IMPROVED VOLATILITY ESTIMATION BASED ON LIMIT ORDER BOOKS *

By Markus Bibinger, Moritz Jirak, Markus Reiss

Humboldt-Universität zu Berlin

For a semi-martingale X_t, which forms a stochastic boundary, a rate-optimal estimator for its quadratic variation $\langle X, X\rangle_t$ is constructed based on observations in the vicinity of X_t. The problem is embedded in a Poisson point process framework, which reveals an interesting connection to the theory of Brownian excursion areas. A major application is the estimation of the integrated squared volatility of an efficient price process X_t from intra-day order book quotes. We derive $n^{-1/3}$ as optimal convergence rate of integrated squared volatility estimation in a high-frequency framework with n observations (in mean). This considerably improves upon the classical $n^{-1/4}$-rate obtained from transaction prices under microstructure noise.

1. Introduction. Consider observations $(\mathcal{Y}_i)$ above a stochastic boundary $(X_t,\, t \in [0,1])$, which is formed by the graph of a continuous semi-martingale. The objective is to optimally recover the driving characteristic $\langle X, X\rangle_t$ of the boundary X_t, given the observations $(\mathcal{Y}_i)$. Such (stochastic) frontier models naturally arise in many applications and a quantification of the information content in these observations is non-trivial. We formulate the problem with an emphasis on the financial context of limit order books. From a microeconomic point of view ask prices will always lie above the efficient market price. Here the underlying latent efficient log-price of a stock $(X_t,\, t \in [0,1])$, observed over a trading period like a day, serves as the boundary, whereas ask prices form the observations $(\mathcal{Y}_i)$. Bid prices can be handled symmetrically and independently, which can be used to validate the model.

Let the continuous Itô semi-martingale

$$X_t = X_0 + \int_0^t a_s\, ds + \int_0^t \sigma_s\, dW_s\,, t \in [0,1], \tag{1.1}$$

*Financial support from the Deutsche Forschungsgemeinschaft via SFB 649 *Ökonomisches Risiko* and FOR 1735 *Structural Inference in Statistics: Adaptation and Efficiency* is gratefully acknowledged.

AMS 2000 subject classifications: Primary 60H30; secondary 60G55

Keywords and phrases: Brownian excursion area, limit order book, integrated volatility, Feynman–Kac, high-frequency data, Poisson point process, nonparametric minimax rate

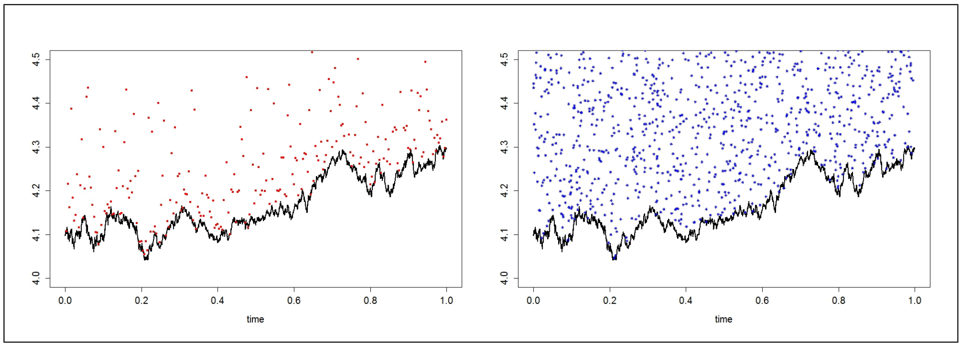

FIG 1. *Left: Microstructure noise model* $Y_i = X_{i/n} + \varepsilon_i, i = 0, \ldots, n = 1000$, *with* $\varepsilon_i \overset{iid}{\sim} Exp(50)$. *Right: Poisson point process model with intensity* $\lambda_{t,y} = 50n\mathbb{1}_{(y \geq X_t)}$ *with* X_t *an Itô process.*

be defined on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$, which satisfies the usual conditions, with W denoting a standard $(\mathcal{F}_t)$-Brownian motion. Its total quadratic variation $\langle X, X\rangle_1 = \int_0^1 \sigma_s^2 \, ds$ is also called integrated squared volatility and forms a central object for risk and portfolio management. A natural continuous-time embedding of the boundary problem is in terms of a Poisson point process (PPP). Conditional on $(X_t,\, t \in [0,1])$ we observe a PPP on $[0,1] \times \mathbb{R}$ with intensity measure

$$\Lambda(A) = \int_0^1 \int_{\mathbb{R}} \mathbb{1}_A(t,y)\lambda_{t,y} \, dt \, dy, \quad \text{where } \lambda_{t,y} = n\lambda \mathbb{1}(y \geq X_t). \tag{1.2}$$

We denote by $(T_j, \mathcal{Y}_j)$ the observations of that point process, which are homogeneously dispersed above the graph of $(X_t,\, t \in [0,1])$. Theoretically and also intuitively, information on the stochastic boundary can only be recovered from the lowest observation points and a homogeneous intensity away from the boundary is assumed for convenience only.

An associated discrete-time regression-type model, which explains well the difference to regular microstructure noise models, is defined by

$$Y_i = X_{t_i^n} + \varepsilon_i \, , \; i = 0, \ldots, n, \; \varepsilon_i \geq 0, \; \varepsilon_i \overset{iid}{\sim} F_\lambda \, , \tag{1.3}$$

with observation times t_i^n and an error distribution function F_λ satisfying

$$F_\lambda(x) = \lambda x\big(1 + \mathcal{O}(1)\big), \text{ as } x \downarrow 0. \tag{1.4}$$

One natural parametric specification is $\varepsilon_i \sim \text{Exp}(\lambda)$. The noise is assumed to be independent of the signal part X. In microstructure noise models for transaction prices it is usually assumed that $\mathbb{E}[\varepsilon_i] = 0$ holds, while here $X_{t_i^n}$

defines the boundary of the support measure for Y_i, which we may interpret as best ask price at time t_i^n. In fact, if the boundary function was piecewise constant, then by standard PPP properties we would obtain the regression-type model (1.3) with exponential noise from the PPP-model (1.2) by taking local minima (on those pieces). Here we show that under so called high-frequency asymptotics, the fundamental quantities in both models exhibit the same asymptotic behaviour, see Proposition 3.2 below. Compare also [17] for the stronger Le Cam−equivalence in the case of smoother boundaries. Both observation models are illustrated in Figure 1.

Mostly, we shall concentrate on the more universal PPP model which also allows for simpler scaling and geometric interpretation. Local minima $m_{n,k}$ of $\mathcal{Y}_j$ for T_j in some small interval $[kh_n, (k+1)h_n) \subseteq [0,1]$ will form the basic quantities to recover the boundary, which by PPP properties leads to the study of

$$\mathbb{P}(m_{n,k} > x) = \mathbb{E}\left[\exp\Big(- \int_{kh_n}^{(k+1)h_n} (X_t + x)_+ \, dt\Big)\right], \, x \in \mathbb{R},$$

where $A_+ = \max(A, 0)$, and its associated moments. For the fundamental case $X_t = \sigma W_t$, this opens an interesting connection to the theory of Brownian excursion areas and also reveals the difficulty of this problem. It is well documented in the literature, see e.g. [16], that no explicit form of the expectation in the expression above is available. Essentially only (double) Laplace transforms and related quantities are known, cf. Proposition 3.3 below and the attached discussion. This makes the recovery of $\langle X, X\rangle_1$ an intricate probabilistic question. Still, we are able to prove that our estimator attains the improved rate $n^{-1/3}$, compared to regular microstructure noise models. What is more, by information-theoretic arguments we are able to derive a lower bound showing that the $n^{-1/3}$-rate is indeed minimax optimal. A more direct proof seems out of reach because the Poisson part from the noise intertwines with the Gaussian martingale part in a way which renders the likelihood and respective Hellinger distances difficult to control, even asymptotically.

The growing finance literature on limit order books so far focusses on modeling and empirical studies. Empirical contributions as [4], [6] and [18] have investigated price and volume distribution, inter-event durations as well as the structure of the order-flow. Probabilistic models proposed for a limit order book include point process models, see [9], [2] and [15], with mutually exciting processes. Other models come from queuing theory, for instance [11], [24] and [10], or stochastic optimal control theory as [7]. The main objective of most modeling approaches is to explain how market prices arise from the

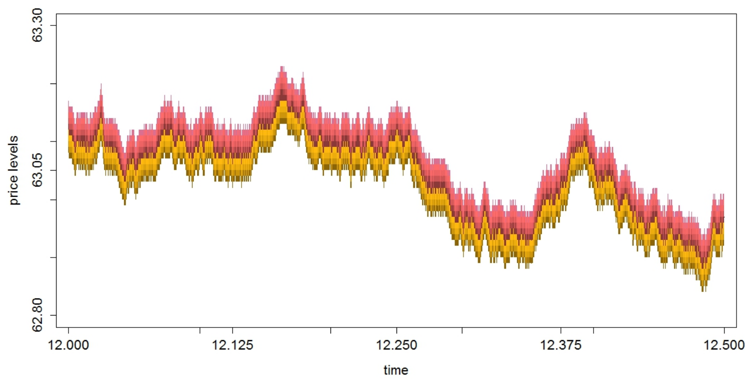

FIG 2. *Order price levels for Facebook asset (NASDAQ) from 12:00 to 12:30 on June 2nd 2014. Colored areas highlight spreads between different bid and ask levels from level 1 up to level 5, bid-ask spread is colored in dark red.*[2]

book. For the financial application, this papers adopts a new course. It is the first work, to the best of our knowledge, with the focus on statistical inference for the volatility based on observations from a limit order book. We have already highlighted the relationship of the suggested model to the regular microstructure noise model which constitutes the standard setup for developing volatility estimators. Let us mention the work by [1], [27], taken up by [3] and [14], among many others, who describe high-frequency intraday trading prices as convolution of discretized observations of the efficient log-price with additive observation noise ascribed to market microstructure. The optimal convergence rate for volatility estimation in this model with Gaussian noise and n observations on an equidistant grid is $n^{-1/4}$, see [13]. Recently, as information from order books become more and more available, researchers and practitioners have sparked the discussion to which kind of observed prices estimation methods should be applied. [12] discuss this point and the possibilities of mid-quotes, executed traded prices or micro-prices which are volume-weighted combinations of bid and ask order levels. None of these observed time series, however, is free from market microstructure corruptions and the idea of an underlying efficient price remains untouched. Figure 2 visualizes the information about the evolution of prices provided by a limit order book for one specific data set. The colored areas highlight differences between the five best bid and five best ask levels, the dark area in the center marking the bid-ask spread between best bid and best ask.

[2]Data provided by LOBSTER academic data – powered by NASDAQ OMX.

The idea is that an efficient price must always lie below the best ask (and symmetrically above the best bid) and that its distance to this stochastic frontier is homogeneous.

The remainder of the paper is organized as follows. In Section 2 we present an estimation approach based on local order statistics whose asymptotic properties are explored in Section 3. In Section 4 we prove the lower bound for the minimax estimation rate. Section 5 concludes. Proofs are provided in the Appendix.

2. Volatility estimation based on local minima. We construct the integrated volatility estimator in both models (1.2) and (1.3). We partition the unit interval into $h_n^{-1} \in \mathbb{N}$ equi-spaced bins $\mathcal{T}_k^n = [kh_n, (k+1)h_n), k = 0, \ldots, h_n^{-1} - 1$, with bin-widths h_n. For simplicity suppose that $nh_n \in \mathbb{N}$. As $n \to \infty$ the bin-width gets smaller $h_n \to 0$, whereas the number of observed values on each bin gets large, $nh_n \to \infty$. If we think of a constant signal locally on a bin observed with one-sided positive errors, classical parametric estimation theory motivates to use the bin-wise minimum as an estimator of the local signal (it then forms a sufficient statistic under exponential noise or equivalently in the PPP model). In the regression-type model (1.3) with equidistant observation times $t_i^n = i/n$, we therefore set

$$m_{n,k} = \min_{i \in \mathcal{I}_k^n} Y_i \ , \ \mathcal{I}_k^n = \{kh_n n, kh_n n + 1, \ldots, (k+1)h_n n - 1\} \,. \tag{2.1}$$

Equally, in the PPP model (1.2) the local minima are given by

$$m_{n,k} = \min_{T_j \in \mathcal{T}_k^n} \mathcal{Y}_j \ , \ \mathcal{T}_k^n = [kh_n, (k+1)h_n) \,. \tag{2.2}$$

The same symbol $m_{n,k}$ is used in both models because the following construction only depends on the $m_{n,k}$. All results and proofs will refer to the concrete model under consideration.

Since $\mathbb{V}\text{ar}(m_{n,k} \,|\, (X_t)) \propto (n\lambda h_n)^{-2}$ holds in both models, the variance is much smaller than for an estimator based on a local mean. Nevertheless, we may continue in the spirit of the pre-averaging paradigm, cf. [14], and interpret $m_{n,k}$ as a proxy for X_t on $\mathcal{T}_k^n$, which in a second step is inserted in the realized variance expression $\sum_{i=1}^{h_n^{-1}} (X_{kh_n} - X_{(k-1)h_n})^2$ without noise. The use of a locally constant signal approximation $X_t = X_{kh_n} + \mathcal{O}_{\mathbb{P}}(h_n^{1/2})$ on $\mathcal{T}_k^n$ is only admissible, however, if h_n is chosen so small that $h_n^{1/2} = o((n\lambda h_n)^{-1})$, which would result in a sub-optimal procedure.

Rate-optimality can be attained if we balance the magnitude $(n\lambda h_n)^{-1}$ of bin-wise minimal errors due to noise with the range $h_n^{1/2}$ of the motion of

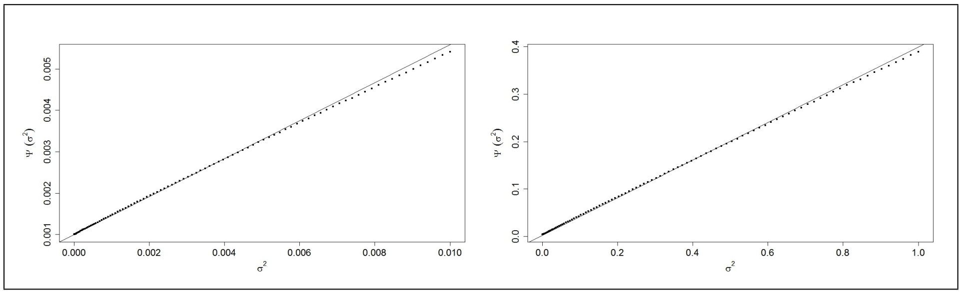

FIG 3. *The points indicate the function $\Psi(\sigma^2)$ with $\mathcal{K} = 31.6$ for small (left) and moderate (right) values of σ^2. The calculation is based on accurate Monte Carlo simulations. The lines show close linear functions for comparison.*

X on the bin. This gives the order

$$h_n \propto (n\lambda)^{-\frac{2}{3}},\ nh_n \propto n^{\frac{1}{3}}\lambda^{-\frac{2}{3}}\ . \tag{2.3}$$

In the PPP model (1.2) this natural choice of the bin-width also follows nicely by a scaling argument: $\bar{W}_t = h_n^{-1/2} W_{h_n t}$ defines a standard Brownian motion for $t \in [0,1]$ based on the values of W on $[0, h_n]$; the correspondingly scaled PPP observations $(\bar{T}_j, \bar{\mathcal{Y}}_j)$ with $\bar{T}_j = h_n^{-1}T_j$, $\bar{\mathcal{Y}}_j = h_n^{-1/2}\mathcal{Y}_j$ have an intensity with density $\bar{\lambda}_{t,y} = n\lambda h_n^{3/2}\mathbb{1}(y \geq \bar{W}_t)$, which becomes independent of n exactly for $h_n = (n\lambda)^{-2/3}$.

In this balanced setup the law of the statistics $m_{n,k}$ depends on the motion of X as well as the error distribution in a non-trivial way. Still, the natural statistics to assess the quadratic variation of the boundary process X are the squared differences $(m_{n,k} - m_{n,k-1})^2$ between consecutive local minima. In the PPP model and with the choice

$$h_n = \mathcal{K}^{\frac{2}{3}}(n\lambda)^{-\frac{2}{3}} \text{ for some constant } \mathcal{K} > 0 \tag{2.4}$$

the law of $h_n^{-1/2} m_{n,k}$ is independent of n, h_n and λ and for $X_t = X_{(k-1)h_n} + \sigma\int_{(k-1)h_n}^t dW_s$ on $\mathcal{T}_{k-1}^n \cup \mathcal{T}_k^n$, we may introduce

$$\Psi\big(\sigma^2\big) = h_n^{-1}\mathbb{E}\Big[\big(m_{n,k} - m_{n,k-1}\big)^2\Big]\,, k = 1, \ldots, h_n^{-1} - 1. \tag{2.5}$$

Below we shall derive theoretical properties of Ψ and in particular we shall see that it is invertible as soon as $\mathcal{K} > 0$ is chosen sufficiently large. Numerically, the function Ψ can be determined by standard Monte Carlo simulations, see Figure 3, and is thus available. This paves the way for a moment-estimator approach. In fact, $\sum_k (m_{n,2k} - m_{n,2k-1})^2$ approximates $\int \Psi(\sigma_t^2)dt$

with corresponding summation and integration intervals. Under regularity assumptions on $t \mapsto \sigma_t^2$ and by the smoothness of Ψ shown below, we have

$$\Psi^{-1}\Bigg(\sum_{k=(l-1)r_n^{-1}/2+1}^{lr_n^{-1}/2}\big(m_{n,2k}-m_{n,2k-1}\big)^2\, 2h_n^{-1}r_n\Bigg) \approx \sigma^2_{lr_n^{-1}h_n},$$

where $r_n^{-1}h_n$ is a coarse grid size with $r_nh_n^{-1}, r_n^{-1} \in 2\mathbb{N}$. This gives rise to the following estimator of integrated volatility $IV = \int_0^1 \sigma_t^2\,dt$ in the PPP model (1.2) with bin-width (2.4):

$$\widetilde{IV}_n^{h_n,r_n} = \sum_{l=1}^{r_nh_n^{-1}} \Psi^{-1}\Bigg(\sum_{k=(l-1)r_n^{-1}/2+1}^{lr_n^{-1}/2}\big(m_{n,2k}-m_{n,2k-1}\big)^2\, 2h_n^{-1}r_n\Bigg)h_nr_n^{-1}. \tag{2.6}$$

In the regression-type model (1.3) the corresponding second moments still depend on n and we write explicitly

$$\Psi_n\big(\sigma^2\big) = h_n^{-1}\mathbb{E}\Big[\big(m_{n,k}-m_{n,k-1}\big)^2\Big]\,, k=1,\ldots,h_n^{-1}-1. \tag{2.7}$$

We shall see below that $\Psi_n \to \Psi$ holds, but a non-asymptotic form of the volatility estimator from regression-type observations is given by

$$\widehat{IV}_n^{h_n,r_n} = \sum_{l=1}^{r_nh_n^{-1}} \Psi_n^{-1}\Bigg(\sum_{k=(l-1)r_n^{-1}/2+1}^{lr_n^{-1}/2}\big(m_{n,2k}-m_{n,2k-1}\big)^2\, 2h_n^{-1}r_n\Bigg)h_nr_n^{-1}. \tag{2.8}$$

For a parametric estimation of $\sigma_t = \sigma = \text{const.}$, we employ an estimator $\widehat{IV}_n^{h_n,h_n}$. This means inversion of the whole sum of squared differences is conducted. In the nonparametric case of varying σ_t instead an estimator $\widehat{IV}_n^{h_n,r_n}$, with $r_n \to 0, r_n^{-1}h_n \to 0$, is applied. A balance between a second order term on each coarse interval of order r_n and an approximation error controlled by a Lipschitz assumption on σ_t of order $r_n^{-1}h_n$ will lead to the choice $r_n \propto h_n^{1/2} \propto (n\lambda)^{-1/3}$.

3. The law of local minima and the convergence rate of the estimator. In order to centralise the local minima, we write

$$m_{n,k}-m_{n,k-1} = \mathcal{R}_{n,k}-\mathcal{L}_{n,k}\;, k=1,\ldots,h_n^{-1}-1\,, \tag{3.1}$$

where $\mathcal{R}_{n,k} = m_{n,k}-X_{kh_n}$ and $\mathcal{L}_{n,k} = m_{n,k-1}-X_{kh_n}$ measure the distances between the minima on bin $\mathcal{T}_k^n$ and $\mathcal{T}_{k-1}^n$, respectively, to the central true value X_{kh_n} between both bins. In our high-frequency framework

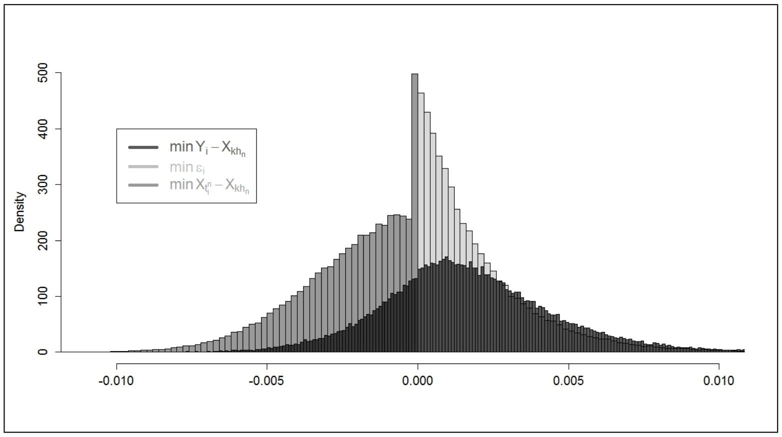

FIG 4. *Distributions of bin-wise minima of the signal process, noise and the convolution. Based on 100000 simulated bins with* $\sigma = 1$, $\varepsilon_i \sim \mathrm{Exp}(5)$, $nh_n = 100$.

the drift is asymptotically negligible and a regular volatility function will be approximated by a piecewise constant function on blocks of the coarse grid. In this setting, where $X_t = X_{kh_n} + \sigma(W_t - W_{kh_n})$, we may invoke time-reversibility of Brownian motion to see that $X_t - X_{kh_n}$, $t \in \mathcal{T}_{k-1}^n$, and $X_t - X_{kh_n}$, $t \in \mathcal{T}_k^n$, form independent Brownian motions of variance σ^2 such that $\mathcal{R}_{n,k}, \mathcal{L}_{n,k}, k = (l-1)r_n^{-1} + 1, \ldots, lr_n^{-1}$, are all identically distributed and there is independence whenever different bins are considered (but $\mathcal{R}_{n,k}$ and $\mathcal{L}_{n,k+1}$ are dependent). From (2.5) and (3.1) we infer

$$\Psi(\sigma_{kh_n}^2)h_n = \mathbb{E}[\mathcal{R}_{n,k}^2] + \mathbb{E}[\mathcal{L}_{n,k}^2] - 2\mathbb{E}[\mathcal{R}_{n,k}]\mathbb{E}[\mathcal{L}_{n,k}] = 2\,\mathbb{V}\mathrm{ar}(\mathcal{R}_{n,k}),$$

and similarly for Ψ_n. The histogram in Figure 4 shows the distribution of $\mathcal{R}_{n,k}$ (equivalently $\mathcal{L}_{n,k}$) in the regression model jointly with the associated histograms for $\min_{i \in \mathcal{I}_k^n} X_{t_i^n} - X_{kh_n}$ and $\min_{i \in \mathcal{I}_k^n} \varepsilon_i$. In this situation the law of $\mathcal{R}_{n,k}$ is given as the convolution between an exponential distribution and the law of the minimum of Brownian motion on the discrete grid $\mathcal{I}_k^n$. The latter converges to the law of the minimum of W on $[0,1]$, but the simulations confirm the known feature that the laws deviate rather strongly around zero for moderate discretisations. Let us state and prove a slightly more general result.

PROPOSITION 3.1. *Choose* h_n *according to* (2.4). *Consider* $t \in \mathcal{T}_k^n$ *for fixed* k *and suppose that* $X_t = X_{kh_n} + \int_{kh_n}^t \sigma\,dW_s, t \in \mathcal{T}_k^n$. *Then in the PPP model* (1.2) *for all* $x \in \mathbb{R}$

$$\mathbb{P}\Big(h_n^{-1/2}\mathcal{R}_{n,k} > x\sigma\Big) = \mathbb{E}\Big[\exp\Big(-\mathcal{K}\sigma\int_0^1 (x + W_t)_+\,dt\Big)\Big]. \tag{3.2}$$

PROOF. By conditioning on the Brownian motion we infer from the PPP properties of $(T_j, \mathcal{Y}_j)$:

$$\mathbb{P}\Big(h_n^{-1/2}\mathcal{R}_{n,k} > x\sigma \,\Big|\, W\Big) = \exp\Bigg(- \int_{\mathcal{T}_k^n} \int_{-\infty}^{x\sigma h_n^{1/2} + X_{kh_n}} \lambda_{t,y} \Bigg)$$
$$= \exp\Bigg(- n\lambda\sigma \int_{\mathcal{T}_k^n} \Big(xh_n^{1/2} - (W_t - W_{kh_n})\Big)_+ \, dt \Bigg).$$

Noting that $\bar{W}_s = h_n^{-1/2}(W_{(k+s)h_n} - W_{kh_n})$, $s \in [0,1]$, is again a Brownian motion, the result follows by rescaling and taking expectations. □

For the regression-type model the survival function is asymptotically of the same form.

PROPOSITION 3.2. *Choose h_n according to (2.4). Suppose that $X_t = X_{kh_n} + \int_{kh_n}^t \sigma \, dW_s$, $t \in \mathcal{T}_k^n$, for a fixed bin number k. Then in the regression-type model (1.3) for all $x \in \mathbb{R}$*

$$\lim_{n\to\infty} \mathbb{P}\Big(h_n^{-1/2}\mathcal{R}_{n,k} > x\sigma\Big) = \mathbb{E}\Big[\exp\Big(- \mathcal{K}\sigma \int_0^1 (x + W_t)_+ \, dt\Big)\Big]. \tag{3.3}$$

The approximation error due to non-constant σ and drift is considered in detail in Appendix A.1 and proved to be asymptotically negligible. This way, the asymptotic analysis of our estimation problem leads into the theory of Brownian excursion areas. Let $\mathcal{R}_t$ be a real random variable distributed as $\lim_{n\to\infty} h_n^{-1/2}\mathcal{R}_{n,\lfloor th_n^{-1}\rfloor}$. The law of $\mathcal{R}_t$ determines $\Psi(\sigma_t^2)$ via

$$\mathbb{V}\mathrm{ar}(\mathcal{R}_t) = \frac{1}{2}\Psi(\sigma_t^2)\,. \tag{3.4}$$

The Feynman–Kac formula gives a connection of the right-hand side in Proposition (3.1) to a parabolic PDE based on the heat semigroup for Brownian motion. We can prove the following explicit result on the Laplace transform which determines the distribution of $(\mathcal{R}_t), t \in [0,1]$.

PROPOSITION 3.3. *The Laplace transform (in t) of*

$$\mathbb{E}\Big[\exp\Big(- \sqrt{2}\vartheta \int_0^t (x + W_s)_+ \, ds\Big)\Big]$$

with $\vartheta \in \mathbb{R}$ satisfies the following identity:

$$\mathbb{E}\left[\int_0^\infty \exp\Big(- st - \sqrt{2}\vartheta \int_0^t (x + W_s)_+ \, ds\Big) \, dt\right] = \vartheta^{-\frac{2}{3}}\zeta_s(x, \vartheta),$$

with $\zeta_s(x,\vartheta)=\zeta_{s,-}(x,\vartheta)\mathbb{1}_{(-\infty,0)}(x)+\zeta_{s,+}(x,\vartheta)\mathbb{1}_{[0,\infty)}(x)$ *defined by the functions*

$$\zeta_{s,+}(x,\vartheta)=\frac{\pi\Big(\vartheta^{1/3}Gi'\big(\vartheta^{-2/3}s\big)-\sqrt{s}\,Gi\big(\vartheta^{-2/3}s\big)\Big)+\vartheta^{2/3}s^{-1/2}}{\sqrt{s}Ai\big(\vartheta^{-2/3}s\big)-\vartheta^{1/3}Ai'\big(\vartheta^{-2/3}s\big)}$$
$$\times\, Ai\big(\sqrt{2}\vartheta^{1/3}x+\vartheta^{-2/3}s\big)+\pi\, Gi\big(\sqrt{2}\vartheta^{1/3}x+\vartheta^{-2/3}s\big)\,,$$
$$\zeta_{s,-}(x,\vartheta)=\left(\frac{\vartheta^{2/3}s^{-1/2}Ai\big(\vartheta^{-2/3}s\big)+\vartheta^{1/3}AI\big(\vartheta^{-2/3}s\big)}{\sqrt{s}Ai\big(\vartheta^{-2/3}s\big)-\vartheta^{1/3}Ai'\big(\vartheta^{-2/3}s\big)}-s^{-1}\vartheta^{2/3}\right)$$
$$\times\exp\big(\sqrt{2s}x\big)+s^{-1}\vartheta^{2/3}\,,$$

where Ai is the Airy function which is bounded on the positive half axis,

$$Ai(x)=\pi^{-1}\int_0^\infty\cos\left(t^3/3+xt\right)dt\,,$$

and Gi is the Scorer function bounded on the positive half axis

$$Gi(x)=\pi^{-1}\int_0^\infty\sin\left(t^3/3+xt\right)dt\,,$$

and we define $AI(x)=\int_x^\infty Ai(y)dy$.

This result generalizes the Laplace transform of the exponential integrated positive part of a Brownian motion derived by [22]. Inserting $x=0$ and setting $\vartheta=1$ renders the result by [22]. An inversion of the Laplace transform in Proposition 3.3 in order to obtain an explicit form of the distribution function and then Ψ appears unfeasible as several experts vainly attempted to solve related problems, see [22] and [16]. Exploiting the strong Markov property of Brownian motion together with hitting times, we are able to circumvent this problem in our study of $\Psi(\sigma^2)$, for details we refer to the Appendix.

We formulate now the main convergence results whose proofs are given in the Appendix. For that we impose some regularity on the drift and diffusion coefficient which are also assumed to be deterministic or more generally independent of the driving Brownian motion W. Moreover, we need that the function Ψ is invertible and sufficiently regular, which by Proposition A.2 below is ensured by a sufficiently large choice of $\mathcal{K}$, but at least numerically seems to be the case for much smaller choices, cf. Figure 3 and [5].

ASSUMPTION 3.4. *The drift* a_s *in* (1.1) *is bounded and Borel-measurable, the volatility* σ_t *in* (1.1) *is a Lipschitz function that does not vanish,* $\sigma_t>0$. *The constant* $\mathcal{K}$ *in the definition* (2.4) *of* h_n *is chosen large enough that Proposition A.2 below applies.*

Theorem 3.5. *Grant Assumption 3.4, choose h_n according to (2.4) and $r_n = \kappa n^{-1/3}$ for some $\kappa > 0$. Then the estimator (2.6) based on observations from the PPP-model satisfies*

$$\Big(\widetilde{IV}_n^{h_n,r_n} - \int_0^1 \sigma_s^2\,ds\Big) = \mathcal{O}_{\mathbb{P}}\Big(n^{-\frac{1}{3}}\Big)\,. \tag{3.5}$$

Based on the same strategy of proof we can obtain an analogous result for the regression-type model.

Corollary 3.6. *Grant Assumption 3.4, choose h_n according to (2.4) and $r_n = \kappa n^{-1/3}$ for some $\kappa > 0$. Then the estimator (2.8) based on observations from the regression-type model satisfies*

$$\Big(\widehat{IV}_n^{h_n,r_n} - \int_0^1 \sigma_s^2\,ds\Big) = \mathcal{O}_{\mathbb{P}}\Big(n^{-\frac{1}{3}}\Big)\,. \tag{3.6}$$

4. Lower bound for the rate of convergence. Consider our PPP-model (1.2). We show that even in the simpler parametric statistical experiment where $X_t = \sigma W_t$, $t \in [0,1]$, and $\sigma > 0$ is unknown the optimal rate of convergence is $n^{-1/3}$ in a minimax sense. This lower bound for the parametric case then serves a fortiori as a lower bound for the general nonparametric case. A lower bound for the discrete regression-type model is obtained in a similar way; in fact the proof is even simpler, replacing the Poisson sampling (T_j^s) below by a deterministic design of distance $n^{-2/3}$.

Theorem 4.1. *We have for any sequence of estimators $\hat{\sigma}_n^2$ of $\sigma^2 \in (0,\infty)$ from the parametric PPP-model for each $\sigma_0^2 > 0$, the local minimax lower bound*

$$\exists \delta > 0 : \liminf_{n\to\infty} \inf_{\hat{\sigma}_n} \max_{\sigma^2 \in \{\sigma_0^2, \sigma_0^2 + \delta n^{-1/3}\}} \mathbb{P}_{\sigma^2}(|\hat{\sigma}_n^2 - \sigma^2| \geq \delta n^{-1/3}) > 0,$$

where the infimum extends over all estimators $\hat{\sigma}_n$ based on the PPP-model (1.2) with $\lambda = 1$ and $X_t = \sigma W_t$. The law of the latter is denoted by $\mathbb{P}_{\sigma^2}$.

The proof falls into three main parts. We first simplify the problem by considering more informative experiments. These reductions are given in the two steps below. Then, in the third step we use bounds for the Hellinger distance. The more technical step 3 is worked out in Appendix B.

1. A PPP with intensity Λ is obtained as the sum of two independent PPPs with intensities Λ_r and Λ_s, respectively, satisfying $\Lambda = \Lambda_r + \Lambda_s$,

see e.g. [20]. Hence, for $b>0$ the experiment of observing $(T_i^r,\mathcal{Y}_i^r)_{i\ge1}$ from a PPP with regularised intensity density

$$\lambda_r(t,y)=n\Big(\big((y-X_t)_+/b\big)^2\wedge 1\Big)$$

and independently $(T_j^s,\mathcal{Y}_j^s)_{j\ge1}$ from a PPP with discontinuous intensity density $\lambda_s=\lambda-\lambda_r$ is more informative. We now provide even more information by replacing $(T_j^s,\mathcal{Y}_j^s)_{j\ge1}$ by $(T_j^s,X_{T_j^s})_{j\ge1}$, the direct observation of the martingale values at the random times (T_j^s). A lower bound proved for observing $(T_i^r,\mathcal{Y}_i^r)_{i\ge1}$ and $(T_j^s,X_{T_j^s})_{j\ge1}$ independently thus also applies to the original (less informative) observations.

2. Due to $\int\int\lambda_s(t,y)dt\,dy=(2/3)nb$, we conclude that the times (T_j^s) are given by a Poisson sampling of intensity $(2/3)nb$ on $[0,1]$ and there are a.s. only finitely many times $(T_j^s)_{j=1,\dots,J}$. Let us first work conditionally on (T_j^s) and put $T_0^s=0$, $T_{J+1}^s=1$. All observations of $(T_i^r,\mathcal{Y}_i^r)_{i\ge1}$ with $T_i^r\in[T_{j-1}^s,T_j^s)$ are transformed via

$$(T_i^r,\mathcal{Y}_i^r)\mapsto\left(T_i^r-T_{j-1}^s,\mathcal{Y}_i^r-\left(X_{T_{j-1}^s}\frac{T_i^r-T_{j-1}^s}{T_j^s-T_{j-1}^s}+X_{T_j^s}\frac{T_j^s-T_i^r}{T_j^s-T_{j-1}^s}\right)\right).$$

Noting that $(B_t-(t/T)B_T,t\in[0,T])$ defines a Brownian bridge $B^{0,T}$ on $[0,T]$, we thus obtain conditionally on (T_j^s) for each $j=1,\dots,J+1$ observations of a PPP on $\big[0,T_j^s-T_{j-1}^s\big]$ with intensity density

$$\lambda^j(t,y)=n\Big(b^{-1}\big(y-\sigma B_t^{0,T_j^s-T_{j-1}^s}\big)_+\wedge 1\Big).$$

The transformation has rendered the family of PPPs with intensity densities $(\lambda^j)_{j=1,\dots,J+1}$ independent by reducing the Brownian motion to piecewise Brownian bridges. Conditionally on (T_j^s) we thus have independent observations of $(T_j^s,X_{T_j^s})_{j=1,\dots,J}$ and independent PPPs with intensity densities $(\lambda^j)_{j=1,\dots,J+1}$.

By using the latter more informative experiment and by choosing $b\propto n^{-1/3}$ we show below that for a Poisson sampling $(T_j^s)_{j=1,\dots,J}$ on $[0,1]$ of intensity $(2/3)nb\propto n^{2/3}$ of direct observations $X_{T_j^s}$ as well as for independent observations of PPPs, generated by σ times a Brownian bridge in-between the sampling points $(T_j^s)_j$, we cannot estimate at a better rate than $n^{-1/3}$. This is accomplished by bounding the Hellinger distance between the experiments for $\sigma^2=\sigma_0^2$ and $\sigma^2=\sigma_0^2+\delta n^{-1/3}$.

5. Conclusion. We have modeled the relationship between limit order book bid and ask quotes and an efficient price process by a stochastic frontier model. The model does not attempt to describe the fine structure of order book dynamics, but is only based on the natural ordering between quotes and prices. This agnostic point of view seems attractive for statistical purposes. Still, we find $n^{-1/3}$ as optimal convergence rate of integrated squared volatility estimators which improves upon the $n^{-1/4}$-rate, known for transaction price models with regular microstructure noise.

The estimation approach uses local order statistics and a coarse-fine grid approximation whose analysis is connected to a Brownian excursion problem. Owing to the nonlinear and implicit definition, finer properties of the estimator like its asymptotic distribution remain open. First numerical results in [5] are in any case promising. An empirical validation, using the three data sets of bid and ask quotes and transaction prices independently, is feasible.

APPENDIX A: PROOFS OF SECTION 3

Proposition 3.2 considers the simplified model where $X_t, t \in \mathcal{T}_k^n$, is approximated by $X_{kh_n} + \int_{kh_n}^t \sigma_{kh_n}\, dW_t$. The resulting approximation error is bounded within Proposition A.1 for the PPP-model and an analogous proof carries over to the regression-type model. From here on $A_n \lesssim B_n$ expresses shortly that $A_n \le K \cdot B_n$ for two sequences A_n, B_n and some real constant $K < \infty$. We write $A_+ = A\mathbb{1}(A \ge 0)$, $A_- = |A|\mathbb{1}(A \le 0)$ and $\|Z\|_p = \mathbb{E}[|Z|^p]^{1/p}, p \ge 1$.

Proof of Proposition 3.2. By law invariance of $\mathcal{R}_{n,k}$ with respect to k for $X_t = X_0 + \sigma W_t$, we can simplify

$$\mathbb{P}(h_n^{-1/2}\mathcal{R}_{n,k} > x\sigma) = \mathbb{P}\Big(h_n^{-1/2} \min_{i=0,\ldots,nh_n-1}(X_{i/n} - X_0 + \varepsilon_i) > x\sigma\Big)$$
$$= \mathbb{P}\Big(\min_{i=0,\ldots,nh_n-1}(W_{i/(nh_n)} + \sigma^{-1}h_n^{-1/2}\varepsilon_i) > x\Big),$$

where we used that $h_n^{1/2} W_{t/h_n}$ is another Brownian motion. We condition on the driving Brownian motion $W = (W_t, t \in [0,1])$ and obtain in terms of the distribution function F_λ of ε_i:

$$\mathbb{P}(h_n^{-1/2}\mathcal{R}_{n,k} > x\sigma) = \mathbb{E}\left[\prod_{i=0}^{nh_n-1} \mathbb{P}\Big(\varepsilon_i > \sigma h_n^{1/2}(x - W_{i/(nh_n)})\,\Big|\, W\Big)\right]$$
$$= \mathbb{E}\Big[\exp\Big(\sum_{i=0}^{nh_n-1} \log\Big(1 - F_\lambda(\sigma h_n^{1/2}(x - W_{i/(nh_n)}))\Big)\Big)\Big].$$

The expansion (1.4) of F_λ together with expanding the logarithm therefore yields

$$\mathbb{P}(h_n^{-1/2}\mathcal{R}_{n,k} > x\sigma) = \mathbb{E}\Big[\exp\Big(-\sigma h_n^{1/2}\lambda \sum_{i=0}^{nh_n-1} (x - W_{i/(nh_n)})_+(1+\mathcal{O}(1))\Big)\Big],$$

where $\mathcal{O}(1)$ is to be understood ω-wise and holds uniformly over i and n whenever $\max_{t\in[0,1]}(x - W_t(\omega))_+$ is bounded. By the choice of h_n we have $h_n^{1/2}\lambda = \mathcal{K}(nh_n)^{-1}$ and the integrand is a Riemann sum tending almost surely to $\exp(-\sigma\mathcal{K}\int_0^1(x-W_t)_+dt)$. Noting that a conditional probability is always bounded by 1, the assertion follows by dominated convergence and use of $-W \stackrel{d}{=} W$. □

Proof of Proposition 3.3. Throughout the proof, we drop the dependence on ϑ in $\zeta_s(x,\vartheta)$, $\zeta_{s,-}(x,\vartheta)$ and $\zeta_{s,+}(x,\vartheta)$ to lighten the notation. We shall apply the Kac formula in the version as in formulae (4.13) and (4.14) of [19]. It connects the considered Laplace transform with the solution of a differential equation which becomes in our case:

$$\frac{d^2\zeta}{dx^2} = 2s\zeta - 2\vartheta^{2/3} \qquad , \; x<0, \tag{A.1a}$$

$$\frac{d^2\zeta}{dx^2} = 2(\sqrt{2}\vartheta x + s)\zeta - 2\vartheta^{2/3} \qquad , \; x>0. \tag{A.1b}$$

Since all assertions necessary to apply the Kac formula are fulfilled, the Laplace transform from above multiplied with a constant Lagrangian $\vartheta^{2/3}$ satisfies

$$\mathbb{E}\left[\int_0^\infty \vartheta^{2/3}\exp\Big(-st - \sqrt{2}\vartheta\int_0^t (x+W_s)_+\,ds\Big)\,dt\right] = \zeta_s(x)\,.$$

The general solution of (A.1a) is given by

$$\zeta_{s,-}(x) = A\exp\left(\sqrt{2s}x\right) + \vartheta^{2/3}s^{-1}\,, \tag{A.2a}$$

with a constant A (depending on s but not on x). Airy's function Ai solves the homogenous differential equation of the type (A.1b), whereas the Scorer function Gi is a particular solution of the inhomogenous equation $\zeta'' - x\zeta = \pi^{-1}$, both being bounded on the positive real line. Hence, a solution ansatz for (A.1b) is given by

$$\zeta_{s,+}(x) = B\mathrm{Ai}\Big(\sqrt{2}\vartheta^{1/3}x + \vartheta^{-2/3}s\Big) + \pi\mathrm{Gi}\Big(\sqrt{2}\vartheta^{1/3}x + \vartheta^{-2/3}s\Big)\,, \tag{A.2b}$$

with a constant B. Continuity conditions on ζ and $d\zeta/dx$ at $x=0$ give rise to

$$B = \frac{\pi\Big(\vartheta^{1/3}\mathrm{Gi}'\big(\vartheta^{-2/3}s\big) - \sqrt{s}\mathrm{Gi}\big(\vartheta^{-2/3}s\big)\Big) + \vartheta^{2/3}s^{-1/2}}{\sqrt{s}\mathrm{Ai}\big(\vartheta^{-2/3}s\big) - \vartheta^{1/3}\mathrm{Ai}'\big(\vartheta^{-2/3}s\big)}\,.$$

In order to express A in a more concise and simple manner, we exploit the following relation for the Wronskian of Ai and Gi:

$$\pi\Big(\mathrm{Gi}'(x)\mathrm{Ai}(x) - \mathrm{Ai}'(x)\mathrm{Gi}(x)\Big) = \mathrm{AI}(x) = \int_x^\infty \mathrm{Ai}(y)\,dy\,. \tag{A.3}$$

A proof of the latter equality can be found in [26]. Thereby, we obtain

$$A = \left(\frac{\vartheta^{2/3}s^{-1/2}\mathrm{Ai}\big(\vartheta^{-2/3}s\big) + \vartheta^{1/3}\mathrm{AI}\big(\vartheta^{-2/3}s\big)}{\sqrt{s}\mathrm{Ai}\big(\vartheta^{-2/3}s\big) - \vartheta^{1/3}\mathrm{Ai}'\big(\vartheta^{-2/3}s\big)} - s^{-1}\vartheta^{2/3}\right)\,.$$

This result concludes the proof. □

A.1. Asymptotic analysis of the estimator. As a first step, we extend Proposition 3.1 by analysing the approximation error due to neglecting the drift and assuming a locally constant volatility. Then we prove Theorem 3.5 exploiting properties of Ψ which are established in Appendix A.2.

PROPOSITION A.1. *Consider h_n in (2.4) and $t \in \mathcal{T}_k^n$ for fixed k. Then*

$$\mathbb{P}\left(\min_{j\in\mathcal{T}_k^n} \mathcal{Y}_j - X_{kh_n} > x\sigma_{kh_n}\sqrt{h_n}\right) = \mathbb{E}\Big[\exp\Big(-\mathcal{K}\sigma_{kh_n}\int_0^1 (x+W_t)_+ dt\Big)\Big] + n\lambda h_n^{5/2}G(x),$$

where $|G(x)| \le C\mathbb{P}\big(|W_1| \ge C_2 x\sigma_{kh_n}/2\big)$, with $C_1, C_2 > 0$.
If σ_t is constant and $a_t = 0$ for $t \in \mathcal{T}_k^n$, then $G(x)=0$.

PROOF. Proposition 3.1 already gives the last statement. Let $\mathcal{A}_z = \mathcal{T}_k^n \times (-\infty, z]$ and $z = x\sigma_{kh_n}\sqrt{h_n}$. Let $\Delta X_t(k) = \int_{kh_n}^t \sigma_s\,dW_s$ and $\Delta A_t(k) = \int_{kh_n}^t a_s\,ds$. Then, using basic properties of a PPP, it follows that

$$\begin{aligned}
\mathbb{P}\left(\min_{j\in\mathcal{T}_k^n} \mathcal{Y}_j - X_{kh_n} > z\right) &= \mathbb{E}\left[\mathbb{P}\left(\Lambda\big(\mathcal{A}_z\big) = 0\middle|X\right)\right] = \mathbb{E}\Big[\exp\Big(-\Lambda(\mathcal{A}_z)\Big)\Big] \\
&= \mathbb{E}\Big[\exp\Big(-n\lambda\int_{\mathcal{A}_z} \mathbb{1}_{\{\Delta X_t(k)+\Delta A_t(k)\le y\}}\,dt\,dy\Big)\Big] \\
&= \mathbb{E}\Big[\exp\Big(-n\lambda\int_{\mathcal{T}_k^n} (z-\Delta X_t(k)-\Delta A_t(k))_+\,dt\Big)\Big].
\end{aligned} \tag{A.4}$$

Introduce

$$T_k = n\lambda \int_{\mathcal{T}_k^n} (z - \Delta X_t(k) - \Delta A_t(k))_+ dt\,,$$
$$V_k = n\lambda \int_{kh_n}^{(k+1)h_n} \Big(z - \sigma_{kh_n}(W_{(k+1)h_n} - W_{kh_n})\Big)_+ dt\,,$$
$$U_k = n\lambda \int_{kh_n}^{(k+1)h_n} \left|\int_{kh_n}^{t} (\sigma_s - \sigma_{kh_n}) dW_s\right| dt \text{ and } A_k = n\lambda h_n^2 \max_{t\in\mathcal{T}_k^n} \Big|a_{kh_n} - a_t\Big|.$$

Then we have the upper and lower bounds

$$V_k - U_k - A_k \le T_k \le V_k + U_k + A_k.$$

By scaling and symmetry properties of Brownian motion, we have that

$$V_k \stackrel{d}{=} n\lambda\sigma_{kh_n} h_n^{3/2} \int_0^1 (x + W_t)_+ dt = \mathcal{K}\sigma_{kh_n} \int_0^1 (W_t + x)_+\, dt.$$

As a first objective, we derive an upper bound for $\mathbb{E}\Big[\exp\Big(-yT_k\Big)\Big]$, $y > 0$. To this end, note that by the Dambis-Dubins-Schwarz Theorem (Thm. 4.6 in [19])

$$\Delta X_t(k) \stackrel{d}{=} W_{\langle \Delta X(k), \Delta X(k)\rangle_t}.$$

Since $\langle \Delta X(k), \Delta X(k)\rangle_t = \int_{kh_n}^t \sigma_s^2\, ds$ for $t \ge kh_n$, we deduce

$$\begin{aligned}\mathbb{P}\Big(\sup_{t\in\mathcal{T}_k^n} |\Delta X_t(k)| \ge z/2\Big) &\le \mathbb{P}\Big(\sup_{0\le t\le 1} |W_t| \ge C\sigma_{kh_n} x\Big)\\ &\le 2\mathbb{P}\Big(|W_1| \ge Cx\sigma_{kh_n}/2\Big)\,,\end{aligned}$$

with some $C > 0$. By the boundedness of a_t, it follows that

$$A_k \lesssim n\lambda h_n^2 \max_{t\in[0,1]} |a_t| \lesssim n^{-1/3}. \tag{A.5}$$

We thus obtain for $y > 0$ the upper bound

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(-yT_k\Big)\Big] &\le 2\mathbb{P}\Big(|W_1| \ge C_1 x\sigma_{kh_n}/2\Big)\\ &\quad + \exp\Big(-C_1 xynh_n\sigma_{kh_n}/2 + C_2 yn^{-1/3}\Big),\end{aligned} \tag{A.6}$$

where $C_1, C_2 > 0$ are finite constants. This also supplies a bound for $\mathbb{E}[\exp(-yV_k)]$. Next, observe that

$$U_k \le nh_n\lambda \sup_{kh_n\le t\le (k+1)h_n} \left|\int_{kh_n}^t (\sigma_s - \sigma_{kh_n}) dW_s\right| \stackrel{def}{=} U_k^+.$$

Using the law of the maximum of (time-transformed) Brownian motion, we see

$$U_k^+ \stackrel{d}{=} nh_n\lambda\sqrt{\int_{kh_n}^{(k+1)h_n}(\sigma_s-\sigma_{kh_n})^2ds}|Z| \text{ with } Z\sim N(0,1).$$

By the Lipschitz property of σ_t, the integral is of order h_n^3. We obtain by Cauchy-Schwarz inequality

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(-T_k\Big)\Big] &\le \mathbb{E}\Big[\exp\Big(-V_k\Big)\Big]+\Big\|\exp\Big(-V_k\Big)\Big\|_2\Big\|\Big(\exp(U_k^+)-1\Big)\Big\|_2\\ &\le \mathbb{E}\Big[\exp\Big(-V_k\Big)\Big]+\Big\|\exp\Big(-V_k\Big)\Big\|_2\sqrt{\exp\Big(C(nh_n^{5/2}\lambda)^2\Big)-1}\\ &\le \mathbb{E}\Big[\exp\Big(-V_k\Big)\Big]+\mathbb{E}\Big[\exp\Big(-2V_k\Big)\Big]^{1/2}C'nh_n^{5/2}\lambda,\end{aligned}$$

with some constants C, C', noting $nh_n^{5/2}\lambda\to 0$. Combining the above with $\mathcal{K}>1$, we thus conclude for some constant C:

$$\begin{aligned}\mathbb{E}\Big[\exp\Big(-T_k\Big)\Big] &\le \mathbb{E}\Big[\exp\Big(-V_k\Big)\Big]+C\Big\|\exp\Big(-V_k\Big)\Big\|_2 n\lambda h_n^{5/2}\\ &\le \mathbb{E}\Big[\exp\Big(-V_k\Big)\Big]+Cn\lambda h_n^{5/2}\mathbb{P}\Big(|W_1|\ge Cx\sigma_{kh_n}/2\Big).\end{aligned}$$

In the same manner, one obtains a lower bound, and the claim follows. □

Proof of Theorem 3.5 and Corollary 3.6. Let $M_{k,n}=\big(m_{n,2k}-m_{n,2k-1}\big)^2 2h_n^{-1}r_n$ and $a_{n,l}=lr_n^{-1}/2$, $s_{n,l}=lh_nr_n^{-1}=2a_{n,l}h_n$. It follows with (A.14) that

$$\Bigg|\sum_{k=a_{n,l-1}+1}^{a_{n,l}}\Big(\mathbb{E}[M_{k,n}]-2r_n\Psi(\sigma^2_{s_{n,l-1}})\Big)\Bigg|\lesssim h_n+\sup_{s_{n,l-1}\le t\le s_{n,l}}\Big|\Psi(\sigma_t^2)-\Psi(\sigma^2_{s_{n,l}})\Big|.$$

Proposition A.2 and the Lipschitz continuity of σ_t yield that

$$\sup_{s_{n,l-1}\le t\le s_{n,l}}\Big|\Psi(\sigma_t^2)-\Psi(\sigma^2_{s_{n,l-1}})\Big|\lesssim \sup_{s_{n,l-1}\le t\le s_{n,l}}\Big|\sigma_t-\sigma_{s_{n,l}}\Big|\lesssim r_n.$$

We thus conclude

$$\text{(A.7)}\qquad \Big|\Delta_{k,n}\Big|\lesssim h_n+r_n \text{ for } \Delta_{k,n}=\sum_{k=a_{n,l-1}+1}^{a_{n,l}}\Big(\mathbb{E}[M_{k,n}]-2r_n\Psi(\sigma^2_{s_{n,l-1}})\Big).$$

Next, we have that $\bar{M}_{k,n}=M_{k,n}-\mathbb{E}\big[M_{k,n}\big]$ is a sequence of independent random variables. Proposition A.1 yields that all moments of $\bar{M}_{k,n}$ exist.

Hence for any index set $\mathcal{J} \subseteq \left\{0, \ldots, h_n^{-1}-1\right\}$, Rosenthal's inequality ensures that for any $p \geq 1$

$$\left\| \sum_{k\in\mathcal{J}} \bar{M}_{k,n} \right\|_p \lesssim r_n \sqrt{\left|\mathcal{J}\right|}, \tag{A.8}$$

where $\left|\mathcal{J}\right|$ is the cardinality of the set $\mathcal{J}$. Let

$$\mathcal{M}_l = \left\{ \sum_{k=a_{n,l-1}+1}^{a_{n,l}} \bar{M}_{k,n} + \Psi\left(\sigma^2_{s_{n,l-1}}\right)/2 > 0 \right\}, \quad \text{and } \mathcal{M} = \bigcap_{l=1}^{h_n^{-1}-1} \mathcal{M}_l.$$

Note that Proposition A.2 yields that $\Psi\left(x^2\right) > 0$ for $x > 0$. Then we obtain from the Markov inequality and (A.8) that

$$\begin{aligned} \mathbb{P}\left(\bigcup_{l=0}^{h_n^{-1}-1} \mathcal{M}_l^c\right) &\leq \sum_{l=0}^{h_n^{-1}-1} \mathbb{P}\left(\mathcal{M}_l^c\right) \lesssim 2^p \sum_{l=0}^{h_n^{-1}-1} \Psi\left(\sigma^2_{s_{n,l-1}}\right)^{-p} \left\| \sum_{k=a_{n,l-1}+1}^{a_{n,l}} \bar{M}_{k,n} \right\|_p^p \\ &\lesssim \sum_{l=0}^{h_n^{-1}-1} r_n^{p/2} = \mathcal{O}\left(1\right), \end{aligned} \tag{A.9}$$

for $p > 4$. We are now ready to proceed to the main proof. From (2.6) it follows that

$$\begin{aligned} &\widetilde{IV}^{h_n,r_n} - \int_0^1 \sigma_t^2\, dt = \sum_{l=1}^{r_n h_n^{-1}} \sigma^2_{s_{n,l-1}} h_n r_n^{-1} - \int_0^1 \sigma_t^2\, dt \\ &+ \sum_{l=1}^{r_n h_n^{-1}} \left(\Psi^{-1}\left(\sum_{k=a_{n,l-1}+1}^{a_{n,l}} (m_{n,2k} - m_{n,2k-1})^2 2\, h_n^{-1} r_n \right) - \sigma^2_{s_{n,l-1}} \right) h_n r_n^{-1}. \end{aligned}$$

Consider first the approximation error in the quadratic variation by setting the volatility locally constant on the blocks of the coarse grid. By the boundedness and Lipschitz continuity of σ_t we can estimate the approximation error:

$$\begin{aligned} \sum_{l=1}^{r_n h_n^{-1}} \left| \int_{s_{n,l-1}}^{s_{n,l}} \left(\sigma^2_{s_{n,l-1}} - \sigma_t^2\right) dt \right| &\leq 2\|\sigma_t\|_\infty \sum_{l=1}^{r_n h_n^{-1}} \int_{s_{n,l-1}}^{s_{n,l}} \left|\sigma_{s_{n,l-1}} - \sigma_t\right| dt \\ &= \mathcal{O}\left(h_n r_n^{-1}\right) = \mathcal{O}\left(n^{-1/3}\right). \end{aligned}$$

In order to bound the remaining estimation error

$$\sum_{l=1}^{r_n h_n^{-1}} \left(\Psi^{-1}\left(\sum_{k=a_{n,l-1}+1}^{a_{n,l}} M_{k,n}\right) - \sigma^2_{s_{n,l-1}}\right) h_n r_n^{-1},$$

we use a Taylor expansion and that the first two derivatives of Ψ^{-1} exist and are bounded according to Proposition A.2 below. More precisely, it follows that

$$\begin{aligned}
&\Psi^{-1}\left(\sum_{k=a_{n,l-1}+1}^{a_{n,l}} M_{k,n} - \Psi\big(\sigma^2_{s_{n,l-1}}\big) + \Psi\big(\sigma^2_{s_{n,l-1}}\big)\right)\\
&= \sigma^2_{s_{n,l-1}} + \big(\Psi^{-1}\big)'\big(\Psi(\sigma^2_{s_{n,l-1}})\big)\left(\sum_{k=a_{n,l-1}+1}^{a_{n,l}} \big(M_{k,n} - \mathbb{E}[M_{k,n}] + \Delta_{k,n}\big)\right)\\
&+ \frac{1}{2}\big(\Psi^{-1}\big)''\big(\xi_l\big)\left(\sum_{k=a_{n,l-1}+1}^{a_{n,l}} \big(M_{k,n} - \mathbb{E}[M_{k,n}] + \Delta_{k,n}\big)\right)^2\\
&\overset{def}{=} \sigma^2_{s_{n,l-1}} + \Delta\big(\Psi\big)_{l,1} + \Delta\big(\Psi\big)_{l,2},
\end{aligned}$$

where $\xi_l \geq \Psi\big(\sigma^2_{s_{n,l-1}}\big)/2 > 0$ on the set $\mathcal{M}_l$. We first deal with $\Delta\big(\Psi\big)_{l,1}$ and set $Z_l = \sum_{k=a_{n,l-1}+1}^{a_{n,l}}\big(M_{k,n} - \mathbb{E}[M_{k,n}]\big)$. Using the independence of Z_l, it follows from (A.8) that

$$\left\|\sum_{l=1}^{r_n h_n^{-1}} \big(\Psi^{-1}\big)'\big(\Psi(\sigma^2_{s_{n,l-1}})\big) Z_l\right\|_2^2 \lesssim \sum_{l=1}^{r_n h_n^{-1}} \big\|Z_l\big\|_2^2 \lesssim r_n^2 h_n^{-1} = \mathcal{O}\big(1\big). \tag{A.10}$$

On the other hand, we obtain from (A.7) that

$$\left|\sum_{l=1}^{r_n h_n^{-1}} \big(\Psi^{-1}\big)'\big(\Psi(\sigma^2_{s_{n,l-1}})\big) \sum_{k=a_{n,l-1}+1}^{a_{n,l}} \Delta_{k,n}\right| \lesssim r_n^2 h_n^{-1} = \mathcal{O}\big(1\big). \tag{A.11}$$

Combining (A.10) and (A.11), we find

$$\left\|\sum_{l=1}^{r_n h_n^{-1}} \Delta\big(\Psi\big)_{l,1} \mathbb{1}_{\mathcal{M}}\right\|_2 = \mathcal{O}\big(1\big). \tag{A.12}$$

In the same manner, but using additionally $\|(\Psi^{-1})''(\xi_l)\mathbb{1}_{\mathcal{M}_l}\|_\infty < \infty$ by Proposition A.2 below, we obtain

$$\left\|\sum_{l=1}^{r_n h_n^{-1}} \Delta\big(\Psi\big)_{l,2} \mathbb{1}_{\mathcal{M}}\right\|_1 \lesssim \sum_{l=1}^{r_n h_n^{-1}} \left(\big\|Z_l\big\|_2^2 + \Big(\sum_{k=a_{n,l-1}+1}^{a_{n,l}} \Delta_{k,n}\Big)^2\right) = \mathcal{O}(1).$$

Since $\mathbb{P}\big(\mathcal{M}^c\big) = \scriptstyle{\mathcal{O}}\big(1\big)$ by (A.9), this suffices to guarantee that

$$\widetilde{IV}^{h_n,r_n} - \int_0^1 \sigma_t^2\,dt = \mathcal{O}_{\mathbb{P}}\big(n^{-1/3}\big)\,.$$

Based on a Taylor expansion for Ψ_n^{-1} and using analogous bounds and Proposition A.3, we obtain likewise

$$\widehat{IV}^{h_n,r_n} - \int_0^1 \sigma_t^2\,dt = \mathcal{O}_{\mathbb{P}}\big(n^{-1/3}\big)\,.$$

and conclude Corollary 3.6. □

A.2. Properties of $\mathbf{\Psi}$. We shall use the following identities for moments of real random variables:

$$\mathbb{E}\big[X\big] = \int_0^\infty \mathbb{P}\big(X > x\big)dx - \int_0^\infty \mathbb{P}\big(-X > x\big)dx\,, \tag{A.13a}$$

$$\mathbb{E}\big[X^2\big] = 2\int_0^\infty x\mathbb{P}\big(X > x\big)dx + 2\int_0^\infty x\mathbb{P}\big(-X > x\big)dx\,. \tag{A.13b}$$

It follows from Proposition A.1 and (A.13a), (A.13b) that for $\tilde{\mathbf{\Psi}}(\sigma) = \Psi(\sigma^2)$

$$h_n^{-1}\mathbb{E}\big[(m_{n,k} - m_{n,k-1})^2\big] = \tilde{\mathbf{\Psi}}(\sigma) + \mathcal{O}\big(nh_n^{5/2}\big). \tag{A.14}$$

Having understood the behaviour of $\tilde{\mathbf{\Psi}}(\sigma)$, analogue properties of $\Psi(\sigma^2)$ readily follow. Let

$$H(x) = \int_0^1 (W_t + x)_+\,dt. \tag{A.15}$$

Then by (A.13a), (A.13b), we derive

$$\begin{aligned}\tilde{\mathbf{\Psi}}\big(\sigma\big) &= 4\sigma^2\int_0^\infty x\bigg(\mathbb{E}\left[e^{-\mathcal{K}\sigma H(x)}\right] + 1 - \mathbb{E}\left[e^{-\mathcal{K}\sigma H(-x)}\right]\bigg)dx \\ &\quad - 2\sigma^2\bigg(\int_0^\infty\bigg(\mathbb{E}\left[e^{-\mathcal{K}\sigma H(x)}\right] - 1 + \mathbb{E}\left[e^{-\mathcal{K}\sigma H(-x)}\right]\bigg)dx\bigg)^2.\end{aligned} \tag{A.16}$$

Next, consider the distribution on the negative half axis. With $x < 0$, we make the decomposition

$$\begin{aligned}&\mathbb{E}\Big[e^{-\sigma\mathcal{K}\int_0^1 (W_t - x)_- dt}\Big] \\ &= \mathbb{E}\Big[e^{-\sigma\mathcal{K}\int_0^1 (W_t - x)_- dt}\mathbb{1}\big(\inf_{0\le t\le 1} W_t \le x\big) + \mathbb{1}\big(\inf_{0\le t\le 1} W_t \ge x\big)\Big] \\ &\stackrel{def}{=} U_1(x) + U_2(x).\end{aligned}$$

Let T_x be the first passage time of W to level x with density

$$f_{T_x}(t) = \frac{|x|}{\sqrt{2\pi t^3}} e^{-x^2/2t}\,, t \geq 0\,,$$

see (6.3) in Section 2.6 of [19]. From $\Big\{T_x \leq 1\Big\} = \Big\{\inf_{0\leq t\leq 1} W_t \leq x\Big\}$ it follows from the strong Markov property of W that

$$\begin{aligned} U_1(x) &= \int_0^1 \mathbb{E}\left[e^{-\sigma\mathcal{K}\int_s^1 (W_t - x)_- dt}\middle| T_x = s\right] f_{T_x}(s)ds \\ &= \int_0^1 \mathbb{E}\left[e^{-\sigma\mathcal{K}\int_0^{1-s} (W_t)_- dt}\right] f_{T_x}(s)ds. \end{aligned}$$

Using a time shift yields

$$U_1(x) = \int_0^1 \mathbb{E}\left[e^{-\sigma\mathcal{K}(1-s)^{3/2}\int_0^1 (W_t)_- dt}\right] f_{T_x}(s)ds.$$

We then obtain that

$$\begin{aligned} \mathbb{E}\left[e^{-\sigma\mathcal{K}\int_0^1 (W_t - x)_- dt}\right] &= \mathbb{P}\Big(\inf_{0\leq t\leq 1} W_t \geq x\Big) + \int_0^1 \mathbb{E}\left[e^{-\sigma\mathcal{K}(1-s)^{3/2}\int_0^1 (W_t)_- dt}\right] f_{T_x}(s)ds \\ &\stackrel{def}{=} \mathbb{P}\Big(\inf_{0\leq t\leq 1} W_t \geq x\Big) + A^-(x), \quad \text{for } x < 0. \end{aligned} \tag{A.17}$$

Let $I(\mathcal{K}\sigma, s) = \mathbb{E}\left[e^{-\mathcal{K}\sigma(1-s)^{3/2}\int_0^1 (W_t)_- dt}\right]$. Then by (A.17)

$$\begin{aligned} &\int_0^\infty x\Big(1 - \mathbb{E}\left[e^{-\mathcal{K}\sigma H(-x)}\right]\Big)dx = \int_0^\infty x\Big(\mathbb{P}\Big(\inf_{0\leq t\leq 1} W_t < -x\Big) - A^-(-x)\Big)dx \\ &= \int_0^\infty x\mathbb{P}(\inf_{0\leq t\leq 1} W_t < -x)dx - \int_0^1 I(\mathcal{K}\sigma, s)\int_0^\infty x f_{T_x}(s)dxds \\ &= \frac{1}{2} - \frac{1}{2}\int_0^1 I(\mathcal{K}\sigma, s)ds, \end{aligned}$$

since $\int_0^\infty x\mathbb{P}(\inf_{0\leq t\leq 1} W_t < -x)dx = \frac{1}{2}$. Likewise, it follows that

$$\begin{aligned} \int_0^\infty \Big(1 - \mathbb{E}\left[e^{-\mathcal{K}\sigma H(-x)}\right]\Big)dx &= \int_0^\infty \mathbb{P}\Big(\inf_{0\leq t\leq 1} W_t < -x\Big)dx - \sqrt{\frac{2}{\pi}}\int_0^1 I(\mathcal{K}\sigma, s)ds \\ &= \sqrt{\frac{2}{\pi}} - \sqrt{\frac{2}{\pi}}\int_0^1 I(\mathcal{K}\sigma, s)ds. \end{aligned} \tag{A.18}$$

We thus obtain

$$\begin{aligned}\tilde{\mathbf{\Psi}}\big(\sigma\big) &= 4\sigma^2\bigg(\int_0^\infty x\mathbb{E}\left[e^{-\mathcal{K}\sigma H(x)}\right]dx + \frac{1}{2} - \frac{1}{2}\int_0^1 I(\mathcal{K}\sigma, s)ds\bigg)\\ &\quad - 2\sigma^2\bigg(\int_0^\infty \mathbb{E}\left[e^{-\mathcal{K}\sigma H(x)}\right]dx - \sqrt{\frac{2}{\pi}} + \sqrt{\frac{2}{\pi}}\int_0^1 I(\mathcal{K}\sigma, s)ds\bigg)^2\\ &= 2\sigma^2\big(2\Lambda_1(\sigma) - \Lambda_2^2(\sigma)\big)\,, \end{aligned} \tag{A.19}$$

with functionals Λ_1, Λ_2. In the sequel, we write $\partial^k f(x) = \partial^k f(x)/\partial^k x$. The further analysis of properties of $\tilde{\mathbf{\Psi}}$ is structured in several lemmas which combined imply the following key proposition.

PROPOSITION A.2. *Suppose that $\sigma \geq \sigma_0 > 0$, $\mathcal{K} \geq C(\sigma_0)$ for $C(\sigma_0)$ sufficiently large (the exact value of $C(\sigma_0)$ follows from (A.24)). Then we have uniformly for $\sigma \geq \sigma_0$*

$$\begin{aligned}\partial\tilde{\mathbf{\Psi}}\big(\sigma\big) &= 4\sigma\left(1 - \frac{2}{\pi}\right) + \mathcal{O}\left(\frac{\sigma^{\frac{2}{3}}}{\mathcal{K}^{\frac{1}{3}}}\right) > 0 \quad \textit{and}\\ \tilde{\mathbf{\Psi}}\big(\sigma\big) &= 2\sigma^2\left(1 - \frac{2}{\pi}\right) + \mathcal{O}\left(\frac{\sigma^{\frac{2}{3}}}{\mathcal{K}^{\frac{1}{3}}}\right) > 0.\end{aligned} \tag{A.20}$$

Moreover, it holds that

$$\sup_{\sigma\geq\sigma_0}\left|\partial^k\tilde{\mathbf{\Psi}}(\sigma)\right| < \infty \quad \textit{for } k = 1, 2. \tag{A.21}$$

Using the relation

$$\partial\tilde{\mathbf{\Psi}}^{-1}(\varrho) = \frac{1}{\partial\tilde{\mathbf{\Psi}}(\sigma)}, \quad \tilde{\mathbf{\Psi}}(\sigma) = \varrho, \tag{A.22}$$

we get that the second derivative is uniformly bounded for $\sigma \geq \sigma_0 = \tilde{\mathbf{\Psi}}^{-1}(\varrho_0)$, i.e.

$$\sup_{\varrho\geq\varrho_0}\left|\partial^2\tilde{\mathbf{\Psi}}^{-1}(\varrho)\right| = \sup_{\sigma\geq\sigma_0}\left|\frac{\partial^2\tilde{\mathbf{\Psi}}(\sigma)}{\big(\partial\tilde{\mathbf{\Psi}}(\sigma)\big)^3}\right| < \infty. \tag{A.23}$$

So far we have focused on results for $\tilde{\mathbf{\Psi}}(\sigma) = \Psi(\sigma^2)$. Essentially the same results are valid for $\Psi_n(\sigma^2)$, which we state now.

PROPOSITION A.3. *Introduce*

$$B_{n,1} = \int_0^\infty x\mathbb{P}\Big(\max_{0\le i\le nh_n-1} W_{i/(nh_n)} \ge x\Big)dx \quad \textit{and}$$
$$B_{n,2} = \int_0^\infty \mathbb{P}\Big(\max_{0\le i\le nh_n-1} W_{i/(nh_n)} \ge x\Big)dx,$$

which satisfy $B_{n,1} \to \frac{1}{2}$, $B_{n,2} \to \sqrt{\frac{2}{\pi}}$. *Then* (A.20) *and* (A.21) *in Proposition A.2 remain valid if we replace* $\tilde{\mathbf{\Psi}}\big(\sigma\big)$ *with* $\Psi_n\big(\sigma^2\big)$ *and* $1-\frac{2}{\pi}$ *with* $2B_{n,1}-B_{n,2}^2$. *Likewise,* (A.23) *also holds.*

PROOF OF PROPOSITION A.2. We write shortly $\Lambda_1 = \Lambda_1(\sigma)$, $\Lambda_2 = \Lambda_2(\sigma)$. We have that

$$\partial\tilde{\mathbf{\Psi}}\big(\sigma\big) = 4\sigma\big(2\Lambda_1 - \Lambda_2^2\big) + 2\sigma^2\left(2\partial\Lambda_1 - 2\Lambda_2\partial\Lambda_2\right).$$

Using Lemmas A.4 and A.7 from below, we obtain with (A.19)

$$\Big|\Lambda_1 - \frac{1}{2}\Big| \le 6(\mathcal{K}\sigma)^{-\frac{2}{3}} + \frac{3}{2}\left(\frac{\mathcal{K}\sigma}{\log(\mathcal{K}\sigma)}\right)^{-\frac{2}{5}} \stackrel{def}{=} R_{\Lambda_1}$$
$$\Big|\Lambda_2 - \sqrt{\frac{2}{\pi}}\Big| \le 2\Big(\sqrt{\frac{2}{\pi}}+1\Big)(\mathcal{K}\sigma)^{-\frac{1}{3}} + 3\sqrt{\frac{2}{\pi}}\left(\frac{\mathcal{K}\sigma}{\log(\mathcal{K}\sigma)}\right)^{-\frac{2}{5}} \stackrel{def}{=} R_{\Lambda_2}.$$

Moreover, applying Lemmas A.5, A.6 and A.7 yields

$$\Big|\partial\Lambda_1\Big| \le 6\big(\mathcal{K}\sigma^3\big)^{-\frac{1}{2}} + \frac{3}{2\sigma}\left(\frac{\mathcal{K}\sigma}{\log(\mathcal{K}\sigma)}\right)^{-\frac{2}{5}} \stackrel{def}{=} R_{\partial\Lambda_1}, \quad \Lambda_2^2 \le \frac{2}{\pi},$$
$$\Big|\Lambda_2\partial\Lambda_2\Big| \le \left(R_{\Lambda_2} + \sqrt{\frac{2}{\pi}}\right)\left(4\big(1+(2\pi)^{-\frac{1}{2}}\big)\mathcal{K}^{-\frac{1}{3}}\sigma^{-\frac{4}{3}} + \sqrt{\frac{2}{\pi}}\frac{3}{\sigma}\left(\frac{\mathcal{K}\sigma}{\log(\mathcal{K}\sigma)}\right)^{-\frac{2}{5}}\right)$$
$$\stackrel{def}{=} R_{\partial\Lambda_2}.$$

We thus obtain from the above that

$$\Big|\partial\tilde{\mathbf{\Psi}} - 4\sigma\big(1-\frac{2}{\pi}\big)\Big| \le 4\sigma\left(2R_{\Lambda_1} + R_{\Lambda_2}\Big(R_{\Lambda_2} + 2\sqrt{\frac{2}{\pi}}\Big)\right)$$
$$+ 4\sigma^2\big(R_{\partial\Lambda_1} + R_{\partial\Lambda_2}\big) = \mathcal{O}\Big(\mathcal{K}^{-\frac{1}{3}}\sigma^{\frac{2}{3}}\Big),$$

$$\Big|\tilde{\mathbf{\Psi}} - 2\sigma^2\Big(1-\sqrt{\frac{2}{\pi}}\Big)\Big| \le 2\sigma^2\left(R_{\Lambda_1} + R_{\Lambda_2}\Big(R_{\Lambda_2} + 2\sqrt{\frac{2}{\pi}}\Big)\right) = \mathcal{O}\Big(\mathcal{K}^{-\frac{1}{3}}\sigma^{\frac{2}{3}}\Big).$$

An explicit sufficient lower bound for $\mathcal{K}$ in terms of σ_0 can be computed from the two conditions

$$1-\frac{2}{\pi} > \left(2R_{\Lambda_1}+R_{\Lambda_2}\left(R_{\Lambda_2}+2\sqrt{\frac{2}{\pi}}\right)\right)+\sigma\big(R_{\partial\Lambda_1}+R_{\partial\Lambda_2}\big),$$

$$1-\sqrt{\frac{2}{\pi}} > R_{\Lambda_1}+R_{\Lambda_2}\left(R_{\Lambda_2}+2\sqrt{\frac{2}{\pi}}\right). \tag{A.24}$$

It remains to show the boundedness property for the first two derivatives of $\tilde{\boldsymbol{\Psi}}$. By Lemma A.8, we have

$$\Big|\partial^k J\big(\sigma\big)\Big| \le C\left(1+\sigma^{-k}\right),\ k=1,2, \tag{A.25}$$

where $J\big(\sigma\big)=2\Lambda_1(\sigma)-\Lambda_2^2(\sigma)$ and C is a constant not depending on σ. Observe that

$$\partial^2\tilde{\boldsymbol{\Psi}}=4J+6\sigma\partial J+\sigma^2\partial^2 J,$$

hence the claim follows. □

Proof of Proposition A.3. The proof can be redirected to Proposition A.2 using Proposition 3.2 and a truncation argument for the integrals over x. The corresponding computations are very similar to those above and the Lemmas given below. We therefore omit the details. □

Lemma A.4. *For $\mathcal{K}>0$, $p\in\mathbb{N}_0$, we obtain the following decay behaviour of the moment integrals:*

$$\int_0^\infty x^p\mathbb{E}\left[e^{-\mathcal{K}\sigma H(x)}\right]dx \le 2^{p+1}\left(\frac{\mathbb{E}[|Z|^{p+1}]}{p+1}+\Gamma(p+1)\right)(\mathcal{K}\sigma)^{-(p+1)/3}$$

with $Z\sim N(0,1)$.

Proof of Lemma A.4. The following useful relation in terms of the $N(0,1)$-distribution function Φ is derived from the law of the minimum of Brownian motion:

$$\mathbb{P}\big(T_x\le l\big)=2(1-\Phi(|x|/\sqrt{l})). \tag{A.26}$$

Then for $0 < l < 1$

$$\int_0^\infty x^p \mathbb{E}\left[e^{-\mathcal{K}\sigma H(x)}\right] dx \le \int_0^\infty x^p \mathbb{E}\left[\mathbb{1}\big(T_{-x/2} \le l\big)\right] dx \\ + \int_0^\infty x^p \mathbb{E}\left[e^{-\mathcal{K}\sigma H(x)} \mathbb{1}\big(T_{-x/2} > l\big)\right] dx \stackrel{def}{=} R_1 + R_2\,.$$

By (A.26) we have

$$R_1 = 2\int_0^\infty x^p(1 - \Phi(x/\sqrt{4l}))\, dx \\ = (4l)^{(p+1)/2} \int_0^\infty 2z^p(1 - \Phi(z))\, dz = (4l)^{(p+1)/2} \frac{\mathbb{E}[|Z|^{p+1}]}{p+1}.$$

We further note that $T_{-x/2} > l$ implies $H(x) \ge \int_0^l (-x/2 + x)_+ dt = lx/2$, such that

$$R_2 \le \int_0^\infty x^p e^{-\mathcal{K}\sigma l x/2} dx = \left(\frac{2}{\mathcal{K}\sigma l}\right)^{p+1} \Gamma\big(p+1\big).$$

Choosing $l = (\mathcal{K}\sigma)^{-2/3}$, we obtain

$$\int_0^\infty x^p \mathbb{E}\left[e^{-\mathcal{K}\sigma H(x)}\right] dx \le \left(2^{p+1}\frac{\mathbb{E}[|Z|^{p+1}]}{p+1} + 2^{p+1}\Gamma\big(p+1\big)\right)(\mathcal{K}\sigma)^{-(p+1)/3},$$

as asserted. □

LEMMA A.5. *Let $\mathcal{K} > 0$. Then*

$$\int_0^\infty \mathbb{E}\left[\mathcal{K}H(x)e^{-\mathcal{K}\sigma H(x)}\right] dx \le 4\big(1 + 1/\sqrt{2\pi}\big)\mathcal{K}^{-1/3}\sigma^{-4/3}.$$

PROOF OF LEMMA A.5. We make the decomposition

$$\int_0^\infty \mathbb{E}\left[\mathcal{K}\sigma H(x)e^{-\mathcal{K}\sigma H(x)}\right] dx = \int_0^\infty \mathbb{E}\left[\mathbb{1}(T_{-x/2} \le l)\mathcal{K}\sigma H(x)e^{-\mathcal{K}\sigma H(x)}\right] dx \\ + \int_0^\infty \mathbb{E}\left[\mathbb{1}(T_{-x/2} > l)\mathcal{K}\sigma H(x)e^{-\mathcal{K}\sigma H(x)}\right] dx\,,$$

with some $l > 0$. Using $ye^{-y} \le 1$ and (A.26), we obtain

$$\int_0^\infty \mathbb{E}\left[\mathbb{1}(T_{-x/2} \le l)\mathcal{K}\sigma H(x)e^{-\mathcal{K}\sigma H(x)}\right] dx \le 2\int_0^\infty (1 - \Phi(x/\sqrt{4l}))\, dx \\ = \sqrt{8l/\pi}.$$

Now, using $ye^{-y} \leq e^{-y/2}$ and $T_{-x/2} > l \Rightarrow H(x) \geq lx/2$, we bound the other term by

$$\int_0^\infty \mathbb{E}\left[\mathbb{1}(T_{-x/2} > l)\mathcal{K}\sigma H(x)e^{-\mathcal{K}\sigma H(x)}\right] dx \leq \int_0^\infty e^{-\mathcal{K}\sigma l x/4} dx = \frac{4}{\mathcal{K}\sigma l}.$$

The choice $l = (\mathcal{K}\sigma)^{-2/3}$ and division by σ yield the claim. ☐

LEMMA A.6. *Let $\mathcal{K} > 0$. Then*

$$\int_0^\infty x\mathbb{E}\left[\mathcal{K}H(x)e^{-\mathcal{K}\sigma H(x)}\right] dx \leq \frac{6}{\mathcal{K}^{1/2}\sigma^{3/2}}.$$

PROOF OF LEMMA A.6. We proceed as in the proof of Lemma A.5 and obtain for any $l > 0$

$$\begin{aligned}\int_0^\infty x\mathbb{E}\left[\mathcal{K}\sigma H(x)e^{-\mathcal{K}\sigma H(x)}\right] dx &\leq \int_0^\infty x\Big(2(1-\Phi(x/\sqrt{4l})) + e^{-\mathcal{K}\sigma l x/4}\Big)\, dx \\ &= 2l + (\mathcal{K}\sigma l/4)^{-1}.\end{aligned}$$

The result follows with $l = (\mathcal{K}\sigma)^{-1/2}$. ☐

LEMMA A.7. *Let $\mathcal{K} \geq \sigma^{-1}$. Then*

$$\int_0^1 I(\mathcal{K}\sigma, s)ds \leq 3\Big(\frac{\mathcal{K}\sigma}{\log(\mathcal{K}\sigma)}\Big)^{-2/5}, \quad \left|\frac{\partial \int_0^1 I(\mathcal{K}\sigma, s)ds}{\partial\sigma}\right| \leq \frac{3}{\sigma}\Big(\frac{\mathcal{K}\sigma}{\log(\mathcal{K}\sigma)}\Big)^{-2/5}.$$

PROOF OF LEMMA A.7. With $\lambda(s) = \mathcal{K}(1-s)^{3/2}$ we obtain for any $T > 0$

$$\begin{aligned}\int_0^1 I(\mathcal{K}\sigma, s)ds &= \int_0^1 \mathbb{E}\Big[e^{-\lambda(s)\sigma\int_0^1 (W_t)_- dt}\Big]\, ds \\ &\leq \int_0^1 \Big(\mathbb{P}\Big(\lambda(s)\sigma\int_0^1 (W_t)_- dt \leq T\Big) + e^{-T}\Big)\, ds.\end{aligned}$$

From $\int_0^1 (W_t)_- dt \geq |Z|$ with $Z = \int_0^1 W_t\, dt \sim N(0, 1/3)$, we deduce $\mathbb{P}(\int_0^1 (W_t)_- dt \leq \varepsilon) \leq \varepsilon$, $\varepsilon > 0$, and thus

$$\left|\frac{\partial \int_0^1 I(\mathcal{K}\sigma, s)ds}{\partial\sigma}\right| \leq \int_0^1 \Big(\Big(T\sigma^{-1}\lambda(s)^{-1}\Big) \wedge 1\Big)\, ds + e^{-T}.$$

Using $(\sigma\lambda(s))^{-1} \leq (\mathcal{K}\sigma/T)^{3/5}$ for $s \leq 1 - (\mathcal{K}\sigma/T)^{-2/5}$ the last integral is bounded by $2(\mathcal{K}\sigma/T)^{-2/5}$. The choice $T = \log(\mathcal{K}\sigma)$ yields the first inequality.

Then using $ye^{-y} \le e^{-y/2}$ we also obtain

$$\left|\frac{\partial \int_0^1 I(\mathcal{K}\sigma, s)ds}{\partial\sigma}\right| = \int_0^1 \mathbb{E}\Big[\lambda(s)\int_0^1 (W_t)_- dt\, e^{-\lambda(s)\sigma\int_0^1 (W_t)_- dt}\Big]\, ds$$
$$\le \sigma^{-1}\int_0^1 \Big(\mathbb{P}\Big(\lambda(s)\sigma\int_0^1 (W_t)_- dt \le T\Big) + e^{-T/2}\Big)\, ds.$$

The previous bounds now apply in the same way. □

LEMMA A.8. *Consider* $J\big(\sigma\big) = 2\Lambda_1(\sigma) - \Lambda_2^2(\sigma)$. *Then there exists a constant* $B = B(\mathcal{K}) > 0$ *only depending on* $\mathcal{K}$ *such that*

$$\Big|\partial^k J\big(\sigma\big)\Big| \le B(1 + \sigma^{-k}), \quad k = 1, 2. \tag{A.27}$$

PROOF OF LEMMA A.8. Without loss of generality, we may assume that $\mathcal{K} = 1$. From the considerations below, existence of the k'th derivative of $J\big(\sigma\big)$ with respect to σ follows. We thus focus on establishing (A.27). First consider $\int_0^\infty x\mathbb{E}\left[e^{-\sigma H(x)}\right] dx$. An application of the Cauchy-Schwarz inequality gives

$$\left|\frac{\partial^k \int_0^\infty x\mathbb{E}\left[e^{-\sigma H(x)}\right] dx}{\partial^k\sigma}\right| = \left|\int_0^\infty x\mathbb{E}\left[\big(-H(x)\big)^k e^{-\sigma H(x)}\right] dx\right|$$
$$\le \int_0^\infty x\mathbb{E}\left[H(x)^{2k}\right]^{1/2} \mathbb{E}\left[e^{-2\sigma H(x)}\right]^{1/2} dx. \tag{A.28}$$

Applying the triangle and Cauchy-Schwarz inequality further yields

$$\mathbb{E}\left[H(x)^{2k}\right]^{1/2} \le \mathbb{E}\left[\left(\int_0^1 |W_s|ds + |x|\right)^{2k}\right]^{1/2} \lesssim 1 \vee x^k. \tag{A.29}$$

The calculations in the proof of Lemma A.4 with $l = \sqrt{x/\sigma}/2$ yield

$$\mathbb{E}[\exp(-2\sigma H(x))] \lesssim \exp(-x^{3/2}\sigma^{1/2}/2). \tag{A.30}$$

Combining (A.29) and (A.30), we deduce that

$$\int_0^\infty x\mathbb{E}\left[H(x)^{2k}\right]^{1/2} \mathbb{E}\left[e^{-2\sigma H(x)}\right]^{1/2} dx$$
$$\lesssim \int_0^\infty \big(x \vee x^{k+1}\big)\exp(-x^{3/2}\sigma^{1/2}/2)\, dx \lesssim \sigma^{-2/3}(1 + \sigma^{-k/3}).$$

This implies that for some $C>0$

$$\left|\frac{\partial^k \int_0^\infty x\mathbb{E}\left[e^{-\sigma H(x)}\right]dx}{\partial^k\sigma}\right| \leq C\sigma^{-2/3}(1+\sigma^{-k/3})\,. \tag{A.31}$$

Arguing in the same manner, one also establishes that

$$\left|\frac{\partial^k \int_0^\infty \mathbb{E}\left[e^{-\sigma H(x)}\right]dx}{\partial^k\sigma}\right| \leq C\sigma^{-1/3}(1+\sigma^{-k/3})\,. \tag{A.32}$$

Moreover, such bounds are also valid for the derivatives of $\int_0^1 I(\sigma,s)ds$. □

APPENDIX B: PROOF OF THEOREM 2

After the reductions of the problem to a simpler and more informative experiment, we now prove Theorem 4.1 using properties of the Hellinger distance $H(P,Q)$ between probability measures, in particular $H^2(P_1\otimes P_2, Q_1\otimes Q_2)\leq H^2(P_1,Q_1)+H^2(P_2,Q_2)$ (subadditivity under independence), $H^2(P,Q)=\mathbb{E}[H^2(P,Q|T)]$ (Hellinger distance conditional on a statistic T) and

$$H^2(PPP(\lambda_1),PPP(\lambda_2))\leq\int(\sqrt{\lambda_1}-\sqrt{\lambda_2})^2$$

(Hellinger bound for PPP measures with intensity densities λ_i, cf. [21]).

Put $\delta_n=\delta\sigma_0^{5/3}n^{-1/3}$. From $H^2(N(0,\sigma_0^2),N(0,\sigma_0^2+\delta_n))\leq 2(\delta_n\sigma_0^{-2})^2$, cf. Appendix in [23], and the independent increments of Brownian motion we infer for the Hellinger distance of the laws of $(X_{T_j^s})_{j=1,\dots,J}$ under σ_0^2 and $\sigma_0^2+\delta_n$

$$H^2\left(P_{\sigma_0^2}^{(X_{T_j^s})},P_{\sigma_0^2+\delta_n}^{(X_{T_j^s})}\,\Big|\,(T_j^s)\right)\leq\sum_{j=1}^J 2\delta_n^2\sigma_0^{-4}=2J\delta_n^2\sigma_0^{-4}.$$

For each PPP with intensity density λ^j we obtain by integral calculations, in terms of $\eta=(\sigma_0^2+\delta_n)^{1/2}-\sigma_0$:

$$\begin{aligned}
&H^2\Big(PPP(\lambda^j(\sigma_0^2)),PPP(\lambda^j(\sigma_0^2+\delta_n))\,\Big|\,(T_j^s),B^{0,T_j^s-T_{j-1}^s}\Big)\\
&\leq n\int_0^{T_j^s-T_{j-1}^s}\int_{\mathbb{R}}\Big(\Big(b^{-1}\big(y-\sigma_0B_t^{0,T_j^s-T_{j-1}^s}\big)_+\Big)\wedge 1\\
&\qquad-\Big(b^{-1}\big(y-(\sigma_0^2+\delta_n)^{1/2}B_t^{0,T_j^s-T_{j-1}^s}\big)_+\Big)\wedge 1\Big)^2dy\,dt\\
&=nb\int_0^{T_j^s-T_{j-1}^s}\int_{\mathbb{R}}\Big(u_+\wedge 1-\big(u-b^{-1}\eta B_t^{0,T_j^s-T_{j-1}^s}\big)_+\wedge 1\Big)^2dy\,dt\\
&\leq nb\int_0^{T_j^s-T_{j-1}^s}b^{-2}\eta^2(B_t^{0,T_j^s-T_{j-1}^s})^2dt.
\end{aligned}$$

Hence, by using the variance of a Brownian bridge we arrive at

$$H^2\Big(PPP(\lambda^j(\sigma_0^2)), PPP(\lambda^j(\sigma_0^2+\delta_n))\,\Big|\,(T_j^s)\Big)$$
$$\le nb^{-1}\eta^2\int_0^{T_j^s-T_{j-1}^s} t(1-(T_j^s-T_{j-1}^s)^{-1}t)\,dt = \frac{n\eta^2}{6b}(T_j^s-T_{j-1}^s)^2.$$

Since conditional on (T_j^s) all observations are independent, the total squared Hellinger distance conditional on (T_j^s) is bounded by

$$2J\delta_n^2\sigma_0^{-4}+\frac{n\eta^2}{6b}\sum_{j=1}^{J+1}(T_j^s-T_{j-1}^s)^2.$$

Taking expectations and using $J\sim \mathrm{Poiss}(2nb/3)$, $T_j^s-T_{j-1}^s\sim \mathrm{Exp}(2nb/3)$ to apply the Wald identity to the second sum, the unconditional total Hellinger distance is bounded by

$$H^2\le \frac{4nb\delta_n^2}{3\sigma_0^4}+\frac{n\eta^2}{6b}(2nb/3)^{-1}(1+o(1)).$$

We have $\eta^2\le \frac{1}{2}\delta_n^2\sigma_0^{-2}$ due to $\sqrt{1+x}\le 1+x/2$ for $x>0$ and thus by choosing $b\propto(\sigma_0^2/n)^{1/3}$ optimally and plugging in δ_n

$$H^2\le \delta^2 n^{1/3}\sigma_0^{4/3}\inf_{b>0}\Big(\frac{2b}{3\sigma_0^2}+\frac{C}{12b^2n}\Big)\le C'\delta^2.$$

From the general lower bound Theorem 2.2(ii) in [25] we thus obtain the result if δ is chosen smaller than $2/C'$.

REFERENCES

[1] Aït-Sahalia, Y., Mykland, P. A. and Zhang, L. (2005). How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*, *18*, 351–416.

[2] Bacry, E. and J.F Muzy (2013) Hawkes model for price and trades high-frequency dynamics. *preprint, arxiv:1301.1135*

[3] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A. and Shephard, N. (2008). Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica 76* (6), 1481–1536.

[4] Biais, B., Hillion, P. and Spatt, C. (1995). An empirical analysis of the limit order book and the order flow in the paris bourse. *The Journal of Finance*, *50* (5), 1655–1689.

[5] Bibinger, M., Jirak, M. and Reiss, M. (2014). Applying volatility estimators based on limit order books, *Technical Report*, available under www.mathematik.hu-berlin.de/for1735/Publ/application.pdf.

[6] Bouchaud, J.P, Mzard, M. and Potters, M. (2002). Statistical properties of stock order books: Empirical results and models. *Quantitative Finance*, *2* (4), 251–256.
[7] Carmona, R. and Lacker, D. (2014). A probabilistic weak formulation of mean field games and applications. *The Annals of Applied Probability,* to appear.
[8] Clément, E., Delattre, S. and Gloter, A. (2013). An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. *Stochastic Processes and their Applications*, *123* (7), 2500–2521.
[9] Cont, R., Stoikov, S. and Talreja, R. (2010). A stochastic model for order book dynamics. *Operations Research*, *58* (3), 549–563.
[10] Cont, R., A. de Larrard (2012). Order Book Dynamics in Liquid Markets: Limit Theorems and Diffusion Approximations. *preprint, SSRN: http://ssrn.com/abstract=1757861*
[11] Foucault, T., Kadan, O. and Kandel, E. (2005). Limit order book as a market for liquidity. *Review of Financial Studies*, *18*, 1171–1217.
[12] Gatheral, J. and R. C. A. Oomen (2010). Zero-Intelligence Realized Variance Estimation. *Finance and Stochastics*, *14* (2), 249–283.
[13] Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors 1 and 2. *ESAIM, Probability and Statistics*, *5*, 225–242.
[14] Jacod, J., Podolskij, M. and Vetter, M. (2010). Limit theorems for moving averages of discretized processes plus noise. *The Annals of Statistics 38* (3), 1478–1545.
[15] T. Jaisson and Rosenbaum, M. (2014). Limit theorems for nearly unstable Hawkes processes. *The Annals of Applied Probability,* to appear.
[16] Janson, S. (2007). Brownian excursion area, Wright's constants in graph enumeration, and other Brownian areas. *Probability Surveys 4*, 80–145.
[17] Meister, A. and Reiss, M. (2013). Asymptotic equivalence for nonparametric regression with non-regular errors. *Probability Theory and Related Fields 155*, (1-2), 201–229.
[18] Naes, R. and J. A. Skjeltorp (2006). Order book characteristics and the volume-volatility relation: Empirical evidence from a limit order market. *Journal of Financial Markets, 9* (4), 408–432.
[19] Karatzas, I. and S. E. Shreve (1991). *Brownian motion and stochastic calculus. Second Edition.* Volume 113 of *Graduate Texts in Mathematics.* New York: Springer-Verlag.
[20] Karr, A. F. (1991) Point processes and their statistical inference. Second edition. *Probability: Pure and Applied*, New York: Marcel Dekker, Inc.
[21] Kutoyants, Y. (1998) *Statistical Inference for Spatial Poisson Processes.* Lecture Notes in Statistics, Springer.
[22] Perman, M. and J. A. Wellner (1996). On the distribution of Brownian areas. *The Annals of Applied Probability 6* (4), 1091–1111.
[23] Reiss, M. (2011). Asymptotic equivalence for inference on the volatility from noisy observations. *The Annals of Statistics*, *2*, 772–802.
[24] Rosu, I. (2009). A dynamic model of the limit order book. *Review of Financial Studies, 22* (11), 4601–4641.
[25] Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation.* New York: Springer-Verlag.
[26] Varlamov, V. (2008). Differential and integral relations involving fractional derivatives of Airy functions and applications. *Journal of Mathematical Analysis and Applications 348* (1), 101–115.
[27] Zhang, L., Mykland, P. A. and Aït-Sahalia, Y. (2005). A tale of two time

scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association 100* (472), 1394–1411.

Markus Bibinger,
Moritz Jirak,
Markus Reiss,
Institut für Mathematik
Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin, Germany