Bernoulli **18**(4), 2012, 1421–1447
DOI: 10.3150/11-BEJ384

Model checks for the volatility under microstructure noise

MATHIAS VETTER* and HOLGER DETTE**

Ruhr-Universität Bochum, Fakultät für Mathematik, 44780 Bochum, Germany.
E-mail: *mathias.vetter@rub.de; **holger.dette@rub.de

We consider the problem of testing the parametric form of the volatility for high frequency data. It is demonstrated that in the presence of microstructure noise commonly used tests do not keep the preassigned level and are inconsistent. The concept of preaveraging is used to construct new tests, which do not suffer from these drawbacks. These tests are based on a Kolmogorov–Smirnov or Cramér–von-Mises functional of an integrated stochastic process, for which weak convergence to a (conditional) Gaussian process is established. The finite sample properties of a bootstrap version of the test are illustrated by means of a simulation study.

Keywords: goodness-of-fit test; heteroscedasticity; microstructure noise; parametric bootstrap; stable convergence

1. Introduction

The volatility is a popular measure of risk in finance with numerous applications including the construction of optimal portfolios, hedging and pricing of options. Therefore, estimating and investigating the volatility and its dynamics is of particular importance in applications and numerous models have been proposed for this purpose (see, e.g., Black and Scholes [6], Vasicek [25], Cox *et al.* [9], Hull and White [17] and Heston [16] among many others). Because the misspecification of the form of the volatility can lead to serious consequences in the subsequent data analysis numerous authors recommend to use goodness-of-fit tests for the postulated model (see, e.g., Ait-Sahalia [3], Corradi and White [8], Dette *et al.* [11], Dette and Podolskij [10] among others).

In the present paper, we consider statistical inference in the case of high frequency data, where for an increasing sample size information about the whole path of the volatility is in principle available. However, in concrete applications the situation is more complicated because of the presence of microstructure noise, which is usually persistent in such data. This additional noise is caused by many sources of the trading process such as discreteness of observations (see, e.g., Harris [14], [15]), bid-ask bounces or special properties of the trading mechanism (see, e.g., Black [5] or Amihud and Mendelson [4]). While microstructure noise has been taken into account for the construction of estimators of the integrated volatility and other related quantities (see, e.g., Zhang *et al.* [26], Jacod *et al.* [19] or Podolskij and Vetter [22], [21]), properties of goodness-of-fit tests in this context have not been investigated so far in the literature.

Consider for example the problem, where the process $\{Z_t\}_{t\in[0,1]}$ is observed at the n time points $0, 1/n, \ldots, 1$. Under the assumption that $Z_t = X_t = \sigma_t \, \mathrm{d}W_t$, Dette and Podolskij [10]

Table 1. Simulated level of the test (1.1) for various choices of ω and θ, where the true volatility function is $\sigma^2(t,x) = \theta + (1-\theta)x^2$ and the noise terms U are normally distributed with mean zero and variance ω^2. In all cases, the sample size is given by $n = 16\,384$

ω	0.01			0.0025			0.000625		
θ/α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
1	0.01	0.02	0.038	0.023	0.058	0.104	0.024	0.047	0.101
0.75	0.004	0.01	0.02	0.004	0.009	0.022	0.003	0.007	0.015
0.5	0.003	0.006	0.013	0.002	0.004	0.014	0.000	0.000	0.002
0.25	0.002	0.004	0.015	0.001	0.002	0.003	0.001	0.003	0.004
0	0.000	0.005	0.019	0.003	0.006	0.015	0.004	0.007	0.016

propose to reject the hypothesis of a constant diffusion coefficient, that is, H_0: $\sigma_t^2 = \sigma^2(t, X_t) = \sigma^2$, whenever

$$\begin{aligned} T_n(Z_1,\dots,Z_n) = \sqrt{n}\sup_{t\in[0,1]}\left|\frac{\sum_{k=1}^{\lfloor nt\rfloor}|Z_{k/n}-Z_{(k-1)/n}|^2 - t\sum_{k=1}^{n}|Z_{k/n}-Z_{(k-1)/n}|^2}{\sqrt{2}\sum_{k=1}^{n}|Z_{k/n}-Z_{(k-1)/n}|^2}\right| \\ > c_{1-\alpha}, \end{aligned} \tag{1.1}$$

where $c_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the supremum of a Brownian Bridge. Now consider the situation, where microstructure noise is present, which is usually modeled by an additional additive component, that is

$$Z_{i/n} = X_{i/n} + U_{i/n}, \qquad i = 1,\dots,n, \tag{1.2}$$

where $\{U_{i/n} \mid i = 1,\dots,n\}$ denotes a triangular array of independent random variables with mean 0 and variance ω^2. In Table 1, we show the finite sample behaviour of the test (1.1) for the hypothesis of a constant volatility if $\sigma_t^2 = \sigma^2(t,x) = \theta + (1-\theta)x^2$ (note that the case $\theta = 1$ corresponds to the null hypothesis). We observe that the test keeps its preassigned level only in the case where ω is rather small. In most cases, the nominal level is clearly underestimated. On the other hand, the test is not able to detect any alternative. An intuitive explanation for this behaviour is that in the presence of microstructure noise the increments $Z_{i/n} - Z_{(i-1)/n} = U_{i/n} - U_{(i-1)/n} + \mathrm{O}_p(1/n)$ are dominated by the noise variables. This leads to inconsistent estimates of the integrated volatility as pointed out in Zhang *et al.* [26]. More precisely, a straightforward calculation shows that under microstructure noise the statistic $T_n(Z_1,\dots,Z_n)$ shows the same asymptotic behavior as the statistic $T_n(U_1,\dots,U_n)$, which converges weakly to $\sqrt{\lambda/2}\sup_{t\in[0,1]}|B_t|$, no matter if the null hypothesis is valid or not. Here B_t denotes a Brownian bridge and $\lambda = E[(U_{k/n}/\omega)^4]$. This means that in the presence of microstructure noise the test (1.1) has asymptotic level α if and only if $\lambda = 2$. In all other cases, the test does not keep its preassigned level. Moreover, because the asymptotic properties under null hypothesis and alternative are the same, the test is not consistent.

The present paper is devoted to the problem of constructing a consistent asymptotic level α test for a general parametric form of the volatility in the presence of microstructure noise.

In Sections 2 and 3, we present the basic model and introduce a stochastic process which can be used to test parametric hypotheses about the form of the volatility in a noisy framework. Our main results are presented in Section 4, where we prove stable convergence of two such processes which form the basis of the proposed goodness-of-fit tests. Section 5 deals with the problem of testing nonlinear hypotheses for the volatility, whereas in Section 6 the finite sample properties of a bootstrap version of the new tests are investigated. All proofs of the results are presented in the Appendix.

2. Testing parametric hypotheses for the volatility

Suppose that the process $X = (X_t)_t$ admits the representation

$$X_t = X_0 + \int_0^t a_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}W_s, \tag{2.1}$$

where $W = (W_t)_t$ is a standard Brownian motion and the drift process a and the volatility process σ satisfy some weak regularity conditions, which will be specified later. Furthermore, we assume that the process can be observed at discrete points on a fixed time interval, say $[0, 1]$.

Various assumptions on the structure of the volatility process have been proposed in the literature. Among such models, a large class involves the case where σ is defined to be a local volatility process, thus merely a function of time and state (see, e.g., Black and Scholes [6], Vasicek [25], Cox *et al.* [9], Chan *et al.* [7], Ait-Sahalia [3] or Ahn and Gao [2] among many others). Because an appropriate modeling of the volatility is of particular importance for the construction of portfolios, hedging and pricing, many authors point out that the postulated model should be validated by an appropriate goodness-of-fit test (see, e.g., Ait-Sahalia [3] or Corradi and White [8]). In several cases, the hypothesis for the parametric form of the volatility is linear and one has to consider the following two situations:

$$\begin{aligned} &H_0\colon\ \sigma_t^2 = \sigma^2(t, X_t) = \sum_{i=1}^d \theta_i \sigma_i^2(t, X_t) \qquad \forall t \text{ a.s.} \quad \text{or} \\ &\bar{H}_0\colon\ \sigma_t = \sigma(t, X_t) = \sum_{i=1}^d \bar{\theta}_i \bar{\sigma}_i(t, X_t) \qquad \forall t \text{ a.s.}, \end{aligned} \tag{2.2}$$

where the functions $\sigma_1, \dots, \sigma_d$ (or $\bar{\sigma}_1, \dots, \bar{\sigma}_d$) are known and the parameters $\theta_1, \dots, \theta_d$ (or $\bar{\theta}_1, \dots, \bar{\theta}_d$) are unknown, but assumed to ensure $\sigma^2(t, X_t) \geq 0$ (or $\sigma(t, X_t) \geq 0$) almost surely. Other models involve volatility functions, where the parameters enter nonlinearly (see Ait-Sahalia [3]) and the corresponding hypotheses will be considered later in Section 5, because the basic concepts are easier to explain in the linear context.

Let us focus on the problem involving H_0 for the moment, as the other testing problem can be treated in the same way. Dette and Podolskij [10] propose to construct a test statistic using an empirical version of the stochastic process

$$N_t = \int_0^t \left\{ \sigma_s^2 - \sum_{j=1}^d \theta_j^{\min} \sigma_j^2(s, X_s) \right\} \mathrm{d}s, \qquad \theta^{\min} = \operatorname*{arg\,min}_{\theta \in \mathbb{R}^d} \int_0^1 \left\{ \sigma_s^2 - \sum_{j=1}^d \theta_j \sigma_j^2(s, X_s) \right\}^2 \mathrm{d}s.$$

Thus, one uses the L^2 distance to determine the best approximation to the unknown volatility process σ^2 by a linear combination of the given functions $\sigma_1^2, \ldots, \sigma_d^2$. It can easily be seen that H_0 is equivalent to $N_t = 0$ $\forall t$ a.s., and a well-known result from Hilbert space theory (see Achieser [1]) implies

$$\theta^{\min} = D^{-1} C, \quad \text{thus} \quad N_t = B_t^0 - B_t^T D^{-1} C, \tag{2.3}$$

where

$$B_t^0 = \int_0^t \sigma_s^2 \,\mathrm{d}s \quad \text{and} \quad B_t^i = \int_0^t \sigma_i^2(s, X_s) \,\mathrm{d}s \qquad \text{for } i = 1, \ldots, d,$$

and D and C denote a $d \times d$-matrix and a d-dimensional vector, respectively, with

$$D_{ij} = \int_0^1 \sigma_i^2(s, X_s) \sigma_j^2(s, X_s) \,\mathrm{d}s \quad \text{and} \quad C_i = \int_0^1 \sigma_s^2 \sigma_i^2(s, X_s) \,\mathrm{d}s.$$

In practice, one does not observe the entire path of the diffusion process $X = (X_t)_t$ and it is therefore necessary to define an empirical version based on appropriate estimators for the quantities in (2.3). Let us briefly discuss the solution to the problem in the case, where X can be observed without further restrictions. Based on the decomposition above, Dette and Podolskij [10] propose to define an empirical version $\tilde{N}_t = \tilde{B}_t^0 - \tilde{B}_t^T \tilde{D}^{-1} \tilde{C}$, where one uses a Riemann approximation of each integral, choosing $n|X_{k/n} - X_{(k-1)/n}|^2$ as a local estimate for $\sigma^2_{(k-1)/n}$. Thus,

$$\tilde{D}_{ij} = \frac{1}{n} \sum_{k=1}^n \sigma_i^2\left(\frac{k}{n}, X_{k/n}\right) \sigma_j^2\left(\frac{k}{n}, X_{k/n}\right) \qquad \text{for } i, j = 1, \ldots, d, \tag{2.4}$$

$$\tilde{C}_i = \sum_{k=1}^n \sigma_i^2\left(\frac{k-1}{n}, X_{(k-1)/n}\right) |X_{k/n} - X_{(k-1)/n}|^2 \qquad \text{for } i = 1, \ldots, d,$$

and the quantities $\tilde{B}_t^0$ and $\tilde{B}_t = (\tilde{B}_t^1, \ldots, \tilde{B}_t^d)^T$ are given by

$$\tilde{B}_t^0 = \sum_{k=1}^{\lfloor nt \rfloor} |X_{k/n} - X_{(k-1)/n}|^2, \qquad \tilde{B}_t^i = \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor} \sigma_i^2\left(\frac{k}{n}, X_{k/n}\right) \qquad \text{for } i = 1, \ldots, d. \tag{2.5}$$

In this context, one can prove a (stable) central limit theorem for the process $(\tilde{N}_t - N_t)_t$ with the optimal rate of convergence $n^{-1/2}$, from which one may assess the distribution of suitable test statistics. For example, if $d = 1$, $\sigma_1^2(t, X_t) = 1$, the hypothesis H_0 reduces to the hypothesis of constant volatility considered in the introduction, and the Kolmogorov–Smirnov statistic (1.1) converges to the supremum of a Brownian bridge.

3. Assumptions and definitions

Since we are dealing with microstructure noise, we have to define a process $Z = (Z_t)_t$ which represents the noisy observations. Typically one relates Z to the underlying Ito semimartingale X through the equation $Z_t = X_t + U_t$ for some noise process U. We restrict ourselves to the case of i.i.d. noise, in which the process $U = (U_t)_t$ is independent of X and satisfies

$$E[U_t] = 0, \qquad E[U_t^2] = \omega^2, \qquad E[U_t^4] < \infty \tag{3.1}$$

with a density having compact support. A precise definition of a proper probability space that accommodates Z can be found in Jacod *et al.* [19]. We assume further that Z is observed at times $0, 1/n, \ldots, 1$. As pointed out in the introduction, the corresponding test based on $\tilde{N}_t$ is not consistent for the hypothesis H_0 in the presence of such microstructure noise. Thus, our aim is to define appropriate estimators for the unknown quantities in (2.3) in this noisy framework, from which a more adequate statistic $\hat{N}_t$ can be constructed. Note that in contrast to the previous setting we do not only need a local estimator for the unknown volatility function σ^2, but also for the (unobservable) path of X itself.

The natural approach in order to construct estimators for the volatility is to use increments of Z as in the no-noise case, even though a single increment does not provide sufficient information about σ^2. This problem can be overcome by applying the idea of pre-averaging, which was invented in Podolskij and Vetter [22] and is based on moving averages of Z. To this end, we choose first a sequence m_n, such that

$$\frac{m_n}{\sqrt{n}} = \kappa + \mathrm{o}(n^{-1/4}) \tag{3.2}$$

for some $\kappa > 0$, and a nonzero real-valued function $g : \mathbb{R} \to \mathbb{R}$, which vanishes outside of the interval $(0, 1)$, is continuous and piecewise C^1 and has a piecewise Lipschitz derivative g'. We associate with g (and n) the following real valued numbers and functions:

$$\begin{cases} g_j^n = g\left(\dfrac{j}{m_n}\right), \qquad g_j'^n = g_j^n - g_{j+1}^n, \qquad \psi_1 = \displaystyle\int_0^1 (g'(s))^2 \,\mathrm{d}s, \qquad \psi_2 = \displaystyle\int_0^1 (g(s))^2 \,\mathrm{d}s, \\ s \in [0,1] \mapsto \phi_1(s) = \displaystyle\int_s^1 g'(u) g'(u-s) \,\mathrm{d}u, \qquad \phi_2(s) = \displaystyle\int_s^1 g(u) g(u-s) \,\mathrm{d}u, \\ i, j = 1, 2: \qquad \Phi_{ij} = \displaystyle\int_0^1 \phi_i(s) \phi_j(s) \,\mathrm{d}s. \end{cases} \tag{3.3}$$

Finally, we define for an arbitrary process V the preaveraged statistic

$$\overline{V}_k^n = \sum_{j=1}^{m_n} g_j^n \Delta_{k+j}^n V, \tag{3.4}$$

where $\Delta_j^n V = V_{j/n} - V_{(j-1)/n}$. Due to the assumptions on g the pre-averaged statistic $\overline{Z}_k^n$ reduces the impact of the noise, but still provides information about the increments of X (and thus locally about σ). Precisely, we have

$$\overline{X}_k^n = \mathrm{O}_p\left(\sqrt{\frac{m_n}{n}}\right) \quad \text{and} \quad \overline{U}_k^n = \mathrm{O}_p\left(\sqrt{\frac{1}{m_n}}\right), \tag{3.5}$$

and by definition of m_n both terms are of the same order. This means in particular that statistics based on $\overline{Z}_k^n$ are in general biased when used for volatility estimation, but it turns out that a larger choice of m_n results in a worse rate of convergence. See Podolskij and Vetter [22] for details.

An estimator for $X_{k/n}$ can be constructed in a similar way: We set

$$\hat{X}_{k/n} = \frac{1}{m_n} \sum_{j=1}^{m_n} Z_{(k+j)/n}, \tag{3.6}$$

and it is easy to see that this procedure reduces the impact of the noise variables around time $\frac{k}{n}$, but still provides information about the latent price $X_{k/n}$, since the path of X is Hölder continuous of any order $\alpha < 1/2$. Also one observes essentially from (3.5) that the auxiliary sequence m_n is chosen in the optimal way, giving the smallest possible size for the approximation error.

As pointed out before, we need additional assumptions on the process X as well as on the given basis functions in H_0 and $\bar{H}_0$, respectively. Since the conditions on σ_i^2 and $\bar{\sigma}_i$ are similar, we will restrict ourselves to the first case only.

It is required that the functions $\sigma_1^2, \ldots, \sigma_d^2$ are linearly independent and that each σ_i^2 is twice continuously differentiable. Moreover, we assume that $E[|\det(D)|^{-\beta}] < \infty$ for some $\beta > 0$.

Regarding the various processes in X, the assumptions are as weak as possible when testing for H_0. We simply have to ensure that the process in (2.1) is well defined, which follows if we assume that a is locally bounded and predictable and that σ is càdlàg (see Jacod and Shiryaev [20] or Revuz and Yor [23]). When working with $\bar{H}_0$ we propose additionally that the true volatility process σ is almost surely positive and that is has a representation of the form (2.1) as well, namely that it satisfies

$$\sigma_t = \sigma_0 + \int_0^t a_s' \,\mathrm{d}s + \int_0^t \sigma_s' \,\mathrm{d}W_s + \int_0^t v_s' \,\mathrm{d}V_s,$$

where a', σ' and v' are adapted càdlàg processes, with a' also being predictable and locally bounded, and V is a second Brownian motion, independent of W. Moreover, a is supposed to be càglàg.

4. Goodness-of-fit tests addressing microstructure noise

We start with the construction of a test for the hypothesis H_0 again. Local estimators for the volatility can now be obtained from $|\overline{Z}_k^n|^2$, but we have seen before that this quantity is not an unbiased estimate for $\sigma^2_{k/n}$ and that it has a different stochastic order than the increments $X_{k/n} - X_{(k-1)/n}$ in the no-noise case. A corrected statistic (see Jacod *et al.* [19]) is given by

$$\hat{\sigma}^2_{k/n} = \frac{n^{1/2}}{\kappa\psi_2}\left(|\overline{Z}^n_k|^2 - n^{-1/2}\frac{\psi_1}{\kappa}\hat{\omega}^2_n\right) \qquad \text{with } \hat{\omega}^2_n = \frac{1}{2n}\sum_{i=1}^{n}|\Delta^n_i Z|^2, \tag{4.1}$$

where the latter term is a consistent estimator for ω^2, see Zhang *et al.* [26]. Mimicking the procedure from the no-noise case presented in Section 2, we set

$$\hat{D}_{ij} = \frac{1}{n}\sum_{k=1}^{n-m_n}\sigma^2_i\left(\frac{k}{n}, \hat{X}_{k/n}\right)\sigma^2_j\left(\frac{k}{n}, \hat{X}_{k/n}\right) \quad \text{and} \quad \hat{C}_i = \frac{1}{n}\sum_{k=1}^{n-m_n}\sigma^2_i\left(\frac{k}{n}, \hat{X}_{k/n}\right)\hat{\sigma}^2_{k/n} \tag{4.2}$$

as well as

$$\hat{B}^0_t = \frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor - m_n}\hat{\sigma}^2_{k/n} \quad \text{and} \quad \hat{B}^i_t = \frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor - m_n}\sigma^2_i\left(\frac{k}{n}, \hat{X}_{k/n}\right) \tag{4.3}$$

for $i, j = 1, \dots, d$. We define at last the process

$$\hat{N}_t = \hat{B}^0_t - \hat{B}^T_t\hat{D}^{-1}\hat{C}, \tag{4.4}$$

which turns out to be an appropriate estimate of the process $\{N_t\}_{t\in[0,1]}$. Our first result specifies the asymptotic properties of the process $\{A_n(t)\}_{t\in[0,1]}$ with $A_n(t) = n^{1/4}(\hat{N}_t - N_t)$.

Theorem 1. *If the assumptions stated in the previous sections are satisfied, the process* $(A_n(t))_{t\in[0,1]}$ *converges weakly in* $D[0,1]$ *to a mean zero process* $(A(t))_{t\in[0,1]}$. *Conditionally on* $\mathcal{F}$ *the limiting process is Gaussian, and its finite dimensional distributions coincide with the conditional (with respect to* $\mathcal{F}$*) finite dimensional distributions of the process*

$$\left\{\gamma_V\left(I\{V \le t\} - B^T_t D^{-1}h(V, X_V)\right) - \left(\int_0^t \gamma_s\,\mathrm{d}s - B^T_t D^{-1}\int_0^1 \gamma_s h(s, X_s)\,\mathrm{d}s\right)\right\}_{t\in[0,1]}, \tag{4.5}$$

where $V \sim \mathcal{U}[0,1]$, $h(s, X_s) = (\sigma^2_1(s, X_s), \dots, \sigma^2_d(s, X_s))^T$ *and*

$$\gamma^2_s = \frac{4}{\psi^2_2}\left(\Phi_{22}\kappa\sigma^4_s + 2\Phi_{12}\frac{\sigma^2_s\omega^2}{\kappa} + \Phi_{11}\frac{\omega^4}{\kappa^3}\right). \tag{4.6}$$

We see from Theorem 3 in the Appendix that the asymptotics is only driven by $\hat{B}^0_t$ and $\hat{C}$. The error due to the estimation of B_t and D is of small order, which explains the particular form of the limiting distribution. Note also that the rate of convergence $n^{-1/4}$ is optimal for this problem, since it is already optimal for the estimation of B^0_t even in a parametric setting (cf. Gloter and Jacod [13]).

In order to construct a test statistic based on Theorem 1, we have to define an appropriate estimator for the conditional variance of the process $\{A(t)\}_{t\in[0,1]}$, which is given by

$$s_t^2 = \int_0^t \gamma_s^2 \,\mathrm{d}s - 2B_t^T D^{-1} \int_0^t \gamma_s^2 g(s, X_s)\,\mathrm{d}s + B_t^T D^{-1} \int_0^1 \gamma_s^2 g(s, X_s) g^T(s, X_s)\,\mathrm{d}s\, D^{-1} B_t.$$

Obviously, we use $\hat{B}_t$ and $\hat{D}$ as the empirical counterparts for B_t and D. In order to obtain estimates for the other random elements of s_t^2, note that γ_s^2 plays a key role in Jacod *et al.* [19] as well, where it is the (local) conditional variance in a central limit theorem for $n^{1/4}(\hat{B}_t^0 - B_t^0)$. Thus, in accordance to that paper we define

$$\begin{aligned}\Gamma_k &= \frac{4\Phi_{22}}{3\kappa\psi_2^4}|\overline{Z}_k^n|^4 + n^{-1/2}\frac{8}{\kappa^2}\left(\frac{\Phi_{12}}{\psi_2^3} - \frac{\Phi_{22}\psi_1}{\psi_2^4}\right)|\overline{Z}_k^n|^2\hat{\omega}^2 \\ &\quad + n^{-1}\frac{4}{\kappa^3}\left(\frac{\Phi_{11}}{\psi_2^2} - \frac{2\Phi_{12}\psi_1}{\psi_2^3} + \frac{\Phi_{22}\psi_1^2}{\psi_2^4}\right)\hat{\omega}^4,\end{aligned}$$

which is a local estimator for the process γ^2 after rescaling. Thus, we set

$$\begin{aligned}\hat{g}_0(t) &= \sum_{k=1}^{\lfloor nt\rfloor - m_n} \Gamma_k \xrightarrow{P} \int_0^t \gamma_s^2\,\mathrm{d}s,\\ g_i(t) &= \sum_{k=1}^{\lfloor nt\rfloor - m_n} \Gamma_k \sigma_i^2\left(\frac{k-1}{n}, \hat{X}_{(k-1)/n}\right) \xrightarrow{P} \int_0^t \gamma_s^2\sigma_i^2(s, X_s)\,\mathrm{d}s,\\ \hat{g}_{ij} &= \sum_{k=1}^{n} \Gamma_k \sigma_i^2\left(\frac{k-1}{n}, \hat{X}_{(k-1)/n}\right)\sigma_j^2\left(\frac{k-1}{n}, \hat{X}_{(k-1)/n}\right) \xrightarrow{P} \int_0^1 \gamma_s^2\sigma_i^2(s, X_s)\sigma_j^2(s, X_s)\,\mathrm{d}s.\end{aligned}$$

Inserting these estimators into the corresponding elements of s_t^2 gives the consistent estimator

$$\hat{s}_t^2 = \hat{g}_0(t) - 2\hat{B}_t^T\hat{D}^{-1}\hat{g}(t) + \hat{B}_t^T\hat{D}^{-1}\hat{G}\hat{D}^{-1}\hat{B}_t, \tag{4.7}$$

where $\hat{g}(t) = (\hat{g}_1(t), \ldots, \hat{g}_d(t))^T$ and $\hat{G} = (\hat{g}_{ij})_{i,j=1}^d$. A consistent test for the hypothesis H_0 is now obtained by rejecting the null hypothesis for large values of Kolmogorov–Smirnov or Cramér–van-Mises functional of the process $\{n^{1/4}\hat{N}_t/\hat{s}_t\}_{t\in[0,1]}$. Note however that the distribution of this process is not feasible in general: even though for each fixed t the statistic $n^{1/4}\hat{N}_t/\hat{s}_t$ converges weakly to a standard normal distribution, the covariance structure of the process typically depends on the entire (unobservable) process $(X_t)_t$. For this reason, we will later use a bootstrap procedure to obtain critical values.

In principle, a similar approach can be used to construct a test for the hypothesis $\bar{H}_0$. However, in this case things change considerably. Dette and Podolskij [10] restate this hypothesis as $M_t = 0$

$\forall t$ a.s., where

$$M_t = \int_0^t \left\{ \sigma_s - \sum_{j=1}^d \bar{\theta}_j^{\min} \bar{\sigma}_j(s, X_s) \right\} \mathrm{d}s,$$
$$\bar{\theta}^{\min} = \underset{\bar{\theta} \in \mathbb{R}^d}{\arg\min} \int_0^1 \left\{ \sigma_s - \sum_{j=1}^d \bar{\theta}_j \bar{\sigma}_j(s, X_s) \right\}^2 \mathrm{d}s. \tag{4.8}$$

Obviously, we have an analogous representation as in (2.3), namely $M_t = R_t^0 - R_t^T Q^{-1} S$, where

$$R_t^0 = \int_0^t \sigma_s \, \mathrm{d}s \quad \text{and} \quad R_t^i = \int_0^t \bar{\sigma}_i(s, X_s) \, \mathrm{d}s \qquad \text{for } i = 1, \ldots, d,$$

and Q and S are a $d \times d$-matrix and a d-dimensional vector, respectively, with

$$Q_{ij} = \int_0^1 \bar{\sigma}_i(s, X_s) \bar{\sigma}_j(s, X_s) \, \mathrm{d}s \quad \text{and} \quad S_i = \int_0^1 \sigma_s \bar{\sigma}_i(s, X_s) \, \mathrm{d}s.$$

However, an appropriate definition of an empirical version of the form $\hat{M}_t = \hat{R}_t^0 - \hat{R}_t^T \hat{Q}^{-1} \hat{S}$ requires some less obvious modifications, because local estimators for σ_s are more difficult to obtain in this setting. Using a preaveraged estimator of the form $|\overline{Z}_k^n|$ again causes an intrinsic bias, but due to the absolute value (instead of the square as in the previous setting) its correction turns out to be impossible at the optimal rate. However, we can see from (3.5) that using in (3.2) a sequence of a larger magnitude than $n^{1/2}$ reduces the impact of the noise terms in $\overline{Z}_k^n$. This modification makes inference about σ_s possible, though resulting in a worse rate of convergence. To be precise, we fix some $\delta > \frac{1}{6}$ and choose l_n such that

$$\frac{l_n}{n^{1/2+\delta}} = \rho + \mathrm{o}\big(n^{-(1/4+\delta/2)}\big)$$

for some $\rho > 0$. Using the sequence l_n instead of m_n, we define all quantities from (3.3) to (3.6) in the straightforward way. Next, we set

$$\bar{\sigma}_{k/n} = n^{1/4-\delta/2} \frac{1}{\sqrt{\rho \psi_2} \mu_1} |\overline{Z}_k^n|$$

as a local estimator for $\sigma_{k/n}$, where μ_1 denotes the first absolute moment of a standard normal distribution. In a similar way as before,

$$\hat{Q}_{ij} = \frac{1}{n} \sum_{k=1}^{n-l_n} \bar{\sigma}_i\left(\frac{k}{n}, \hat{X}_{k/n}\right) \bar{\sigma}_j\left(\frac{k}{n}, \hat{X}_{k/n}\right) \quad \text{and} \quad \hat{S}_i = \frac{1}{n} \sum_{k=1}^{n-l_n} \bar{\sigma}_i\left(\frac{k}{n}, \hat{X}_{k/n}\right) \bar{\sigma}_{k/n}$$

as well as

$$\hat{R}_t^0 = \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor - l_n} \bar{\sigma}_{k/n} \quad \text{and} \quad \hat{R}_t^i = \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor - l_n} \bar{\sigma}_i\left(\frac{k}{n}, \hat{X}_{k/n}\right)$$

for $i, j = 1, \ldots, d$. Finally, we define $B_n(t) = n^{1/4-\delta/2}(\hat{M}_t - M_t)$ for any $t \in [0,1]$ and obtain the following result.

Theorem 2. *If the assumptions stated in the previous sections are satisfied, the process* $(B_n(t))_{t\in[0,1]}$ *converges weakly in* $D[0,1]$ *to a mean zero process* $(B(t))_{t\in[0,1]}$. *Conditionally on* $\mathcal{F}$ *the limiting process is Gaussian, and its finite dimensional distributions coincide with the conditional* (*with respect to* $\mathcal{F}$) *finite dimensional distributions of the process*

$$\left\{\bar{\gamma}_V\big(I\{V \le t\} - R_t^T Q^{-1}\bar{h}(V, X_V)\big) - \left(\int_0^t \bar{\gamma}_s\,\mathrm{d}s - R_t^T Q^{-1}\int_0^1 \bar{\gamma}_s \bar{h}(s, X_s)\,\mathrm{d}s\right)\right\}_{t\in[0,1]}, \tag{4.9}$$

where $V \sim \mathcal{U}[0,1]$, $\bar{h}(s, X_s) = (\bar{\sigma}_1(s, X_s), \ldots, \bar{\sigma}_d(s, X_s))^T$ *and*

$$\begin{aligned}\bar{\gamma}_s^2 &= \frac{2\rho\Xi}{\mu_1^2}\sigma_s^2, \qquad \Xi = \int_0^1 \xi(s)\,\mathrm{d}s, \qquad \xi(s) = f\left(\frac{\phi_2(s)}{\psi_2}\right),\\ f(u) &= \frac{2}{\pi}\big(u\arcsin(u) + \sqrt{1-u^2} - 1\big).\end{aligned} \tag{4.10}$$

The estimation of the conditional variance of the process $\{B(t)\}_{t\in[0,1]}$,

$$r_t^2 = \int_0^t \bar{\gamma}_s^2\,\mathrm{d}s - 2R_t^T Q^{-1}\int_0^t \bar{\gamma}_s^2\bar{h}(s, X_s)\,\mathrm{d}s + R_t^T Q^{-1}\int_0^1 \bar{\gamma}_s^2 \bar{h}(s, X_s)\bar{g}^T(s, X_s)\,\mathrm{d}s\, Q^{-1}R_t,$$

becomes easier in this context, as the order of l_n is chosen in such a way that no characteristics of U are involved anymore. A natural estimator for $\sigma_{k/n}^2$ becomes

$$\bar{\Gamma}_k = n^{-(1/2+\delta)}\frac{2\Xi}{\psi_2\mu_1^2}|\overline{Z}_k^n|^2,$$

thus

$$\begin{aligned}\hat{h}_0(t) &= \sum_{k=1}^{\lfloor nt\rfloor - l_n}\bar{\Gamma}_k \xrightarrow{P} \int_0^t \bar{\gamma}_s^2\,\mathrm{d}s,\\ \hat{h}_i(t) &= \sum_{k=1}^{\lfloor nt\rfloor - l_n}\bar{\Gamma}_k\bar{\sigma}_i\left(\frac{k-1}{n}, \hat{X}_{(k-1)/n}\right) \xrightarrow{P} \int_0^t \bar{\gamma}_s^2\bar{\sigma}_i(s, X_s)\,\mathrm{d}s,\\ \hat{h}_{ij} &= \sum_{k=1}^{n}\bar{\Gamma}_k\bar{\sigma}_i\left(\frac{k-1}{n}, \hat{X}_{(k-1)/n}\right)\bar{\sigma}_j\left(\frac{k-1}{n}, \hat{X}_{(k-1)/n}\right) \xrightarrow{P} \int_0^1 \bar{\gamma}_s^2\bar{\sigma}_i(s, X_s)\bar{\sigma}_j(s, X_s)\,\mathrm{d}s,\end{aligned}$$

and consequently a consistent estimator $\hat{r}_t^2$ for the conditional variance is given by

$$\hat{r}_t^2 = \hat{h}_0(t) - 2\hat{R}_t^T\hat{Q}^{-1}\hat{h}(t) + \hat{R}_t^T\hat{Q}^{-1}\hat{H}\hat{Q}^{-1}\hat{R}_t, \tag{4.11}$$

where $\hat{h}(t) = (\hat{h}_1(t), \ldots, \hat{h}_d(t))^T$ and $\hat{H} = (\hat{h}_{ij})_{i,j=1}^d$. A consistent test for the hypothesis $\bar{H}_0$ is now obtained by rejecting the null hypothesis for large values of the Kolmogorov–Smirnov or Cramér–van-Mises functional of the process $\{n^{1/4-\delta/2}\hat{M}_t/\hat{r}_t\}_{t\in[0,1]}$.

Note that one knows from previous work that it is neither necessary to define X to be an Ito semimartingale with continuous paths as in (2.1) nor to model the noise terms U as being independent and identically distributed to obtain similar results as in Theorems 1 and 2. In fact, for an underlying Ito semimartingale exhibiting jumps one can use bipower-type estimators as discussed in Podolskij and Vetter [21] in order to define an estimator closely related to $\hat{B}_t^0$. Moreover, it has been argued in Jacod *et al.* [19] that even for a noise process with a càdlàg variance a similar theory as presented in this paper applies.

5. Nonlinear hypotheses

In this section, we briefly discuss the case of a nonlinear hypothesis

$$H_0\colon\ \sigma_t^2 = \sigma^2(t, X_t) = \sigma^2(t, X_t, \theta) \qquad \forall t \text{ a.s.}, \tag{5.1}$$

where $\theta \in \Theta \subset \mathbb{R}^d$ denotes the unknown parameter and σ^2 satisfies some differentiability assumption. As before, we restate H_0 as $N_t = 0\ \forall t$ a.s., where N_t is the difference between the true integrated volatility and its best L^2-approximation from the parametric class. Therefore, we set $N_t = B_t^0 - B_t(\theta_0)$ with B_t^0 from above and $B_t(\theta) = \int_0^t \sigma^2(s, X_s, \theta)\,\mathrm{d}s$. We have $\theta_0 = \arg\min_{\theta\in\Theta} f(\theta)$ with

$$f(\theta) = \int_0^t \{\sigma_s^2 - \sigma^2(s, X_s, \theta)\}^2\,\mathrm{d}s.$$

In order to obtain some $\hat{N}_t$, we use $\hat{B}_t^0$ from (4.3) and need estimates for $B_t(\theta)$ and $f(\theta)$. We set

$$\hat{B}_t(\theta) = \frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor - m_n} \sigma^2\left(\frac{k}{n}, \hat{X}_{k/n}, \theta\right) \quad \text{and} \quad f_n(\theta) = \frac{1}{n}\sum_{k=1}^{n-m_n}\left\{\hat{\sigma}_{k/n}^2 - \sigma^2\left(\frac{k}{n}, \hat{X}_{k/n}, \theta\right)\right\}^2, \tag{5.2}$$

and with $\hat{\theta} = \arg\min_{\theta\in\Theta} f_n(\theta)$ we define $\hat{N}_t = \hat{B}_t^0 - \hat{B}_t(\hat{\theta})$.

When deriving the asymptotic distribution of $n^{1/4}(\hat{N}_t - N_t)$, the difference compared to the previous section regards only $\hat{B}_t(\theta_0) - B_t(\hat{\theta})$. In the following, we will give some hints that explain why that discrepancy is actually quite small. In fact, we will show that

$$\hat{B}_t(\hat{\theta}) - B_t(\theta_0) = -\int_0^t \left(\frac{\partial}{\partial\theta}\sigma^2(s, X_s, \theta)\Big|_{\theta=\theta_0}\right)^T \mathrm{d}s\,(f''(\theta_0))^{-1} f_n'(\theta_0) + \mathrm{o}_p(n^{-1/4}) \tag{5.3}$$

holds. Thus there is a one-to-one correspondence to the linear case, as the first two quantities are analogues of B_t^T and D^{-1}, whereas $-f_n'(\theta_0)$ plays the role of $\hat{C} - C$. Consequently, the process $n^{1/4}(\hat{N}_t - N_t)$ exhibits a similar asymptotic behavior as in the linear case.

In order to prove (5.3), note from similar arguments as in the proof of Theorem 3 that

$$\hat{B}_t(\hat{\theta}) - B_t(\theta_0) = \int_0^t \{\sigma^2(t, X_t, \hat{\theta}) - \sigma^2(t, X_t, \theta_0)\}\, \mathrm{d}s + \mathrm{o}_p(n^{-1/4}). \tag{5.4}$$

Under common regularity conditions for nonlinear regression (see Gallant [12] or Seber and Wild [24]), θ_0 is the unique minimum of f and attained at an interior point of Θ. It is easy to see that $\hat{\theta} \to \theta_0$ in probability in this case, and thus we can assume that $\hat{\theta}$ satisfies $f_n'(\hat{\theta}) = 0$. This implies

$$0 = f_n'(\hat{\theta}) = f_n'(\theta_0) + f_n''(\tilde{\theta})(\hat{\theta} - \theta_0) \quad \Leftrightarrow \quad \hat{\theta} - \theta_0 = -(f_n''(\tilde{\theta}))^{-1} f_n'(\theta_0)$$

for an appropriate choice of $\tilde{\theta}$. We have $\tilde{\theta} \to \theta_0$ in probability as well, and therefore it can be assumed that the $d \times d$-dimensional matrix $f_n''(\tilde{\theta})$ is positive definite and that the difference $\| f_n''(\tilde{\theta}) - f_n''(\theta_0)\|$ is small. Furthermore, $f_n''(\theta_0)$ takes the form

$$f_n''(\theta_0) = 2\left(\frac{1}{n} S^T S - \frac{1}{n} \sum_{k=1}^{n-m_n} \left\{\hat{\sigma}^2_{k/n} - \sigma^2\left(\frac{k}{n}, \hat{X}_{k/n}, \theta_0\right)\right\} H_k\right),$$

where the $(n - m_n) \times d$ matrix S and the Hessian H_k are given by

$$S = \left(\frac{\partial}{\partial\theta}\sigma^2\left(\frac{k}{n}, \hat{X}_{k/n}, \theta\right)\bigg|_{\theta=\theta_0}\right)_{k=1,\ldots,n-m_n} \quad \text{and} \quad H_k = \frac{\partial^2}{\partial\theta^2}\sigma^2\left(\frac{k}{n}, X_{k/n}, \theta\right)\bigg|_{\theta=\theta_0}.$$

From the same arguments that lead to (5.4), we have $f_n''(\theta_0) = f''(\theta_0) + \mathrm{O}_p(n^{-1/4})$, where

$$\begin{aligned} f''(\theta_0) = {} & 2\int_0^1 \left(\frac{\partial}{\partial\theta}\sigma^2(s, X_s, \theta)\bigg|_{\theta=\theta_0}\right)^T \left(\frac{\partial}{\partial\theta}\sigma^2(s, X_s, \theta)\bigg|_{\theta=\theta_0}\right) \mathrm{d}s \\ & - 2\int_0^1 \left\{\sigma_s^2 - \sigma^2(s, X_s, \theta_0)\right\} \frac{\partial^2}{\partial\theta^2}\sigma^2(s, X_s, \theta)\bigg|_{\theta=\theta_0} \mathrm{d}s \end{aligned}$$

is positive definite. Note that the second term in this sum vanishes, when either the hypothesis is linear (since the Hessian is zero) or the null hypothesis is valid (since σ_s^2 equals $\sigma^2(s, X_s, \theta_0)$). In these cases the matrix $f''(\theta_0)$ takes precisely the same form as D in the linear setting. In any case, $f''(\theta_0)$ is of order $\mathrm{O}_p(1)$.

Regarding $f_n'(\theta_0)$, a similar calculation as given in the Appendix plus the definition of θ_0 yield

$$\begin{aligned} -f_n'(\theta_0) = {} & 2\left(\frac{1}{n}\sum_{k=1}^{n-m_n} \hat{\sigma}^2_{k/n} \frac{\partial}{\partial\theta}\sigma^2\left(\frac{k}{n}, \hat{X}_{k/n}, \theta\right)\bigg|_{\theta=\theta_0} - \int_0^1 \sigma_s^2 \frac{\partial}{\partial\theta}\sigma^2(s, X_s, \theta)\bigg|_{\theta=\theta_0} \mathrm{d}s\right) \\ & + \mathrm{o}_p(n^{-1/4}), \end{aligned}$$

and thus $f_n'(\theta_0)$ is of order $\mathrm{O}_p(n^{-1/4})$, just as $\hat{C} - C$. We conclude that $\hat{\theta} - \theta_0 = \mathrm{O}_p(n^{-1/4})$ as well, and a Taylor expansion gives (5.3).

Table 2. Simulated level of the bootstrap test proposed by Dette and Podolskij [10], where the volatility function equals H_0: $\sigma^2(t,x) = \theta x^2$, but the observations are corrupted with normally distributed noise having variance ω^2

n	256			1024		
ω/α	0.025	0.05	0.1	0.025	0.05	0.1
0.001	0.033	0.062	0.111	0.333	0.415	0.512
0.002	0.158	0.243	0.324	0.810	0.862	0.907
0.004	0.392	0.518	0.650	0.993	0.996	0.998
0.005	0.497	0.628	0.742	0.991	0.994	0.998
0.01	0.596	0.754	0.873	0.987	0.998	0.999

6. Simulation study

We have indicated in the introduction that the original test for a constant volatility from the noise-free model loses its asymptotic properties in the presence of noise. Unsurprisingly, for a smaller variance of the noise variables, the data look more like observations from a continuous semimartingale and thus the test statistics behaves roughly in the same way as before, provided that the sample size is not too large. On the other hand, for a large variance of the error terms these are dominating, and thus the whole procedure breaks down even for small sample sizes. The same problem arises if the variance of the error is small but the sample size is large (see the discussion in the Introduction). We start with a further example simulating the level of the bootstrap test proposed by Dette and Podolskij [10] for a parametric hypothesis, assessing its quality for various sample sizes n and different variances ω^2.

Precisely, we have used that test for testing the hypothesis H_0: $\sigma^2(t,x) = \theta x^2$, where $b(t,x) = 0.1x$. The results are obtained from 1000 simulation runs and 500 bootstrap replications and displayed in Table 2 for various sample sizes and standard deviations ω of the noise process. We observe that for $n = 256$ and a (small) standard deviation of $\omega = 0.001$ the test does roughly keep its asymptotic level, whereas it cannot be used at all when the variance becomes larger. Moreover, even if the variance is small but the sample size is increased, the test does not keep its pre-assigned level (see the results for $\omega = 0.001$ and $n = 1024$ in Table 2). Thus, in practice the application of testing procedures addressing the problem of microstructure noise is strictly recommended.

In the following section, we illustrate the finite sample properties of a bootstrap version of the Kolmogorov–Smirnov test based on the processes investigated in Sections 4 and 5. Since the stochastic order of $|\Delta_i^n Z|$ is basically determined by the maximum of $n^{-1/2}$ and ω (which are the orders of $|\Delta_i^n X|$ and $|\Delta_i^n U|$, respectively), we kept $n\omega^2 = 0.1024$ fixed in order to have comparable results for different sample sizes n. The regularisation parameters κ and ρ were set to be $1/2$ each. All simulation results presented in the following paragraphs are based on 1000 simulation runs and 500 bootstrap replications (if the bootstrap is applied to estimate critical values).

For all testing problems discussed below, we have not used exactly the statistics $\hat{N}_t$ and $\hat{M}_t$, but related versions accounting for finite sample adjustments. Following Jacod *et al.* [19], where

Table 3. Simulated nominal level of the test, which rejects the null hypothesis of homoscedasticity for a large value of $\sup|A_n(t)/\hat{s}_t|$, using the critical values from the asymptotic theory. The variance of the noise process is defined by $n\omega^2 = 0.1024$

n/α	0.025	0.05	0.1
256	0.008	0.022	0.058
1024	0.007	0.023	0.062
4096	0.013	0.029	0.079
16 384	0.017	0.038	0.077

it has been shown that finite sample corrections improve the behaviour of the estimate $\hat{B}_t^0$ (and presumably of $\hat{C}$ as well) substantially, we have replaced the quantities ψ_i and Φ_{ij} in (3.3) by certain numbers ψ_i^n and Φ_{ij}^n, which constitute the "true" quantities for finite samples, but are replaced by their limits ψ_i and Φ_{ij} in the asymptotics. See Jacod *et al.* [19] for details.

6.1. Testing for homoscedasticity

In the problem of testing for homoscedasticity the limiting process $A(t)_{t\in[0,1]}$ has an extremely simple form, when the null hypothesis of a constant volatility holds. In fact, the finite dimensional distributions of the process $(A(t))_{t\in[0,1]}$ coincide with those of a rescaled Brownian bridge, thus $(A_n(t)/\hat{s}_t)_{t\in[0,1]}$ converges weakly to $(B_t)_{t\in[0,1]}$. We have investigated the properties of the Kolmogorov–Smirnov test for different sample sizes n, where the noise satisfies $U \sim \mathcal{N}(0,\omega^2)$ and the drift function is again given by $b(t,x) = 0.1x$. A similar test can be constructed using Theorem 2, but the corresponding results are omitted for the sake of brevity as the rate of convergence in this case becomes worse.

In Table 3, we present the simulated level of the Kolmogorov–Smirnov test using the critical values from the asymptotic distribution. It can be seen that the asymptotic level of the test is slightly underestimated. This effect becomes less visible for a larger sample size, but even then it is still apparent. Note that these findings are in line with previous simulations on noisy observations and it is likely that they are due to the fact the rate of convergence for most testing problems is only $n^{-1/4}$.

6.2. Testing general hypotheses

For a general null hypothesis in (2.2), the distribution of the limiting process $(A(t))_{t\in[0,1]}$ depends on the path of the underlying semimartingale $(X_t)_{t\in[0,1]}$ and on the volatility $(\sigma_t)_{t\in[0,1]}$, and thus we cannot use it directly for the calculation of critical values. For this reason, we propose the application of the parametric bootstrap in order to obtain simulated critical values. First, we compute the global estimators $\hat{\omega}^2$ and $\hat{\theta} = \hat{D}^{-1}\hat{C}$ as well as each $n^{1/4}\hat{N}_t$ and $\hat{s}_t^2$ from the observed data. Under the null hypothesis N_t equals zero, and thus it is intuitively clear that the null

Table 4. Simulated level of the bootstrap test based on the standardised Kolmogorov–Smirnov functional of $(\hat{N}_t)$ for various hypotheses. The variance of the noise process is defined by $n\omega^2 = 0.1024$

$\sigma_1^2(t,x)$	1			x^2		
n/α	0.025	0.05	0.1	0.025	0.05	0.1
256	0.019	0.046	0.113	0.03	0.066	0.118
1024	0.02	0.049	0.099	0.034	0.07	0.119
4096	0.021	0.04	0.072	0.022	0.048	0.090

hypothesis has to be rejected for large values of the standardised Kolmogorov–Smirnov statistic $Y_n = \sup_{t\in[0,1]} |n^{1/4}\hat{N}_t/\hat{s}_t|$.

In a second step we generate bootstrap data $Z^{*(j)}_{1/n} = X^{*(j)}_{1/n} + U^{*(j)}_{1/n}$, where the $X^{*(j)}_{i/n}$ are realisations of the process in (2.1) with $b_s \equiv 0$ and $\sigma_s^2 = \sigma^2(s, X_s) = \sum_{k=1}^d \hat{\theta}_k \sigma_k^2(s, X_s)$ (corresponding to the null hypothesis) and each $U^{*(j)}_{i/n}$ is normally distributed with mean zero and variance $\hat{\omega}^2$. Using these data, we calculate the corresponding bootstrap statistics $Y_n^{*(j)}$ and use these to compute the quantiles of the bootstrap distribution. Finally, the null hypothesis is rejected if Y_n is larger than its $(1-\alpha)$-quantile.

In order to investigate the approximation of the nominal level we consider the hypothesis of constant volatility and the hypothesis H_0: $\sigma^2(t,x) = \theta x^2$. The data is generated under the null hypothesis with drift function $b(t,x) = 0.1x$ and the rejection probabilities are depicted in Table 4. These results show that the bootstrap approximation works well even for a small n. In particular, we see that in the case of homoscedasticity the exact asymptotic test using the weak convergence of Y_n to the supremum of a standard Brownian bridge is outperformed (compare with Table 3). In the case of testing, the parametric hypothesis H_0: $\sigma^2(t,x) = x^2$ we observe a slight overestimation of the nominal level by the bootstrap test.

As an example for testing the hypothesis $\bar{H}_0$, we have chosen $\sigma(t,x) = \theta|x|$ and investigated the properties of the analogues of Y_n and $Y_n^{*(j)}$ from above, where we have replaced $n^{1/4}\hat{N}_t$ and $\hat{s}_t$ by $n^{1/4-\delta/2}\hat{M}_t$ and $\hat{r}_t$, respectively. In this case, we chose $\delta = \frac{1}{4}$, corresponding to $l_n = \mathrm{O}(n^{-3/4})$ and a rate of convergence $n^{-1/8}$. Note that in this particular situation there is no need for stating the hypothesis in terms of $\bar{H}_0$ as it is equivalent to $\sigma^2(t,x) = \theta|x|^2$, but nevertheless it gives a reasonable impression on how well the bootstrap approximation works for testing hypotheses of the form $\bar{H}_0$.

We observe from the results in Table 5 that even though the rate of convergence in Theorem 2 is worse than in Theorem 1, there is no substantial difference in the approximation of the nominal level by the bootstrap test for both types of hypotheses: The nominal level is slightly overestimated, but in general the parametric bootstrap yields a satisfactory and reliable approximation of the nominal level.

Finally, Table 6 contains the rejection probabilities of the bootstrap test under the alternative. The null hypothesis is given by H_0: $\sigma^2(t,x) = \theta|x|^2$, and we discuss two local volatility alternatives, namely $\sigma^2(t,x) = 1$ and $\sigma^2(t,x) = 1 + |x|$, and one alternative coming from a stochastic

Table 5. Simulated level of the bootstrap test based on the standardised Kolmogorov–Smirnov functional of $(\hat{M}_t)$ for $\sigma(t,x) = \theta|x|$. The variance of the noise process is defined by $n\omega^2 = 0.1024$

n/α	0.025	0.05	0.1
256	0.040	0.076	0.136
1024	0.032	0.057	0.119

volatility model is considered. For this case, we chose the Heston model, that is,

$$X_t = X_0 + \int_0^t (\mu - \nu_s/2)\,\mathrm{d}s + \int_0^t \sigma_s\,\mathrm{d}W_t$$
$$\text{with } \nu_t = \nu_0 + \delta \int_0^t (\alpha - \nu_s)\,\mathrm{d}s + \gamma \int_0^1 \nu_s^{1/2}\,\mathrm{d}B_s,$$

where $\nu_t = \sigma_t^2$ and $\mathrm{Corr}(W, B) = \eta$ and the parameters were chosen as $\mu = 0.05/252, \delta = 5/252, \alpha = 0.04/252, \gamma = 0.05/252$ and $\rho = -0.5$.

We observe from the results depicted in Table 6 that the bootstrap test indicates in all cases that the null hypothesis is not satisfied. It is also remarkable that it is more difficult to detect the local volatility alternatives than the one coming from the Heston model. In the latter case, the rejection probabilities are extremely large even for a small sample size, contrary to the first two situations.

Appendix: Proof of Theorem 1

We will only prove the Theorem 1, as similar methods show Theorem 2 as well. We start with a typical localisation argument, which allows us to assume that several quantities are bounded. Recall first that a and σ are locally bounded by assumption, from which is follows that X is locally bounded as well. Thus we can conclude along the lines of Jacod [18] that we may assume without loss of generality that each of these processes is actually bounded. Since further each

Table 6. Simulated rejection probabilities of the bootstrap test based on the standardised Kolmogorov–Smirnov functional of $(\hat{N}_t)$ for various alternatives. The data is simulated with $\sigma^2(t,x) = \theta|x|^2$ and the variance of the noise process is defined by $n\omega^2 = 0.1024$

alt	1			$1+\|x\|$			Heston		
n/α	0.025	0.05	0.1	0.025	0.05	0.1	0.025	0.05	0.1
256	0.057	0.128	0.237	0.073	0.152	0.263	0.722	0.870	0.941
1024	0.170	0.230	0.329	0.224	0.326	0.465	0.975	0.980	0.985

σ_i^2 is continuous and because U has a compact support, we may conclude that both (s, X_t) and $(s, \hat{X}_{k/n})$ (for arbitrary s, t, k and n) are living on a compact set, and thus $\sigma_i^2(s, X_t)$ and $\sigma_i^2(s, \hat{X}_{k/n})$ are also bounded, the latter one uniformly in n. Similar results hold for the first two derivatives of σ_i^2 as well as for any of the functions $\bar{\sigma}_i$. Constants are denoted by K throughout this section.

The proof of Theorem 1 is based on several preliminary results, and we start with two results determining the rate of convergence of the quantities $\hat{B}_t^i - B_t^i$ and $\hat{D}_{ij} - D_{ij}$ defined in (2.5) and (2.4), respectively. The following result ensures that the (conditional) variance in a limit theorem for $\hat{N}_t - N_t$ will not depend on $\hat{B}_t^i$ and $\hat{D}_{ij}$, since the rate of convergence is $n^{-1/4}$. Thus, we will focus in the following on the behavior of $\hat{C}_i$ and $\hat{B}_t^0$.

Theorem 3. *Under the assumptions from Section 3 we have*

$$\begin{aligned} \hat{B}_t^i - B_t^i &= \mathrm{o}_p(n^{-1/4}) \qquad \text{for } i = 1, \dots, d, \\ \hat{D}_{ij} - D_{ij} &= \mathrm{o}_p(n^{-1/4}) \qquad \text{for } i, j = 1, \dots, d, \end{aligned} \tag{A.1}$$

where the first result holds uniformly with respect to $t \in [0, 1]$.

Proof. For a proof of the first estimate, we use for a fixed index i the decomposition

$$\begin{aligned} \hat{B}_t^i - B_t^i &= \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor - m_n} \left(\sigma_i^2\left(\frac{k}{n}, \hat{X}_{k/n}\right) - \sigma_i^2\left(\frac{k}{n}, X_{k/n}\right) \right) \\ &\quad + \left(\frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor - m_n} \sigma_i^2\left(\frac{k}{n}, X_{k/n}\right) - \int_0^t \sigma_i^2(s, X_s)\, \mathrm{d}s \right). \end{aligned}$$

Regarding the first term in this sum, note that

$$\hat{X}_{k/n} - X_{k/n} = \frac{1}{m_n} \sum_{j=1}^{m_n} \left(U_{(k+j)/n} + \int_{k/n}^{(k+j)/n} \sigma_u \, \mathrm{d}W_u \right) + \mathrm{O}_p(n^{-1/2})$$

and thus $\hat{X}_{k/n} - X_{k/n} = \mathrm{O}_p(n^{-1/4})$. A Taylor expansion and boundedness of the second derivative of the function σ^2 give

$$\frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor - m_n} \left(\sigma_i^2\left(\frac{k}{n}, \hat{X}_{k/n}\right) - \sigma_i^2\left(\frac{k}{n}, X_{k/n}\right) \right) = \frac{1}{n} \sum_{k=1}^{\lfloor nt \rfloor - m_n} A_{k,n} + \mathrm{O}_p(n^{-1/2})$$

with

$$A_{k,n} = \frac{1}{m_n} \sum_{j=1}^{m_n} \frac{\partial}{\partial y} \sigma_i^2\left(\frac{k}{n}, X_{k/n}\right) \left(U_{(k+j)/n} + \int_{k/n}^{(k+j)/n} \sigma_s \, \mathrm{d}W_s \right).$$

However, we have $E[A_{k,n}A_{l,n}] = \mathrm{O}(n^{-1/2})$ for arbitrary k and l as well as $E[A_{k,n}A_{k+l,n}] = 0$ for $l \geq m_n$ by conditioning on $\mathcal{F}_{(k+l)/n}$. This yields

$$E\left[\left(\frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor - m_n} A_{k,n}\right)^2\right] = \frac{1}{n^2}\sum_{k=m_n}^{\lfloor nt\rfloor - 2m_n}\sum_{l=-m_n}^{m_n} E[A_{k,n}A_{k+l,n}] + \mathrm{O}\left(\frac{m_n}{n^2}\right) = \mathrm{O}\left(\frac{1}{n}\right),$$

which is small enough. For the second term in the decomposition of $\hat{B}_t^i - B_t^i$ it holds that

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{\lfloor nt\rfloor - m_n} \sigma_i^2\left(\frac{k}{n}, X_{k/n}\right) - \int_0^t \sigma_i^2(s, X_s)\,\mathrm{d}s \\
&\quad = \sum_{k=1}^{\lfloor nt\rfloor} \int_{(k-1)/n}^{k/n} \left(\sigma_i^2\left(\frac{k-1}{n}, X_{(k-1)/n}\right) - \sigma_i^2\bigl(s, X_{(k-1)/n}\bigr)\right. \\
&\qquad\qquad \left. + \sigma_i^2\bigl(s, X_{(k-1)/n}\bigr) - \sigma_i^2(s, X_s)\right)\mathrm{d}s + \mathrm{O}_p(n^{-1/2}).
\end{aligned}$$

By differentiability in both components and from a similar expansion as above the claim follows. The result on $\hat{D}_{ij} - D_{ij}$ can be shown in the same way. □

The following result specifies the convergence of the finite dimensional distributions of the processes, which are used for the construction of $\{\hat{N}_t\}_{t\in[0,1]}$. Below we use the notation $G_n \xrightarrow{\mathcal{D}_{st}} G$ to indicate stable convergence of a sequence of random variables (G_n) to a limiting variable G, which is defined on an appropriate extension of the original probability space. For details on stable convergence see Jacod and Shiryaev [20].

Theorem 4. *Define for any fixed $t_1, \ldots, t_k \in [0, 1]$ the matrix $\Sigma_{t_1,\ldots,t_k}(s, X_s) = \gamma_s^2 \ell(s, X_s)\ell^T(s, X_s)$ where $\ell(s, X_s) = (1_{[0,t_1]}(s), \ldots, 1_{[0,t_k]}(s), h^T(s, X_s))^T$. Then we have*

$$n^{1/4}(\hat{B}_{t_1}^0 - B_{t_1}^0, \ldots, \hat{B}_{t_k}^0 - B_{t_k}^0, \hat{C}_1 - C_1, \ldots, \hat{C}_d - C_d)^T \xrightarrow{\mathcal{D}_{st}} \int_0^1 \Sigma_{t_1,\ldots,t_k}^{1/2}(s, X_s)\,\mathrm{d}W_s',$$

where W' is another Brownian motion, which is independent of the σ-algebra $\mathcal{F}$.

Proof. Since $\omega^2 - \hat{\omega}_n^2 = \mathrm{O}_p(n^{-1/2})$, one obtains

$$\begin{aligned}
\hat{C}_i &= \frac{1}{n}\sum_{k=1}^{n-m_n} \sigma_i^2\left(\frac{k}{n}, X_{k/n}\right)\hat{\sigma}_{k/n}^2 + \frac{1}{n}\sum_{k=1}^{n-m_n}\left(\sigma_i^2\left(\frac{k}{n}, \hat{X}_{k/n}\right) - \sigma_i^2\left(\frac{k}{n}, X_{k/n}\right)\right)\hat{\sigma}_{k/n}^2 \\
&\quad + \mathrm{O}_p(n^{-1/2}).
\end{aligned}$$

From similar arguments as given in the proof of Theorem 3 we find that the second term is of order $\mathrm{o}_p(n^{-1/4})$ and thus asymptotically negligible as well. Therefore, we are left to focus on

$F_{in} = \frac{1}{n}\sum_{k=1}^{n-m_n} \sigma_i^2(\frac{k}{n}, X_{k/n})\hat{\sigma}^2_{k/n}$. Due to the dependence structure of the summands in F_{in} it will be convenient to use a "small-blocks–big-blocks"-technique as in Jacod *et al.* [19] in order to prove Theorem 4. To this end, we choose an integer p, which eventually goes to infinity, and partition the n observations into several subsets: We define $b_k(p) = k(p+1)m_n$ and $c_k(p) = k(p+1)m_n + pm_n$ and denote by $j_n(p)$ the largest integer k such that $c_k(p) \le n - m_n$ holds. Moreover, we use the notation $i_n(p) = (j_n(p)+1)pm_n$, and introduce for each $0 \le k \le j_n(p)$ and any p the following random variables:

$$G(k,p)_1^n = \frac{1}{n}\sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right)\sum_{j=b_k(p)}^{c_k(p)-1}\hat{\sigma}^2_{k/n},$$

$$G(k,p)_2^n = \frac{1}{n}\sigma_i^2\left(\frac{c_k(p)}{n}, X_{c_k(p)/n}\right)\sum_{j=c_k(p)}^{b_{k+1}(p)-1}\hat{\sigma}^2_{k/n}.$$

The remainder terms from $i_n(p)$ to $n - m_n$ are gathered in some $G(p)_3^n$. Note that each of these quantities depends on i, although it does not appear in the notation.

The main intuition behind these quantities is that the terms $G(k,p)_1^n$ are defined on non-overlapping intervals, which means that the intervals on which each $\overline{Z}_j^n$ within $G(k,p)_1^n$ lives are disjoint from those of any $\overline{Z}_j^n$ within any other $G(l,p)_1^n$. This is sufficient to ensure some type of conditional independence, which will be used in order to prove Theorem 4. The variables $G(k,p)_2^n$ and $G(p)_3^n$ are filling the gaps between $G(k,p)_1^n$ and $G(l,p)_1^n$ and can be shown to be asymptotically negligible.

An important tool will be the following decomposition of $|\overline{Z}_j^n|^2$. We set

$$V_s^j = \int_{j/n}^{j/n+s} g_n\left(u - \frac{j}{n}\right)a_u \,\mathrm{d}u + \int_{j/n}^{j/n+s} g_n\left(u - \frac{j}{n}\right)\sigma_u \,\mathrm{d}W_u,$$

and obtain by an application of Ito's formula

$$\begin{aligned}
|\overline{Z}_j^n|^2 &= |\overline{X}_j^n|^2 + |\overline{U}_j^n|^2 + 2\overline{X}_j^n\overline{U}_j^n \\
&= 2\int_{j/n}^{(j+m_n)/n} V_s^j g_n\left(s - \frac{j}{n}\right)a_s \,\mathrm{d}s + 2\int_{j/n}^{(j+m_n)/n} V_s^j g_n\left(s - \frac{j}{n}\right)\sigma_s \,\mathrm{d}W_s \\
&\quad + \int_{j/n}^{(j+m_n)/n} g_n^2\left(s - \frac{j}{n}\right)\sigma_s^2 \,\mathrm{d}s + |\overline{U}_j^n|^2 + 2\overline{U}_j^n \int_{j/n}^{(j+m_n)/n} g_n\left(s - \frac{j}{n}\right)a_s \,\mathrm{d}s \\
&\quad + 2\overline{U}_j^n \int_{j/n}^{(j+m_n)/n} g_n\left(s - \frac{j}{n}\right)\sigma_s \,\mathrm{d}W_s \\
&= \sum_{l=1}^{6} D(j)_l^n,
\end{aligned}$$

where the last identity defines the quantities $D(j)_l^n$ in an obvious manner.

For $b_k(p) \le j < c_k(p)$ we introduce approximations for the quantities $D(j)_2^n$ and $D(j)_6^n$, namely

$$\tilde{D}(k,j,p)_2^n = 2\sigma_{b_k(p)/n}^2 \int_{j/n}^{(j+m_n)/n} \left(\int_{j/n}^{j/n+s} g_n\left(u - \frac{j}{n}\right) \mathrm{d}W_u \right) g_n\left(s - \frac{j}{n}\right) \mathrm{d}W_s,$$

$$\tilde{D}(k,j,p)_6^n = 2\sigma_{b_k(p)/n} \overline{U}_j^n \int_{j/n}^{(j+m_n)/n} g_n\left(s - \frac{j}{n}\right) \mathrm{d}W_s.$$

Additionally, we set $H(k,p)^n = \sigma_i^2(\frac{b_k(p)}{n}, X_{b_k(p)/n}) Y(k,p)^n$, where

$$Y(k,p)^n = \frac{1}{\kappa\psi_2} n^{-1/2} \sum_{j=b_k(p)}^{c_k(p)-1} \left\{ \tilde{D}(k,j,p)_2^n + \tilde{D}(k,j,p)_6^n + \left(D(j)_4^n - n^{-1/2} \frac{\psi_1}{\kappa} \omega^2 \right) \right\}. \quad \text{(A.2)}$$

Finally, we define

$$\chi(p)_k^n = E\left[\left(\sup_{s,t \in [b_k(p)/n, c_k(p)/n]} |a_s - a_t| + |\sigma_s - \sigma_t| \right)^2 \Big| \mathcal{F}_{b_k(p)/n} \right]^{1/2}.$$

The main part of the proof of Theorem 1 are two auxiliary results which specify the asymptotic properties of F_{in}.

Lemma 1. *We have*

$$\lim_{p\to\infty} \limsup_{n\to\infty} n^{1/4} \left\{ \left(\sum_{k=0}^{j_n(p)} \left(G(k,p)_1^n + G(k,p)_2^n \right) + G(p)_3^n - C_i \right) - \sum_{k=0}^{j_n(p)} H(k,p) \right\} = 0.$$

Proof. The proof goes through a rather large number of steps and makes extensive use of the decomposition in (A.2). We will show first that the influence of the random variables $D(j)_1^n$ and $D(j)_5^n$ within $G(k,p)_1^n$ (and analogously for $G(k,p)_2^n$ and $G(p)_3^n$) is asymptotically negligible, that is

$$\lim_{p\to\infty} \limsup_{n\to\infty} n^{-1/4} \sum_{k=0}^{j_n(p)} \sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n} \right) \sum_{j=b_k(p)}^{c_k(p)-1} \left(D(j)_1^n + D(j)_5^n \right) = 0. \quad \text{(A.3)}$$

For a proof of (A.3), assume without loss of generality that $b_k(p) \le j < c_k(p)$. One obtains

$$\begin{aligned} D(j)_1^n &= 2a_{b_k(p)/n} \int_{j/n}^{(j+m_n)/n} \left(\int_{j/n}^{j/n+s} g_n\left(u - \frac{j}{n}\right) \sigma_u \,\mathrm{d}W_u \right) g_n\left(s - \frac{j}{n}\right) \mathrm{d}s \\ &\quad + 2 \int_{j/n}^{(j+m_n)/n} \left(\int_{j/n}^{j/n+s} g_n\left(u - \frac{j}{n}\right) \sigma_u \,\mathrm{d}W_u \right) g_n\left(s - \frac{j}{n}\right) \left(a_s - a_{b_k(p)/n} \right) \mathrm{d}s \\ &\quad + \mathrm{O}_p\left(\frac{1}{n}\right) \end{aligned}$$

and from the martingale property of a stochastic integral with respect to Brownian motion and the Cauchy–Schwarz inequality we derive that $|E[D(j)_1^n|\mathcal{F}_{b_k(p)/n}]| \le Kn^{-3/4}\chi(p)_k^n$. Thus, with the notation $\delta(k,p)_1^n = \sum_{j=b_k(p)}^{c_k(p)-1} D(j)_1^n$ we conclude

$$\big|E\big[\delta(k,p)_1^n|\mathcal{F}_{b_k(p)/n}\big]\big| \le Kpn^{-1/4}\chi(p)_k^n \quad \text{and} \quad E\big[(\delta(k,p)_1^n)^2|\mathcal{F}_{b_k(p)/n}\big] \le Kp^2n^{-1/2},$$

and for $k>l$ it follows

$$\begin{aligned}&\left|E\left\{\sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right)\sigma_i^2\left(\frac{b_l(p)}{n}, X_{b_l(p)/n}\right)\delta(l,p)_1^n E\big[\delta(k,p)_1^n|\mathcal{F}_{b_k(p)/n}\big]\right\}\right| \\ &\quad\le Kp^2n^{-1/2}E[\chi(p)_k^n].\end{aligned}$$

Since $j_n(p)$ is of order $n^{1/2}/p$, we obtain

$$\begin{aligned}&E\left[\left(n^{-1/4}\sum_{k=0}^{j_n(p)}\sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right)\sum_{j=b_k(p)}^{c_k(p)-1} D(j)_1^n\right)^2\right] \\ &\quad\le K\left(pn^{-1/2} + \sum_{k>l}^{j_n(p)} p^2n^{-1}E[\chi(p)_k^n]\right).\end{aligned}$$

From Lemma 5.4. in Jacod *et al.* [19] it follows that $\lim_{n\to\infty} n^{-1/2}\sum_{k=1}^{j_n(p)} E[\chi(p)_k^n] = 0$ for any p, which gives that the first term in the sum (A.3) converges to 0. The second term in (A.3) converges to zero from the independence of X and U and a standard martingale argument.

The next step is devoted to the analysis of the term $D(j)_2^n$. We prove

$$\lim_{p\to\infty}\limsup_{n\to\infty} n^{-1/4}\sum_{k=0}^{j_n(p)}\sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right)\sum_{j=b_k(p)}^{c_k(p)-1}\big(D(j)_2^n - \tilde{D}(k,j,p)_2^n\big) = 0 \tag{A.4}$$

as well as

$$\lim_{p\to\infty}\limsup_{n\to\infty} n^{-1/4}\sum_{k=0}^{j_n(p)}\sigma_i^2\left(\frac{c_k(p)}{n}, X_{c_k(p)/n}\right)\sum_{j=c_k(p)}^{b_{k+1}(p)-1} D(j)_2^n = 0, \tag{A.5}$$

$$\lim_{p\to\infty}\limsup_{n\to\infty} n^{-1/4}\sigma_i^2\left(\frac{i_n(p)}{n}, X_{i_n(p)/n}\right)\sum_{j=i_n(p)}^{n-m_n} D(j)_2^n = 0. \tag{A.6}$$

Set $b_k(p) \le j < c_k(p)$ again. A martingale argument as before allows us to focus on

$$D''(j)_2^n = 2\int_{j/n}^{(j+m_n)/n}\left(\int_{j/n}^{j/n+s} g_n\left(u-\frac{j}{n}\right)\sigma_u\,\mathrm{d}W_u\right)g_n\left(s-\frac{j}{n}\right)\sigma_s\,\mathrm{d}W_s$$

only. We have $E[D''(j)_2^n|\mathcal{F}_{b_k(p)/n}]=0$ and $E[|D''(j)_2^n D''(l)_2^n||\mathcal{F}_{b_k(p)/n}]\le Kn^{-1}$, thus (A.6) follows easily. For (A.5), note that $E[(\sum_{j=c_k(p)}^{b_{k+1}(p)-1} D''(j)_2^n)^2]\le K$, which gives (recall the definition of $j_n(p)$, $b_k(p)$ and $c_k(p)$)

$$\begin{aligned} &n^{-1/2}\sum_{k=0}^{j_n(p)} E\left[\sigma_i^4\left(\frac{c_k(p)}{n}, X_{c_k(p)/n}\right)\left(\sum_{j=c_k(p)}^{b_{k+1}(p)-1} D''(j)_2^n\right)^2\right] \\ &\quad\le Kn^{-1/2}\frac{n^{1/2}}{p}=K\frac{1}{p}, \end{aligned}$$

converging to zero as p tends to infinity. We are thus left to prove

$$\lim_{p\to\infty}\limsup_{n\to\infty} n^{-1/4}\sum_{k=0}^{j_n(p)}\sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right)\sum_{j=b_k(p)}^{c_k(p)-1}\big(D''(j)_2^n-\tilde{D}(k,j,p)_2^n\big)=0.$$

This time, we have $E[D''(j)_2^n-\tilde{D}(k,j,p)_2^n|\mathcal{F}_{b_k(p)/n}]=0$ and

$$E\big[\big|\big(D''(j)_2^n-\tilde{D}(k,j,p)_2^n\big)\big(D''(l)_2^n-\tilde{D}(k,l,p)_2^n\big)\big||\mathcal{F}_{b_k(p)/n}\big]\le Kn^{-1}(\chi(p)_k^n)^2.$$

Thus,

$$\begin{aligned} &E\left[\left\{n^{-1/4}\sum_{k=0}^{j_n(p)}\sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right)\sum_{j=b_k(p)}^{c_k(p)-1}\big(D(j)_2^n-\tilde{D}(k,j,p)_2^n\big)\right\}^2\right] \\ &\quad\le Kp^2 n^{-1/2}\sum_{k=0}^{j_n(p)} E[(\chi(p)_k^n)^2], \end{aligned}$$

and with a similar argument as in the proof of (A.3) we are done. Proving that $D(j)_6^n$ can be replaced by $\tilde{D}(k,j,p)_6^n$ works analogously, thus we finish the proof of Lemma 1 showing

$$\begin{aligned} \lim_{p\to\infty}\limsup_{n\to\infty} n^{1/4}\Bigg\{\frac{1}{\kappa\psi_2}n^{-1/2}\Bigg(&\sum_{k=0}^{j_n(p)}\Bigg(\sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right)\sum_{j=b_k(p)}^{c_k(p)-1} D(j)_3^n \\ &+\sigma_i^2\left(\frac{c_k(p)}{n}, X_{c_k(p)/n}\right)\sum_{j=c_k(p)}^{b_{k+1}(p)-1} D(j)_3^n\Bigg) \qquad \text{(A.7)} \\ &+\sum_{j=i_n(p)}^{n-m_n} D(j)_3^n\Bigg)-C_i\Bigg\}=0. \end{aligned}$$

We start with the following proposition:

$$\lim_{p\to\infty}\limsup_{n\to\infty} n^{1/4}\Biggl\{\Biggl(\sum_{k=0}^{j_n(p)}\Biggl(\int_{b_k(p)/n}^{c_k(p)/n}\sigma_i^2\Bigl(\frac{b_k(p)}{n}, X_{b_k(p)/n}\Bigr)\sigma_s^2\,\mathrm{d}s + \int_{c_k(p)/n}^{b_{k+1}(p)/n}\sigma_i^2\Bigl(\frac{c_k(p)}{n}, X_{c_k(p)/n}\Bigr)\sigma_s^2\,\mathrm{d}s\Biggr) + \int_{\frac{i_n(p)}{n}}^{1}\sigma_i^2\Bigl(\frac{i_n(p)}{n}, X_{i_n(p)/n}\Bigr)\sigma_s^2\,\mathrm{d}s\Biggr) - C_i\Biggr\} = 0. \tag{A.8}$$

As in the proof of Theorem 3, we have

$$\begin{aligned}&\sigma_i^2(s, X_s) - \sigma_i^2\Bigl(\frac{b_k(p)}{n}, X_{b_k(p)/n}\Bigr)\\ &\quad = \frac{\partial}{\partial y}\sigma_i^2\bigl(s, X_{b_k(p)/n}\bigr)\Bigl(\int_{b_k(p)/n}^{s}\sigma_u\,\mathrm{d}W_u\Bigr) + \mathrm{O}_p\Bigl(\frac{pm_n}{n}\Bigr),\end{aligned} \tag{A.9}$$

thus

$$\begin{aligned}&\int_{b_k(p)/n}^{c_k(p)/n}\Bigl(\sigma_i^2(s, X_s) - \sigma_i^2\Bigl(\frac{b_k(p)}{n}, X_{b_k(p)/n}\Bigr)\Bigr)\sigma_s^2\,\mathrm{d}s\\ &\quad = \delta'(k,p)_3^n + \delta''(k,p)_3^n + \mathrm{O}_p\Bigl(\frac{p^2m_n^2}{n^2}\Bigr),\end{aligned} \tag{A.10}$$

where

$$\delta'(k,p)_3^n = \sigma_{b_k(p)/n}^3\int_{b_k(p)/n}^{c_k(p)/n}\frac{\partial}{\partial y}\sigma_i^2\bigl(s, X_{b_k(p)/n}\bigr)\Bigl(\int_{b_k(p)/n}^{s}\mathrm{d}W_u\Bigr)\,\mathrm{d}s$$

and $\delta''(k,p)_3^n$ is defined implicitly by equation (A.10). We obtain

$$\lim_{p\to\infty}\limsup_{n\to\infty} n^{1/2}E\Biggl[\Biggl(\sum_{k=0}^{j_n(p)}\delta'(k,p)_3^n\Biggr)^2\Biggr] = 0$$

from the usual martingale argument and also

$$\lim_{p\to\infty}\limsup_{n\to\infty} n^{1/4}\sum_{k=0}^{j_n(p)}E\bigl[|\delta''(k,p)_3^n|\bigr] \le \lim_{p\to\infty}\limsup_{n\to\infty} Kp^{3/2}n^{-1/2}\sum_{k=0}^{j_n(p)}E\bigl[\chi(p)_k^n\bigr] = 0$$

as before. The corresponding results for the other summands in (A.8) can be shown analogously.

To finish the proof of Lemma 1, we have to show

$$\lim_{p\to\infty}\limsup_{n\to\infty} n^{1/4}\Bigg\{\sum_{k=0}^{j_n(p)}\Bigg(\sigma_i^2\bigg(\frac{b_k(p)}{n}, X_{b_k(p)/n}\bigg)\bigg(\frac{1}{\kappa\psi_2}n^{-1/2}\sum_{j=b_k(p)}^{c_k(p)-1} D(j)_3^n - \int_{b_k(p)/n}^{c_k(p)/n}\sigma_s^2\,\mathrm{d}s\bigg)$$
$$+\sigma_i^2\bigg(\frac{c_k(p)}{n}, X_{c_k(p)/n}\bigg)$$
$$\times\bigg(\frac{1}{\kappa\psi_2}n^{-1/2}\sum_{j=c_k(p)}^{b_{k+1}(p)-1} D(j)_3^n - \int_{c_k(p)/n}^{b_{k+1}(p)/n}\sigma_s^2\,\mathrm{d}s\bigg)\Bigg)$$
$$+\sigma_i^2\bigg(\frac{i_n(p)}{n}, X_{i_n(p)/n}\bigg)$$
$$\times\bigg(\frac{1}{\kappa\psi_2}n^{-1/2}\sum_{j=i_n(p)}^{n-m_n} D(j)_3^n - \int_{i_n(p)/n}^{1}\sigma_s^2\,\mathrm{d}s\bigg)\Bigg\} = 0,$$

The last term is negligible, and the main idea for the tedious proof of the remaining terms is to fix k for a moment and to prove a representation of the form

$$\frac{1}{\kappa\psi_2}n^{-1/2}\sum_{j=b_k(p)}^{c_k(p)-1} D(j)_3^n = \int_{b_k(p)/n}^{b_{k+1}(p)/n} h_{n,p}\bigg(s-\frac{b_k(p)}{n}\bigg)\sigma_s^2\,\mathrm{d}s \tag{A.11}$$

for a suitable function $h_{n,p}(s)$, using the definition of $D(j)_3^n$. A similar expression can be found for the sum from $c_k(p)$ to $b_{k+1}(p)$ with some $\bar{h}_{n,p}(s)$. A careful computation shows that $h_{n,p}(s)$ is either close to one (for s in the center of the corresponding interval) or that $h_{n,p}(s)$ and $\bar{h}_{n,p}(s)$ sum up to one (on its boundary). Then a Taylor expansion as in the proof of (A.8) gives the result. □

Lemma 2. *We have*

$$\lim_{p\to\infty}\limsup_{n\to\infty} n^{1/4}\Bigg\{F_{in} - \Bigg(\sum_{k=0}^{j_n(p)}\big(G(k,p)_1^n + G(k,p)_2^n\big) + G(p)_3^n\Bigg)\Bigg\} = 0.$$

Proof. Without loss of generality, is suffices to show

$$\lim_{p\to\infty}\limsup_{n\to\infty} n^{-1/4}\sum_{k=0}^{j_n(p)}\sum_{j=b_k(p)}^{c_k(p)-1}\bigg(\sigma_i^2(s,X_s) - \sigma_i^2\bigg(\frac{b_k(p)}{n}, X_{b_k(p)/n}\bigg)\bigg)\bigg(|\overline{Z}_j^n|^2 - n^{-1/2}\frac{\psi_1}{\kappa}\omega^2\bigg)$$
$$= 0.$$

The proof of this claim is tedious again. Essentially one simplifies the expression above by the Taylor expansion from (A.9) and a similar decomposition as in (A.2) for $|\overline{Z}_j^n|^2$ and discusses each term separately. □

Note that we have completely analogous results for a decomposition of $\hat{B}_t^0 - B_t^0$. Thus, we end up with

$$\lim_{p\to\infty} \limsup_{n\to\infty} n^{1/4} \left\{ (\hat{B}_t^0 - B_t^0) - \sum_{k=0}^{j_n(p)} Y(k,p) 1_{\{c_k(p)/n \le t\}} \right\} = 0, \tag{A.12}$$

$$\lim_{p\to\infty} \limsup_{n\to\infty} n^{1/4} \left\{ (\hat{C}_i - C_i) - \sum_{k=0}^{j_n(p)} \sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right) Y(k,p) \right\} = 0,$$

where $Y(k,p)$ was defined in (A.2). Since

$$nE\bigl[(Y(k,p))^2 | \mathcal{F}_{b_k(p)/n}\bigr] = p\kappa\gamma^2_{b_k(p)/n} + \mathrm{o}_p(1) \quad \text{and} \quad E\bigl[Y(k,p)|\mathcal{F}_{b_k(p)/n}\bigr] = 0$$

as in Jacod *et al.* [19], we conclude

$$\lim_{p\to\infty}\lim_{n\to\infty} n^{1/2} \sum_{k=0}^{j_n(p)} E\bigl[Y(k,p)^2 1_{\{c_k(p)/n \le t_i \wedge t_j\}} | \mathcal{F}_{b_k(p)/n}\bigr]$$
$$= \int_0^1 \gamma_s^2 1_{[0,t_i\wedge t_j]}(s)\,\mathrm{d}s,$$

$$\lim_{p\to\infty}\lim_{n\to\infty} n^{1/2} \sum_{k=0}^{j_n(p)} E\left[Y(k,p)^2 1_{\{c_k(p)/n \le t_i\}} \sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right) | \mathcal{F}_{b_k(p)/n}\right]$$
$$= \int_0^1 \gamma_s^2 1_{[0,t_i]}(s) \sigma_j^2(s, X_s)\,\mathrm{d}s,$$

$$\lim_{p\to\infty}\lim_{n\to\infty} n^{1/2} \sum_{k=0}^{j_n(p)} E\left[Y(k,p)^2 \sigma_i^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right) \sigma_j^2\left(\frac{b_k(p)}{n}, X_{b_k(p)/n}\right) | \mathcal{F}_{b_k(p)/n}\right]$$
$$= \int_0^1 \gamma_s^2 \sigma_i^2(s, X_s)\sigma_j^2(s, X_s)\,\mathrm{d}s.$$

Theorem 4 follows now from Theorem IX 7.28 in Jacod and Shiryaev [20], since the missing conditions can be shown in the same way as in Jacod *et al.* [19]. □

The convergence of the finite dimensional distributions follows from the delta method for stably converging sequences, since we have

$$n^{1/4}(\hat{N}_{t_1} - N_{t_1}, \ldots, \hat{N}_{t_k} - N_{t_k})^T \xrightarrow{\mathcal{D}_{st}} Y \int_0^1 \Sigma^{1/2}_{t_1,\ldots,t_k}(s, X_s)\,\mathrm{d}W_s,$$

where the $k \times (d+k)$-dimensional matrix Y has the form

$$Y = (I_{k\times k} \quad -Y^*), \qquad Y^* = (B_{t_1}^T D^{-1} \quad \cdots \quad B_{t_k}^T D^{-1})^T.$$

A straightforward calculation shows that the conditional covariance coincides with the one of the finite dimensional distributions of the process defined in (4.2). We are left to prove the tightness of the process $n^{1/4}(\hat{N}_t - N_t)$, and this can be done by an application of Theorem VI.4.5 in Jacod and Shiryaev [20], using the boundedness of the processes involved as well as $E[|\det(D)|^{-\beta}] < \infty$. □

Acknowledgements

The authors would like to thank Martina Stein, who typed parts of this manuscript with considerable technical expertise. This work has been supported in part by the Collaborative Research Center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG). The authors are also grateful to the referees and the associate editor for constructive comments on an earlier version of this paper.

References

[1] Achieser, N.J. (1956). *Theory of Approximation*. New York: Dover Publications Inc.
[2] Ahn, D. and Gao, B. (1999). A parametric nonlinear model of term structure dynamics. *Review of Financial Studies* **12** 721–762.
[3] Ait-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies* **9** 385–426.
[4] Amihud, Y. and Mendelson, H. (1987). Trading mechanisms and stock returns: An empirical investigation. *J. Finance* **42** 533–553.
[5] Black, F. (1986). Noise. *J. Finance* **41** 529–543.
[6] Black, F. and Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* **81** 637–659.
[7] Chan, K.C., Karolyi, G.A., Longstaff, F.A. and Sanders, A.B. (1992). An empirical comparison of alternative models of the short-term interest rate. *J. Finance* **47** 1209–1227.
[8] Corradi, V. and White, H. (1999). Specification tests for the variance of a diffusion. *J. Time Ser. Anal.* **20** 253–270. MR1693173
[9] Cox, J.C., Ingersoll, J.E. Jr. and Ross, S.A. (1985). A theory of the term structure of interest rates. *Econometrica* **53** 385–407. MR0785475
[10] Dette, H. and Podolskij, M. (2008). Testing the parametric form of the volatility in continuous time diffusion models—a stochastic process approach. *J. Econometrics* **143** 56–73. MR2384433
[11] Dette, H., Podolskij, M. and Vetter, M. (2006). Estimation of integrated volatility in continuous-time financial models with applications to goodness-of-fit testing. *Scand. J. Statist.* **33** 259–278. MR2279642
[12] Gallant, A.R. (1987). *Nonlinear Statistical Models*. *Wiley Series in Probability and Mathematical Statistics*: *Applied Probability and Statistics*. New York: Wiley. MR0921029
[13] Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Statist.* **5** 243–260 (electronic). MR1875673
[14] Harris, L. (1990). Estimation of stock variance and serial covariance from discrete observations. *Journal of Financial and Quantitative Analysis* **25** 291–306.
[15] Harris, L. (1991). Stock price clustering and discreteness. *Review of Financial Studies* **4** 389–415.

[16] Heston, S.L. (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies* **6** 327–343.

[17] Hull, J. and White, A. (1987). The Pricing of Options on Assets with Stochastic Volatilities. *J. Finance* **42** 281–300.

[18] Jacod, J. (2008). Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Process. Appl.* **118** 517–559. MR2394762

[19] Jacod, J., Li, Y., Mykland, P.A., Podolskij, M. and Vetter, M. (2009). Microstructure noise in the continuous case: The pre-averaging approach. *Stochastic Process. Appl.* **119** 2249–2276. MR2531091

[20] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems for Stochastic Processes*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **288**. Berlin: Springer. MR1943877

[21] Podolskij, M. and Vetter, M. (2009). Bipower-type estimation in a noisy diffusion setting. *Stochastic Process. Appl.* **119** 2803–2831. MR2554029

[22] Podolskij, M. and Vetter, M. (2009). Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* **15** 634–658. MR2555193

[23] Revuz, D. and Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **293**. Berlin: Springer. MR1725357

[24] Seber, G.A.F. and Wild, C.J. (1989). *Nonlinear Regression*. *Wiley Series in Probability and Mathematical Statistics*: *Probability and Mathematical Statistics*. New York: Wiley. MR0986070

[25] Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics* **5** 177–188.

[26] Zhang, L., Mykland, P.A. and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.* **100** 1394–1411. MR2236450

Received November 2009 and revised January 2011