HAL
open science

EFFICIENT VOLATILITY ESTIMATION IN A TWO-FACTOR MODEL

Olivier Féron, Pierre Gruet, Marc Hoffmann

To cite this version:

Olivier Féron, Pierre Gruet, Marc Hoffmann. EFFICIENT VOLATILITY ESTIMATION IN A TWO-FACTOR MODEL. 2018. hal-01966340

HAL Id: hal-01966340

https://hal.science/hal-01966340v1

Preprint submitted on 28 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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.* [9], 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.* [10], 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 [19] 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.* [17], a one-factor model is designed for each maturity date, having correlations between the Brownian motions for distinct dates. In Kiesel *et al.* [18], 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

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

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

$$T \leq T_1 < \ldots < T_d$$

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

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

Asymptotics are taken as $n \to \infty$. 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 [6, 7, 11, 12, 13] 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 [20] and Jacod [14] 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 $n^{1/2}$-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.* [4] 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.

1.2. **Main results and organization of the paper.** In Section 2.1, we provide an estimator of ϑ, based on quadratic variation, in the above observation scheme. We will explain that 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 [14, 20], we show that

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

stably in law, where, conditional on $\mathcal{F}$, the random variable $\mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma}))$ is centred Gaussian, with conditional variance $V_\vartheta(\sigma, \overline{\sigma})$, and possibly defined on an extension of the original probability space. When $d \geq 3$, the model is somehow degenerate, as it had been reported by Jeffrey *et al.* [16] in a similar context, because the d marginal components of the process are driven by less than d Brownian motions. The remaining source of randomness is the drift process, and while we shall find a Δ_n^{-1}-consistent estimator $\hat{\vartheta}_{d,n}$ for ϑ, we will need that b has some integrated regularity in expectation, reminiscent of the so-called Besov regularity, as will be made precise by Assumption 1, in order to establish a satisfying limit theorem for $\Delta_n^{-1}(\hat{\vartheta}_{3,n} - \vartheta)$, namely that it converges in probability to some $\mathcal{F}$-measurable random variable. All our results in dimension $d = 2, 3$ will be stated in Theorem 1. In Section 2.2, we perform a relatively classical nonparametric estimation procedure to get point estimates of σ_t^2 and $\overline{\sigma}_t^2$ when $d = 2$. Note however that this is not an usual nonparametric problem, since

(1) σ and $\overline{\sigma}$ are random themselves, so that we do not estimate them pointwise, instead we estimate the trajectories $\left(\sigma_t^2(\omega)\right)_t$ and $\left(\overline{\sigma}_t^2(\omega)\right)_t$ pointwise, as realisations of the volatility processes;
(2) an increment $\Delta_i^n X$ is the sum of two stochastic integrals, in which the volatility processes have different regularities.

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 shall then derive estimators $\widehat{\sigma}_n^2$ and $\widehat{\overline{\sigma}}_n^2$ of σ^2 and $\overline{\sigma}^2$ and in Theorem 2, adding Assumption 2 stating that the volatility processes are Hölder in expectation, it will be shown that each of those point estimators is $\Delta_n^{-\alpha/(2\alpha+1)}$-consistent, where α is the lowest of two values of the Hölder regularities of σ^2 and $\overline{\sigma}^2$. In Section 2.3, referring to the theory of semiparametric estimation, reported for instance in [21], we compute a lower bound $V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})$ for the limit variance while estimating ϑ with $d=2$ observed processes, for deterministic volatility functions, in Theorem 3. As soon as $\overline{\sigma}$ is not constant, this bound is lower than $V_\vartheta(\sigma,\overline{\sigma})$. Then, we derive an estimator $\tilde{\vartheta}_{2,n}$ such that

$$\Delta_n^{-1/2}(\tilde{\vartheta}_{2,n}-\vartheta)\to\mathcal{N}(0,V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma}))$$

stably in law, where conditional on $\mathcal{F}$, the random variable $\mathcal{N}(0,V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma}))$ is centred Gaussian with conditional variance $V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})$. This estimator is efficient in the sense that it achieves the minimal conditional variance $V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})$ among all possible $\sqrt{n}$-consistent estimators that are asymptotically centred mixed normal. In Section 3, we discuss the case $d>3$ and also present a possible extension of our setting, that puts additive model error terms in the observables in order to break the degeneration of the model. We perform some numerical experiments in Section 4, 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 5.

2. Construction of the estimators and convergence results

2.1. Rate-optimal estimation of ϑ.

2.1.1. *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 $t\rightsquigarrow\overline{\sigma}_t$ are unknown. Indeed X has the same law under the choice of $(\vartheta,\sigma,\overline{\sigma})$ and $(\vartheta+1,e^{T_1-\cdot}\sigma,\overline{\sigma})$.

2.1.2. *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

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

and

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

in probability, we also obtain the convergence of the ratio

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

in probability. The function $\vartheta\rightsquigarrow\psi_{T_1,T_2}(\vartheta)$ maps $(0,\infty)$ onto $(-1,0)$ and this leads to a first $\Delta_n^{-1/2}$-consistent estimation strategy by setting

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

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

2.1.3. *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! 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, so that we could expect to find Δ_n^{-1}-consistent estimators instead of $\Delta_n^{-1/2}$-consistent ones. We then have

$$\Psi^n_{T_1,T_2,T_3} = \frac{\sum_{i=1}^n (\Delta_i^n X^3 - \Delta_i^n X^2)^2}{\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2} \to \Big(\frac{e^{-\vartheta T_3} - e^{-\vartheta T_2}}{e^{-\vartheta T_2} - e^{-\vartheta T_1}}\Big)^2 = \psi_{T_1,T_2,T_3}(\vartheta),$$

say. The function $\vartheta \rightsquigarrow \psi_{T_1,T_2,T_3}(\vartheta)$ maps $(0,\infty)$ onto $\big(0, \big(\frac{T_3-T_2}{T_2-T_1}\big)^2\big)$ and is also invertible (see Lemma 3), leading to the estimator

$$\hat\vartheta_{3,n} = \psi^{-1}_{T_1,T_2,T_3}\big(\Psi^n_{T_1,T_2,T_3}\big)$$

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

2.1.4. *Convergence results.* We need some assumption about the regularity of the processes b, σ and $\overline\sigma$. For a random process $X = (X_t)_{0\le t\le T}$, introduce the following modulus of continuity:

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

Assumption 1. *The processes σ and $\overline\sigma$ are almost surely positive. Moreover, for some $s > 1/2$, we have $\sup_{t\in[0,T]} t^{-s}\omega(b^j)_t < \infty$ for every $j = 1,\dots,d$.*

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

$$\overline b_t = 2(e^{-\vartheta T_2} - e^{-\vartheta T_1})(e^{-\vartheta T_3} - e^{-\vartheta T_2})\big((e^{-\vartheta T_2} - e^{-\vartheta T_1})(b_t^3 - b_t^2) - (e^{-\vartheta T_3} - e^{-\vartheta T_2})(b_t^2 - b_t^1)\big)$$

and

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

Define also

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

Theorem 1. *Work under Assumption 1.*

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

$$\Delta_n^{-1/2}(\hat\vartheta_{2,n} - \vartheta) \to \mathcal{N}\big(0, V_\vartheta(\sigma,\overline\sigma)\big)$$

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

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

(2) *For $d = 3$ we have*

$$\Delta_n^{-1}(\hat\vartheta_{3,n} - \vartheta) \to \frac{\widetilde b_T + \int_0^T \bar b_t e^{\vartheta t}\sigma_t dB_t}{2(e^{-\vartheta T_2} - e^{-\vartheta T_1})D_3 \int_0^T e^{2\vartheta t}\sigma_t^2 dt}$$

in probability as $n\to\infty$.

2.2. **Rate-optimal estimation of the volatility processes.**

2.2.1. *Construction of an estimator.* We start with the observation that for any sufficiently regular test function $g : [0,T] \to \mathbb{R}$, we have, for any $j = 1,\dots,d$,

$$\sum_{i=1}^n g((i-1)\Delta_n)\big(\Delta_i^n X^j\big)^2 \to \int_0^T g(s) d\langle X^j\rangle_s = \int_0^T g(s)\big(e^{-2\vartheta(T_j - s)}\sigma_s^2 + \overline\sigma_s^2\big) ds \tag{2}$$

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

$$e^{-2\vartheta(T_1-t)}\sigma_t^2 + \overline\sigma_t^2 \ \text{ and } \ e^{-2\vartheta(T_2-t)}\sigma_t^2 + \overline\sigma_t^2$$

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 $\overline{\sigma}_t^2$ as well by inverting a 2×2 linear system, namely

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

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

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

The bandwidth h_n is set below to balance both bias and variance, while $\varpi_n > 0$ garantees the well-posedness of the estimator.

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

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

$$\mathbb{E}\big[(\sigma_t^2 - \sigma_s^2)^2\big] + \mathbb{E}\big[(\overline{\sigma}_t^2 - \overline{\sigma}_s^2)^2\big] \le c|t-s|^{2\alpha}. \tag{4}$$

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

$$\Delta_n^{-\alpha/(2\alpha+1)}\big(\widehat{\sigma}_{n,t}^2 - \sigma_t^2\big) \ \text{ and } \ \Delta_n^{-\alpha/(2\alpha+1)}\big(\widehat{\overline{\sigma}}_{n,t}^2 - \overline{\sigma}_t^2\big)$$

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

2.3. **Efficient estimation of ϑ when $d = 2$.** We 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 $(\vartheta, \sigma, \overline{\sigma})$ since $t \rightsquigarrow \sigma_t$ and $t \rightsquigarrow \overline{\sigma}_t$ are random processes themselves. In order to bypass this difficulty, we first restrict our attention to the case where σ and $\overline{\sigma}$ 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 $V_\vartheta^{\mathrm{opt}}(\sigma, \overline{\sigma})$. In a second step, allowing σ and $\overline{\sigma}$ 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 $V_\vartheta^{\mathrm{opt}}(\sigma, \overline{\sigma})$, *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

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

with parameter $(\vartheta, \sigma, \overline{\sigma}) \in \Theta \times \Sigma(c, \tilde{c})$, with $\Theta = (0, \infty)$ and $\Sigma(c, \tilde{c})$ being the space of positive (deterministic) functions $(\sigma, \overline{\sigma})$ defined on $[0, T]$, satisfying (4) of Assumption 2 with constant c and satisfying moreover $\tilde{c} \le \inf_t \sigma_t \le \sup_t \sigma_t \le c$ for some $\tilde{c} > 0$.

Theorem 3. *Let $\hat{\vartheta}_n$ be an estimator of ϑ in the experiment $\mathcal{E}^n$ such that $\Delta_n^{-1/2}(\hat{\vartheta}_n - \vartheta)$ converges to $\mathcal{N}\big(0, V_\vartheta(\sigma, \overline{\sigma})\big)$ in distribution as $n \to \infty$. Then*

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

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

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

Let us further denote by $f^{i,n}_{\vartheta,\sigma,\overline{\sigma}}$ 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

$$\hat{\vartheta}_n = \hat{\vartheta}_{2,n} + \frac{\sum_{i=1}^n \ell^i_{\vartheta=\hat{\vartheta}_{2,n},\sigma,\overline{\sigma}}(\Delta_i^n X^1, \Delta_i^n X^2)}{\sum_{i=1}^n \big(\ell^i_{\vartheta=\hat{\vartheta}_{2,n},\sigma,\overline{\sigma}}(\Delta_i^n X^1, \Delta_i^n X^2)\big)^2}$$

where

$$\ell^{i,n}_{\vartheta,\sigma,\overline{\sigma}}(\Delta_i^n X^1, \Delta_i^n X^2) = \partial_\vartheta \log f^{i,n}_{\vartheta,\sigma,\overline{\sigma}}(\Delta_i^n X^1, \Delta_i^n X^2) \tag{6}$$

is the score function associated to $(\Delta_i^n X^1, \Delta_i^n X^2)$, see for instance Section 8.9 in [21]. However, this oracle procedure is not achievable and we need to invoke the theory of semiparametric efficiency (see for instance Chapter 25 of [21]). In the presence of an extra nuisance parameter $(\sigma, \overline{\sigma})$, we consider instead the so-called efficient score

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

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 $\widetilde{\ell}^{i,n}_{\vartheta,\sigma,\overline{\sigma}}$ which enables us to derive a one-step correction formula using $\widetilde{\ell}^{i,n}_{\vartheta,\sigma,\overline{\sigma}}$ and plug-in estimators in order to achieve the optimal bound.

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

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

for $i = 1, \dots, n$ and $(\vartheta, \mu) \in (0, \infty) \times [\tilde{c}, c]$.

Theorem 4. *Work under Assumptions 1 and 2 with $\alpha > 1/2$. For $i = 1, \dots, n$, the efficient score for the parameter ϑ associated to $(\Delta_i^n X^1, \Delta_i^n X^2)$ is given by $\widetilde{\ell}^{i,n}_{\vartheta,\Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2 dt}(\Delta_i^n X^1, \Delta_i^n X^2)$. Moreover, the estimator $\tilde{\vartheta}_{2,n}$ defined by*

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

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

$$\Delta_n^{-1/2}\big(\tilde{\vartheta}_{2,n} - \vartheta\big) \to \mathcal{N}\big(0, V_\vartheta^{\mathrm{opt}}(\sigma, \overline{\sigma})\big)$$

in distribution as $n \to \infty$. Finally, the result is still valid if σ and $\overline{\sigma}$ are random processes such that $\mathbb{P}\big((\sigma, \overline{\sigma}) \in \Sigma(c, \tilde{c})\big) = 1$. In that case, the limiting distribution is, conditionally on $\mathcal{F}$, centred Gaussian with (conditional) variance $V_\vartheta^{\mathrm{opt}}(\sigma, \overline{\sigma})$.

This result shows that the lower bound $V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})$ can be attained, and therefore that efficient estimation can be performed (which has a sense only for deterministic volatility functions). 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 $\Delta_n^{-1/2}(\hat{\vartheta}_{2,n}-\vartheta)$ if and only if $\overline{\sigma}$ is constant over the interval $[0,T]$. Otherwise, efficient estimation is more accurate than the one in the first part of Theorem 1.

3. Discussion on possibly richer observation schemes

3.1. **Discussion on the case $d>3$.** In Section 2.1, we built estimators of ϑ for $d=2,3$. When $d>3$, we meet the same problem of degeneracy as when $d=3$: the d processes are driven by 2 Brownian motions only. We may therefore build an estimator similar to the one with three processes. We have

$$\Psi^n_{T_{1..d}}=\sum_{j=3}^d\frac{\sum_{i=1}^n(\Delta_i^nX^j-\Delta_i^nX^{j-1})^2}{\sum_{i=1}^n(\Delta_i^nX^2-\Delta_i^nX^1)^2}\to\sum_{j=3}^d\Big(\frac{e^{-\vartheta T_j}-e^{-\vartheta T_{j-1}}}{e^{-\vartheta T_2}-e^{-\vartheta T_1}}\Big)^2=\psi_{T_{1..d}}(\vartheta).$$

The function $\vartheta\rightsquigarrow\psi_{T_{1..d}}(\vartheta)$ maps $(0,\infty)$ onto $\big(0,\sum_{j=3}^d\big(\frac{T_j-T_{j-1}}{T_2-T_1}\big)^2\big)$ and is invertible as the sum of $d-2$ monotone functions (see Lemma 3). We can thus propose the estimator

$$\hat{\vartheta}_{d,n}=\psi^{-1}_{T_{1..d}}\big(\Psi^n_{T_{1..d}}\big)$$

whenever $\Psi^n_{T_{1..d}}\in\big(0,\sum_{j=3}^d\big(\frac{T_j-T_{j-1}}{T_2-T_1}\big)^2\big)$ and 0 otherwise.

Using the notation

$$\begin{aligned}\overline{b}_t^d=&2(e^{-\vartheta T_2}-e^{-\vartheta T_1})\sum_{j=3}^d(e^{-\vartheta T_j}-e^{-\vartheta T_{j-1}})\big[(e^{-\vartheta T_2}-e^{-\vartheta T_1})(b_t^j-b_t^{j-1})\\&\qquad-(e^{-\vartheta T_j}-e^{-\vartheta T_{j-1}})(b_t^2-b_t^1)\big],\\\widetilde{b}_T^d=&(e^{-\vartheta T_2}-e^{-\vartheta T_1})^2\int_0^T\sum_{j=3}^d(b_t^j-b_t^{j-1})^2dt-\sum_{j=3}^d(e^{-\vartheta T_j}-e^{-\vartheta T_{j-1}})^2\int_0^T(b_t^2-b_t^1)^2dt,\\D_d=&\sum_{j=3}^d(e^{-\vartheta T_j}-e^{-\vartheta T_{j-1}})\big[(e^{-\vartheta T_j}-e^{-\vartheta T_{j-1}})(T_2e^{-\vartheta T_2}-T_1e^{-\vartheta T_1})\\&\qquad-(e^{-\vartheta T_2}-e^{-\vartheta T_1})(T_je^{-\vartheta T_j}-T_{j-1}e^{-\vartheta T_{j-1}})\big],\end{aligned}$$

it is possible to prove that under Assumption 1, we have

$$\Delta_n^{-1}(\hat{\vartheta}_{d,n}-\vartheta)\to\frac{\widetilde{b}_T^d+\int_0^T\overline{b}_t^de^{\vartheta t}\sigma_tdB_t}{2(e^{-\vartheta T_2}-e^{-\vartheta T_1})D_d\int_0^Te^{2\vartheta t}\sigma_t^2dt}$$

in probability as $n\to\infty$.

A natural question arises while defining this new estimator: are we able to determine if using $d>3$ processes is better than using $d=3$ processes only? As the convergence rate is the same, the criterion should be the comparison of the limits in probability of $\Delta_n^{-1}(\hat{\vartheta}_{3,n}-\vartheta)$ and $\Delta_n^{-1}(\hat{\vartheta}_{d,n}-\vartheta)$. The question of finding sufficient conditions so that one of those limits is closer to zero than the other one remains theoretically open. In numerical experiments however, we shall compute all the estimators and compare them.

3.2. **Incorporating model errors.** Since the HJM model certainly cannot fit a dataset perfectly, we may consider including model errors that somehow fill the gap between the observables and the outputs of the mathematical model. Indeed and as reported earlier, as soon as the dimension of the process is greater than two, the model is degenerate: this is a common feature of HJM models, introduced in Heath *et al.* [9]. Jeffrey *et al.* [16] calls this phenomenon *stochastic singularity*. In the absence of drift processes, arbitrage would be possible as some linear combination of processes would be zero; this is not a feature of empirical data. The classical approach to avoid it is to add another source of randomness, as is done in Jeffrey *et al.* [16], Bhar and Chiarella [2] and Bhar *et al.* [3]. In the latter, estimation bearing on the prices of interest rates

products is performed (in a parametric setting) with the addition of a measurement error to face stochastic singularity. We may introduce a shrinking noise, standing for model errors. We consider again the context of Section 1.1, as we are modeling a multidimensional diffusion process driven by two independent Brownian motions, with the same volatility structure. Yet, at time $t_i = i\Delta_n$, we have observations $Y^j_{t_i}$, $j = 1, \dots, d$, with

$$Y^j_{t_i} = X^j_{t_i} + \kappa^n_j \epsilon^j_i,$$

where $\kappa^n_j > 0$ are deterministic, and ϵ^j_i are iid centred random variables.
Asymptotics are again taken as $n \to \infty$, and in this high-frequency framework, we would like to estimate ϑ and the random components $t \rightsquigarrow \sigma^2_t$ and $t \rightsquigarrow \overline{\sigma}^2_t$. Depending on the structure of the error terms $\kappa^n_j \epsilon^j_i$, the properties of the estimators we have derived may change, in terms of asymptotic behaviour.

The estimation of processes contaminated by noise has been paid some attention, inspired in finance by the issues arising in microstructure noise modelling (although the context is quite different here). See for instance Zhang *et al.* [23, 22]. Jacod and Protter, in Chapter 16 of [15] obtain laws of large numbers and central limit theorems under a very general specification of error terms. In the present setting, while considering noise around the diffusion process, we would like to extend our previous results in order to be able to perform estimation of the parametric and nonparametric components of the volatility when data are noisy; to do so, we need the noise not to be asymptotically bigger than the process of interest, in the sense that we want it to be $O_{\mathbb{P}}(n^{-1/2})$. Such a specification should allow us to give simple extensions of the previous results, based on approximation of quadratic variation. Estimation at the rate $n^{1/2}$ should be possible, while the best rate should be lower when errors are bigger and tools based on quadratic variation are usually not suitable. See Gloter and Jacod [8] for the attainable rate in a simple model with shrinking errors.

4. Numerical implementation

4.1. **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. Table 1 shows an example of available forward contracts in the French Market on May 23rd, 2015. For example, the contract called "June 2015" will deliver to the holder 1 MWh of electricity, with a constant power, between the first hour of June 1st (this is T_s) and the last hour of June 30th of 2015 (T_e). This table also introduces the contracts of relative maturity (denoted by the "ahead" formulation). A "ahead" contract is a contract with constant delivery period but with changing delivery dates. For example, the 2-month-ahead contract is the forward contract "July 2015" when it is quoted on May 31th, 2015 (2 months ahead from the quotation date), and becomes the forward contract "August 2015" on June 1st, 2015 (a jump of contract to stay 2 months ahead from the quotation date). In this study we only consider the 6 observable monthly contracts (*i.e.* $T_e - T_s = 1$ month) to estimate ϑ and the volatility processes σ and $\overline{\sigma}$. 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.

4.2. **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 [18], the volatility processes are constant, and the estimated values are $\sigma = 0.37$ y$^{-1/2}$ and $\overline{\sigma} = 0.15$ y$^{-1/2}$. Here we use a CIR-like model (the Cox-Ingersoll-Ross model for interest rates has been introduced in [5], 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 [10]). Our parameters are

$$b^j_t = 3.65 \cdot 10^{-1}(\log(30) - X^j_t), \sigma_t = 0.37\Sigma^d_t \text{ and } \overline{\sigma}_t = 0.15\Sigma^d_t,$$

Product	Example: May 23rd, 2015		
	Name of the product	Begin of delivery	End of delivery
1 Month Ahead	June 2015	2015-06-01	2015-06-30
2 Month Ahead	July 2015	2015-07-01	2015-07-31
3 Month Ahead	August 2015	2015-08-01	2015-08-31
4 Month Ahead	September 2015	2015-09-01	2015-09-30
5 Month Ahead	October 2015	2015-10-01	2015-10-31
6 Month Ahead	November 2015	2015-11-01	2015-11-30
1 Quarter Ahead	3rd quarter 2015	2015-07-01	2015-09-30
2 Quarter Ahead	4th quarter 2015	2015-10-01	2015-12-31
3 Quarter Ahead	1st quarter 2016	2016-01-01	2016-03-31
1 Year Ahead	Year 2016	2016-01-01	2016-12-31

TABLE 1. Data available on a daily basis.

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 y^{-1}): 1.4, 10, 20, 40. The first value is the estimated parameter shown in [18] 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 $\hat{\vartheta}_{j,n}$ the estimator of ϑ from the configuration where j processes are observed, and also by $\tilde{\vartheta}_{2,n}$ 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 $\tilde{\vartheta}_{2,n}$ 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 2, 3 and 4 give the estimation results for $\vartheta = 1.4$, 20 and 40 y^{-1}, respectively. In each configuration, these tables give the number of converging instances[1] of the estimator and their average, and the empirical confidence interval at 95% (issued from taking the quantiles of the sample of estimated values). We observe that the estimators perform quite well: except in three lines in Table 4, the true value of ϑ is always in the confidence 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 confidence interval,

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

and it may shift it away from the true value of ϑ. Notice also that the one-step correction from $\hat{\vartheta}_{2,n}$ to $\tilde{\vartheta}_{2,n}$ never led to very different values.

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

TABLE 2. Results of the estimation on simulated data with $\vartheta = 1.4$ y^{-1}.

Processes	Estimator	Instances that converged	Average	Quantile interval
2	$\hat{\vartheta}_{2,n}$	85,677	21.145	[10.677,47.124]
2	$\tilde{\vartheta}_{2,n}$	85,677	21.130	[10.673,47.047]
3	$\hat{\vartheta}_{3,n}$	100,000	19.672	[18.198,20.247]
4	$\hat{\vartheta}_{4,n}$	100,000	19.567	[18.233,20.204]
5	$\hat{\vartheta}_{5,n}$	100,000	19.518	[18.247,20.202]
6	$\hat{\vartheta}_{6,n}$	100,000	19.699	[18.739,20.362]

TABLE 3. Results of the estimation on simulated data with $\vartheta = 20$ y^{-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 4. Results of the estimation on simulated data with $\vartheta = 40$ y^{-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 and visual inspection: as the number of data is quite poor, the empirical criterion to be minimised in the cross validation method does not always admit a minimum. We therefore retain a value of h_n close to the values that are given by cross validation when the minimisation is well defined, and we check that it does not lead to obvious under- or oversmoothing. The retained value is 14 days. We also set $\varpi_n = 3.65 \cdot 10^{-2}$ y^{-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 $\vartheta = 10$ y^{-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 [18], that is $\sigma = 0.37$ y$^{-1/2}$ and $\overline{\sigma} = 0.15$ y$^{-1/2}$. A deterministic specification allows us to compare the curve of point estimates with the deterministic function that was used to simulate the processes. Remember that the nonparametric estimation result, Theorem 2, gives convergence uniformly on $[h_n, T]$. Therefore we expect that the fit is not good for values ot t being less than h_n. We perform simulation and estimation 10,000 times, and then take the average and the quantiles of the 10,000 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 10,000 occurrences of $\widehat{\sigma}^2_{n,t}$ and $\widehat{\overline{\sigma}}^2_{n,t}$). Figure 1 gives the square of the estimated equivalent volatility function

$$e^{-2\hat{\vartheta}_{2,n}(T_1-t)}\widehat{\sigma}^2_{n,t} + \widehat{\overline{\sigma}}^2_{n,t},$$

together with the true function $e^{-2\vartheta(T_1-t)}\sigma_t^2 + \overline{\sigma}_t^2$. It shows a good estimation of this equivalent volatility, the error (between the average of the 10,000 estimators and the true value) being maximal in the two ends of the curve. The estimation of $\overline{\sigma}_t^2$, given in Figure 2, also performs well. However, we can observe in Figure 3 a bad performance of estimation of σ_t^2, especially for large values of $T-t$, even when $t > h_n$. This can be explained by the fact that, due to the presence of the exponential term $e^{-\vartheta(T.-t)}$, the short term factor $e^{-2\vartheta(T.-t)}\sigma_t^2$ is low when $T-t$ is large. Also, if ϑ happens to be overestimated, the estimator of σ_t^2 has to take a very high value so that the product $\sigma_t^2 e^{-2\vartheta(T.-t)}$ may fit the curve. Therefore, 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.

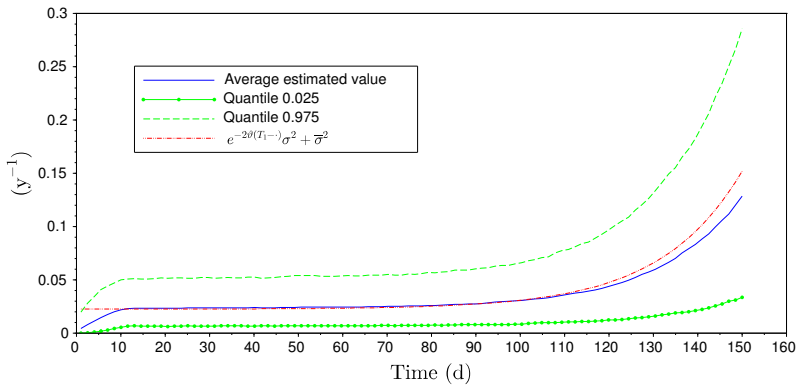

FIGURE 1. Quantiles for the square of the equivalent volatility, with 2 processes, $\vartheta = 10$ y^{-1} and deterministic constant volatilities

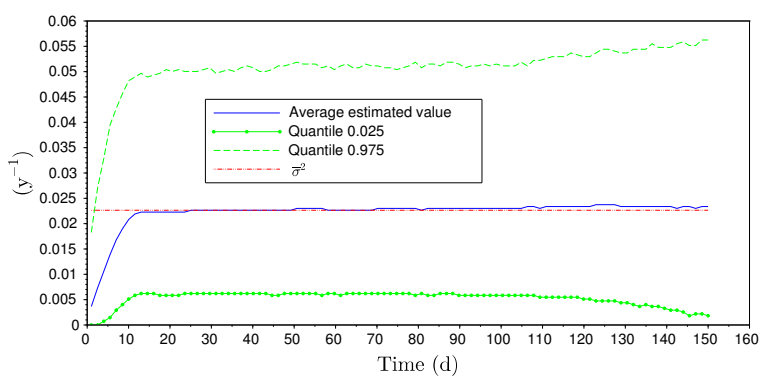

FIGURE 2. Quantiles for the square of the long-term volatility, with 2 processes, $\vartheta = 10$ y^{-1} and deterministic constant volatilities

Now, we are back to the specification $\sigma_t = 0.37\Sigma_t^d, \overline{\sigma}_t = 0.15\Sigma_t^d$. As the volatility processes depend on the path of X, we cannot compare visually the real volatility and its point estimators. Yet, we plot the average

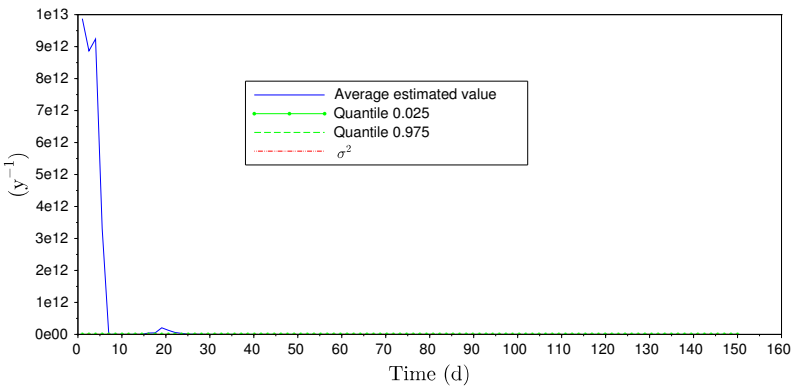

FIGURE 3. Quantiles for the square of the short-term volatility, with 2 processes, $\vartheta = 10$ y^{-1} and deterministic constant volatilities

and the quantile curves of the 10,000 estimators for the two volatility processes and for the equivalent square volatility process, in Figures 4, 5 and 6. The behaviours of the series of point estimators are very similar to the ones we described while considering deterministic volatility functions.

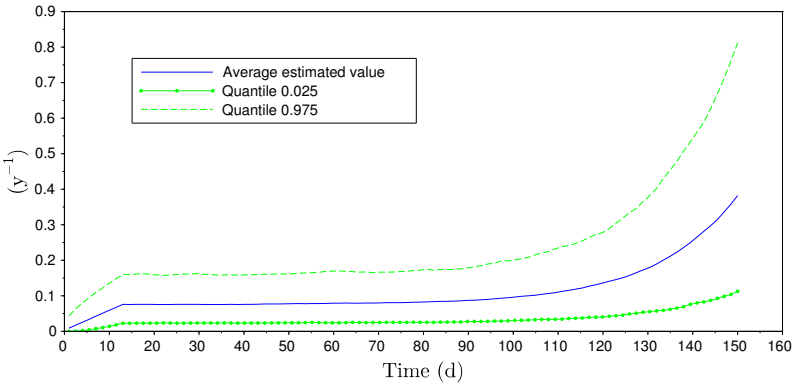

FIGURE 4. Quantiles for the square of the equivalent volatility, with 2 processes, $\vartheta = 10$ y^{-1} and the CIR-like specification for volatility processes

4.3. **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 December 31st, 2013. On this history, we get 125 periods of 1 month ($n \simeq 20$) where 6 processes (the 6 month-ahead contracts) are jointly observed with no missing data, whereas we get 141 periods of 5 months ($n \simeq 100$) where 2 processes (the 1 month-ahead and the 2 month-ahead contracts) are jointly observed. These numbers of periods are given in Table 5 for all the configurations described in Section 4.2. In the same column, Table 5 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. The main message about these results is that, 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, the estimators from 3 to 6 processes present

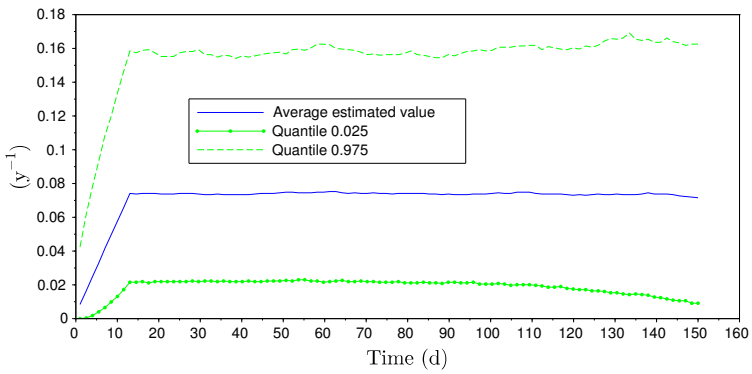

FIGURE 5. Quantiles for the square of the long-term volatility, with 2 processes, $\vartheta = 10$ y^{-1} and the CIR-like specification for volatility processes

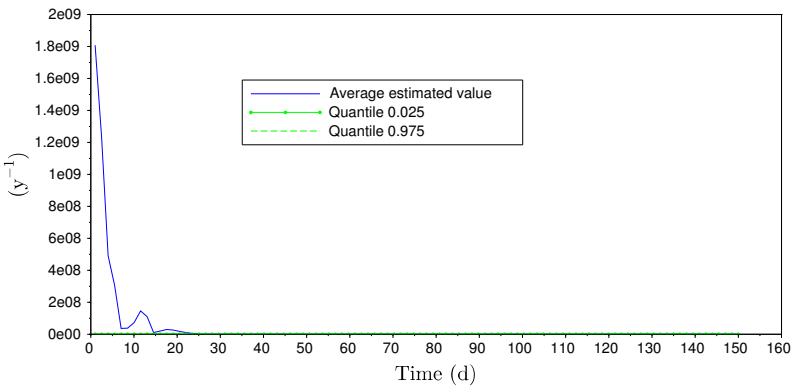

FIGURE 6. Quantiles for the square of the short-term volatility, with 2 processes, $\vartheta = 10$ y^{-1} and the CIR-like specification for volatility processes

Estimator	Per. with convergence/ Number of per.	Average	Standard deviation
$\hat{\vartheta}_{2,n}$	49/141	26.065	11.788
$\tilde{\vartheta}_{2,n}$	49/141	26.081	11.779
$\hat{\vartheta}_{3,n}$	100/142	4.3707	3.5329
$\hat{\vartheta}_{4,n}$	111/143	3.1333	2.6758
$\hat{\vartheta}_{5,n}$	111/139	2.0936	2.4969
$\hat{\vartheta}_{6,n}$	105/125	3.3881	2.8221

TABLE 5. Estimators of ϑ on real data in France (unit: y).

a theoretical bias, of which value is unknown: this was stated in Theorem 1 and in Section 3.1. Second, these differences may be due to the presence of errors linked to measurement or to the model.

5. PROOFS

5.1. **Preliminaries: localisation.** With no loss of generality, we may (and will) assume that the processes b^j, σ and $\overline{\sigma}$ are bounded, relying on a so-called localisation argument. For an integer $p \geq 1$, introduce the

stopping time $\tau_p = \inf\{t \in [0,T], \min(b^j_t, \sigma_t, \overline{\sigma}_t) > p\}$. Replacing X^j_t by $X_{t\wedge\tau_p}$, we have bounded processes b^j, σ and $\overline{\sigma}$. Moreover since these processes are at least locally bounded, we have $\mathbb{P}(\tau_p > T) \to 1$ as $p \to \infty$. We refer to Section 3.6.3 in [14] for details.

5.2. **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

$$\zeta^n_i = (\Delta^n_i X^2)^2 - (\Delta^n_i X^1)^2$$

and

$$\xi^n_i = \big(\Delta^n_i X^2 - \Delta^n_i X^1\big)^2.$$

Clearly

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

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

$$\zeta^n_i = \frac{1}{\psi_{T_1,T_2}(\vartheta)}\xi^n_i + \chi^n_i. \tag{7}$$

By standard convergence of the quadratic variation (see for instance Section 2.1.5 in [20]),

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

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

$$\begin{aligned}&\mathbb{E}\Big[\Big(\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t}\sigma_t dB_t \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t d\overline{B}_t\Big)^2\Big]\\ =&\sum_{i=1}^n \mathbb{E}\big[\big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t}\sigma_t dB_t\big)^2\big]\mathbb{E}\big[\big(\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t d\overline{B}_t\big)^2\big] \le \Delta_n e^{2\vartheta T} M^2 \to 0.\end{aligned}$$

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

$$\Psi^n_{T_1,T_2} = \frac{\sum_{i=1}^n \big(\Delta^n_i X^2 - \Delta^n_i X^1\big)^2}{\sum_{i=1}^n (\Delta^n_i X^2)^2 - (\Delta^n_i X^1)^2} = \frac{\sum_{i=1}^n \xi^n_i}{\sum_{i=1}^n \zeta^n_i} \to \psi_{T_1,T_2}(\vartheta)$$

in probability. We derive the convergence

$$\psi_{T_1,T_2}(\hat{\vartheta}_{2,n}) \to \psi_{T_1,T_2}(\vartheta)$$

in probability on the event $\{\Psi^n_{T_1,T_2} \in (-1,0)\}$, hence the convergence $\hat{\vartheta}_{2,n} \to \vartheta$ in probability as well since $\{\Psi^n_{T_1,T_2} \in (-1,0)\}$ 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

$$\Delta_n^{-1/2}\big(\Psi^n_{T_1,T_2} - \psi_{T_1,T_2}(\vartheta)\big) = \Delta_n^{-1/2}\Big(\frac{\sum_{i=1}^n \xi^n_i}{\sum_{i=1}^n \zeta^n_i} - \psi_{T_1,T_2}(\vartheta)\Big) = -\psi_{T_1,T_2}(\vartheta)\frac{\Delta_n^{-1/2}\sum_{i=1}^n \chi^n_i}{\sum_{i=1}^n \zeta^n_i}.$$

Consider next the sequence of 1-dimensional processes

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

where $Y_t = (Y_t^1, Y_t^2) = \big(\int_0^t e^{\vartheta s}\sigma_s dB_s, \int_0^t \overline{\sigma}_s d\overline{B}_s\big)$. By Theorem 3.21, p. 231 in [14] 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

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

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 $\chi_n \to \chi$ holds stably in law and the convergence

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

in probability, we derive

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

in distribution. Conditional on $\mathcal{F}$, the limiting variable is centred Gaussian, with conditional variance $v_\vartheta(\sigma,\overline{\sigma}) = 4\frac{(e^{-\vartheta T_2}-e^{-\vartheta T_1})^2}{(e^{-\vartheta T_2}+e^{-\vartheta T_1})^4}\frac{\int_0^T e^{2\vartheta t}\sigma_t^2\overline{\sigma}_t^2 dt}{(\int_0^T e^{2\vartheta t}\sigma_t^2 dt)^2}$.

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

$$\Delta_n^{-1/2}\big(\hat{\vartheta}_{2,n} - \vartheta\big) = \Delta_n^{-1/2}\big(\Psi^n_{T_1,T_2} - \psi_{T_1,T_2}(\vartheta)\big)\partial_\vartheta\psi^{-1}_{T_1,T_2}(Z_n)$$

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

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

together with the fact that $\{\Psi^n_{T_1,T_2} \in (-1,0)\}$ 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. Then, by Girsanov theorem, we apply a change of measure which is $\mathcal{F}$-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 [20] for a simple explanation)). Finally, relaxing the boundedness assumption on $\sigma, \overline{\sigma}$ and b^j is standard, see Section 5.1 above.

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

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

By standard convergence of the quadratic variation

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

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 $\psi_{T_1,T_2,T_3}\big(\hat{\vartheta}_{n,3}\big) \to \psi_{T_1,T_2,T_3}(\vartheta)$ in probability on the event $\big\{\Psi_{T_1,T_2,T_3} \in \big(0, \big(\frac{T_3-T_2}{T_2-T_1}\big)^2\big)\big\}$ which has asymptotically probability 1, hence the convergence $\hat{\vartheta}_{n,3} \to \vartheta$ in probability.

Step 2. We further have

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

with

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

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

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

and

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

We will need the following lemma, proof of which is relatively straightforward yet technical and given in Section 6.1.

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*

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

and

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

in probability.

We successively have

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

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

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

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

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

in probability.

Step 3. Finally, we have

$$\Delta_n^{-1}\big(\hat{\vartheta}_{3,n} - \vartheta\big) = \Delta_n^{-1}\big(\Psi^n_{T_1,T_2,T_3} - \psi_{T_1,T_2,T_3}(\vartheta)\big)\partial_\vartheta\psi^{-1}_{T_1,T_2,T_3}(Z_n),$$

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

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

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

5.3. **Proof of Theorem 2.** We may (and will) assume that $b^j = 0$. For ease of notation, we write $\hat{\vartheta}_{2,n}$ for $\max(\hat{\vartheta}_{2,n}, \varpi_n)$ 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

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

with

$$I = \Big(\frac{1}{e^{-2\hat{\vartheta}_{2,n}(T_1-t)} - e^{-2\hat{\vartheta}_{2,n}(T_2-t)}} - \frac{1}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}}\Big) \times \sum_{i=1}^n K_{h_n}\big(t - t_{i-1}\big)\big((\Delta^n_i X^1)^2 - (\Delta^n_i X^2)^2\big)$$

and

$$II = \frac{\sum_{i=1}^n K_{h_n}\big(t - t_{i-1}\big)\big((\Delta^n_i X^1)^2 - (\Delta^n_i X^2)^2\big)}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}} - \sigma^2_t.$$

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

$$\mathbb{E}\Big[\Big|\sum_{i=1}^n K_{h_n}(t - t_{i-1})\big((\Delta^n_i X^1)^2 - (\Delta^n_i X^2)^2\big)\Big|\Big] \lesssim \sum_{i=1}^n K_{h_n}(t - t_{i-1})\Delta_n \lesssim 1$$

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 $\sum_{i=1}^n K_{h_n}\big(t - t_{i-1}\big)\big((\Delta^n_i X^1)^2 - (\Delta^n_i X^2)^2\big)$ 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 $II = (e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)})^{-1}\big(B_n(t) + V_n(t)\big)$, having

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

and

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

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

$$\sup_{t\in[h_n,T]} \mathbb{E}\Big[\Big(\sum_{i=1}^n K_{h_n}(t - t_{i-1})\big((\Delta^n_i X^1)^2 - (\Delta^n_i X^2)^2 - \mathbb{E}\big[(\Delta^n_i X^1)^2 - (\Delta^n_i X^2)^2\big|\mathcal{F}_{i-1}\big]\big)\Big)^2\Big]$$
$$= \sup_{t\in[h_n,T]} h_n^{-2}\sum_{i=1}^n K\Big(\frac{t - t_{i-1}}{h_n}\Big)^2 \mathbb{E}\Big[\big((\Delta^n_i X^1)^2 - (\Delta^n_i X^2)^2 - \mathbb{E}\big[(\Delta^n_i X^1)^2 - (\Delta^n_i X^2)^2\big|\mathcal{F}_{i-1}\big]\big)^2\Big]$$

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

$$\mathbb{E}\Big[\big((\Delta^n_i X^1)^2 - (\Delta^n_i X^2)^2 - \mathbb{E}\big((\Delta^n_i X^1)^2 - (\Delta^n_i X^2)^2\big|\mathcal{F}_{i-1}\big)\big)^2\Big] \lesssim \Delta_n^2,$$

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

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

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

where

$$III = \int_0^T h_n^{-1}K\Big(\frac{t-u}{h_n}\Big)e^{-2\vartheta(t-u)}\sigma^2_u du - \sigma^2_t$$

and

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

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

$$\begin{aligned}\mathbb{E}\big[III^2\big] &= \mathbb{E}\Big[\Big(\int_{\frac{t-T}{h_n}}^{\frac{t}{h_n}} K(x)e^{-2\vartheta h_n x}\sigma^2_{t-h_n x}dx - \sigma_t^2\Big)^2\Big]\\ &= \mathbb{E}\Big[\Big(\int_{\mathrm{supp}(K)} K(x)\big(e^{-2\vartheta h_n x}\sigma^2_{t-h_n x}-\sigma_t^2\big)dx\Big)^2\Big]\\ &\le \int_{\mathrm{supp}(K)} K(x)^2\mathbb{E}\big[\big(e^{-2\vartheta h_n x}\sigma^2_{t-h_n x}-\sigma_t^2\big)^2\big]dx\end{aligned}$$

by Jensen inequality since $\mathrm{supp}(K)\subset\big[\frac{t-T}{h},\frac{t}{h}\big]$ that integrates to 1. By convexity,

$$(e^{-2\vartheta h_n x}\sigma^2_{t-h_n x}-\sigma_t^2)^2 \le 2(e^{-2\vartheta h_n x}\sigma^2_{t-h_n x}-\sigma^2_{t-h_n x})^2 + 2(\sigma^2_{t-h_n x}-\sigma_t^2)^2$$

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

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

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

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

since the term

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

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

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

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

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

and the difference $K\Big(\frac{t-t_{j-1}}{h_n}\Big)-K\Big(\frac{t-u}{h_n}\Big)$ 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,

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

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

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

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

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 $\widehat{\overline{\sigma}}^2$. We split $\widehat{\overline{\sigma}}^2_{n,t} - \overline{\sigma}^2_t$ as follows

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

and proceed analogously. The proof of Theorem 2 is complete.

5.4. **Proof of Theorem 3.** With no loss of generality, we may (and will) assume that $b^1 = b^2 = 0$.

Preliminaries on efficient semiparametric estimation. We refer to Sections 25.3–25.4 of [21] for a comprehensive presentation of efficient semiparametric estimation, that we need to adapt to our framework. Assuming $(\sigma, \overline{\sigma})$ to be deterministic, the data $(\Delta_i^n X^1, \Delta_i^n X^2), i = 1, \dots, n$ generate a product experiment $\mathcal{E}^n = \otimes_{i=1}^n \mathcal{P}^{i,n}$, where

$$\mathcal{P}^{i,n} = \big\{f^{i,n}_{\vartheta,\sigma,\overline{\sigma}}, (\vartheta, \sigma, \overline{\sigma}) \in [0,\infty) \times \Sigma(c, \tilde{c}\big\},$$

where $f^{i,n}_{\vartheta,\sigma,\overline{\sigma}}$ 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.

For $\vartheta \in (0,\infty)$, let $\varepsilon > 0$ and $\iota \in \mathbb{R}$ be such that $\vartheta + \iota\varepsilon > 0$. Pick a path

$$\big(t \rightsquigarrow (\eta^u_t, \overline{\eta}^u_t)\big)_{0\le u\le \varepsilon}, \text{ such that } (\eta^0_t, \overline{\eta}^0_t) = (\sigma_t, \overline{\sigma}_t) \text{ for every } t \in [0,T],$$

for two regular functions $t \rightsquigarrow \eta^u_t$ and $t \rightsquigarrow \overline{\eta}^u_t$ satisfying moreover

$$\tilde{c} < \inf_{t\in[0,T], u\in[0,\varepsilon]} \min(\eta^u_t, \overline{\eta}^u_t) \le \sup_{t\in[0,T], u\in[0,\varepsilon]} \max(\eta^u_t, \overline{\eta}^u_t) < c.$$

We obtain a parametric submodel of $\mathcal{P}^{i,n}$ around $\big(t \rightsquigarrow (\sigma^2_t, \overline{\sigma}^2_t)\big)$ by setting

$$\mathcal{P}^{i,n}_0 = \{f^{i,n}_{\vartheta+\iota u, \eta^u, \overline{\eta}^u}, 0 \le u \le \varepsilon\},$$

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

$$g^{i,n}_{\iota,\eta,\overline{\eta}} = \iota \ell^{i,n}_{\vartheta,\sigma,\overline{\sigma}} + g^{i,n}_{\eta,\overline{\eta}}, \tag{9}$$

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

Completion of proof of Theorem 3. From (6), and the explicit representation

$$f^{i,n}_{\vartheta,\sigma,\overline{\sigma}}(x,y)=\frac{\exp\Big(-\frac{\big(\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2dt\big)(x-y)^2+\big(\int_{(i-1)\Delta_n}^{i\Delta_n}e^{-2\vartheta(T_1-t)}\sigma_t^2dt\big)\big(y-e^{-\vartheta(T_2-T_1)}x\big)^2}{2\big(\int_{(i-1)\Delta_n}^{i\Delta_n}e^{-2\vartheta(T_1-t)}\sigma_t^2dt\big)\big(\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2dt\big)\big(1-e^{-\vartheta(T_2-T_1)}\big)^2}\Big)}{2\pi\big(\int_{(i-1)\Delta_n}^{i\Delta_n}e^{-2\vartheta(T_1-t)}\sigma_t^2dt\big)^{1/2}\big(\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2dt\big)^{1/2}\big(1-e^{-\vartheta(T_2-T_1)}\big)},$$

we derive

$$\begin{aligned}\partial_\vartheta\log f^{i,n}_{\vartheta,\sigma,\overline{\sigma}}(\Delta_i^nX^1,\Delta_i^nX^2)&=\frac{\int_{(i-1)\Delta_n}^{i\Delta_n}(T_1-t)e^{-2\vartheta(T_1-t)}\sigma_t^2dt}{\int_{(i-1)\Delta_n}^{i\Delta_n}e^{-2\vartheta(T_1-t)}\sigma_t^2dt}-\frac{e^{-\vartheta(T_2-T_1)}(T_2-T_1)}{1-e^{-\vartheta(T_2-T_1)}}\\&\quad+(\Delta_i^nX^2-\Delta_i^nX^1)^2\frac{\int_{(i-1)\Delta_n}^{i\Delta_n}(e^{-\vartheta(T_2-T_1)}(T_2-t)-(T_1-t))e^{-2\vartheta(T_1-t)}\sigma_t^2dt}{\big(\int_{(i-1)\Delta_n}^{i\Delta_n}e^{-2\vartheta(T_1-t)}\sigma_t^2dt\big)^2(1-e^{-\vartheta(T_2-T_1)})^3}\\&\quad+(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)^2\frac{(T_2-T_1)e^{-\vartheta(T_2-T_1)}}{\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2dt(1-e^{-\vartheta(T_2-T_1)})^3}\\&\quad-\Delta_i^nX^1(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)\frac{e^{-\vartheta(T_2-T_1)}(T_2-T_1)}{\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2dt(1-e^{-\vartheta(T_2-T_1)})^2}.\end{aligned}$$

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

$$g^{i,n}_{\iota,k,\overline{k}}=\iota\ell^i_{\vartheta,\sigma,\overline{\sigma}}+g^{i,n}_{k,\overline{k}}.$$

according to (9) and parametrised by $(k,\overline{k})$. Formally,

$$\iota\ell^{i,n}_{\vartheta,\sigma,\overline{\sigma}}=\big(\tfrac{d}{du}\log f^{i,n}_{\vartheta+\iota u,\sigma,\overline{\sigma}}\big)_{u=0}\ \text{ and }\ g^{i,n}_{k,\overline{k}}=\big(\tfrac{d}{du}\log f^{i,n}_{\vartheta,\eta^u,\overline{\eta}^u}\big)_{u=0},$$

so that

$$\begin{aligned}g^{i,n}_{k,\overline{k}}&=-\frac{\int_{(i-1)\Delta_n}^{i\Delta_n}e^{-2\vartheta(T_1-t)}\sigma_t^2k(t)dt}{\int_{(i-1)\Delta_n}^{i\Delta_n}e^{-2\vartheta(T_1-t)}\sigma_t^2dt}-\frac{\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2\overline{k}(t)dt}{\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2dt}\\&\quad+(\Delta_i^nX^2-\Delta_i^nX^1)^2\frac{\int_{(i-1)\Delta_n}^{i\Delta_n}e^{-2\vartheta(T_1-t)}\sigma_t^2k(t)dt}{\big(\int_{(i-1)\Delta_n}^{i\Delta_n}e^{-2\vartheta(T_1-t)}\sigma_t^2dt\big)^2(1-e^{-\vartheta(T_2-T_1)})^2}\\&\quad+(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)^2\frac{\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2\overline{k}(t)dt}{\big(\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2dt\big)^2(1-e^{-\vartheta(T_2-T_1)})^2}.\end{aligned}$$

Introduce the orthogonal projection Π onto (the closure of) of $\mathrm{Span}\{g^{i,n}_{k,\overline{k}},\text{for all admissible }(k,\overline{k})\}$. Then $\widetilde{\ell}^{i,n}_{\vartheta,\sigma,\overline{\sigma}}=\ell^{i,n}_{\vartheta,\sigma,\overline{\sigma}}-\Pi\ell^{i,n}_{\vartheta,\sigma,\overline{\sigma}}$ is the efficient score for ϑ and $\widetilde{I}^{i,n}_{\vartheta,\sigma,\overline{\sigma}}=\mathbb{E}_{\vartheta,\sigma,\overline{\sigma}}\big[(\widetilde{\ell}^{i,n}_{\vartheta,\sigma,\overline{\sigma}})^2\big]$ is the best achievable information bound, see Sections 25.3–25.4 of [21] for details. By orthogonality,

$$\mathbb{E}_{\vartheta,\sigma,\overline{\sigma}}\big[(\ell^{i,n}_{\vartheta,\sigma,\overline{\sigma}}-\Pi\ell^{i,n}_{\vartheta,\sigma,\overline{\sigma}})g^{i,n}_{k,\overline{k}}\big]=0,\ \text{ for all admissible }\ (k,\overline{k}),$$

and anticipating further the representation $\Pi\ell^{i,n}_{\vartheta,\sigma,\overline{\sigma}}=g^{i,n}_{k^\star,\overline{k}^\star}$ for some admissible $(k^\star,\overline{k}^\star)$, it suffices to solve

$$0=\mathbb{E}_{\vartheta,\sigma,\overline{\sigma}}\big[(l^{i,n}_{\vartheta,\sigma,\overline{\sigma}}-g^{i,n}_{k^\star,\overline{k}^\star})g^{i,n}_{k,\overline{k}}\big]\quad\text{for all admissible }\ (k,\overline{k}).$$

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

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

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

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

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

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

as $n\to\infty$. Taking the inverse and dividing by Δ_n we obtain the desired bound. The proof of Theorem 3 is complete.

5.5. **Proof of Theorem 4.** The first assertion was obtained in Section 5.4 in the course of the proof of Theorem 3. With no loss of generality, we may (and will) assume that $b^i = 0$ and that $\sup_t \max(\sigma_t, \overline{\sigma}_t) \le M$. 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

$$\Delta_n^{1/2}\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}^{i,n}_{\vartheta_n,\overline{\sigma}} - \widetilde{\ell}^{i,n}_{\vartheta_n,\widehat{\overline{\sigma}}_n}\big)\to 0 \tag{10}$$

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

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

with

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

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

$$\begin{aligned}
&\mathbb{E}\big[\big|\Delta_n^{-1/2}(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\frac{\Delta_n\overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2 dt}{\overline{\sigma}^2_{t_{i-1}}\int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2 dt}\big|\big]\\
&\le \Delta_n^{-3/2}\tilde{c}^{-4}\mathbb{E}\big[\big((\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\big)^2\big]^{1/2}\mathbb{E}\big[\big(\Delta_n\overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2 dt\big)^2\big]^{1/2}.
\end{aligned}$$

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

$$\begin{aligned}
&\mathbb{E}\big[\big((\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\big)^2\big] \lesssim \Delta_n^2,\\
&\mathbb{E}\big[\big(\Delta_n\overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2 dt\big)^2\big] \lesssim \Delta_n^{2(1+\alpha)}.
\end{aligned}$$

We infer $\mathbb{E}[|I|] \lesssim \sum_{i\in\mathcal{I}_n}\Delta_n^{-3/2}\Delta_n\Delta_n^{1+\alpha} \lesssim \Delta_n^{\alpha-1/2}\to 0$ 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 $\widehat{\overline{\sigma}}^2_{n,t_{i-1}}$ is

$\mathcal{F}_{i-1}$-measurable. Conditioning on $\mathcal{F}_{i-1}$, we set

$$\begin{aligned}\chi_i^n &= \mathbb{E}\big[\Delta_n^{-1/2}(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\tfrac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\big|\mathcal{F}_{i-1}\big]\\ &= \Delta_n^{-1/2}(e^{-\vartheta(T_2-T_1)} - e^{-\vartheta_n(T_2-T_1)})(e^{-\vartheta(T_2-T_1)}-1)\xi_i^n,\end{aligned}$$

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

$$\sum_{i=1}^n \mathbb{E}\big[|\xi_i^n|\big|\mathcal{F}_{i-1}\big] \le \Delta_n M^2\tilde{c}^{-4}\sum_{i=1}^n \sup_{i\in\mathcal{I}_n}\big|\widehat{\overline{\sigma}}^2_{n,t_{i-1}} - \overline{\sigma}^2_{t_{i-1}}\big| \to 0$$

in probability by Theorem 2. Since $\Delta_n^{-1/2}\big(e^{-\vartheta(T_2-T_1)} - e^{-\vartheta_n(T_2-T_1)}\big)$ is bounded, we use Lemma 3.4 in [14] applied to variables ξ_i^n to conclude $\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\mathbb{E}[\chi_i^n\,|\,\mathcal{F}_{i-1}] \to 0$ in probability, locally uniformly in t. Moreover,

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

which is of order $\Delta_n^{-1}\Delta_n^2\Delta_n^{\alpha/(2\alpha+1)}$ so that $\sum_{i\in\mathcal{I}_n}\mathbb{E}\big[\big(\chi_i^n\big)^2\big|\mathcal{F}_{i-1}\big] \lesssim \Delta_n^{\alpha/(2\alpha+1)} \to 0$ in probability. Applying Lemma 3.4 in [14] to the sequence χ_i^n enables us to conclude that II converges to 0 in probability and (10) follows.

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

$$\sqrt{n}\big(\widetilde{\ell}^{i,n}_{\vartheta_n,\overline{\sigma}} - \widetilde{\ell}^{i,n}_{\vartheta,\overline{\sigma}}\big) - \sqrt{n}(\vartheta_n - \vartheta)\partial_\vartheta\widetilde{\ell}^{i,n}_{\vartheta,\overline{\sigma}} \to 0 \tag{11}$$

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

$$\Delta_n\sum_{i\in\mathcal{I}_n}\partial_\vartheta\widetilde{\ell}^{i,n}_{\vartheta,\overline{\sigma}} \sim -\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}} = -\frac{(T_2-T_1)^2}{(e^{\vartheta T_2}-e^{\vartheta T_1})^2}\int_0^T\frac{e^{2\vartheta t}\sigma_t^2}{\overline{\sigma}_t^2}dt \tag{12}$$

in probability under $\mathbb{P}_{\vartheta,\sigma,\overline{\sigma}}$, that is the total Fisher information associated to the efficient scores $\widetilde{\ell}^{i,n}_{\vartheta,\overline{\sigma}}$. Summing each term in (11) for $i\in\mathcal{I}_n$, using (12) and the fact that Δ_n^{-1} and n are of the same order, we further infer

$$\Delta_n^{1/2}\Big(\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}^{i,n}_{\vartheta_n,\widehat{\overline{\sigma}}_n} - \widetilde{\ell}^{i,n}_{\vartheta,\overline{\sigma}}\big) + \Delta_n^{-1}(\vartheta_n-\vartheta)\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}\Big) \to 0 \tag{13}$$

in probability, using (10) in order to substitute $\overline{\sigma}$ by $\widehat{\overline{\sigma}}_n$ in the first term. Moreover, (13) 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 [21] for the details.

Step 3. We establish

$$\Delta_n\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}^{i,n}_{\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}_n}\big)^2 \to \widetilde{I}_{\vartheta,\sigma,\overline{\sigma}} \tag{14}$$

in probability under $\mathbb{P}_{\vartheta,\sigma,\overline{\sigma}}$. The computations are similar to Step 1, combining Theorems 1 and 2. We briefly give the mains steps. Observe that

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

with

$$I_i = \Delta_n \frac{(\Delta_i^n X^2 - \Delta_i^n X^1)^2}{\left(\Delta_n \widehat{\overline{\sigma}}^2_{n,t_{i-1}}\right)^2} \Big(\frac{(\Delta_i^n X^2 - e^{-\vartheta_n (T_2-T_1)} \Delta_i^n X^1)^2 e^{-2\vartheta_n (T_2-T_1)}}{(1-e^{-\vartheta_n (T_2-T_1)})^6} - \frac{(\Delta_i^n X^2 - e^{-\vartheta (T_2-T_1)} \Delta_i^n X^1)^2 e^{-2\vartheta (T_2-T_1)}}{(1-e^{-\vartheta (T_2-T_1)})^6} \Big),$$

$$II_i = \Delta_n (\Delta_i^n X^2 - \Delta_i^n X^1)^2 \frac{(\Delta_i^n X^2 - e^{-\vartheta (T_2-T_1)} \Delta_i^n X^1)^2 e^{-2\vartheta (T_2-T_1)}}{(1-e^{-\vartheta (T_2-T_1)})^6} \Big(\frac{1}{\left(\Delta_n \widehat{\overline{\sigma}}^2_{n,t_{i-1}}\right)^2} - \frac{1}{\left(\Delta_n \overline{\sigma}^2_{t_{i-1}}\right)^2} \Big),$$

$$III_i = \Delta_n (\Delta_i^n X^2 - \Delta_i^n X^1)^2 \frac{(\Delta_i^n X^2 - e^{-\vartheta (T_2-T_1)} \Delta_i^n X^1)^2 e^{-2\vartheta (T_2-T_1)}}{(1-e^{-\vartheta (T_2-T_1)})^6} \Big(\frac{1}{\left(\Delta_n \overline{\sigma}^2_{t_{i-1}}\right)^2} - \frac{1}{\left(\int_{t_{i-1}}^{t_i} \overline{\sigma}_t^2 dt\right)^2} \Big).$$

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

$$\Delta_n \sum_{i\in\mathcal{I}_n} \big(\widetilde{\ell}^{i,n}_{\vartheta_n,\widehat{\overline{\sigma}}_n}\big)^2 - \Delta_n \sum_{i\in\mathcal{I}_n} \big(\widetilde{\ell}^{i,n}_{\vartheta,\overline{\sigma}}\big)^2 \to 0$$

in probability and the result remains true with $\hat{\vartheta}_{2,n}$ in place of ϑ_n. Since $\Delta_n \sum_{i\in\mathcal{I}_n} \big(\widetilde{\ell}^{i,n}_{\vartheta,\overline{\sigma}}\big)^2 \sim \widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}$ in probability under $\mathbb{P}_{\vartheta,\sigma,\overline{\sigma}}$ we obtain (14).

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

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

in probability under $\mathbb{P}_{\vartheta,\sigma,\overline{\sigma}}$ thanks to (13) and (14) established in the two previous steps. We further write

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

and we claim that

$$\Delta_n^{1/2} \sum_{i=1}^{\lfloor h_n \Delta_n^{-1} \rfloor} \widetilde{\ell}^{i,n}_{\vartheta,\overline{\sigma}} \to 0 \tag{15}$$

in probability. We conclude the proof using the following limit theorem, proof of which is delayed until Appendix 6.2.

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

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

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

In view of Lemma 2 we obtain Theorem 4.

Step 5. It remains to prove (15). Write $\Delta_n^{1/2} \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} = \frac{(T_2-T_1)e^{-\vartheta(T_2-T_1)}}{(1-e^{-\vartheta(T_2-T_1)})^3}(I_i + II_i)$, with

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

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

We readily have

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

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

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

since $\alpha \geq 1/2$, and (15) follows. The proof of Theorem 4 is complete.

6. Appendix

6.1. 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

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

with

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

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

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

as soon as $\Delta \leq \eta$ which is true for large enough n. Thus $I \to 0$ 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

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

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,

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

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

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

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

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

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

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

It follows that

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

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

6.2. **Proof of Lemma 2.** First we establish the result for

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

by applying Lemma 3.7 in [14]. To do so, we check conditions (3.43)–(3.46) in [14] for χ_i^n. We keep up with the notation of [14]. First, we have $\mathbb{E}[\chi_i^n|\mathcal{F}_{i-1}] = 0$, which ensures (3.43) with $A_t = 0$. Next,

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

so that

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

in probability. This is Condition (3.44) in [14] with $C_t = \int_0^t\frac{e^{-2\vartheta(T_1-s)}\sigma_s^2}{\overline{\sigma}_s^2}ds$. It follows that

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

by independence of the two Wiener integrals. Therefore $\sum_{i=1}^n\mathbb{E}[(\chi_i^n)^4|\mathcal{F}_{i-1}] \to 0$ in probability. This is condition (3.45) in [14]. Finally, $\mathbb{E}\big[\chi_i^n\Delta_i^nB\big|\mathcal{F}_{i-1}\big] = \mathbb{E}\big[\chi_i^n\Delta_i^n\overline{B}\big|\mathcal{F}_{i-1}\big] = 0$ by independence which ensures condition (3.46) in [14]. We subsequently apply Lemma 3.7 in [14] 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 $\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}$. In order to complete the proof, we write

$$\Delta_n^{1/2}\sum_{i=1}^n \widetilde{\ell}^{i,n}_{\vartheta,\overline{\sigma}} = \Delta_n^{1/2}\sum_{i=1}^n \Big(\widetilde{\ell}^{i,n}_{\vartheta,\overline{\sigma}} - \frac{T_2-T_1}{e^{\vartheta(T_2-T_1)}-1}\sum_{i=1}^n \chi_i^n\Big) + \frac{T_2-T_1}{e^{\vartheta(T_2-T_1)}-1}\Delta_n^{1/2}\sum_{i=1}^n \chi_i^n$$

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

6.3. **Further technical results.**

Lemma 3. *Let $0<\alpha<\beta<\gamma$. The functions $f(x)=\frac{e^{-\beta x}-1}{e^{-\alpha x}-1}$ and $g(x)=\frac{e^{-\gamma x}-e^{-\beta x}}{e^{-\beta x}-e^{-\alpha x}}$ are decreasing on $(0,\infty)$.*

Proof. Both f and g are smooth. For $x>0$, we have $f'(x)=\frac{-\beta e^{-\beta x}(e^{-\alpha x}-1)+\alpha e^{-\alpha x}(e^{-\beta x}-1)}{(e^{-\alpha x}-1)^2}$ and $f'(x)=0$ is equivalent to

$$\tfrac{\beta}{\alpha}e^{-(\beta-\alpha)x} = f(x).$$

Assume on the contrary that f is increasing on a given subinterval of $(0,\infty)$. Since $\lim_{x\to 0_+} f'(x)=\frac{\beta}{\alpha}(-\beta+\alpha)<0$, there exists $0<x_1<x_2$ such that $f(x_1)\le f(x_2)$, and $f'(x_1)=f'(x_2)=0$. Thus, writing the previous equation for $x=x_1$ and $x=x_2$ and substracting the first equation to the second one leads to

$$\tfrac{\beta}{\alpha}\big(e^{-(\beta-\alpha)x_2}-e^{-(\beta-\alpha)x_1}\big) = f(x_2)-f(x_1).$$

The LHS is negative and RHS is positive, a contradiction. It follows that $g(x)=\frac{e^{-\gamma x}-e^{-\alpha x}}{e^{-\beta x}-e^{-\alpha x}}-1=\frac{e^{-(\gamma-\alpha)x}-1}{e^{-(\beta-\alpha)x}-1}-1$. Using the first part of the lemma we obtain that g is decreasing too. □

References

[1] Fred Espen Benth and Steen Koekebakker. Stochastic modeling of financial electricity contracts. *Energy Economics*, 30(3):1116–1157, 2008.

[2] Ramaprasad Bhar and Carl Chiarella. Interest rate futures: estimation of volatility parameters in an arbitrage-free framework. *Applied Mathematical Finance*, 4(4):181–199, 1997.

[3] Ramaprasad Bhar, Carl Chiarella, and Thuy-Duong Tô. A maximum likelihood approach to estimation of Heath-Jarrow-Morton models. Technical Report 80, Quantitative Finance Research Centre, University of Technology, Sydney, 2002.

[4] 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.

[5] 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.

[6] Valentine Genon-Catalot and Jean Jacod. On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Instit. H. Poincaré*, 29(1):119–151, 1993.

[7] 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.

[8] Arnaud Gloter and Jean Jacod. Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Statist.*, 5:225–242, 2001.

[9] 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.

[10] Juri Hinz, Lutz von Grafenstein, Michel Verschuere, and Martina Wilhelm. Pricing electricity risk by interest rate methods. *Quantitative Finance*, 5(1):49–60, 2005.

[11] Marc Hoffmann. Minimax estimation of the diffusion coefficient through irregular samplings. *Statistics & Probability Letters*, 32(1):11–24, 1997.

[12] Marc Hoffmann. Adaptive estimation in diffusion processes. *Stochastic Process. Appl.*, 79(1):135–163, 1999.

[13] Marc Hoffmann, Axel Munk, and Johannes Schmidt-Hieber. Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. *Ann. Instit. H. Poincaré*, 48(4):1186–1216, 2012.

[14] 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.

[15] Jean Jacod and Philip Protter. *Discretization of processes*, volume 67 of *Stochastic Modelling and Applied Probability*. Springer, Heidelberg, 2012.

[16] 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.

[17] 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.

[18] Rüdiger Kiesel, Gero Schindlmayr, and Reik H. Börger. A two-factor model for the electricity forward market. *Quantitative Finance*, 9(3):279–287, 2009.
[19] Steen Koekebakker and Fridthjof Ollmar. Forward curve dynamics in the nordic electricity market. *Managerial Finance*, 31(6):73–94, 2005.
[20] 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.
[21] Aad Van der Vaart. *Asymptotic Statistics*. University of Cambridge, 1998.
[22] Lan Zhang. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12(6):1019–1043, 2006.
[23] Lan Zhang, Per A. Mykland, and Yacine Aït-Sahalia. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.

Olivier Féron, EDF Lab Paris-Saclay and FiME, Laboratoire de Finance des Marchés de l'Energie, 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'Energie, 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