal{P}^{i,n} = \{f_{\vartheta, \sigma, \bar{\sigma}}^{i,n}, (\vartheta, \sigma, \bar{\sigma}) \in [0, \infty) \times \Sigma(c, \tilde{c})\},$$

where $f_{\vartheta, \sigma, \bar{\sigma}}^{i, n}$ is the density on $\mathbb{R}^2$ of the Gaussian vector $(\Delta_i^n X^1, \Delta_i^n X^2)$, see Section 2.3. Let $\varepsilon > 0$ considered as a small perturbation parameter. For every $u \in [0, \varepsilon]$, let us be given moreover two regular functions $t \rightsquigarrow \eta_t^u$ and $t \rightsquigarrow \bar{\eta}_t^u$ such that

$$(9) \quad \tilde{c} < \inf_{t \in [0, T], u \in [0, \varepsilon]} \min(\eta_t^u, \bar{\eta}_t^u) \leq \sup_{t \in [0, T], u \in [0, \varepsilon]} \max(\eta_t^u, \bar{\eta}_t^u) < c.$$

We call $(t \rightsquigarrow (\eta_t^u, \bar{\eta}_t^u))_{0 \leq u \leq \varepsilon}$ a perturbation path if we have (9) together with the property

$$(\eta_t^0, \bar{\eta}_t^0) = (\sigma_t, \bar{\sigma}_t) \text{ for every } t \in [0, T].$$

Let ϑ and $\iota \in \mathbb{R}$ be such that $\vartheta + \iota\varepsilon > 0$. Remember that $f_{\vartheta, \sigma, \bar{\sigma}}^{i, n}$ is the density on $\mathbb{R}^2$ of the Gaussian vector $(\Delta_i^n X^1, \Delta_i^n X^2)$ extracted from (5). We obtain a parametric submodel of $\mathcal{P}^{i, n}$ around $(t \rightsquigarrow (\sigma_t^2, \bar{\sigma}_t^2))$ by setting

$$\mathcal{P}_0^{i, n} = \{f_{\vartheta + \iota u, \eta^u, \bar{\eta}^u}^{i, n}, 0 \leq u \leq \varepsilon\},$$

noting that $\mathcal{P}_0^{i, n}$ passes through the true distribution at $u = 0$. We consider only submodels that are differentiable in quadratic mean at $u = 0$, with score function $g_{i, \iota, \eta, \bar{\eta}} \in L^2(\mathbb{P}_{\vartheta, \sigma, \bar{\sigma}})$. If we let $\mathcal{P}_0^{i, n}$ range over all admissible submodels as $(\eta, \bar{\eta})$ varies among perturbation paths, we obtain a collection of score functions that define in turn the tangent set $\dot{\mathcal{P}}_{\vartheta, \sigma, \bar{\sigma}}^{i, n}$ of the model $\mathcal{P}^{i, n}$ at the true distribution. Any score function $g_{i, \eta, \bar{\eta}}^{i, n} \in \dot{\mathcal{P}}_{\vartheta, \sigma, \bar{\sigma}}^{i, n}$ admits the representation

$$(10) \quad g_{i, \eta, \bar{\eta}}^{i, n} = \ell_{\vartheta, \sigma, \bar{\sigma}}^{i, n} + g_{\eta, \bar{\eta}}^{i, n},$$

where $\ell_{\vartheta, \sigma, \bar{\sigma}}^{i, n}$ is the score function of the original model defined in (6) when σ and $\bar{\sigma}$ are known, and $g_{\eta, \bar{\eta}}^{i, n}$ is the score function obtained from a parametric submodel at $\iota = 0$, to be interpreted as the score relative to the nuisance parameter, while $\ell_{\vartheta, \sigma, \bar{\sigma}}^{i, n}$ corresponds to the score relative to the parameter of interest ϑ .

Completion of proof of Theorem 2. From (6), and the explicit representation

$$f_{\vartheta, \sigma, \bar{\sigma}}^{i, n}(x, y) = \frac{\exp\left(-\frac{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 dt\right)(x-y)^2 + \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 dt\right)\left(y - e^{-\vartheta(T_2-T_1)}x\right)^2}{2\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 dt\right)\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 dt\right)\left(1 - e^{-\vartheta(T_2-T_1)}\right)^2}\right)}{2\pi\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 dt\right)^{1/2}\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 dt\right)^{1/2}\left(1 - e^{-\vartheta(T_2-T_1)}\right)},$$

we derive

$$\begin{aligned} \partial_{\vartheta} \log f_{\vartheta, \sigma, \bar{\sigma}}^{i, n}(\Delta_i^n X^1, \Delta_i^n X^2) &= \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} (T_1 - t) e^{-2\vartheta(T_1-t)} \sigma_t^2 dt}{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 dt} - \frac{e^{-\vartheta(T_2-T_1)}(T_2 - T_1)}{1 - e^{-\vartheta(T_2-T_1)}} \\ &\quad + (\Delta_i^n X^2 - \Delta_i^n X^1)^2 \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} (e^{-\vartheta(T_2-T_1)}(T_2 - t) - (T_1 - t)) e^{-2\vartheta(T_1-t)} \sigma_t^2 dt}{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 dt\right)^2 (1 - e^{-\vartheta(T_2-T_1)})^3} \\ &\quad + (\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)} \Delta_i^n X^1)^2 \frac{(T_2 - T_1) e^{-\vartheta(T_2-T_1)}}{\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 dt (1 - e^{-\vartheta(T_2-T_1)})^3} \\ &\quad - \Delta_i^n X^1 (\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)} \Delta_i^n X^1) \frac{e^{-\vartheta(T_2-T_1)}(T_2 - T_1)}{\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 dt (1 - e^{-\vartheta(T_2-T_1)})^2}. \end{aligned}$$

We pick a path $(\eta_t^u, \bar{\eta}_t^u)_{0 \leq u \leq \varepsilon}$ of the form $\eta_t^u = (1 + u\bar{k}(t))\sigma_t$ and $\bar{\eta}_t^u = (1 + u\bar{k}(t))\bar{\sigma}_t$ so that $(\eta^u, \bar{\eta}^u) \in \Sigma(c, \tilde{c})$. The submodel is differentiable in quadratic mean at $u = 0$, with score function having representation

$$g_{i, k, \bar{k}}^{i, n} = \ell_{\vartheta, \sigma, \bar{\sigma}}^{i, n} + g_{k, \bar{k}}^{i, n}.$$

according to (10) and parametrised by $(k, \bar{k})$. Formally,

$$\ell_{\vartheta, \sigma, \bar{\sigma}}^{i, n} = \left(\frac{d}{du} \log f_{\vartheta + \iota u, \sigma, \bar{\sigma}}^{i, n}\right)_{u=0} \text{ and } g_{k, \bar{k}}^{i, n} = \left(\frac{d}{du} \log f_{\vartheta, \eta^u, \bar{\eta}^u}^{i, n}\right)_{u=0},$$

so that

$$\begin{aligned}
g_{k,\bar{k}}^{i,n} = & -\frac{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 k(t) dt}{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 dt} - \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 \bar{k}(t) dt}{\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 dt} \\
& + (\Delta_i^n X^2 - \Delta_i^n X^1)^2 \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 k(t) dt}{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 dt\right)^2 (1 - e^{-\vartheta(T_2-T_1)})^2} \\
& + (\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)} \Delta_i^n X^1)^2 \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 \bar{k}(t) dt}{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 dt\right)^2 (1 - e^{-\vartheta(T_2-T_1)})^2}.
\end{aligned}$$

Introduce the orthogonal projection Π onto (the closure of) of $\text{Span}\{g_{k,\bar{k}}^{i,n}, \text{ for all admissible } (k, \bar{k})\}$. Then $\tilde{\ell}_{\vartheta,\sigma,\bar{\sigma}}^{i,n} = \ell_{\vartheta,\sigma,\bar{\sigma}}^{i,n} - \Pi \ell_{\vartheta,\sigma,\bar{\sigma}}^{i,n}$ is the efficient score for ϑ and $\tilde{I}_{\vartheta,\sigma,\bar{\sigma}}^{i,n} = \mathbb{E}_{\vartheta,\sigma,\bar{\sigma}}[(\tilde{\ell}_{\vartheta,\sigma,\bar{\sigma}}^{i,n})^2]$ is the best achievable information bound, see Sections 25.3–25.4 of [17] for details. By orthogonality,

$$\mathbb{E}_{\vartheta,\sigma,\bar{\sigma}}[(\ell_{\vartheta,\sigma,\bar{\sigma}}^{i,n} - \Pi \ell_{\vartheta,\sigma,\bar{\sigma}}^{i,n}) g_{k,\bar{k}}^{i,n}] = 0, \text{ for all admissible } (k, \bar{k}),$$

and anticipating further the representation $\Pi \ell_{\vartheta,\sigma,\bar{\sigma}}^{i,n} = g_{k^*,\bar{k}^*}^{i,n}$ for some admissible $(k^*, \bar{k}^*)$, it suffices to solve

$$0 = \mathbb{E}_{\vartheta,\sigma,\bar{\sigma}}[(\ell_{\vartheta,\sigma,\bar{\sigma}}^{i,n} - g_{k^*,\bar{k}^*}^{i,n}) g_{k,\bar{k}}^{i,n}] \text{ for all admissible } (k, \bar{k}).$$

Elementary computations yield $k^*(t) = \frac{(T_2-t)e^{-\vartheta(T_2-T_1)}t(T_1-t)}{1-e^{-\vartheta(T_2-T_1)}}$ and $\bar{k}^*(t) = 0$. We conclude

$$\tilde{\ell}_{\vartheta,\sigma,\bar{\sigma}}^{i,n} = \tilde{\ell}_{\vartheta,\bar{\sigma}}^{i,n} = \frac{(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)} \Delta_i^n X^1) e^{-\vartheta(T_2-T_1)} (T_2 - T_1)}{(1 - e^{-\vartheta(T_2-T_1)})^3 \int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 dt}$$

and is independent of σ . Furthermore, the best achievable information bound becomes

$$\tilde{I}_{\vartheta,\sigma,\bar{\sigma}}^{i,n} = \frac{(T_2 - T_1)^2}{(e^{\vartheta(T_2-T_1)} - 1)^2} \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} \sigma_t^2 dt}{\int_{(i-1)\Delta_n}^{i\Delta_n} \bar{\sigma}_t^2 dt}.$$

By independence of the increments $(\Delta_i^n X^1, \Delta_i^n X^2)$, it remains to piece together the results for each $\mathcal{P}^{i,n}$ using the product structure of $\mathcal{E}^n$. We find the asymptotically equivalent bound

$$\sum_{i=1}^n \tilde{I}_{\vartheta,\sigma,\bar{\sigma}}^{i,n} \sim n \frac{(T_2 - T_1)^2}{T(e^{\vartheta(T_2-T_1)} - 1)^2} \int_0^T \frac{e^{-2\vartheta(T_1-t)} \sigma_t^2 dt}{\bar{\sigma}_t^2}$$

as $n \rightarrow \infty$. Taking the inverse and dividing by Δ_n we obtain the desired bound. The proof of Theorem 2 is complete.

4.5. Proof of Theorem 3. The first assertion was obtained in Section 4.4 in the course of the proof of Theorem 2. With no loss of generality, we may (and will) assume that $b^j = 0$, the case $b^j \neq 0$ being obtained exactly in the same line as in Step 4 of the proof of Theorem 1 (1). We also assume with no loss of generality that $\sup_t \max(\sigma_t, \bar{\sigma}_t) \leq M$. We further abbreviate $i\Delta_n$ by t_i .

Step 1. Let ϑ_n be a deterministic sequence such that $\sqrt{n}(\vartheta_n - \vartheta)$ is bounded. We first prove

$$(11) \quad \Delta_n^{1/2} \sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\vartheta_n, \bar{\sigma}}^{i,n} - \tilde{\ell}_{\vartheta_n, \hat{\sigma}_n}^{i,n}) \rightarrow 0$$

in probability, as $n \rightarrow \infty$. We have

$$\begin{aligned}
& \Delta_n^{1/2} \sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\vartheta_n, \bar{\sigma}}^{i,n} - \tilde{\ell}_{\vartheta_n, \hat{\sigma}_n}^{i,n}) \\
&= \Delta_n^{1/2} \sum_{i \in \mathcal{I}_n} (\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1) \left(\frac{1}{\int_{t_{i-1}}^{t_i} \bar{\sigma}_t^2 dt} - \frac{1}{\Delta_n \hat{\sigma}_{n,t_{i-1}}^2} \right) \\
&= I + II,
\end{aligned}$$

with

$$I = \Delta_n^{-1/2} \sum_{i \in \mathcal{I}_n} (\Delta_i^n X^2 - \Delta_i^n X^1) (\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1) \frac{\Delta_n \bar{\sigma}_{t_{i-1}}^2 - \int_{t_{i-1}}^{t_i} \bar{\sigma}_t^2 dt}{\bar{\sigma}_{t_{i-1}}^2 \int_{t_{i-1}}^{t_i} \bar{\sigma}_t^2 dt},$$

$$II = \Delta_n^{-1/2} \sum_{i \in \mathcal{I}_n} (\Delta_i^n X^2 - \Delta_i^n X^1) (\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1) \frac{\hat{\sigma}_{n,t_{i-1}}^2 - \bar{\sigma}_{n,t_{i-1}}^2}{\bar{\sigma}_{t_{i-1}}^2 \hat{\sigma}_{n,t_{i-1}}^2}.$$

For the term I , by Cauchy-Schwarz inequality and the fact that $\mathbb{P}((\sigma, \bar{\sigma}) \in \Sigma(c, \tilde{c})) = 1$, we have

$$\mathbb{E} \left[\left| \Delta_n^{-1/2} (\Delta_i^n X^2 - \Delta_i^n X^1) (\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1) \frac{\Delta_n \bar{\sigma}_{t_{i-1}}^2 - \int_{t_{i-1}}^{t_i} \bar{\sigma}_t^2 dt}{\bar{\sigma}_{t_{i-1}}^2 \int_{t_{i-1}}^{t_i} \bar{\sigma}_t^2 dt} \right| \right]$$

$$\leq \Delta_n^{-3/2} \tilde{c}^{-4} \mathbb{E} \left[\left((\Delta_i^n X^2 - \Delta_i^n X^1) (\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1) \right)^2 \right]^{1/2} \mathbb{E} \left[\left(\Delta_n \bar{\sigma}_{t_{i-1}}^2 - \int_{t_{i-1}}^{t_i} \bar{\sigma}_t^2 dt \right)^2 \right]^{1/2}.$$

Combining Cauchy-Schwarz and Burkholder-Davis-Gundy inequalities and the smoothness Assumption 2 we successively obtain

$$\mathbb{E} \left[\left((\Delta_i^n X^2 - \Delta_i^n X^1) (\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1) \right)^2 \right] \lesssim \Delta_n^2,$$

$$\mathbb{E} \left[\left(\Delta_n \bar{\sigma}_{t_{i-1}}^2 - \int_{t_{i-1}}^{t_i} \bar{\sigma}_t^2 dt \right)^2 \right] \lesssim \Delta_n^{2(1+\alpha)}.$$

We infer $\mathbb{E}[|I|] \lesssim \sum_{i \in \mathcal{I}_n} \Delta_n^{-3/2} \Delta_n \Delta_n^{1+\alpha} \lesssim \Delta_n^{\alpha-1/2} \rightarrow 0$ since $\alpha > 1/2$ by assumption. For the term II , since the kernel K used for the nonparametric estimation has support included in $[0, \infty)$, we have that $\hat{\sigma}_{n,t_{i-1}}^2$ is $\mathcal{F}_{i-1}$ -measurable. Conditioning on $\mathcal{F}_{i-1}$, we set

$$\chi_i^n = \mathbb{E} \left[\Delta_n^{-1/2} (\Delta_i^n X^2 - \Delta_i^n X^1) (\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1) \frac{\hat{\sigma}_{n,t_{i-1}}^2 - \bar{\sigma}_{n,t_{i-1}}^2}{\bar{\sigma}_{t_{i-1}}^2 \hat{\sigma}_{n,t_{i-1}}^2} \middle| \mathcal{F}_{i-1} \right]$$

$$= \Delta_n^{-1/2} (e^{-\vartheta(T_2-T_1)} - e^{-\vartheta_n(T_2-T_1)}) (e^{-\vartheta(T_2-T_1)} - 1) \xi_i^n,$$

say, with $\xi_i^n = \mathbb{E} \left[\int_{t_{i-1}}^{t_i} e^{-2\vartheta(T_1-t)} \sigma_t^2 dt \middle| \mathcal{F}_{i-1} \right] \frac{\hat{\sigma}_{n,t_{i-1}}^2 - \bar{\sigma}_{n,t_{i-1}}^2}{\bar{\sigma}_{t_{i-1}}^2 \hat{\sigma}_{n,t_{i-1}}^2}$. It follows that

$$\sum_{i=1}^n \mathbb{E} [|\xi_i^n| | \mathcal{F}_{i-1}] \leq \Delta_n M^2 \tilde{c}^{-4} \sum_{i=1}^n \sup_{i \in \mathcal{I}_n} |\hat{\sigma}_{n,t_{i-1}}^2 - \bar{\sigma}_{t_{i-1}}^2| \rightarrow 0$$

in probability by Theorem 1. Since $\Delta_n^{-1/2} (e^{-\vartheta(T_2-T_1)} - e^{-\vartheta_n(T_2-T_1)})$ is bounded, we use Lemma 3.4 in [11] applied to variables ξ_i^n to conclude $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E} [\chi_i^n | \mathcal{F}_{i-1}] \rightarrow 0$ in probability, locally uniformly in t . Moreover,

$$\mathbb{E} [(\chi_i^n)^2 | \mathcal{F}_{i-1}] \lesssim \Delta_n^{-1} \mathbb{E} [((\Delta_i^n X^2 - \Delta_i^n X^1) (\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1))^2 | \mathcal{F}_{i-1}] \sup_{i \in \mathcal{I}_n} |\hat{\sigma}_{n,t_{i-1}}^2 - \bar{\sigma}_{t_{i-1}}^2|$$

which is of order $\Delta_n^{-1} \Delta_n^2 \Delta_n^{\alpha/(2\alpha+1)}$ so that $\sum_{i \in \mathcal{I}_n} \mathbb{E} [(\chi_i^n)^2 | \mathcal{F}_{i-1}] \lesssim \Delta_n^{\alpha/(2\alpha+1)} \rightarrow 0$ in probability. Applying Lemma 3.4 in [11] to the sequence χ_i^n enables us to conclude that II converges to 0 in probability and (11) follows.

Step 2. Since $\vartheta \rightsquigarrow \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i,n}$ is smooth (at least twice differentiable) and $\vartheta_n - \vartheta$ is of order $n^{-1/2}$ a second-order Taylor expansion of $\tilde{\ell}_{\vartheta, \bar{\sigma}}^{i,n}$ at ϑ implies that

$$(12) \quad \sqrt{n} (\tilde{\ell}_{\vartheta_n, \bar{\sigma}}^{i,n} - \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i,n}) - \sqrt{n} (\vartheta_n - \vartheta) \partial_{\vartheta} \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i,n} \rightarrow 0$$

in probability. Since $|\mathcal{I}_n| \sim n(1 - h_n) \sim n$, it is not difficult to check that

$$(13) \quad \Delta_n \sum_{i \in \mathcal{I}_n} \partial_{\vartheta} \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i,n} \sim -\tilde{I}_{\vartheta, \sigma, \bar{\sigma}} = -\frac{(T_2 - T_1)^2}{(e^{\vartheta T_2} - e^{\vartheta T_1})^2} \int_0^T \frac{e^{2\vartheta t} \sigma_t^2}{\bar{\sigma}_t^2} dt$$

in probability under $\mathbb{P}_{\vartheta, \sigma, \bar{\sigma}}$, that is the total Fisher information associated to the efficient scores $\tilde{\ell}_{\vartheta, \bar{\sigma}}^{i, n}$. Summing each term in (12) for $i \in \mathcal{I}_n$, using (13) and the fact that Δ_n^{-1} and n are of the same order, we further infer

$$(14) \quad \Delta_n^{1/2} \left(\sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\vartheta_n, \hat{\sigma}_n}^{i, n} - \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i, n}) + \Delta_n^{-1} (\vartheta_n - \vartheta) \tilde{I}_{\vartheta, \sigma, \bar{\sigma}} \right) \rightarrow 0$$

in probability, using (11) in order to substitute $\bar{\sigma}$ by $\hat{\sigma}_n$ in the first term. Moreover, (14) remains true if we replace ϑ_n by the discretised version $\hat{\vartheta}_{2, n}$ of $\hat{\vartheta}_n$, using moreover that $\sqrt{n}(\hat{\vartheta}_n - \vartheta)$ is bounded in probability thanks to Theorem 1. We refer to the proof of Theorem 5.48 in [17] for the details.

Step 3. We establish

$$(15) \quad \Delta_n \sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\hat{\vartheta}_{2, n}, \hat{\sigma}_n}^{i, n})^2 \rightarrow \tilde{I}_{\vartheta, \sigma, \bar{\sigma}}$$

in probability under $\mathbb{P}_{\vartheta, \sigma, \bar{\sigma}}$. The computations are similar to Step 1, combining Theorems 1 and 1. We briefly give the mains steps. Observe that

$$\Delta_n \sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\vartheta_n, \hat{\sigma}_n}^{i, n})^2 - \Delta_n \sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\vartheta, \bar{\sigma}}^{i, n})^2 = (T_2 - T_1) \sum_{i \in \mathcal{I}_n} (I_i + II_i + III_i),$$

with

$$\begin{aligned} I_i &= \Delta_n \frac{(\Delta_i^n X^2 - \Delta_i^n X^1)^2}{(\Delta_n \hat{\sigma}_{n, t_{i-1}}^2)^2} \left(\frac{(\Delta_i^n X^2 - e^{-\vartheta_n(T_2 - T_1)} \Delta_i^n X^1)^2 e^{-2\vartheta_n(T_2 - T_1)}}{(1 - e^{-\vartheta_n(T_2 - T_1)})^6} - \frac{(\Delta_i^n X^2 - e^{-\vartheta(T_2 - T_1)} \Delta_i^n X^1)^2 e^{-2\vartheta(T_2 - T_1)}}{(1 - e^{-\vartheta(T_2 - T_1)})^6} \right), \\ II_i &= \Delta_n (\Delta_i^n X^2 - \Delta_i^n X^1)^2 \frac{(\Delta_i^n X^2 - e^{-\vartheta(T_2 - T_1)} \Delta_i^n X^1)^2 e^{-2\vartheta(T_2 - T_1)}}{(1 - e^{-\vartheta(T_2 - T_1)})^6} \left(\frac{1}{(\Delta_n \hat{\sigma}_{n, t_{i-1}}^2)^2} - \frac{1}{(\Delta_n \bar{\sigma}_{t_{i-1}}^2)^2} \right), \\ III_i &= \Delta_n (\Delta_i^n X^2 - \Delta_i^n X^1)^2 \frac{(\Delta_i^n X^2 - e^{-\vartheta(T_2 - T_1)} \Delta_i^n X^1)^2 e^{-2\vartheta(T_2 - T_1)}}{(1 - e^{-\vartheta(T_2 - T_1)})^6} \left(\frac{1}{(\Delta_n \bar{\sigma}_{t_{i-1}}^2)^2} - \frac{1}{\left(\int_{t_{i-1}}^{t_i} \bar{\sigma}_t^2 dt \right)^2} \right). \end{aligned}$$

For the term I_i we use a Taylor expansion of ϑ_n near ϑ in order to obtain $\mathbb{E}[|\mathcal{T}_{i, n}^1|] \lesssim |\vartheta_n - \vartheta| \Delta_n^{-1} \Delta_n^2$ and in turn $\mathbb{E}[|\sum_{i \in \mathcal{I}_n} I_i|] \lesssim |\vartheta_n - \vartheta|$. For the term II_i we use the convergence of $\hat{\sigma}_{n, t_{i-1}}$ and the conditioning argument in a similar way as in Step 1 to obtain $\sum_{i \in \mathcal{I}_n} II_i \rightarrow 0$ in probability. For the term III_i , an analysis of the convergence of $\hat{\sigma}_{n, t_{i-1}}$ using Assumption 2 shows that $\mathbb{E}[\sum_{i \in \mathcal{I}_n} |III_i|] \lesssim \Delta_n^\alpha$. This proves

$$\Delta_n \sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\vartheta_n, \hat{\sigma}_n}^{i, n})^2 - \Delta_n \sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\vartheta, \bar{\sigma}}^{i, n})^2 \rightarrow 0$$

in probability and the result remains true with $\hat{\vartheta}_{2, n}$ in place of ϑ_n . Since $\Delta_n \sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\vartheta, \bar{\sigma}}^{i, n})^2 \sim \tilde{I}_{\vartheta, \sigma, \bar{\sigma}}$ in probability under $\mathbb{P}_{\vartheta, \sigma, \bar{\sigma}}$ we obtain (15).

Step 4. By definition of $\hat{\vartheta}_{2, n}$, we have

$$\Delta_n^{-1/2} (\hat{\vartheta}_{2, n} - \vartheta) \tilde{I}_{\vartheta, \sigma, \bar{\sigma}} = \Delta_n^{-1/2} (\hat{\vartheta}_{2, n} - \vartheta) \tilde{I}_{\vartheta, \sigma, \bar{\sigma}} + \Delta_n^{-1/2} \frac{\Delta_n \tilde{I}_{\vartheta, \sigma, \bar{\sigma}} \sum_{i \in \mathcal{I}_n} \tilde{\ell}_{\hat{\vartheta}_{2, n}, \hat{\sigma}_n}^{i, n}}{\Delta_n \sum_{i \in \mathcal{I}_n} (\tilde{\ell}_{\hat{\vartheta}_{2, n}, \hat{\sigma}_n}^{i, n})^2} \sim \Delta_n^{1/2} \sum_{i \in \mathcal{I}_n} \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i, n}$$

in probability under $\mathbb{P}_{\vartheta, \sigma, \bar{\sigma}}$ thanks to (14) and (15) established in the two previous steps. We further write

$$\Delta_n^{1/2} \sum_{i \in \mathcal{I}_n} \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i, n} = \Delta_n^{1/2} \sum_{i=1}^n \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i, n} - \Delta_n^{1/2} \sum_{i=1}^{\lfloor h_n \Delta_n^{-1} \rfloor} \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i, n},$$

and we claim that

$$(16) \quad \Delta_n^{1/2} \sum_{i=1}^{\lfloor h_n \Delta_n^{-1} \rfloor} \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i, n} \rightarrow 0$$

in probability. We conclude the proof using the following limit theorem, proof of which is delayed until Appendix 5.2.

Lemma 2. *Work under Assumptions 1 and 2 with $\alpha > 1/2$. Then*

$$\Delta_n^{1/2} \sum_{i=1}^n \tilde{\ell}_{\vartheta, \sigma, \bar{\sigma}}^{i,n} \rightarrow \mathcal{N}(0, \tilde{I}_{\vartheta, \sigma, \bar{\sigma}})$$

stably in law, where, conditional on $\mathcal{F}$, the random variable $\mathcal{N}(0, \tilde{I}_{\vartheta, \sigma, \bar{\sigma}})$ is centred Gaussian with conditional variance $\tilde{I}_{\vartheta, \sigma, \bar{\sigma}}$.

In view of Lemma 2 we obtain Theorem 3.

Step 5. It remains to prove (16). Write $\Delta_n^{1/2} \tilde{\ell}_{\vartheta, \sigma, \bar{\sigma}}^i = \frac{(T_2 - T_1)e^{-\vartheta(T_2 - T_1)}}{(1 - e^{-\vartheta(T_2 - T_1)})^3} (I_i + II_i)$, with

$$I_i = \Delta_n^{1/2} \frac{(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta(T_2 - T_1)} \Delta_i^n X^1)}{\Delta_n \bar{\sigma}_{t_{i-1}}^2},$$

$$II_i = \Delta_n^{1/2} (\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta(T_2 - T_1)} \Delta_i^n X^1) \frac{\int_{t_{i-1}}^{t_i} (\bar{\sigma}_{t_{i-1}}^2 - \bar{\sigma}_t^2) dt}{\Delta_n \bar{\sigma}_{t_{i-1}}^2 \int_{t_{i-1}}^{t_i} \bar{\sigma}_t^2 dt}.$$

We readily have

$$\mathbb{E}[I_i^2] \leq \tilde{c}^{-2} \Delta_n^{-1} \mathbb{E}((\Delta_i^n X^2 - \Delta_i^n X^1)^2 (\Delta_i^n X^2 - e^{-\vartheta(T_2 - T_1)} \Delta_i^n X^1)^2) \lesssim \Delta_n,$$

so that $\mathbb{E}[(\sum_{i=1}^{\lfloor h_n \Delta_n^{-1} \rfloor} I_i)^2] = \sum_{i=1}^{\lfloor h_n \Delta_n^{-1} \rfloor} \mathbb{E}[I_i^2] \lesssim h_n \Delta_n^{-1} \Delta_n = h_n \rightarrow 0$. The term II_i is similar to the term I in Step 1. We readily obtain $\mathbb{E}[|II_i|] \lesssim \Delta_n^{1/2+\alpha}$. It follows that

$$\mathbb{E}\left[\left|\sum_{i=1}^{\lfloor h_n \Delta_n^{-1} \rfloor} II_i\right|\right] \lesssim h_n \Delta_n^{-1} \Delta_n^{1/2+\alpha} = h_n \Delta_n^{\alpha-1/2} \rightarrow 0$$

since $\alpha \geq 1/2$, and (16) follows. The proof of Theorem 3 is complete.

5. APPENDIX

5.1. Proof of Lemma 1.

The first part of the result. Since $\sup_{t \in [0, T]} t^{-s} \omega(Y)_t < \infty$ for some $s > 1/2$, we have that Y is continuous in probability on $[0, T]$. Write

$$\Delta_n^{-1} \sum_{i=1}^n (\bar{\Delta}_i^n(Y))^2 - \int_0^T Y_t^2 dt = I + II + III,$$

with

$$I = \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} (Y_{(i-1)\Delta_n}^2 - Y_t^2) dt,$$

$$II = \Delta_n^{-1} \sum_{i=1}^n \left(\int_{(i-1)\Delta_n}^{i\Delta_n} (Y_t - Y_{(i-1)\Delta_n}) dt \right)^2,$$

$$III = 2 \sum_{i=1}^n Y_{t_{i-1}} \int_{(i-1)\Delta_n}^{i\Delta_n} (Y_t - Y_{(i-1)\Delta_n}) dt.$$

First, fix $\epsilon > 0$. There exists some $\eta > 0$ such that $\mathbb{E}[|Y_t - Y_s|] < \epsilon$ as soon as $|t - s| < \eta$. Moreover, by localisation we may (and will) assume that there is some $M > 0$ such that $\sup_t |Y_t| \leq M$. It follows that

$$\mathbb{E}[|I|] \leq \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}[|Y_{(i-1)\Delta_n}^2 - Y_t^2|] dt \leq 2M \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}[|Y_{(i-1)\Delta_n} - Y_t|] dt \leq 2TM\epsilon$$

as soon as $\Delta \leq \eta$ which is true for large enough n . Thus $I \rightarrow 0$ in probability. The proof is similar for II and III .

The second part of the result. Write

$$\Delta_n^{-1} \sum_{i=1}^n \overline{\Delta}_i^n(Y) \int_{(i-1)\Delta_n}^{i\Delta_n} Z_t dB_t - \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} Y_t Z_t dB_t = I + II,$$

with

$$I = \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} (Y_{(i-1)\Delta_n} - Y_t) Z_t dB_t$$

$$II = \Delta_n^{-1} \sum_{i=1}^n (\overline{\Delta}_i^n(Y) - \Delta_n Y_{(i-1)\Delta_n}) \int_{(i-1)\Delta_n}^{i\Delta_n} Z_t dB_t.$$

Fix $\epsilon > 0$ and $\eta > 0$ such that $\mathbb{E}[|Y_t - Y_s|] < \epsilon$ as soon as $|t - s| < \eta$. By localisation, we may assume that Z is such that $\sup_t \max(|Z_t|, |Y_t|) \leq M$. By the martingale property,

$$\mathbb{E}[I^2] = \sum_{i=1}^n \mathbb{E}[(\int_{(i-1)\Delta_n}^{i\Delta_n} (Y_{(i-1)\Delta_n} - Y_t) Z_t dB_t)^2] = \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}[(Y_{(i-1)\Delta_n} - Y_t)^2 Z_t^2] dt \leq 2TM^3\epsilon$$

as soon as $\Delta_n \leq \eta$ which is true for large enough n . For II , by Cauchy-Schwarz inequality,

$$II \leq (\sum_{i=1}^n (\Delta_n^{-1} \int_{t_{i-1}}^{t_i} (Y_t - Y_{t_{i-1}}) dt)^2)^{1/2} (\sum_{i=1}^n (\int_{t_{i-1}}^{t_i} Z_t dB_t)^2)^{1/2} \lesssim (\sum_{i=1}^n (\Delta_n^{-1} \int_{t_{i-1}}^{t_i} (Y_t - Y_{t_{i-1}}) dt)^2)^{1/2}$$

in probability. Let $\varphi(x) = \mathbf{1}_{[0,1]}$ be the Haar function, and $\varphi_{j,k}(x) = 2^{j/2} \varphi(2^j x - k) = 2^{j/2} \mathbf{1}_{[k2^{-j}, (k+1)2^{-j}]}$ for any $j \geq 0, k \in \mathbb{Z}$. We prove the result under the restriction that $n = 2^j$ and that the t_i are of the form $k2^{-j}$. The general case of a regular mesh $t_i = i\Delta_n$ is slightly more intricate but follows the same ideas. We have

$$\sum_{i=1}^n (\Delta_n^{-1} \int_{t_{i-1}}^{t_i} (Y_t - Y_{t_{i-1}}) dt)^2 = \frac{1}{T^2} \sum_{k=1}^{2^j} (2^j \int_{t_{k-1}}^{t_k} (Y_t - Y_{t_{k-1}}) dt)^2 = \sum_{k=0}^{2^j-1} ((P_j(Y_T)(k2^{-j}) - Y_{k2^{-j}T})^2$$

where $P_j(f) = \sum_{k=0}^{2^j-1} (\int \varphi_{j,k} f) \varphi_{j,k}$ is the orthogonal projection on $\text{Span}\{\varphi_{j,k}, k = 1, \dots, 2^j\}$. For large enough j , there exists some constant $C > 0$ such that

$$\mathbb{E}[\sum_{k=0}^{2^j-1} (P_j(Y_T)(\frac{k}{2^j}) - Y_{\frac{k}{2^j}T})^2] \leq C2^j \mathbb{E}[\int_0^1 (P_j(Y_T)(u) - Y_{uT})^2 du].$$

It follows that

$$\begin{aligned} \mathbb{E}[\sum_{i=1}^n (\Delta_n^{-1} \int_{t_{i-1}}^{t_i} (Y_t - Y_{t_{i-1}}) dt)^2] &\leq C2^j \mathbb{E}[\int_0^1 (P_j(Y_T)(u) - Y_{uT})^2 du] \\ &\lesssim 2^j \mathbb{E}[\int_0^1 (\int_0^1 2^j \mathbf{1}_{|u-y| \in [0, 2^{-j}]} (Y_{yT} - Y_{uT}) dy)^2 du] \\ &\lesssim 2^j (\int_0^1 \omega_{2^{-j}xT}(Y) dx)^2 \lesssim 2^{j(1-2s)} \end{aligned}$$

and this term converges to 0 since $s > 1/2$ by assumption. The proof of Lemma 1 is complete.

5.2. Proof of Lemma 2. First we establish the result for

$$\frac{T_2 - T_1}{e^{\vartheta(T_2 - T_1)} - 1} \Delta_n^{1/2} \sum_{i=1}^n \chi_i^n, \quad \text{with } \chi_i^n = \frac{\sigma_{t_{i-1}} \Delta_i^n \overline{B} \int_{t_{i-1}}^{t_i} e^{-\vartheta(T_1 - t)} dB_t}{\Delta_n \overline{\sigma}_{(i-1)\Delta_n}},$$

by applying Lemma 3.7 in [11]. To do so, we check conditions (3.43)–(3.46) in [11] for χ_i^n . We keep up with the notation of [11]. First, we have $\mathbb{E}[\chi_i^n | \mathcal{F}_{i-1}] = 0$, which ensures (3.43) with $A_t = 0$. Next,

$$\begin{aligned}\mathbb{E}[(\chi_i^n)^2 | \mathcal{F}_{i-1}] &= \Delta_n \frac{\sigma_{(i-1)\Delta_n}^2}{\Delta_n^2 \bar{\sigma}_{(i-1)\Delta_n}^2} \mathbb{E}[(\Delta_i^n \bar{B})^2 | \mathcal{F}_{i-1}] \mathbb{E}[(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)} dB_t)^2 | \mathcal{F}_{i-1}] \\ &= \frac{\sigma_{(i-1)\Delta_n}^2}{\bar{\sigma}_{(i-1)\Delta_n}^2} \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} dt,\end{aligned}$$

so that

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}[(\chi_i^n)^2 | \mathcal{F}_{i-1}] \rightarrow \int_0^t \frac{e^{-2\vartheta(T_1-s)} \sigma_s^2}{\bar{\sigma}_s^2} ds$$

in probability. This is Condition (3.44) in [11] with $C_t = \int_0^t \frac{e^{-2\vartheta(T_1-s)} \sigma_s^2}{\bar{\sigma}_s^2} ds$. It follows that

$$\begin{aligned}\mathbb{E}[(\chi_i^n)^4 | \mathcal{F}_{i-1}] &= \Delta_n^2 \frac{\sigma_{(i-1)\Delta_n}^4}{\Delta_n^4 \bar{\sigma}_{(i-1)\Delta_n}^4} \mathbb{E}[(\Delta_i^n \bar{B})^4 | \mathcal{F}_{i-1}] \mathbb{E}[(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)} dB_t)^4 | \mathcal{F}_{i-1}] \\ &\leq 9 \Delta_n^2 \frac{\sigma_{(i-1)\Delta_n}^4}{\bar{\sigma}_{(i-1)\Delta_n}^4} \left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)} dt \right)^2\end{aligned}$$

by independence of the two Wiener integrals. Therefore $\sum_{i=1}^n \mathbb{E}[(\chi_i^n)^4 | \mathcal{F}_{i-1}] \rightarrow 0$ in probability. This is condition (3.45) in [11]. Finally, $\mathbb{E}[\chi_i^n \Delta_i^n B | \mathcal{F}_{i-1}] = \mathbb{E}[\chi_i^n \Delta_i^n \bar{B} | \mathcal{F}_{i-1}] = 0$ by independence which ensures condition (3.46) in [11]. We subsequently apply Lemma 3.7 in [11] to conclude that

$$\frac{T_2 - T_1}{e^{\vartheta(T_2 - T_1)} - 1} \Delta_n^{1/2} \sum_{i=1}^n \chi_i^n$$

converges stably in law to a random variable, which, conditional on $\mathcal{F}$, is Gaussian with variance $\tilde{I}_{\vartheta, \sigma, \bar{\sigma}}$. In order to complete the proof, we write

$$\Delta_n^{1/2} \sum_{i=1}^n \tilde{\ell}_{\vartheta, \bar{\sigma}}^{i,n} = \Delta_n^{1/2} \sum_{i=1}^n \left(\tilde{\ell}_{\vartheta, \bar{\sigma}}^{i,n} - \frac{T_2 - T_1}{e^{\vartheta(T_2 - T_1)} - 1} \sum_{i=1}^n \chi_i^n \right) + \frac{T_2 - T_1}{e^{\vartheta(T_2 - T_1)} - 1} \Delta_n^{1/2} \sum_{i=1}^n \chi_i^n$$

and it remains to show the convergence of $\Delta_n^{1/2} \sum_{i=1}^n \left(\tilde{\ell}_{\vartheta, \bar{\sigma}}^{i,n} - \frac{T_2 - T_1}{e^{\vartheta(T_2 - T_1)} - 1} \sum_{i=1}^n \chi_i^n \right) \rightarrow 0$ in probability. This is done using similar arguments as used in the proof of Theorem 1. We omit the details.

Acknowledgements. *We are grateful to the suggestions and comments of two anonymous referees that helped to considerably improve a former version of the manuscript.*

REFERENCES

- [1] Fred Espen Benth and Steen Koekebakker. Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157, 2008.
- [2] Emmanuelle Clément, Sylvain Delattre, and Arnaud Gloter. An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. *Stochastic Processes and their Applications*, 123(7):2500–2521, 2013.
- [3] John C. Cox, Jonathan E. Ingersoll Jr., and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407, 1985.
- [4] Valentine Genon-Catalot and Jean Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Institut. H. Poincaré*, 29(1):119–151, 1993.
- [5] Valentine Genon-Catalot, Catherine Laredo, and Dominique Picard. Nonparametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics*, 19(4):317–335, 1992.
- [6] David Heath, Robert Jarrow, and Andrew Morton. Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica*, 60(1):77–105, 1992.
- [7] Juri Hinz, Lutz von Grafenstein, Michel Verschuere, and Martina Wilhelm. Pricing electricity risk by interest rate methods. *Quantitative Finance*, 5(1):49–60, 2005.
- [8] Marc Hoffmann. Minimax estimation of the diffusion coefficient through irregular samplings. *Statistics & Probability Letters*, 32(1):11–24, 1997.
- [9] Marc Hoffmann. Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79(1):135–163, 1999.
- [10] Marc Hoffmann, Axel Munk, and Johannes Schmidt-Hieber. Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. *Ann. Institut. H. Poincaré*, 48(4):1186–1216, 2012.

- [11] Jean Jacod. Statistics and high-frequency data. In Mathieu Kessler, Alexander Lindner, and Michael Sørensen, editors, *Statistical Methods for Stochastic Differential Equations*, pages 191–310. Chapman and Hall/CRC, 2012.
- [12] Andrew Jeffrey, Oliver Linton, Thong Nguyen, and Peter C.B. Phillips. Nonparametric estimation of a multifactor Heath-Jarrow-Morton model: An integrated approach. Technical report, Cowles Foundation for Research in Economics, 2001.
- [13] Jussi Keppo, Nicolas Audet, Pirja Heiskanen, and Iivo Vehviläinen. Modeling electricity forward curve dynamics in the nordic market. In D.W. Bunn, editor, *Modelling Prices in Competitive Electricity Markets*, pages 251–264. Wiley Series in Financial Economics, 2004.
- [14] Rüdiger Kiesel, Gero Schindlmayr, and Reik H. Börger. A two-factor model for the electricity forward market. *Quantitative Finance*, 9(3):279–287, 2009.
- [15] Steen Koekebakker and Fridthjof Ollmar. Forward curve dynamics in the nordic electricity market. *Managerial Finance*, 31(6):73–94, 2005.
- [16] Per Mykland and Lan Zhang. The econometrics of high-frequency. In Mathieu Kessler, Alexander Lindner, and Michael Sørensen, editors, *Statistical Methods for Stochastic Differential Equations*, pages 109–190. Chapman and Hall/CRC, 2012.
- [17] Aad Van der Vaart. *Asymptotic Statistics*. University of Cambridge, 1998.

OLIVIER FÉRON, EDF LAB PARIS-SACLAY AND FIME, LABORATOIRE DE FINANCE DES MARCHÉS DE L'ÉNERGIE, 91120 PALAISEAU, FRANCE.

E-mail address: `olivier-2.feron@edf.fr`

PIERRE GRUET, EDF LAB PARIS-SACLAY AND FIME, LABORATOIRE DE FINANCE DES MARCHÉS DE L'ÉNERGIE, 91120 PALAISEAU, FRANCE.

E-mail address: `pierre.gruet@edf.fr`

MARC HOFFMANN, UNIVERSITÉ PARIS-DAUPHINE, CEREMADE, 75016 PARIS, FRANCE

E-mail address: `hoffmann@ceremade.dauphine.fr`

Finance for Energy Market Research Centre

Institut de Finance de Dauphine, Université Paris-Dauphine

1 place du Maréchal de Lattre de Tassigny

75775 PARIS Cedex 16

www.fime-lab.org