Bernoulli **21**(4), 2015, 2393–2418
DOI: 10.3150/14-BEJ648

Estimation of integrated volatility of volatility with applications to goodness-of-fit testing

MATHIAS VETTER

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

In this paper, we are concerned with nonparametric inference on the volatility of volatility process in stochastic volatility models. We construct several estimators for its integrated version in a high-frequency setting, all based on increments of spot volatility estimators. Some of those are positive by construction, others are bias corrected in order to attain the optimal rate $n^{-1/4}$. Associated central limit theorems are proven which can be widely used in practice, as they are the key to essentially all tools in model validation for stochastic volatility models. As an illustration we give a brief idea on a goodness-of-fit test in order to check for a certain parametric form of volatility of volatility.

Keywords: central limit theorem; goodness-of-fit testing; high-frequency observations; model validation; stable convergence; stochastic volatility model

1. Introduction

Nowadays, stochastic volatility models are standard tools in the continuous-time modelling of financial time series. Typically, the underlying (log) price process is assumed to follow a diffusion process of the form

$$X_t = X_0 + \int_0^t \mu_s \, \mathrm{d}s + \int_0^t \sigma_s \, \mathrm{d}W_s, \tag{1.1}$$

where μ and σ can be quite general stochastic processes themselves. A classical case is where the volatility $\sigma_s^2 = \sigma^2(s, X_s)$ is a function of time and state – a situation referred to as the one of a local volatility model. It has turned out in empirical finance that such models do not fit the data very well, as some stylised facts such as the leverage effect or volatility clustering cannot be explained using local volatility only. Stochastic volatility models, however, are able to reproduce such features, as they bear an additional source of randomness. In these models, the volatility process is a diffusion process itself, and we focus on a rather general situation, namely

$$\sigma_t^2 = \sigma_0^2 + \int_0^t \nu_s \, \mathrm{d}s + \int_0^t \beta_s \, \mathrm{d}W_s + \int_0^t \eta_s \, \mathrm{d}W_s', \tag{1.2}$$

where ν, β and η again are suitable stochastic processes and W' is another Brownian motion, independent of W. This model obviously includes the widely used special case of a volatility

1350-7265

with only one driving Brownian motion, which is $d\sigma_t^2 = \nu_t\,dt + \tau_t\,dV_t$, where V and W are jointly Brownian with some correlation ρ.

Stochastic volatility models are typically parametric ones, and probably the prime example among those is the Heston model of [14], given by

$$X_t = X_0 + \int_0^t \left(\beta - \frac{\sigma_s^2}{2}\right) ds + \int_0^t \sigma_s\,dW_s,$$
$$\sigma_t^2 = \sigma_0^2 + \kappa \int_0^t \left(\alpha - \sigma_s^2\right) ds + \xi \int_0^t \sigma_s\,dV_s,$$

for some parameters β, κ, α and ξ, and with $\mathrm{Corr}(W, V) = \rho$. Here, the volatility process follows a Cox–Ingersoll–Ross model, that means it is mean-reverting with mean α and speed κ, and both diffusion coefficients are proportional with parameter ξ. Particularly the latter property appears to be rather typical for stochastic volatility models, and in this sense the Heston model can be regarded as prototypic. Popular alternatives are for example coming from the more general (but again parametric) class of (one factor) CEV models, where the diffusion coefficient τ of σ^2 becomes a general power function of σ, whereas the drift part of the volatility remains in principle the same. See [21] for a survey.

For this reason, statistical inference for stochastic volatility models has focused on parametric methods for most times, and usually the authors provide tools for a specific class of models. However, one is faced with two severe problems: First, it is in most cases impossible to assess the distribution of X (or its increments), which makes standard maximum likelihood theory unavailable. Second, the volatility process σ^2 is not observable, and many statistical concepts have in common that they propose to reproduce the unknown volatility process from observed option prices, typically by using proxies based on implied volatility. A survey on early estimation methods in this context can be found in [8]. One remarkable exception where stock price data only is used is the paper of [7] who construct a GMM estimator for the parameters of the Heston model from increments of realised variance. But also in a general setting with no specific model in mind, the focus has been on parametric approaches. An early approach on parameter estimation when σ^2 is ergodic is the work of [12], optimal rates are discussed in [15] and [13], and a maximum likelihood approach based on proxies for the volatility can be found in [1]. Even nonparametric concepts have been used to identify parameters of a stochastic volatility model; see, for example, [3] or [25].

Genuine nonparametric inference for stochastic volatility models has typically focused on function estimation. Both [24] and [9] discuss techniques for the estimation of f and g, when the volatility process satisfies $d\sigma_t^2 = f(\sigma_t^2)\,dt + g(\sigma_t^2)\,dV_t$. In the more general model-free context of (1.2), only [4] and [28] have discussed estimation of functionals of volatility of volatility. While the latter focus on estimation of a kind of leverage effect which involves the volatility of volatility process(es), the work of [4] provides a consistent estimator for integrated volatility of volatility $\int_0^t \tau_s^2\,ds$ in the one-factor case. Their approach is inspired by the asymptotic behaviour of realised variance, which states that the sum of squared increments of σ^2 converges in probability to the quantity of interest. Since σ^2 is not observable, the authors use spot volatility estimators instead.

We will pursue their approach and discuss in detail the asymptotic behaviour of several estimators for integrated volatility of volatility, all based on increments of spot volatility estimators, thus using observations of X only. It turns out that in order to attain the optimal rate of convergence in this context, it is necessary to conduct a certain bias correction which destroys positivity of the estimator – a feature which is well known from the related problem of volatility estimation under microstructure noise. Several stable central limit theorems are provided, and by defining appropriate estimators for the asymptotic (conditional) variance we obtain feasible versions as well. The latter results are of theoretical interest on one hand, but are extremely important from an applied point of view as well, as they make model validation for stochastic volatility models possible. Given the tremendous number of such models with entirely different qualitative behaviours, there is a lack of techniques that help deciding whether a certain model fits the data appropriately or not.

As a first approach to model validation in this framework, we give a brief idea on how to do goodness-of-fit testing, but our method is by no means limited to it. Related procedures can be used to test for example, whether a Brownian component or jumps are present in the volatility process and what in general the structure of the jump part is. Such problems have been solved for the price process X in recent years (see [18] for an overview), and in principle the methods are all based on the estimation of plain integrated volatility $\int_0^t \sigma_s^2 \,\mathrm{d}s$ and further quantities, such as truncated versions or bipower variation. Using our main results, these concepts can be translated to the stochastic volatility case by using estimators for integrated volatility of volatility instead, but usually with the slower rate of convergence $n^{-1/4}$.

The paper is organised as follows: In Section 2, we introduce our estimators and state the central limit theorems, whereas Section 3 is on goodness-of-fit testing in stochastic volatility models. Some Monte Carlo results can be found in Section 4, followed by some concluding remarks in Section 5. An overview on some proofs plus a couple of details can be found in the Appendix, whereas large parts of them have been relegated to a supplementary article [26].

2. Main results

Let us start with some conditions on the processes involved. All of these are rather mild and covered by a variety of (stochastic) volatility models used. The only major restriction is that we will assume most processes to be continuous for a while and only discuss briefly later how possible adjustments in order to handle jumps in price and volatility could look like.

Assumption 2.1. *Suppose that the process X is given by* (1.1), *where W is a standard Brownian motion and the drift process μ is left continuous. We assume further that the volatility process σ^2 is a continuous Itô semimartingale itself, having the representation* (1.2). *ν is assumed to be left continuous as well, whereas β satisfies the regularity condition*

$$\beta_s^2 = \beta_0^2 + \int_0^t \omega_s \,\mathrm{d}s + \int_0^t \vartheta_s^{(1)} \,\mathrm{d}W_s + \int_0^t \vartheta_s^{(2)} \,\mathrm{d}W_s', \tag{2.1}$$

where ω is locally bounded and each $\vartheta^{(l)}$ is left continuous, $l = 1, 2$. A similar condition is assumed to hold for η as well. Finally, all processes are defined on the same probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, and all coefficients are specified in such a way that σ^2 is almost surely positive and that β^2 and η^2 are either almost surely positive or vanishing identically, respectively.

As noted in the Introduction, (1.2) covers a large class of volatility models used. For $\eta \equiv 0$ we are essentially in the case of a local volatility model, whereas $\beta_s = \rho\tau_s$ and $\eta_s = \sqrt{1-\rho^2}\tau_s$ for some process τ and $\rho \in (-1, 1)$ refers to the setting of the typical stochastic volatility models mentioned before, in which both driving Brownian motions are correlated with ρ. The model in (1.2) is even more flexible, and it is straight-forward to extend all results to the case of a multi-factor model driven by more than two independent Brownian motions as well.

Our aim in the following is to draw inference on the integrated volatility of volatility up to time t, which becomes $\int_0^t (\beta_s^2 + \eta_s^2)\,\mathrm{d}s$ in our context. Any statistical inference will be based on high-frequency observations of X, and we assume that the data is recorded at equidistant times. Thus, without loss of generality let the process be defined on the interval [0, 1] and observed at the time points i/n, $i = 0, \dots, n$.

Before we discuss several concepts to assess integrated volatility of volatility in detail, let us recall the principles of estimation of standard integrated volatility $\int_0^1 \sigma_s^2\,\mathrm{d}s$. The usual estimator in the general model-free setting of (1.1) is realised volatility, given by

$$RV_t^n = \sum_{i=1}^{\lfloor nt \rfloor} \left(\Delta_i^n X\right)^2,$$

where we set $\Delta_i^n Z = Z_{i/n} - Z_{(i-1)/n}$ for any process Z. This estimator is optimal in several respects, even though Itô formula proves

$$\left(\Delta_i^n X\right)^2 = \int_{(i-1)/n}^{i/n} \sigma_s^2\,\mathrm{d}s + 2\int_{(i-1)/n}^{i/n} (X_s - X_{(i-1)/n})\,\mathrm{d}X_s \tag{2.2}$$

only, from which it is simple to see that each squared increment $(\Delta_i^n X)^2$ is only on average equal to integrated volatility over the corresponding time interval, but not consistent for it. (Realised volatility, the sum of the squared increments, however, *is* consistent for the entire integrated volatility, which is basically due to a martingale argument.)

Our estimators for integrated volatility of volatility will be based on a similar intuition: Define statistics via sums of increments such that each summand is on average equal to integrated volatility of volatility over the corresponding time interval, but not necessarily consistent. As before, one would like to build those estimators upon increments of σ^2. These are in general not observable, so a proxy for them is needed. Since we are in a model-free world, a natural estimator for spot volatility $\sigma_{i/n}^2$ is given by

$$\hat{\sigma}_{i/n}^2 = \frac{n}{k_n} \sum_{j=1}^{k_n} \left(\Delta_{i+j}^n X\right)^2, \qquad i = 0, \dots, n - k_n,$$

for some auxiliary (integer-valued) sequence k_n. See [2] or [25] for details on the asymptotic behaviour of this estimator. Itô formula again gives

$$\hat{\sigma}_{i/n}^2 = \frac{n}{k_n} \sum_{j=1}^{k_n} 2\int_{(i+j-1)/n}^{(i+j)/n} (X_s - X_{(i+j-1)/n})\,\mathrm{d}X_s + \frac{n}{k_n}\int_{i/n}^{(i+k_n)/n} \sigma_s^2\,\mathrm{d}s =: A_i^n + B_i^n, \tag{2.3}$$

so that $\hat{\sigma}^2_{i/n} - \sigma^2_{i/n}$ consists of two sources of error. From the proofs later on, we see that $A^n_i = \mathrm{O}_p(\sqrt{1/k_n})$, whereas $B^n_i - \sigma^2_{i/n} = \mathrm{O}_p(\sqrt{k_n/n})$. Therefore, it appears natural to choose k_n to be of the order $n^{1/2}$ in order to minimize the error of the spot volatility estimator (and we will see later that this is indeed the best thing to do), but we will keep this sequence arbitrary in order to allow for other estimators as well.

While the choice of the spot volatility estimators depends on the auxiliary sequence k_n, we will introduce a second sequence of integers l_n which governs the length of the intervals over which increments of $\hat{\sigma}^2$ are computed. Thus, the basic element of our final estimators will be $(\hat{\sigma}^2_{(i+l_n)/n} - \hat{\sigma}^2_{i/n})^2$, which can be decomposed as

$$\left(\hat{\sigma}^2_{(i+l_n)/n} - \hat{\sigma}^2_{i/n}\right)^2 = \left(A^n_{i+l_n} - A^n_i\right)^2 + \left(B^n_{i+l_n} - B^n_i\right)^2 + 2\left(A^n_{i+l_n} - A^n_i\right)\left(B^n_{i+l_n} - B^n_i\right).$$

The average behaviour of the terms above is discussed in the following lemma, and it depends crucially on the size of both k_n and l_n.

Lemma 2.2. *Suppose that Assumption 2.1 holds and let $\mathbb{E}^n_i[Z]$ denote conditional expectation of some variable Z with respect to $\mathcal{F}_{i/n}$. Set also $M_n = \max(k_n, l_n)$ and $m_n = \min(k_n, l_n)$. Then we have*

$$\mathbb{E}^n_i\left[\left(A^n_{i+l_n} - A^n_i\right)^2\right] = 4l_n(k_n M_n)^{-1}\sigma^4_{i/n}\left(1 + \mathrm{O}_p\left(M_n^{1/2}n^{-1/2}\right)\right),$$
$$\mathbb{E}^n_i\left[\left(B^n_{i+l_n} - B^n_i\right)^2\right] = l_n m_n(M_n - m_n/3)(k_n M_n n)^{-1}\left(\beta^2_{i/n} + \eta^2_{i/n}\right)\left(1 + \mathrm{O}_p\left(M_n^{1/2}n^{-1/2}\right)\right).$$

The previous lemma gives us several hints on how to obtain an estimator for integrated volatility of volatility via sums over $(\hat{\sigma}^2_{(i+l_n)/n} - \hat{\sigma}^2_{i/n})^2$. First, information about $\beta^2_{i/n} + \eta^2_{i/n}$ is contained in increments over the B^n_i only. Therefore, it appears to be reasonable to choose k_n and l_n later on in such a way that these terms are at least not smaller than the bias terms due to increments of A^n_i. Or in other words, the condition becomes that $n \le C k_n l_n$ for some generic $C > 0$.

Also, there are basically two ways to construct an estimator. Either, pick k_n and l_n such that the bias due to increments of A^n_i is negligible even after dividing by the rate of convergence. This concept will lead to the estimator

$$\hat{T}^n_t = \sum_{i=0}^{\lfloor nt\rfloor - (k_n+l_n)} k_n M_n\left(l_n m_n(M_n - m_n/3)\right)^{-1}\left(\hat{\sigma}^2_{(i+l_n)/n} - \hat{\sigma}^2_{i/n}\right)^2$$

which is positive by construction. As noted in the Introduction, this is the kind of estimator [4] were looking at. Alternatively, one can use a bias correction and subtract an estimator for the local quarticity $\sigma^4_{i/n}$. In this case one loses positivity, but we will see later that the rate of convergence is much faster in this situation.

Let us pursue the first path for a moment, however. In order to understand what the rate of convergence for estimation of integrated volatility of volatility will be, the next result is extremely helpful, as it gives the central limit theorem for the "oracle" estimator

$$\hat{S}_t^n = \sum_{i=0}^{\lfloor nt\rfloor-(k_n+l_n)} k_n M_n \big(l_n m_n (M_n - m_n/3)\big)^{-1} \big(B_{i+l_n}^n - B_i^n\big)^2$$

which depends on the unobservable increments of B_i^n only. All results in this section will be pointwise in t, even though it is likely that functional versions hold as well.

Proposition 2.3. *Suppose that Assumption 2.1 holds and that both $k_n \sim cn^\alpha$ and $l_n \sim dn^\beta$ hold for some $\alpha, \beta \in (0,1)$ and $c, d > 0$. Let also M_n and m_n be defined as before.*

(a) *If $\alpha \neq \beta$, we have*

$$\sqrt{\frac{n}{M_n}}\left(\hat{S}_t^n - \int_0^t \big(\beta_s^2 + \eta_s^2\big)\,\mathrm{d}s\right) \stackrel{\mathcal{L}-(s)}{\longrightarrow} \sqrt{4/3}\int_0^t \big(\beta_s^2 + \eta_s^2\big)\,\mathrm{d}\overline{W}_s.$$

(b) *For $k_n = l_n$ we have*

$$\sqrt{\frac{n}{M_n}}\left(\hat{S}_t^n - \int_0^t \big(\beta_s^2 + \eta_s^2\big)\,\mathrm{d}s\right) \stackrel{\mathcal{L}-(s)}{\longrightarrow} \sqrt{151/70}\int_0^t \big(\beta_s^2 + \eta_s^2\big)\,\mathrm{d}\overline{W}_s.$$

In both cases, $\overline{W}$ is a Brownian motion defined on an extension of the original probability space and independent of $\mathcal{F}$ and the convergence in (2.6) is $\mathcal{F}$-stable in law.

Remark 2.4. It is obvious from Proposition 2.3 that the rate of convergence becomes faster the smaller M_n is chosen. On the other hand, the condition $n \le Ck_nl_n$ forces M_n to be at least of the order $n^{1/2}$. In this case, the rate of convergence in Proposition 2.3 becomes $n^{-1/4}$, and this rate is known to be optimal for this statistical problem. Indeed, a related parametric setting has been discussed in [15] a decade ago, and it was shown therein that this rate is optimal in the special case, where β vanishes identically and η is a function of time and state, known up to a parameter θ.

Our first main theorem specifies conditions for a central limit theorem for $\hat{T}_t^n$ and is a simple consequence of Lemma 2.2 and Proposition 2.3.

Theorem 2.5. *Suppose that all the assumptions of Proposition 2.3 hold true. If further $n^{3/2}M_n^{-3/2}m_n^{-1} \to 0$ and $\alpha \neq \beta$, then the stable central limit theorem*

$$\sqrt{\frac{n}{M_n}}\left(\hat{T}_t^n - \int_0^t \big(\beta_s^2 + \eta_s^2\big)\,\mathrm{d}s\right) \stackrel{\mathcal{L}-(s)}{\longrightarrow} \sqrt{4/3}\int_0^t \big(\beta_s^2 + \eta_s^2\big)\,\mathrm{d}\overline{W}_s. \tag{2.4}$$

holds true.

The optimal rate of convergence in this case is obtained for the choice of $M_n = \mathrm{O}(n^{3/5+\varepsilon})$ and $m_n = \mathrm{O}(n^{3/5})$ and approaches $n^{-1/5}$ for $\varepsilon \to 0$. This proves also that it is no restriction to assume $\alpha \neq \beta$ above.

In order to obtain an estimator with the optimal rate of convergence, we choose l_n and k_n to be both the same and of the order $n^{1/2}$, but as noted above we need a bias correction then. Therefore, we define with a slight abuse of notation

$$\hat{R}_t^n = \sum_{i=0}^{\lfloor nt \rfloor - 2k_n} \left(\frac{3}{2k_n} \left(\hat{\sigma}^2_{(i+k_n)/n} - \hat{\sigma}^2_{i/n} \right)^2 - 6 \frac{1}{k_n^2} \hat{\sigma}^4_{i/n} \right), \tag{2.5}$$

where $\hat{\sigma}^4_{i/n} = \frac{n^2}{3k_n} \sum_{j=1}^{k_n} |\Delta^n_{i+j} X|^4$ is in general different from $(\hat{\sigma}^2_{i/n})^2$. Its asymptotic behaviour is discussed in the following theorem.

Theorem 2.6. *Suppose that Assumption 2.1 holds and let* $k_n = cn^{1/2} + \mathrm{o}(n^{1/4})$ *for some* $c > 0$. *Then*

$$\sqrt{\frac{n}{k_n}} \left(\hat{R}_t^n - \int_0^t \left(\beta_s^2 + \eta_s^2 \right) \mathrm{d}s \right) \xrightarrow{\mathcal{L}-(s)} U_t \tag{2.6}$$

for all $t > 0$, *where the limiting variable has the representation*

$$U_t = \int_0^t \alpha_s \, \mathrm{d}\overline{W}_s, \qquad \alpha_s^2 = \frac{48}{c^4} \sigma_s^8 + \frac{12}{c^2} \sigma_s^4 \left(\beta_s^2 + \eta_s^2 \right) + \frac{151}{70} \left(\beta_s^2 + \eta_s^2 \right)^2. \tag{2.7}$$

Remark 2.7. The situation encountered above has an interesting connection to the problem of eliminating microstructure noise, as we face similar problems regarding optimal rates of convergence and positivity of the estimators. Whereas the optimal rate of convergence for estimating integrated volatility in the noisy setting is $n^{-1/4}$, standard estimators attaining this rate are not always positive. To ensure positivity, one typically accepts a drop in the rate of convergence to $n^{-1/5}$ as well. See, for example, [6] for a thorough discussion in a general multivariate setting.

Remark 2.8. Recently, [19] discussed efficient estimation of $\int_0^t g(\sigma_s^2) \, \mathrm{d}s$ for general functions g. It turned out that Riemann sums based on $g(\hat{\sigma}^2_{i/n})$ indeed attain the optimal rate of convergence $n^{-1/2}$ in this context, but again the choice of k_n affects the limiting distribution. The optimal $k_n \sim n^{1/2}$ leads to additional bias terms in their setting, and at least some of these can be avoided by choosing k_n in a different way.

The limiting distribution in Theorem 2.5 and Theorem 2.6 is mixed normal, and in order to obtain a feasible central limit theorem we have to introduce consistent estimators for the respective conditional variances. These are constructed using the same intuition as before, and precisely we obtain the following theorem.

Theorem 2.9. (a) *Under the conditions of Theorem 2.5, we have*

$$\hat{Q}_t^n = \sum_{i=0}^{\lfloor nt \rfloor - (k_n + l_n)} \frac{4nk_n^2 M_n^2}{9(l_n m_n (M_n - m_n/3))^2} \left(\hat{\sigma}^2_{(i+l_n)/n} - \hat{\sigma}^2_{i/n} \right)^4 \xrightarrow{\mathbb{P}} \int_0^t \frac{4}{3} \left(\beta_s^2 + \eta_s^2 \right)^2 \mathrm{d}s.$$

(b) *In the situation of Theorem 2.6, we have*

$$G_{t,n}^{(1)} = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor - k_n} \left(\hat\sigma_{i/n}^4\right)^2 \xrightarrow{\mathbb{P}} \int_0^t \sigma_s^8 \,\mathrm{d}s,$$

$$G_{t,n}^{(2)} = \sum_{i=1}^{\lfloor nt\rfloor - 2k_n} \left(\frac{3}{2k_n}\left(\hat\sigma_{(i+k_n)/n}^2 - \hat\sigma_{i/n}^2\right)^2 - 6\frac{1}{k_n^2}\hat\sigma_{i/n}^4\right)\hat\sigma_{i/n}^4 \xrightarrow{\mathbb{P}} \int_0^t \sigma_s^4\left(\beta_s^2+\eta_s^2\right)\mathrm{d}s,$$

$$G_{t,n}^{(3)} = \sum_{i=1}^{\lfloor nt\rfloor - 2k_n} \frac{n}{k_n^2}\left(\hat\sigma_{(i+k_n)/n}^2 - \hat\sigma_{i/n}^2\right)^4 \xrightarrow{\mathbb{P}} \int_0^t \left(\frac{48}{c^4}\sigma_s^8 + \frac{16}{c^2}\sigma_s^4\left(\beta_s^2+\eta_s^2\right) + \frac{4}{3}\left(\beta_s^2+\eta_s^2\right)^2\right)\mathrm{d}s.$$

Therefore, as a consequence

$$\hat P_t^n = \frac{453}{280}G_{t,n}^{(3)} - \frac{n}{k_n^2}\frac{486}{35}G_{t,n}^{(2)} - \frac{n^2}{k_n^4}\frac{1038}{35}G_{t,n}^{(1)} \xrightarrow{\mathbb{P}} \int_0^t \alpha_s^2\,\mathrm{d}s.$$

Remark 2.10. Theorem 2.9 shows that a consistent estimator for $\int_0^t (\beta_s^2+\eta_s^2)^2\,\mathrm{d}s$ is for example, given by

$$\frac{3}{4}G_{t,n}^{(3)} - 12\frac{n}{k_n^2}G_{t,n}^{(2)} - 36\frac{n^2}{k_n^4}G_{t,n}^{(1)},$$

and its proof suggests that a central limit theorem holds with the same rate of convergence as before. In general, it is quite likely that this methods provides estimates for arbitrary even powers of integrated volatility of volatility. A precise theory is left for future research.

The properties of stable convergence guarantee that dividing by the square root of a consistent estimator for the conditional variance gives a feasible central limit theorem for the estimation of integrated volatility of volatility. See, for example, [23] for details. Therefore, the following corollary can be concluded easily.

Corollary 2.11. (a) *Under the assumptions of Theorem 2.5, we have for all $t > 0$*

$$\sqrt{\frac{n}{M_n}}\left(\hat T_t^n - \int_0^t \left(\beta_s^2+\eta_s^2\right)\mathrm{d}s\right)\left(\hat Q_t^n\right)^{-1/2} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{2.8}$$

(b) *Under the assumptions of Theorem 2.6, we have for all $t > 0$*

$$\sqrt{\frac{n}{k_n}}\left(\hat R_t^n - \int_0^t \left(\beta_s^2+\eta_s^2\right)\mathrm{d}s\right)\left(\hat P_t^n\right)^{-1/2} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1). \tag{2.9}$$

Remark 2.12. So far we have only discussed the case where both processes have continuous paths. Extensions to the situation of additional jumps in the price process seem to be possible, but are already quite involved. The following observation is useful: Whenever there is a jump within the interval $[i/n, (i+2k_n)/n]$, it appears squared and blown up by n^2/k_n^2 within $(\hat{\sigma}^2_{(i+k_n)/n} - \hat{\sigma}^2_{i/n})^2$. This is a much larger order than the usual k_n/n in the continuous case. For this reason, it appears as if the truncation method due to [22] can be applied, and a similar intuition holds for the bias correction as well. Note, however, that the raw statistics in this context are sums of squared increments of X rather than plain increments of X as for the power variations encountered in [22]. Therefore, the required techniques are different than the standard ones in this area.

The case of jumps in the volatility appears to be even more complicated, as these come into play via

$$B^n_{i+k_n} - B^n_i = \frac{n}{k_n} \int_{i/n}^{(i+k_n)/n} \left(\sigma^2_{s+k_n/n} - \sigma^2_s\right) \mathrm{d}s,$$

and therefore the amount to which each increment is affected by a jump depends crucially on the time at which the jump occurs. Thus, plain truncation might not be sufficient in this case and an entirely different estimator was necessary. Both topics are left for future research.

3. Model checks for stochastic volatility models

In this section, we propose a first approach to goodness-of-fit testing for stochastic volatility models. Assume we have representation (1.1) for the log price process X, whereas the volatility process satisfies $\mathrm{d}\sigma_t^2 = \nu_t\,\mathrm{d}t + \tau_t\,\mathrm{d}V_t$ as in typical SV models. There is still a lot of freedom in the modelling of σ^2, and the various proposals in the literature typically differ in the representation of its diffusion part τ. As noted in the Introduction, a quite general class of stochastic volatility models is given by the so-called CEV models, in which $\tau_s^2 = \theta(\sigma_s^2)^\gamma$ for some nonnegative γ and an unknown parameter θ, and the most popular among these is the Heston model from [14], corresponding to $\gamma = 1$.

In order to construct a test whether a certain functional relationship between σ and τ is present, we employ a technique which was already used in [10] or [27] when dealing with local volatility models. Suppose we are interested in testing for $\tau_s^2 = \tau^2(s, X_s, \sigma_s^2, \theta)$, where τ^2 is a given function and θ is some unknown (in general multidimensional) parameter. For simplicity, we will focus on the one-dimensional linear case only, that is

$$H_0 : \tau_s^2 = \theta\tau^2\left(s, X_s, \sigma_s^2\right) \qquad \text{for all } s \in [0, 1] \text{ (a.s.)}$$

Extensions to the general case follow along the lines of Section 5 in [27].

A test for the null hypothesis will be based on the observation that H_0 is equivalent to $N_t = 0$ for all $t \in [0, 1]$ (a.s.), where the process N_t is given by

$$N_t = \int_0^t \left(\tau_s^2 - \theta_{\min}\tau^2\left(s, X_s, \sigma_s^2\right)\right) \mathrm{d}s,$$

$$\theta_{\min} = \arg\min_\theta \int_0^1 \left(\tau_s^2 - \theta\tau^2\left(s, X_s, \sigma_s^2\right)\right)^2 \mathrm{d}s.$$

Assume that the function τ^2 is bounded away from zero. Then a standard argument from Hilbert space theory shows that $\theta_{\min} = D^{-1}C$ (and therefore $N_t = R_t - B_t D^{-1} C$), where we have set $R_t = \int_0^t \tau_s^2 \,\mathrm{d}s$ and

$$B_t = \int_0^t \tau^2\big(s, X_s, \sigma_s^2\big)\,\mathrm{d}s,$$

$$D = \int_0^1 \tau^4\big(s, X_s, \sigma_s^2\big)\,\mathrm{d}s,$$

$$C = \int_0^1 \tau_s^2 \tau^2\big(s, X_s, \sigma_s^2\big)\,\mathrm{d}s.$$

To define estimators let k_n as before and recall (2.5). We set

$$\hat{\tau}_{i/n}^2 = 3n(2k_n)^{-1}\big(\hat{\sigma}_{(i+k_n)/n}^2 - \hat{\sigma}_{i/n}^2\big)^2 - 6nk_n^{-2}\hat{\sigma}_{i/n}^4 \tag{3.1}$$

and also $\hat{N}_t^n = \hat{R}_t^n - \hat{B}_t^n (\hat{D}^n)^{-1} \hat{C}^n$ with $\hat{R}_t^n$ from the previous section, whereas we denote

$$\hat{B}_t^n = \frac{1}{n} \sum_{i=0}^{\lfloor nt \rfloor - k_n} \tau^2\left(\frac{i}{n}, X_{i/n}, \hat{\sigma}_{i/n}^2\right),$$

$$\hat{D}^n = \frac{1}{n} \sum_{i=0}^{n-k_n} \tau^4\left(\frac{i}{n}, X_{i/n}, \hat{\sigma}_{i/n}^2\right),$$

$$\hat{C}^n = \frac{1}{n} \sum_{i=0}^{n-2k_n} \hat{\tau}_{i/n}^2 \tau^2\left(\frac{i}{n}, X_{i/n}, \hat{\sigma}_{i/n}^2\right).$$

In the sequel, we will prove weak convergence of $\hat{N}_t^n - N_t$, up to a suitable normalisation. Theorem 2.6 suggests that $\sqrt{n/k_n}$ is a reasonable choice, and the following claim proves that two of the estimators converge at a faster speed, at least if we impose an additional smoothness condition on the function τ^2.

Lemma 3.1. *Suppose that the function τ^2 has continuous partial derivatives of second order. Then we have*

$$\hat{B}_t^n - B_t = \mathrm{o}_p\big(n^{-1/4}\big), \qquad \hat{D}^n - D = \mathrm{o}_p\big(n^{-1/4}\big),$$

the first result holding uniformly in $t \in [0, 1]$.

The above claim indicates that we have to focus on the terms involving $\hat{\tau}_{i/n}^2$ only, which is familiar ground due to the results of Section 2. We start with a proposition on the joint asymptotic behaviour of $\hat{R}_t^n$ and $\hat{C}^n$.

Lemma 3.2. *Let d be an integer and $t_1, \dots, t_d$ be arbitrary in* $[0, 1]$. *Set*

$$\Sigma_{t_1,\dots,t_d}\big(s, X_s, \sigma_s^2\big) = \alpha_s^2 h_{t_1,\dots,t_d}\big(s, X_s, \sigma_s^2\big) h_{t_1,\dots,t_d}\big(s, X_s, \sigma_s^2\big)^T$$

with $h_{t_1,\dots,t_d}(s, X_s, \sigma_s^2) = (1_{[0,t_1]}, \dots, 1_{[0,t_d]}, \tau^2(s, X_s, \sigma_s^2))^T$ *and* α_s^2 *as in Theorem* 2.6. *Under the previous assumptions we have the stable convergence*

$$\sqrt{\frac{n}{k_n}}\big(\hat{R}_{t_1}^n - R_{t_1}, \dots, \hat{R}_{t_d}^n - R_{t_d}, \hat{C}^n - C\big)^T \xrightarrow{\mathcal{L}-(s)} \int_0^1 \Sigma_{t_1,\dots,t_d}^{1/2}\big(s, X_s, \sigma_s^2\big)\,\mathrm{d}\overline{W}_s,$$

where $\overline{W}$ is a $(d + 1)$-dimensional standard Brownian motion defined on an extension of the original space and independent of $\mathcal{F}$.

We are interested in the asymptotics of the process $A_n(t) = \sqrt{n/k_n}(\hat{N}_t^n - N_t)$, and the preceding lemma basically leads to its finite dimensional convergence. The entire result on weak convergence of A_n reads as follows.

Theorem 3.3. *Assume that the previous assumptions hold. Then the process $(A_n(t))_{t\in[0,1]}$ converges weakly to a mean zero process $(A(t))_{t\in[0,1]}$, which is Gaussian conditionally on $\mathcal{F}$ and whose conditional covariance equals the one of the process*

$$\big\{\alpha_U\big(1_{\{U\le t\}} - B_t D^{-1}\tau^2\big(U, X_U, \sigma_U^2\big)\big)\big\}_{t\in[0,1]}$$

where $U \sim \mathcal{U}[0, 1]$, independent of $\mathcal{F}$.

As indicated before, convergence of the finite dimensional distributions is a direct consequence of Lemma 3.2, using the Delta method for stable convergence (see, e.g., [11]). Tightness follows from Theorem VI. 4.5 in [20] with a minimal amount of work.

Recall that $N_t = 0$ for all t under the null hypothesis. Therefore Theorem 3.3 shows that a consistent test is obtained by rejecting the null hypothesis for large values of a suitable functional of the process $\{\sqrt{n/k_n}\hat{N}_t^n\}_{t\in[0,1]}$. If we choose the Kolmogorov–Smirnov functional $K_n = \sup_{t\in[0,1]} \sqrt{n/k_n}|\hat{N}_t^n|$ for example, we have weak convergence under the null to $\sup_{t\in[0,1]} |A_t|$ as a consequence of Theorem 3.3. The distribution of the latter statistic is extremely difficult to assess, as it typically depends on the entire process (X, σ^2). We therefore propose to obtain critical values via a simple bootstrap procedure, which will be introduced in the next section.

To end this section, we define an appropriate estimator for the conditional variance of $A(t)$, which is given by

$$s_t^2 = \int_0^t \alpha_s^2\,\mathrm{d}s - 2B_t D^{-1} \int_0^t \alpha_s^2 \tau^2\big(s, X_s, \sigma_s^2\big)\,\mathrm{d}s + B_t^2 D^{-2} \int_0^t \alpha_s^2 \tau^4\big(s, X_s, \sigma_s^2\big)\,\mathrm{d}s,$$

due to Theorem 3.3. Empirical counterparts for B_t and D are obviously defined by the statistics $\hat{B}_t$ and $\hat{D}$, whereas Theorem 2.9 suggests that a local estimator for $\alpha^2_{i/n}$ is given by

$$\hat{\alpha}^2_{i/n} = \frac{n^2}{k_n^2}\left(\frac{453}{280}\left(\hat{\sigma}^2_{(i+k_n)/n} - \hat{\sigma}^2_{i/n}\right)^4 - \frac{486}{35}\hat{\tau}^2_{i/n}\hat{\sigma}^4_{i/n}\right) - \frac{n^6}{k_n^5}\frac{346}{1225}\sum_{j=1}^{k_n}\left|\Delta^n_{i+j}X\right|^8.$$

We obtain the following result, which can be proven in the same way as Theorem 2.9.

Theorem 3.4. *Let t be arbitrary and set*

$$\left(\hat{s}^n_t\right)^2 = \frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor - 2k_n}\hat{\alpha}^2_{i/n} - 2\hat{B}_t\hat{D}^{-1}\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor - 2k_n}\hat{\alpha}^2_{i/n}\tau^2\left(\frac{i}{n}, X_{i/n}, \hat{\sigma}^2_{i/n}\right)$$
$$+ \hat{B}_t^2\hat{D}^{-2}\frac{1}{n}\sum_{i=1}^{\lfloor nt\rfloor - 2k_n}\hat{\alpha}^2_{i/n}\tau^2\left(\frac{i}{n}, X_{i/n}, \hat{\sigma}^2_{i/n}\right).$$

Then $(\hat{s}^n_t)^2$ is consistent for s^2_t.

As a consequence, each statistic $\sqrt{n/k_n}\hat{N}^n_t/\hat{s}^n_t$ converges weakly to a normal distribution. This result will be used to construct a feasible bootstrap statistic in the following.

4. Simulation study

Let us start with a simulation study concerning the performance of the rate-optimal $\hat{R}^n_t$ as an estimator for integrated volatility of volatility. Throughout this section, we will work with the Heston model only, and the parameters are chosen as follows: $\beta = 0.3$, $\kappa = 5$, $\alpha = 0.2$ and $\xi = 0.5$. Furthermore, we set $X_0 = 0$ and $\sigma^2_0 = \alpha$. Note that the Feller condition $2\kappa\alpha \geq \xi^2$ is satisfied, which ensures that the process σ^2 is almost surely positive as requested. So does τ^2, and it is obvious that (2.1) holds as well. Therefore all conditions from Section 2 are satisfied.

We discuss the finite sample properties of $\hat{R}^n_t$ for different choices of the correlation parameter ρ and the number of observations n, and for comparability only we take n to be a square number and k_n equal to $n^{1/2}$ in all cases, so we have $c = 1$. Theorem 2.6 suggests that such a medium size of c is reasonable for finite samples, and additional results not reported here also point towards the fact that k_n should be chosen close to $n^{1/2}$. Finally, we set $t = 1$. Tables 1–6 below are based on 10 000 simulations.

Table 1 shows the performance for $\rho = 0$, for which we see that it takes quite some time for the asymptotics to kick in. Apparent is a slight overestimation of the lower tails of the distribution, which seems to originate from the relation of the estimators $\hat{R}_1$ and $G^{(3)}_{1,n}$. By construction, in cases where $\hat{R}_1$ is underestimating the true quantity, it is typically the case that increments of $\hat{\sigma}^2$ are relatively small. As these increments occur in $G^{(3)}_{1,n}$ as well, most likely the asymptotic variance is underestimated as well, which explains a too large negative standardised statistic. The

Table 1. Mean/variance and simulated quantiles of the feasible test statistic (2.9) for $\rho = 0$. The last column gives the relative amount of negative estimates

n	Mean	Variance	0.025	0.05	0.1	0.9	0.95	0.975	Neg.
400	−0.397	0.856	0.0497	0.0968	0.1756	0.9754	0.9946	0.9989	0.3522
2 500	−0.287	0.965	0.0526	0.0932	0.1619	0.9572	0.9862	0.9965	0.1963
10 000	−0.170	1.023	0.0449	0.0799	0.1425	0.9325	0.9757	0.9928	0.0933
22 500	−0.112	1.002	0.0404	0.0696	0.1253	0.9271	0.9722	0.9914	0.0510
40 000	−0.073	1.029	0.0401	0.0703	0.1235	0.9203	0.9690	0.9874	0.03
52 900	−0.031	1.022	0.0368	0.0653	0.1157	0.9154	0.9633	0.9872	0.0221

same effect is visible for the upper quantiles as well (but resulting in an overestimation), and this simple explanation is supported by a detailed look at simulation results not reported here which reveal that the estimation of the asymptotic variance is extremely accurate for moderate sizes of $\hat{R}_1 - \int_0^1 \tau_s^2 \,\mathrm{d}s$, but becomes worse when the deviation is rather large. Similar conclusions can be drawn for the case of a moderately negative $\rho = -0.2$.

We proceed with the finite sample behaviour of the statistics $\hat{T}_t^n$, for which we have a lot of freedom in choosing k_n and l_n. However, in order for both M_n to be rather small and the condition $n^{3/2} M_n^{-3/2} m_n^{-1} \to 0$ to be satisfied, we choose $M_n = \lfloor n^{3/4} \rfloor$ and $m_n = n^{1/2}$, resulting in a rate of convergence of about $n^{-1/8}$. Also, we restrict ourselves to $\rho = 0$.

As expected, the approximation of the nominal level is rather poor in this situation, both when reproducing mean/variance and the quantiles in the tails. Empirically the results do not improve for other choices of k_n and l_n. Note from Table 3 and Table 4 that results do not differ very much when choosing either k_n or l_n large, apart from the remarkable expection of a larger l_n and $n = 10\,000$. But even in this case, the results are not better than for the rate-optimal $\hat{R}_t^n$, which is why we recommend to choose this one rather than $\hat{T}_t^n$, even though only the latter estimator is ensured to be positive.

As an example for an application in goodness-of-fit testing, we have constructed a test for a Heston-like volatility structure via a bootstrap procedure as follows: Based on the observation that for each t, $\sqrt{n/k_n}\hat{N}_t^n/\hat{s}_t^n$ converges weakly to a standard normal distribution if the

Table 2. Mean/variance and simulated quantiles of the feasible test statistic (2.9) for $\rho = -0.2$. The last column gives the relative amount of negative estimates

n	Mean	Variance	0.025	0.05	0.1	0.9	0.95	0.975	Neg.
400	−0.386	0.874	0.0491	0.0967	0.1816	0.9724	0.9942	0.9989	0.3528
2 500	−0.295	0.971	0.0552	0.0963	0.1614	0.9559	0.9864	0.9962	0.1996
10 000	−0.176	1.013	0.0464	0.0808	0.1427	0.9369	0.9770	0.9940	0.0954
22 500	−0.226	0.987	0.0480	0.0840	0.1476	0.9436	0.9776	0.9932	0.0557
40 000	−0.075	1.001	0.0410	0.0673	0.1217	0.9254	0.9713	0.9904	0.0310
52 900	−0.040	1.019	0.0396	0.0677	0.1171	0.9180	0.9663	0.9879	0.0246

Table 3. Mean/variance and simulated quantiles of the feasible test statistic (2.8) for $k_n = \lfloor n^{3/4} \rfloor, l_n = n^{1/2}$ and $\rho = 0$

n	Mean	Variance	0.025	0.05	0.1	0.9	0.95	0.975
400	−0.490	1.261	0.0966	0.1304	0.1848	0.9990	1	1
2 500	−0.320	0.837	0.0548	0.0837	0.1390	0.9963	0.9999	1
10 000	−0.291	0.806	0.0514	0.0800	0.1332	0.9941	1	1
22 500	−0.259	0.920	0.0537	0.0873	0.1424	0.9739	0.9948	0.9996
40 000	−0.215	1.040	0.0658	0.0920	0.1380	0.9654	0.9970	1
52 900	−0.164	1.016	0.0594	0.0826	0.1274	0.9690	0.9988	1

null is satisfied, it seems reasonable to reject the hypothesis for large values of the standardised Kolmogorov–Smirnov statistic $Y_n = \sup_{i \le n-2k_n} |\sqrt{n/k_n} \hat{N}^n_{i/n} / \hat{s}^n_{i/n}|$. Since its (asymptotic) distribution is in general hard to assess, we used bootstrap quantiles instead, and precisely we have generated bootstrap data $X^{*(b)}_{i/n}$, $b = 1, \dots, B$, following the equation

$$X^*_t = \int_0^t \sigma^*_s \, \mathrm{d}W^*_s, \qquad (\sigma^*_t)^2 = \hat{\alpha} + \int_0^t \hat{\kappa} \big(\hat{\alpha} - (\sigma^*_s)^2\big) \, \mathrm{d}s + \hat{\xi} \int_0^t \sigma^*_s \, \mathrm{d}V^*_s.$$

Here, W^* and V^* are independent Brownian motions, and we have identified $\hat{\alpha}$ with the realised volatility of the original data (which is a measure for the average volatility over [0, 1]) and defined $\hat{\xi} = \hat{\theta}^{1/2}$, since both quantities coincide under the null. Finally, we have simply set $\hat{\kappa} = 5\hat{\theta}/\hat{\alpha}$ such that Feller's condition is satisfied. Setting $B = 200$, we have run 500 simulations each.

Table 5 shows that the simulated levels are rather close to the expected ones, irrespectively of n. We have tested two alternatives from the class of CEV models, namely

$$\sigma^2_t = \sigma^2_0 + \kappa \int_0^t \big(\alpha - \sigma^2_s\big) \, \mathrm{d}s + V_t$$

Table 4. Mean/variance and simulated quantiles of the feasible test statistic (2.8) for $l_n = \lfloor n^{3/4} \rfloor, k_n = n^{1/2}$ and $\rho = 0$

n	Mean	Variance	0.025	0.05	0.1	0.9	0.95	0.975
400	−0.476	1.255	0.0976	0.1316	0.1889	0.9983	0.9999	1
2 500	−0.311	0.817	0.0505	0.0779	0.1322	0.9950	0.9996	1
10 000	0.149	1.196	0.0450	0.0657	0.1005	0.8784	0.9589	0.9904
22 500	−0.276	0.812	0.0460	0.0728	0.1234	0.9886	0.9989	1
40 000	−0.217	1.033	0.0648	0.0914	0.1354	0.9647	0.9974	1
52 900	−0.306	0.824	0.0494	0.0829	0.1456	0.9882	0.9981	1

Table 5. Simulated level of the bootstrap test based on the standardised Kolmogorov–Smirnov statistic Y_n

n	0.01	0.025	0.05	0.1	0.2
400	0.004	0.012	0.024	0.064	0.172
2500	0.018	0.040	0.064	0.120	0.216
10000	0.010	0.018	0.040	0.084	0.194
22500	0.016	0.024	0.034	0.088	0.194
40000	0.020	0.038	0.068	0.128	0.220
52900	0.010	0.020	0.052	0.118	0.200

and

$$\sigma_t^2 = \sigma_0^2 + \kappa \int_0^t \left(\alpha - \sigma_s^2\right) \mathrm{d}s + \sqrt{\kappa} \int_0^t \sigma_s^2 \,\mathrm{d}V_s,$$

corresponding to $\gamma = 0$ and $\gamma = 2$, respectively, and using the parameters from above. We see from the simulation results that the rejection probabilities are much larger for the second alternative than for the first, which can partially explained from two observations: First, the Vasicek model does not satisfy the assumptions from the previous sections since the volatility may become negative (in which case it is set to zero); second, our choice of $\hat{\kappa}$ is responsible for a large speed of mean reversion in the bootstrap algorithm which makes it difficult to distinguish between a Heston-like volatility of volatility and a constant one. It is expected that the power improves for an entirely data-driven choice of $\hat{\kappa}$.

5. Conclusion

In this paper, we have discussed a nonparametric method to estimate the integrated volatility of volatility process in stochastic volatility models. Our concept is based on spot volatility estimators, and just as for standard realised volatility we use sums of squares of these spot volatility

Table 6. Simulated rejection probabilities of the bootstrap test based on the standardised Kolmogorov–Smirnov functional statistic Y_n for various alternatives

Alt	$\gamma = 0$					$\gamma = 2$				
n	0.01	0.025	0.05	0.1	0.2	0.01	0.025	0.05	0.1	0.2
400	0.032	0.072	0.124	0.192	0.292	0.056	0.080	0.128	0.204	0.320
2500	0.028	0.052	0.082	0.134	0.262	0.044	0.090	0.156	0.248	0.372
10000	0.032	0.048	0.086	0.138	0.260	0.036	0.084	0.176	0.284	0.396
22500	0.024	0.042	0.068	0.138	0.302	0.032	0.086	0.162	0.284	0.432
40000	0.028	0.046	0.094	0.196	0.426	0.028	0.064	0.120	0.310	0.482
52900	0.026	0.040	0.082	0.174	0.422	0.024	0.058	0.144	0.320	0.488

estimators to obtain a global estimator for integrated volatility of volatility. Two classes of estimators have been investigated – one consisting of positive estimators with a slow rate of convergence, the other one being bias corrected but converging at the optimal rate $n^{-1/4}$. In both cases, central limit theorems are provided, and we also discuss briefly why a truncated version could be useful when there are additional jumps in the price process.

Given the variety of stochastic volatility models (in continuous time) which are used to describe financial data, there is a severe lack in tools on model validation. Our results fill this gap to a first extent, as we provide a bootstrap method for goodness-of-fit testing in such models which investigates whether a specific parametric model for volatility of volatility is appropriate given the data or not. A rigorous proof that the proposed procedure keeps the asymptotic level and is consistent against a large class of alternatives has not been provided, however, and is left for future research.

A different issue to take microstructure issues into account which are likely to be present when data is observed at high-frequency. Again it is promising to combine filtering methods for noisy diffusions with the method proposed in this paper to obtain an estimator for integrated volatility of volatility in such models as well, but the rate of convergence is expected to drop further. Precise statements are beyond the scope of the paper as well.

Appendix

Note first that every left-continuous process is locally bounded, thus all processes appearing are. Second, standard localisation procedures as in [5] or [17] allow us to assume that any locally bounded process is actually bounded, and that almost surely positive processes can be regarded as bounded away from zero. Universal constants are denoted by C or C_r, the latter if we want to emphasise dependence on some additional parameter r.

Within the main corpus, we give the proof of Theorem 2.6 only, which is the by far most complicated result of this work. Analogues of Lemma 2.2 and Proposition 2.3 for the special case of $l_n = k_n$ are of course parts of it, and it is not difficult to generalise the proofs in order for both claims to be covered as well. Therefore, these results are not shown explicitly. Let us start with a brief sketch of what we will be doing. In general, $\mathcal{F}$-stable convergence of a sequence Z_n to some limiting variable Z defined on an extension $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$ of the original space is equivalent to

$$\mathbb{E}\big[h(Z_n)Y\big] \to \widetilde{\mathbb{E}}\big[h(Z)Y\big] \tag{A.1}$$

for any bounded Lipschitz function h and any bounded $\mathcal{F}$-measurable Y. For details, see, for example, [20] and related work. Suppose now that there are additional variables $Z_{n,p}$ and Z_p (the latter defined on the same extension as Z) such that

$$\lim_{p\to\infty} \limsup_{n\to\infty} \mathbb{E}\big[|Z_n - Z_{n,p}|\big] = 0, \tag{A.2}$$

$$Z_{n,p} \stackrel{\mathcal{L}-(s)}{\longrightarrow} Z_p \qquad \text{for all } p, \tag{A.3}$$

$$\lim_{p\to\infty} \widetilde{\mathbb{E}}\big[|Z_p - Z|\big] = 0, \tag{A.4}$$

hold. Then the desired stable convergence $Z_n \stackrel{\mathcal{L}-(s)}{\longrightarrow} Z$ follows. Indeed, let $\varepsilon > 0$. Then there exists a $\delta > 0$ such that $|x-y| < \delta$ implies $|h(x)-h(y)| < \varepsilon$. Thus we have

$$\begin{aligned}
&\big|\mathbb{E}\big[h(Z_n)Y\big] - \mathbb{E}\big[h(Z_{n,p})Y\big]\big| \\
&\quad \le C\big(\mathbb{E}\big[h(Z_n) - h(Z_{n,p})|1_{\{|Z_n - Z_{n,p}| \ge \delta\}}\big] + \mathbb{E}\big[\big|h(Z_n) - h(Z_{n,p})\big|1_{\{|Z_n - Z_{n,p}| < \delta\}}\big]\big) \\
&\quad \le C\big(P\big(|Z_n - Z_{n,p}| \ge \delta\big) + \varepsilon\big).
\end{aligned}$$

We have $\lim_{p\to\infty}\limsup_{n\to\infty}|\mathbb{E}[h(Z_n)Y] - \mathbb{E}[h(Z_{n,p})Y]| = 0$ from Markov inequality, (A.2) and as ε was arbitrary. $\lim_{p\to\infty}|\widetilde{\mathbb{E}}[h(Z_p)Y] - \widetilde{\mathbb{E}}[h(Z)Y]| = 0$ can be shown similarly using (A.4), and (A.3) is by definition equivalent to $\lim_{n\to\infty}|\mathbb{E}[h(Z_{n,p})Y] - \widetilde{\mathbb{E}}[h(Z_p)Y]| = 0$. Putting the latter three claims together (plus the triangle inequality and the fact that all three limiting conditions on p and n are actually the same) gives (A.1).

Our aim in this proof is to employ a certain blocking technique, which allows us to make use of a type of conditional independence between the summands within $\hat{R}_t^n$. To this end, we apply the above methodology, so we have to define an appropriate double sequence $U_t^{n,p}$, which will correspond to an approximated version of $\hat{R}_t^n$ where we sum over the big blocks only. Some additional notation is necessary. Let $p \in \mathbb{N}$ be arbitrary. We set

$$\begin{aligned}
a_\ell(p) &= (\ell-1)(p+2)k_n, \\
b_\ell(p) &= a_\ell(p) + pk_n, \\
c(p) &= J_n(p)(p+2)k_n + 1,
\end{aligned}$$

the first two for any $\ell = 1, \ldots, J_n(p)$ with $J_n(p) = \lfloor \lfloor nt - 2k_n \rfloor / ((p+2)k_n) \rfloor$. These numbers depend on n as well, even though it does not show up in the notation. We define further $H_i^n = \int_{(i-1)/n}^{i/n}(W_s - W_{(i-1)/n})\,\mathrm{d}W_s$. In order to exploit the afore-mentioned conditional independence, we need approximations for A_i^n and B_i^n from (2.3). For the sake of brevity, we will only state the approximated increments explicitly, which are given by

$$\begin{aligned}
\widetilde{A}_{(i+k_n)/n} - \widetilde{A}_{i/n} &:= \frac{n}{k_n}\sum_{j=1}^{k_n} 2\sigma^2_{a_\ell(p)/n}\big(H^n_{i+j+k_n} - H^n_{i+j}\big) \\
&= \frac{n}{k_n}\sigma^2_{a_\ell(p)/n}\sum_{j=1}^{k_n}\big(\big(\Delta^n_{i+k_n+j}W\big)^2 - \big(\Delta^n_{i+j}W\big)^2\big),
\end{aligned} \tag{A.5}$$

where the latter identity is a consequence of Itô formula, and

$$\begin{aligned}
&\widetilde{B}_{(i+k_n)/n} - \widetilde{B}_{i/n} \\
&\quad := \frac{n}{k_n}\int_{i/n}^{(i+k_n)/n}\big(\beta_{a_\ell(p)/n}(W_{s+k_n/n} - W_s) + \eta_{a_\ell(p)/n}\big(W'_{s+k_n/n} - W'_s\big)\big)\,\mathrm{d}s.
\end{aligned} \tag{A.6}$$

These quantities are defined for $i = a_\ell(p), \dots, b_\ell(p) - 1$, thus over the big blocks. For later reasons, we introduce similar approximations over the small blocks. Set

$$\widetilde{C}_{(i+k_n)/n} - \widetilde{C}_{i/n} = \frac{n}{k_n}\sigma^2_{b_\ell(p)/n}\sum_{j=1}^{k_n}\bigl((\Delta^n_{i+k_n+j}W)^2 - (\Delta^n_{i+j}W)^2\bigr),$$

$$\widetilde{D}_{(i+k_n)/n} - \widetilde{D}_{i/n} = \frac{n}{k_n}\int_{i/n}^{(i+k_n)/n}\bigl(\beta_{b_\ell(p)/n}(W_{s+k_n/n} - W_s) + \eta_{b_\ell(p)/n}\bigl(W'_{s+k_n/n} - W'_s\bigr)\bigr)\,\mathrm{d}s)\,\mathrm{d}s,$$

both for $i = b_\ell(p), \dots, a_{\ell+1}(p) - 1$. Then the following claim holds, whose proof is postponed to the supplemental file [26].

Lemma A.1. *We have*

$$\mathbb{E}\bigl[\bigl|A_{(i+k_n)/n} - A_{i/n} - (\widetilde{A}_{(i+k_n)/n} - \widetilde{A}_{i/n})\bigr|^r\bigr] \le C_r\bigl(pn^{-1}\bigr)^{r/2},$$
$$\mathbb{E}\bigl[\bigl|B_{(i+k_n)/n} - B_{i/n} - (\widetilde{B}_{(i+k_n)/n} - \widetilde{B}_{i/n})\bigr|^r\bigr] \le C_r\bigl(pn^{-1}\bigr)^{r/2},$$

as well as $\mathbb{E}[|A_{(i+k_n)/n} - A_{i/n}|^r] \le C_r n^{-r/4}$ *and* $\mathbb{E}[|B_{(i+k_n)/n} - B_{i/n}|^r] \le C_r n^{-r/4}$ *for every* $r > 0$. *The latter bounds hold also for the approximated versions, and the same results are true for the approximation via increments of* $\widetilde{C}$ *and* $\widetilde{D}$ *over the small blocks.*

Up to a different standardisation, the role of $Z_{n,p}$ in this proof will be played by $U_t^{n,p} = \sum_{\ell=1}^{J_n(p)} U_\ell^{n,p}$, where

$$U_\ell^{n,p} = \sum_{i=a_\ell(p)}^{b_\ell(p)-1}\frac{3}{2k_n}\bigl((\widetilde{A}_{(i+k_n)/n} - \widetilde{A}_{i/n}) + (\widetilde{B}_{(i+k_n)/n} - \widetilde{B}_{i/n})\bigr)^2 - \frac{pk_n}{n}\left[\frac{6n}{k_n^2}\sigma^4_{a_\ell(p)/n} + \bigl(\beta^2_{a_\ell(p)/n} + \eta^2_{a_\ell(p)/n}\bigr)\right] \tag{A.7}$$

involves quantities from the big blocks only. The $U_\ell^{n,p}$ can be shown to be martingale differences, and the most involved part in the proof is to use Lemma A.1 to obtain

$$\lim_{p\to\infty}\limsup_{n\to\infty}\sqrt{\frac{n}{k_n}}\mathbb{E}\left[\left|\left(\hat{R}_t^n - \int_0^t\bigl(\beta_s^2 + \eta_s^2\bigr)\,\mathrm{d}s\right) - U_t^{n,p}\right|\right] = 0, \tag{A.8}$$

which is the analogue of (A.2). Let us focus on the remaining two steps as well. We set

$$U_t^p = \int_0^t \alpha(p)_s\,\mathrm{d}\overline{W}_s,$$

$$\alpha(p)_s^2 = \frac{p}{p+2}\left(\frac{48p + d_1}{pc^4}\sigma_s^8 + \frac{12p + d_2}{pc^2}\sigma_s^4\bigl(\beta_s^2 + \eta_s^2\bigr) + \frac{151p + d_3}{70p}\bigl(\beta_s^2 + \eta_s^2\bigr)^2\right)$$

for certain unspecified constants d_l, $l = 1, 2, 3$. In order to prove the stable convergence

$$\sqrt{\frac{n}{k_n}} U_t^{n,p} \xrightarrow{\mathcal{L}-(s)} U_t^p \tag{A.9}$$

we use a well-known result for triangular arrays of martingale differences, which is due to Jacod [16]. In particular, the following three conditions have to be checked.

$$\frac{n}{k_n} \sum_{\ell=1}^{J_n(p)} \mathbb{E}^n_{a_\ell(p)}\big[\big(U_\ell^{n,p}\big)^2\big] \xrightarrow{\mathbb{P}} \int_0^t \alpha(p)_s^2 \, \mathrm{d}s, \tag{A.10}$$

$$\frac{n^2}{k_n^2} \sum_{\ell=1}^{J_n(p)} \mathbb{E}^n_{a_\ell(p)}\big[\big(U_\ell^{n,p}\big)^4\big] \xrightarrow{\mathbb{P}} 0, \tag{A.11}$$

$$\sqrt{\frac{n}{k_n}} \sum_{\ell=1}^{J_n(p)} \mathbb{E}^n_{a_\ell(p)}\big[U_\ell^{n,p}(N_{a_{\ell+1}(p)/n} - N_{a_\ell(p)/n})\big] \xrightarrow{\mathbb{P}} 0, \tag{A.12}$$

where N is any component of (W, W') or a bounded martingale orthogonal to both W and W'. The final step $\lim_{p\to\infty} \widetilde{E}|U_t^p - U_t| = 0$ is obvious.

A.1. Proof of (A.8)

For simplicity, we set $\eta \equiv 0$ and $\vartheta^{(2)} \equiv 0$ from now on, as otherwise the proof is exactly the same. In a brief first step, we replace $\hat{R}_t^n$ by a version in which the unknown bias and not the estimator for it is subtracted, that is we introduce

$$U_t^n = \sum_{i=0}^{\lfloor nt \rfloor - 2k_n} \frac{3}{2k_n}\big(\hat{\sigma}^2_{(i+k_n)/n} - \hat{\sigma}^2_{i/n}\big)^2 - \frac{6}{c^2}\int_0^t \sigma_s^4 \, \mathrm{d}s - \int_0^t \beta_s^2 \, \mathrm{d}s.$$

Theorem 2.1 in [5] shows that integrals over σ can be estimated with rate $n^{-1/2}$, so the assumption on k_n and a standard argument regarding boundary terms prove that

$$\sqrt{\frac{n}{k_n}} \mathbb{E}\bigg[\bigg|\bigg(\hat{R}_t^n - \int_0^t \beta_s^2 \, \mathrm{d}s\bigg) - U_t^n\bigg|\bigg] = \mathrm{o}(1),$$

uniformly in t. A simple consequence of Lemma A.1 is that the remainder terms in U_t^n are negligible, that is

$$\lim_{p\to\infty} \limsup_{n\to\infty} \sqrt{\frac{n}{k_n}} \mathbb{E}\Bigg[\Bigg|\sum_{i=c(p)}^{\lfloor nt \rfloor - 2k_n} \frac{3}{2k_n}\big(\hat{\sigma}^2_{(i+k_n)/n} - \hat{\sigma}^2_{i/n}\big)^2 - \frac{6}{c^2}\int_{c(p)/n}^t \sigma_s^4 \, \mathrm{d}s - \int_{c(p)/n}^t \beta_s^2 \, \mathrm{d}s\Bigg|\Bigg] = 0,$$

using also boundedness of the processes on the right hand side and the definition of $c(p)$. Therefore, we are left to show

$$\lim_{p\to\infty}\limsup_{n\to\infty}\sqrt{n/k_n}\,\mathbb{E}\big[\big|\widetilde{U}_t^{n,p}-U_t^{n,p}\big|\big]=0 \tag{A.13}$$

with

$$\begin{aligned}\widetilde{U}_t^{n,p}&=\sum_{\ell=1}^{J_n(p)}\Bigg(\sum_{i=a_\ell(p)}^{b_\ell(p)-1}+\sum_{i=b_\ell(p)}^{a_{\ell+1}(p)-1}\Bigg)\big(\hat{\sigma}^2_{(i+k_n)/n}-\hat{\sigma}^2_{i/n}\big)^2\\&\quad-\frac{6}{c^2}\int_0^{c(p)/n}\sigma_s^4\,\mathrm{d}s-\int_0^{c(p)/n}\beta_s^2\,\mathrm{d}s.\end{aligned} \tag{A.14}$$

For the integrals within (A.14), recall that these are replaced by approximated versions in $U_t^{n,p}$. Therefore we have to show for example,

$$\lim_{p\to\infty}\limsup_{n\to\infty}\sqrt{\frac{n}{k_n}}\mathbb{E}\Bigg[\Bigg|\sum_{\ell=1}^{J_n(p)}\int_{a_\ell(p)/n}^{b_\ell(p)/n}\big(\beta_s^2-\beta^2_{a_\ell(p)/n}\big)\,\mathrm{d}s\Bigg|\Bigg]=0. \tag{A.15}$$

For its proof, recall (2.1). The result above follows from

$$\mathbb{E}\Bigg[\Bigg|\sum_{\ell=1}^{J_n(p)}\int_{a_\ell(p)/n}^{b_\ell(p)/n}\int_{a_\ell(p)/n}^{s}\omega_r\,\mathrm{d}r\,\mathrm{d}s\Bigg|\Bigg]\le C\frac{n}{pk_n}\bigg(\frac{pk_n}{n}\bigg)^2\le Cpn^{-1/2}$$

and

$$\mathbb{E}\Bigg(\sum_{\ell=1}^{J_n(p)}\int_{a_\ell(p)/n}^{b_\ell(p)/n}\int_{a_\ell(p)/n}^{s}\vartheta_r^{(1)}\,\mathrm{d}W_r\,\mathrm{d}s\Bigg)^2=\sum_{\ell=1}^{J_n(p)}\mathbb{E}\bigg(\int_{a_\ell(p)/n}^{b_\ell(p)/n}\int_{a_\ell(p)/n}^{s}\vartheta_r^{(1)}\,\mathrm{d}W_r\,\mathrm{d}s\bigg)^2\le Cp^2n^{-1}.$$

Of course, the similar claim

$$\lim_{p\to\infty}\limsup_{n\to\infty}\sqrt{\frac{n}{k_n}}\mathbb{E}\Bigg[\Bigg|\sum_{\ell=1}^{J_n(p)}\int_{a_\ell(p)/n}^{b_\ell(p)/n}\big(\sigma_s^4-\sigma^4_{a_\ell(p)/n}\big)\,\mathrm{d}s\Bigg|\Bigg]=0 \tag{A.16}$$

holds for the same reasons. We have further

$$\lim_{p\to\infty}\limsup_{n\to\infty}\sqrt{\frac{n}{k_n}}\mathbb{E}\Bigg[\Bigg|\sum_{\ell=1}^{J_n(p)}\frac{pk_n}{n}\bigg(\frac{n}{k_n^2}-\frac{1}{c^2}\bigg)\sigma^4_{a_\ell(p)/n}\Bigg|\Bigg]=0, \tag{A.17}$$

which by boundedness of σ amounts to prove $n^{-3/4}(k_n^2-nc^2)=\mathrm{o}(1)$, and the latter is satisfied by definition of k_n. Note that analogues of (A.15), (A.16) and (A.17) are satisfied over the small blocks as well.

The latter claims prove that we are left to show the approximation over the big blocks, which is

$$\lim_{p\to\infty}\limsup_{n\to\infty}\sqrt{\frac{n}{k_n}}\mathbb{E}\Bigg[\Bigg|\sum_{\ell=1}^{J_n(p)}\sum_{i=a_\ell(p)}^{b_\ell(p)-1}\frac{3}{2k_n}\Big(\big((A_{(i+k_n)/n}-A_{i/n})+(B_{(i+k_n)/n}-B_{i/n})\big)^2 - \big((\widetilde{A}_{(i+k_n)/n}-\widetilde{A}_{i/n})+(\widetilde{B}_{(i+k_n)/n}-\widetilde{B}_{i/n})\big)^2\Big)\Bigg|\Bigg]=0, \tag{A.18}$$

and the negligibility of the small blocks, that is

$$\lim_{p\to\infty}\limsup_{n\to\infty}\sqrt{\frac{n}{k_n}}\mathbb{E}\Bigg[\Bigg|\sum_{\ell=1}^{J_n(p)}\Bigg(\sum_{i=b_\ell(p)}^{a_{\ell+1}(p)-1}\frac{3}{2k_n}\big(\hat{\sigma}^2_{(i+k_n)/n}-\hat{\sigma}^2_{i/n}\big)^2 - \frac{2k_n}{n}\bigg[\frac{6n}{k_n^2}\sigma^4_{(b_\ell(p))/n}+\beta^2_{b_\ell(p)/n}\bigg]\Bigg)\Bigg|\Bigg]=0 \tag{A.19}$$

to obtain (A.8).

To prove (A.18), the binomial theorem tells us that we can discuss the approximation for B, the one for A and the mixed part separately. Using further $x^2-y^2=2y(x-y)+(x-y)^2$ and $xx'-yy'=(x-y)y'+y(x'-y')+(x-y)(x'-y')$, we see from Lemma A.1 and the growth conditions that (A.18) follows from $\lim_{p\to\infty}\limsup_{n\to\infty}\sum_{r=1}^4\mathbb{E}[|L^{(j)}_{n,p}|]=0$ with

$$L^{(1)}_{n,p}=\sqrt{\frac{n}{k_n}}\sum_{\ell=1}^{J_n(p)}\sum_{i=a_\ell(p)}^{b_\ell(p)-1}\frac{1}{k_n}\big((B_{(i+k_n)/n}-B_{i/n}) - (\widetilde{B}_{(i+k_n)/n}-\widetilde{B}_{i/n})\big)(\widetilde{B}_{(i+k_n)/n}-\widetilde{B}_{i/n}), \tag{A.20}$$

$$L^{(2)}_{n,p}=\sqrt{\frac{n}{k_n}}\sum_{\ell=1}^{J_n(p)}\sum_{i=a_\ell(p)}^{b_\ell(p)-1}\frac{1}{k_n}\big((B_{(i+k_n)/n}-B_{i/n}) - (\widetilde{B}_{(i+k_n)/n}-\widetilde{B}_{i/n})\big)(\widetilde{A}_{(i+k_n)/n}-\widetilde{A}_{i/n}), \tag{A.21}$$

$$L^{(3)}_{n,p}=\sqrt{\frac{n}{k_n}}\sum_{\ell=1}^{J_n(p)}\sum_{i=a_\ell(p)}^{b_\ell(p)-1}\frac{1}{k_n}\big((A_{(i+k_n)/n}-A_{i/n}) - (\widetilde{A}_{(i+k_n)/n}-\widetilde{A}_{i/n})\big)(\widetilde{A}_{(i+k_n)/n}-\widetilde{A}_{i/n}), \tag{A.22}$$

$$L^{(4)}_{n,p}=\sqrt{\frac{n}{k_n}}\sum_{\ell=1}^{J_n(p)}\sum_{i=a_\ell(p)}^{b_\ell(p)-1}\frac{1}{k_n}\big((A_{(i+k_n)/n}-A_{i/n}) - (\widetilde{A}_{(i+k_n)/n}-\widetilde{A}_{i/n})\big)(\widetilde{B}_{(i+k_n)/n}-\widetilde{B}_{i/n}). \tag{A.23}$$

Proofs of these claims can be found in the supplementary material [26].

Finally, to obtain (A.19), we compute the conditional expectation of the approximated increments, and we will do this for the $\widetilde{A}$ and $\widetilde{B}$ terms only. We have

$$\begin{aligned}
&\mathbb{E}^n_{a_\ell(p)}\big[(\widetilde{A}_{(i+k_n)/n} - \widetilde{A}_{i/n})^2\big] \\
&= \frac{n^2}{k_n^2}\sigma^4_{a_\ell(p)/n}\sum_{j=1}^{k_n}\mathbb{E}\big[\big(\big(\Delta^n_{i+k_n+j}W\big)^2 - \big(\Delta^n_{i+k_n}W\big)^2\big)^2\big] \qquad \text{(A.24)}\\
&= \frac{4}{k_n}\sigma^4_{a_\ell(p)/n}
\end{aligned}$$

as well as

$$\begin{aligned}
&\mathbb{E}^n_{a_\ell(p)}\big[(\widetilde{B}_{(i+k_n)/n} - \widetilde{B}_{i/n})^2\big] \\
&= 2\frac{n^2}{k_n^2}\beta^2_{a_\ell(p)/n}\int_{i/n}^{(i+k_n)/n}\int_{i/n}^{s}\mathbb{E}\big[(W_{s+k_n/n} - W_s)(W_{r+k_n/n} - W_r)\big]\,\mathrm{d}r\,\mathrm{d}s \\
&= 2\frac{n^2}{k_n^2}\beta^2_{a_\ell(p)/n}\int_{i/n}^{(i+k_n)/n}\int_{i/n}^{s}\left(r + k_n/n - s\right)\mathrm{d}r\,\mathrm{d}s \\
&= \frac{2k_n}{3n}\beta^2_{a_\ell(p)/n}.
\end{aligned}$$

The expectation of the mixed part is zero. Obviously, we have

$$\begin{aligned}
\mathbb{E}^n_{b_\ell(p)}\Bigg[\Bigg(\sum_{i=b_\ell(p)}^{a_{\ell+1}(p)-1}\frac{3}{2k_n}\big((\widetilde{C}_{(i+k_n)/n} - \widetilde{C}_{i/n}) + (\widetilde{D}_{(i+k_n)/n} - \widetilde{D}_{i/n})\big)^2 \\
- \frac{2k_n}{n}\left[\frac{6n}{k_n^2}\sigma^4_{b_\ell(p)/n} + \beta^2_{b_\ell(p)/n}\right]\Bigg)\Bigg] = 0
\end{aligned}$$

as well. (A.19) then follows from the fact that

$$\begin{aligned}
\frac{n}{k_n}\sum_{\ell=1}^{J_n(p)}\mathbb{E}\Bigg[\Bigg(\sum_{i=b_\ell(p)}^{a_{\ell+1}(p)-1}\frac{3}{2k_n}\big((\widetilde{C}_{(i+k_n)/n} - \widetilde{C}_{i/n}) + (\widetilde{D}_{(i+k_n)/n} - \widetilde{D}_{i/n})\big)^2 \\
- \frac{2k_n}{n}\left[\frac{6n}{k_n^2}\sigma^4_{b_\ell(p)/n} + \beta^2_{b_\ell(p)/n}\right]\Bigg)^2\Bigg]
\end{aligned}$$

is bounded by a constant times p^{-1}, using Lemma A.1.

A.2. Proof of (A.9)

Let us check the conditions for stable convergence in this step, where particularly the proof of (A.10) is tedious. Write $U_\ell^{n,p} = \sum_{s=1}^3 U_\ell^{n,p,s}$ with

$$U_\ell^{n,p,1} = \sum_{i=a_\ell(p)}^{b_\ell(p)-1} \frac{3}{2k_n}\left((\widetilde{A}_{(i+k_n)/n} - \widetilde{A}_{i/n})^2 - \frac{4}{k_n}\sigma^4_{a_\ell(p)/n}\right),$$

$$U_\ell^{n,p,2} = \sum_{i=a_\ell(p)}^{b_\ell(p)-1} \frac{3}{2k_n}\left((\widetilde{B}_{(i+k_n)/n} - \widetilde{B}_{i/n})^2 - \frac{2k_n}{3n}\beta^2_{a_\ell(p)/n}\right),$$

$$U_\ell^{n,p,3} = \sum_{i=a_\ell(p)}^{b_\ell(p)-1} \frac{3}{k_n}(\widetilde{A}_{(i+k_n)/n} - \widetilde{A}_{i/n})(\widetilde{B}_{(i+k_n)/n} - \widetilde{B}_{i/n}).$$

We have seen in the final step above that these terms are indeed martingale differences, and it turns out that only the $(U_\ell^{n,p,s})^2$ terms are responsible for the conditional variance, whereas the remaining mixed ones are of small order each. To summarize, the following lemma holds which is proven in the supplementary material [26].

Lemma A.2. *We have*

$$\mathbb{E}^n_{a_\ell(p)}\big[\big(U_\ell^{n,p,1}\big)^2\big] = \sigma^8_{a_\ell(p)/n}\frac{48p + d_1}{k_n^2} + \mathrm{O}_P\big(pn^{-3/2}\big),$$

$$\mathbb{E}^n_{a_\ell(p)}\big[\big(U_\ell^{n,p,2}\big)^2\big] = \beta^4_{a_\ell(p)/n}\left(\frac{151}{70}p + d_2\right)\frac{k_n^2}{n^2} + \mathrm{O}_P\big(pn^{-3/2}\big),$$

$$\mathbb{E}^n_{a_\ell(p)}\big[\big(U_\ell^{n,p,3}\big)^2\big] = \sigma^4_{a_\ell(p)/n}\beta^2_{a_\ell(p)/n}\frac{12p + d_3}{n} + \mathrm{O}_P\big(pn^{-3/2}\big),$$

for certain unspecified constants d_m, $m = 1, 2, 3$, as well as for each $r \neq s$

$$\mathbb{E}^n_{a_\ell(p)}\big[U_\ell^{n,p,r}U_\ell^{n,p,s}\big] = \mathrm{O}_P\big(pn^{-3/2}\big).$$

We use Lemma A.2 to obtain

$$\begin{aligned}
&\frac{n}{k_n}\sum_{\ell=1}^{J_n(p)} \mathbb{E}^n_{a_\ell(p)}\big[\big(U_\ell^{n,p}\big)^2\big] \\
&\quad = \frac{pk_n}{n}\sum_{\ell=1}^{J_n(p)}\left(\frac{n^2}{k_n^4}\left(48 + \frac{d_1}{p}\right)\sigma^8_{a_\ell(p)/n}\right. \\
&\qquad\qquad \left. + \frac{n}{k_n^2}\left(12 + \frac{d_2}{p}\right)\sigma^4_{a_\ell(p)/n}\beta^2_{a_\ell(p)/n} + \left(\frac{151}{70} + \frac{d_3}{p}\right)\beta^4_{a_\ell(p)/n}\right) + \mathrm{O}_p\left(\frac{1}{n^{1/2}}\right),
\end{aligned}$$

thus (A.10) holds using $k_n \sim cn^{1/2}$. Simpler to obtain is (A.11), as Lemma A.1 gives

$$\frac{n^2}{k_n^2}\sum_{\ell=1}^{J_n(p)} \mathbb{E}^n_{a_\ell(p)}\big[\big(U_\ell^{n,p}\big)^4\big] \le C\frac{n^3}{pk_n^3}p^4n^{-2},$$

which converges to zero in the usual sense. Finally, one can prove

$$\mathbb{E}^n_{a_\ell(p)}\Bigg[\sum_{i=a_\ell(p)}^{b_\ell(p)-1}\frac{3}{2k_n}\big((\widetilde{A}_{(i+k_n)/n}-\widetilde{A}_{i/n}) + (\widetilde{B}_{(i+k_n)/n}-\widetilde{B}_{i/n})\big)^2(N_{a_{\ell+1}(p)/n}-N_{a_\ell(p)/n})\Bigg]=0, \tag{A.25}$$

where N is either W or W' or when N is a bounded martingale, orthogonal to (W, W'). Focus on the first case and decompose $((\widetilde{A}_{(i+k_n)/n} - \widetilde{A}_{i/n}) + (\widetilde{B}_{(i+k_n)/n} - \widetilde{B}_{i/n}))^2$ via the binomial theorem. For the pure $\widetilde{A}$ and the pure $\widetilde{B}$ term, the claim follows immediately from properties of the normal distribution upon using that $\sigma_{a_\ell(p)/n}$ or $\beta_{a_\ell(p)/n}$ are $\mathcal{F}_{a_\ell(p)/n}$ measurable. For the mixed term, one has to use the special form of $\widetilde{A}_{(i+k_n)/n} - \widetilde{A}_{i/n}$ as a difference of two sums, and a symmetry argument proves (A.25) in this case. For an orthogonal N, we use standard calculus. By Itô formula, both $(\widetilde{A}_{(i+k_n)/n} - \widetilde{A}_{i/n})^2$ and $(\widetilde{B}_{(i+k_n)/n} - \widetilde{B}_{i/n})^2$ are a measurable variable times the sum of a constant and a stochastic integral with respect to W and W', respectively. Thus (A.25) holds. In the mixed case, we use integration by parts formula to reduce $(\widetilde{A}_{(i+k_n)/n} - \widetilde{A}_{i/n})(\widetilde{B}_{(i+k_n)/n} - \widetilde{B}_{i/n})$ to the sum of a constant, a $\mathrm{d}W$- and a $\mathrm{d}W'$-integral. Then the same argument applies. Altogether, this gives (A.12).

Acknowledgements

The author is grateful for financial support through the collaborative research center "Statistical modeling of nonlinear dynamic processes" (SFB 823) of the German Research Foundation (DFG). Special thanks go to two anonymous referees for their valuable comments on earlier versions of this paper.

Supplementary Material

Additional proofs for claims made in the article (DOI: 10.3150/14-BEJ648SUPP; .pdf). We provide several proofs for either theorems from the main corpus or additional steps discussed in the Appendix.

References

[1] Aït-Sahalia, Y. and Kimmel, R. (2007). Maximum likelihood estimation of stochastic volatility models. *J. Financial Economics* **134** 507–551.

[2] Alvarez, A., Panloup, F., Pontier, M. and Savy, N. (2012). Estimation of the instantaneous volatility. *Stat*. *Inference Stoch*. *Process*. **15** 27–59. MR2892587

[3] Bandi, F. and Renò, R. (2008). Nonparametric stochastic volatility. Technical report.

[4] Barndorff-Nielsen, O. and Veraart, A. (2009). Stochastic volatility of volatility in continuous time. Technical report.

[5] Barndorff-Nielsen, O.E., Graversen, S.E., Jacod, J., Podolskij, M. and Shephard, N. (2006). A central limit theorem for realised power and bipower variations of continuous semimartingales. In *From Stochastic Calculus to Mathematical Finance* 33–68. Berlin: Springer. MR2233534

[6] Barndorff-Nielsen, O.E., Hansen, P.R., Lunde, A. and Shephard, N. (2011). Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading. *J*. *Econometrics* **162** 149–169. MR2795610

[7] Bollerslev, T. and Zhou, H. (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *J*. *Econometrics* **109** 33–65. MR1899692

[8] Chernov, M. and Ghysels, E. (2000). Estimation of stochastic volatility models for the purpose of option pricing. In *Computational Finance* 1999 (Y. Abu-Mostafa, B. LeBaron, A. Lo and A. Weigend, eds.) 567–581. Cambridge: MIT Press.

[9] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2010). Nonparametric estimation for a stochastic volatility model. *Finance Stoch*. **14** 49–80. MR2563205

[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*. *Stat*. **33** 259–278. MR2279642

[12] Genon-Catalot, V., Jeantheau, T. and Laredo, C. (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **5** 855–872. MR1715442

[13] Gloter, A. (2007). Efficient estimation of drift parameters in stochastic volatility models. *Finance Stoch*. **11** 495–519. MR2335831

[14] Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bonds and currency options. *Rev*. *Financial Studies* **6** 327–343.

[15] Hoffmann, M. (2002). Rate of convergence for parametric estimation in a stochastic volatility model. *Stochastic Process*. *Appl*. **97** 147–170. MR1870964

[16] Jacod, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. In *Séminaire de Probabilités XXXI*. *Lecture Notes in Math*. **1655** 232–246. Berlin: Springer. MR1478732

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

[18] Jacod, J. and Protter, P. (2012). *Discretization of Processes*. *Stochastic Modelling and Applied Probability* **67**. Heidelberg: Springer. MR2859096

[19] Jacod, J. and Rosenbaum, M. (2013). Quarticity and other functionals of volatility: Efficient estimation. *Ann*. *Statist*. **41** 1462–1484. MR3113818

[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] Jones, C.S. (2003). The dynamics of stochastic volatility: Evidence from underlying and options markets. *J*. *Econometrics* **116** 181–224. Frontiers of financial econometrics and financial engineering. MR2002525

[22] Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand*. *J*. *Stat*. **36** 270–296. MR2528985

[23] Podolskij, M. and Vetter, M. (2010). Understanding limit theorems for semimartingales: A short survey. *Stat*. *Neerl*. **64** 329–351. MR2683464

[24] Renò, R. (2006). Nonparametric estimation of stochastic volatility models. *Econom. Lett.* **90** 390–395. MR2212176
[25] Vetter, M. (2012). Estimation of correlation for continuous semimartingales. *Scand. J. Stat.* **39** 757–771. MR3000847
[26] Vetter, M. (2014). Supplement to "Estimation of integrated volatility of volatility with applications to goodness-of-fit testing." DOI:10.3150/14-BEJ648SUPP.
[27] Vetter, M. and Dette, H. (2012). Model checks for the volatility under microstructure noise. *Bernoulli* **18** 1421–1447. MR2995803
[28] Wang, C.D. and Mykland, P.A. (2014). The estimation of leverage effect with high-frequency data. *J. Amer. Statist. Assoc.* **109** 197–215. MR3180557

Received July 2012 and revised March 2014