Ann Finance (2012) 8:235–258
DOI 10.1007/s10436-010-0152-8

SYMPOSIUM

A Gaussian calculus for inference from high frequency data

Per A. Mykland

Received: 6 May 2009 / Accepted: 23 March 2010 / Published online: 30 April 2010

Abstract In the econometric literature of high frequency data, it is often assumed that one can carry out inference conditionally on the underlying volatility processes. In other words, conditionally Gaussian systems are considered. This is often referred to as the assumption of "no leverage effect". This is often a reasonable thing to do, as general estimators and results can often be conjectured from considering the conditionally Gaussian case. The purpose of this paper is to try to give some more structure to the things one can do with the Gaussian assumption. We shall argue in the following that there is a whole treasure chest of tools that can be brought to bear on high frequency data problems in this case. We shall in particular consider approximations involving locally constant volatility processes, and develop a general theory for this approximation. As applications of the theory, we develop an ANOVA for processes with multiple regressors, and give an estimator for error bars on the Hayashi–Yoshida estimator of quadratic covariation. Other applications are considered in other papers.

Keywords Asynchronous observation · Consistency · Cumulants · Contiguity · Continuity · Discrete observation · Efficiency · High frequency data · Itô process · Likelihood inference · Realized volatility · Stable convergence

JEL Classification C02 · C13 · C14 · C22 · D52 · D81

1 Introduction

Recent years have seen an explosion of literature in the area of estimating volatility on the basis of high frequency data. The concepts go back to stochastic calculus,

---

P. A. Mykland (✉)
Department of Statistics, University of Chicago, Chicago, IL 60637, USA
e-mail: mykland@pascal.uchicago.edu

see, for example, Karatzas and Shreve (1991) (Section 1.5), Jacod and Shiryaev (2003) (Theorem I.4.47 on page 52), and Protter (2004) (Theorem II-22 on page 66). An early econometric discussion of this relationship can be found in Andersen et al. (2000). Recent work both from the probabilistic and econometric side give the mixed normal distribution of the error in approximation. References include Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002), Zhang (2001) and Mykland and Zhang (2006).

Further econometric literature includes, in particular, Gallant et al. (1999), Chernov and Ghysels (2000), Andersen et al. (2001, 2003), Dacorogna et al. (2001), and Gonçalves and Meddahi (2009). Problems that are attached to the estimation of covariations between two processes are discussed in Hayashi and Yoshida (2005) and Zhang (2010). Estimating instantaneous volatility at each point in time goes back to Foster and Nelson (1996) and Comte and Renault (1998), see also Mykland and Zhang (2008), but this has not caught on quite as much in the econometric application. There is also an emerging literature on that happens in the presence of observation error, but we are not planning to address this question here.

In the econometric literature, it is often assumed that one can carry out inference conditionally on the underlying volatility processes. In other words, conditionally Gaussian systems are considered. This is often referred to as the assumption of "no leverage effect". This is often a reasonable thing to do, as general estimators and results can often be conjectured from considering the conditionally Gaussian case.

The purpose of this paper is to try to give some more structure to the things one can do with the Gaussian assumption. We shall argue in the following that there is a whole treasure chest of tools that can be brought to bear on high frequency data problems in this case. After setting up the structure in Sect. 2, we do a warm-up in Sect. 3 to show that likelihood (parametric inference) and cumulant methods can be used to define and analyze estimators. It will become clear that there is mileage in considering systems that have locally constant volatility, and we approach this systematically in Sect. 4, culminating in our main Theorems 1 and 2.

As applications of the theory, we first (Sect. 5.1) revisit the ANOVA problem from Zhang (2001) and Mykland and Zhang (2006), this time in the setting of several regressors. We shall see that with the theory in hand, one can use classical regression theory of carry our an ANOVA. In fact, the amount of smoothing needed is over a *finite* number of observations. This is a proposition which, I think, would have attracted long odds if not announced as a theorem, and it is evidence of the power of the theorems from Sect. 4. Second, we discuss the problem of setting error bars on the Hayashi–Yoshida estimator of quadratic covariation (Sect. 5.2). References to other applications, including how to improve efficiency, are given at the beginning of Sect. 5.

Finally, some disclaimers. First of all, we are not claiming to have invented the conditional Gaussian assumption; it is used by a big fraction of the theory. Second, this paper is conceptually similar to Mykland and Zhang (2009) in that we consider approximate measures where the volatility is locally constant. The approximations used in the two papers are, however, quite different. For one thing, in the case of one step approximations, the "approximation" in the current paper is actually exact, thus permitting us to consider (exact) nonparametric maximum likelihood estimators in Sect. 2. Furthermore, if one compares Theorem 1 with Theorem 3 in Mykland and Zhang (2009), one can see that, also in the multi step case, the approximation in this paper is much closer than the one in the other paper. Third, this is not an attempt at a comprehensive study of what one can to with the Gaussian case; there are much too many tools available and open problems for that. In particular, we do not consider the case where observations have error. Our hope is that this study will encourage further use in high frequency data of the ideas and results that are available for Gaussian situations.

2 The model, and some immediate conclusions

In general, we shall work with a p-variate Itô process $(X_t^{(1)}, \dots, X_t^{(p)})$, given by the system

$$dX_t^{(k)} = \mu_t^{(k)}\, dt + \sigma_t^{(k)}\, dW_t^{(k)},\ k = 1, \dots, p, \tag{1}$$

where $\mu_t^{(k)}$ and $\sigma^{(k)}$ are adapted *càdlàg* random processes, and the $W_t^{(k)}$ are Brownian motions that are not necessarily independent. We shall suppose that the process $X_t^{(k)}$ is observed at times $0 = t_{k,0} < t_{k,1} < \cdots < t_{k,n_k} = T$. If $p = 1$, we may sometimes suppress the "k". The underlying filtration will be called $(\mathcal{F}_t)$.

Assumption 1 (*Sampling times*) In addition, when doing asymptotics, we suppose that there is an index N, so that $t_{k,i} = t_{N,k,i}$ (the additional subscript will normally be suppressed). The grids $\{0 = t_{N,k,0} < t_{N,k,1} < \cdots < t_{N,k,n_{N,k}} = T\}$ will not be assumed to be nested when N varies. To get a concrete example of what one can take as N, one can use

$$N = n_{N,1} + \cdots + n_{N,p}. \tag{2}$$

We then do asymptotics as $N \to \infty$. The basic assumption is that

$$\max_{1 \le i \le n_{N,k}} |t_{N,k,i} - t_{N,k,i-1}| = O(N^{-1}) \tag{3}$$

for each k, $1 \le k \le p$. We emphasize that p is a fixed number which does not vary with N.

We now describe the setting for parametric inference.

Assumption 2 (*A conditionally Gaussian system*) We let P be a probability distribution on the form Eq. 1 for which $\mu_t^{(k)} = 0$ for all k. We assume that we can take the quadratic variations and covariations $\langle X^{(k)}, X^{(l)} \rangle_t$ to be $\mathcal{F}_0$-measurable. As is customary, in this case, we call $(X_t^{(1)}, \dots, X_t^{(p)})$ a Brownian martingale. Note that "nonrandom" is a special case of "$\mathcal{F}_0$-measurable". We let P_ω denote the regular conditional probability distribution given $\mathcal{F}_0$, and note that the (X_t) process is Gaussian under P_ω, for (almost) every ω. We also suppose that the observation times $t_{k,i}$ are nonrandom, but irregular.

We emphasize that the quadratic variations and covariations will normally themselves be taken to be Itô-processes. It will be clear from Sect. 4 that this is desirable in predicting results. See Remark 2 just after Theorem 2.

Remark 1 (*If the drift μ is not zero*) Our asymptotic results in Sect. 4 and onwards will remain valid, under mild regularity conditions, even when drift $\mu_t^{(k)}$ in Eq. 1 is nonzero. This is explained in Sect. 4.3

Under Assumption 2, the set of observations $(X_{t_{k,i}}^{(k)} - X_{t_{k,i-1}}^{(k)}, 1 \le i \le n_k, 1 \le k \le p)$ is, conditionally on $\mathcal{F}_0$, simply a multivariate normal vector with mean zero, and with covariances given by

$$\kappa_{k,i;l,j} = \mathrm{Cov}_\omega(X_{t_{k,i}}^{(k)} - X_{t_{k,i-1}}^{(k)}, X_{t_{l,j}}^{(l)} - X_{t_{l,j-1}}^{(l)}) \\ = \begin{cases} \langle X^{(k)}, X^{(l)} \rangle_{t_{k,i} \wedge t_{l,j}} - \langle X^{(k)}, X^{(l)} \rangle_{t_{k,i-1} \vee t_{l,j-1}} & \text{if } (t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset \\ 0 \quad \text{otherwise} \end{cases} \tag{4}$$

where, as usual, $x \wedge y = \min(x, y)$ and $x \vee y = \max(x, y)$.

The log likelihood is then, as usual, given by

$$\ell(\kappa) = -\frac{1}{2} \ln \det(\kappa) - \frac{1}{2} \sum_{k,i,l,j} \kappa^{k,i;l,j} (X_{t_{k,i}}^{(k)} - X_{t_{k,i-1}}^{(k)})(X_{t_{l,j}}^{(l)} - X_{t_{l,j-1}}^{(l)}) - \frac{N}{2} \ln(2\pi), \tag{5}$$

where $\kappa^{k,i;l,j}$ are the elements of the matrix inverse of $(\kappa_{k,i;l,j})$, and N is given by Eq. 2. κ is the $N \times N$ matrix of all the $\kappa_{k,i;l,j}$.

We are now in a position to show that the by now classical estimates of volatility and covariation are, in fact, likelihood estimates under Assumption 2.

Note first that by standard considerations, the MLEs of the parameters are given by

$$\hat{\kappa}_{k,i;l,j} = \begin{cases} \left(X_{t_{k,i}}^{(k)} - X_{t_{k,i-1}}^{(k)}\right)\left(X_{t_{l,j}}^{(l)} - X_{t_{l,j-1}}^{(l)}\right) & \text{if } (t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

(cf., for example, the derivation in Chapter 4 of Mardia et al. 1979). Thus, two immediate conclusions.

Example 1 (*The classical estimate of quadratic variation*) The MLE of $\langle X^{(k)}, X^{(k)} \rangle_T$ is given by

$$\widehat{\langle X^{(k)}, X^{(k)} \rangle}_T = \sum_i \hat{\kappa}_{k,i;k,i} = \sum_i \left(X_{t_{k,i}}^{(k)} - X_{t_{k,i-1}}^{(k)}\right)^2. \tag{7}$$

This is, of course, the estimate which has been commonly used in the literature, cf. the references in the Introduction.

Example 2 (*The MLE for covariation*) The MLE of $\langle X^{(k)}, X^{(l)}\rangle_T$ is similarly given by

$$
\begin{aligned}
\widehat{\langle X^{(k)}, X^{(l)}\rangle}_T &= \sum_{i,j:(t_{k,i-1},t_{k,i})\cap(t_{l,j-1},t_{l,j})\neq\emptyset} \hat{\kappa}_{k,i;l,j} \\
&= \sum_{i,j:(t_{k,i-1},t_{k,i})\cap(t_{l,j-1},t_{l,j})\neq\emptyset} \left(X^{(k)}_{t_{k,i}} - X^{(k)}_{t_{k,i-1}}\right)\left(X^{(l)}_{t_{l,j}} - X^{(l)}_{t_{l,j-1}}\right). \qquad (8)
\end{aligned}
$$

This coincides with the Hayashi and Yoshida (2005)-estimator.

Example 1 is, of course, the reinvention of a long known estimator. In the case of Example 2, however, we are dealing with a procedure which only dates back a couple of years. Thus, we are already close to the research frontier. And there is more in the following.

Note that even if estimators have been derived under Assumption 2, it has earlier been shown by the authors cited in the Introduction that these estimators have reasonable properties also in the more general case. Thus, likelihood in the conditionally Gaussian case is a useful way of generating estimators.

3 Warm-up: the quantification of error in the estimators

It is customary in likelihood inference to use the Fisher information to quantify variance. We shall here see that this leads to interesting conclusions. There is some insight in the following lemma.

Lemma 1 (Covariance and expected information) *Under Assumption 2, if* $(t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset$ *and* $(t_{m,g-1}, t_{m,g}) \cap (t_{n,h-1}, t_{n,h}) \neq \emptyset$

$$
\begin{aligned}
\mathrm{Cov}_\omega\left(\hat{\kappa}_{k,i;l,j}, \hat{\kappa}_{m,g;n,h}\right) &= E_\omega\left(-\frac{\partial^2 \ell}{\partial\kappa_{k,i;l,j}\partial\kappa_{m,g;n,h}}\right)^{-1} \\
&= \kappa_{k,i;n,h}\kappa_{l,j;m,g} + \kappa_{k,i;m,g}\kappa_{l,j;n,h} \qquad (9)
\end{aligned}
$$

The lemma is a direct consequence of the exponential family structure, and the fact that if $Z_1, \ldots, Z_4$ are jointly normal, then $\mathrm{Cov}(Z_1 Z_2, Z_3 Z_4) = \mathrm{Cov}(Z_1, Z_3)\mathrm{Cov}(Z_2, Z_4) + \mathrm{Cov}(Z_1, Z_4)\mathrm{Cov}(Z_2, Z_3)$.

Again, there are some immediate consequences of this.

Example 1 (continued) For the estimate Eq. 7 of quadratic variation one obtains

$$
\begin{aligned}
\mathrm{Var}_\omega\left(\widehat{\langle X^{(k)}, X^{(k)}\rangle}_T\right) &= 2\sum_i \kappa^2_{k,i;k,i} \\
&= 2\sum_i \left(E_\omega\left(X^{(k)}_{t_{k,i}} - X^{(k)}_{t_{k,i-1}}\right)^2\right)^2 \\
&= \frac{2}{3}\sum_i E_\omega\left(\left(X^{(k)}_{t_{k,i}} - X^{(k)}_{t_{k,i-1}}\right)^4\right), \qquad (10)
\end{aligned}
$$

since for a mean zero normal random variable Z, $E(Z^4) = 3E(Z^2)^2$. This quantity is naturally estimated by

$$\frac{2}{3}\sum_i \left(X^{(k)}_{t_{k,i}} - X^{(k)}_{t_{k,i-1}}\right)^4. \tag{11}$$

This, of course, is the quarticity estimate of Barndorff-Nielsen and Shephard (2002). Again, the estimator follows from a simple conditionally Gaussian likelihood. Its asymptotic validity has been shown under more general Itô process assumptions by Barndorff-Nielsen and Shephard (2002) and Mykland and Zhang (2006) (in the latter, see Remark 2 (p. 1944) and the proof on p. 1952).

The question now arises whether we can do anything new with the setup we have given. That is what the rest of the paper is about.

A natural first question to ask is whether we can provide the error of the Hayashi and Yoshida (2005) estimator, a.k.a. (Eq. 8). The simple part of this is that by Lemma 1,

$$\begin{aligned}
&\mathrm{Var}_\omega\left(\widehat{\langle X^{(k)}, X^{(l)}\rangle}_T\right)\\
&= \sum_{i,j:(t_{k,i-1},t_{k,i})\cap(t_{l,j-1},t_{l,j})\neq\emptyset}\ \sum_{g,h:(t_{k,g-1},t_{k,g})\cap(t_{l,h-1},t_{l,h})\neq\emptyset} \mathrm{Cov}_\omega\left(\hat\kappa_{k,i;l,j}, \hat\kappa_{k,g;l,h}\right)\\
&= \sum_{i,j:(t_{k,i-1},t_{k,i})\cap(t_{l,j-1},t_{l,j})\neq\emptyset}\ \sum_{g,h:(t_{k,g-1},t_{k,g})\cap(t_{l,h-1},t_{l,h})\neq\emptyset} \kappa_{k,i;k,g}\kappa_{l,j;l,h} + \kappa_{k,i;l,h}\kappa_{k,g;l,j}\\
&= \sum_{i,j:(t_{k,i-1},t_{k,i})\cap(t_{l,j-1},t_{l,j})\neq\emptyset} \kappa_{k,i;k,i}\kappa_{l,j;l,j} + \sum_{(i,j,g,h)\in A_n} \kappa_{k,i;l,h}\kappa_{k,g;l,j},
\end{aligned} \tag{12}$$

where

$$\begin{aligned}
A_n = \{(i,j,g,h) :\ & (t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset, (t_{k,g-1}, t_{g,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset,\\
& (t_{k,i-1}, t_{k,i}) \cap (t_{l,h-1}, t_{l,h}) \neq \emptyset,\ \text{and } (t_{k,g-1}, t_{k,g}) \cap (t_{l,h-1}, t_{l,h}) \neq \emptyset\}.
\end{aligned} \tag{13}$$

A perhaps slightly more interpretable expression is to write

$$\begin{aligned}
\mathrm{Var}_\omega\left(\widehat{\langle X^{(k)}, X^{(l)}\rangle}_T\right) = &\sum_{i,j:(t_{k,i-1},t_{k,i})\cap(t_{l,j-1},t_{l,j})\neq\emptyset} (\kappa_{k,i;k,i}\kappa_{l,j;l,j} + (\kappa_{k,i;l,j})^2)\\
&+ \sum_{(i,j,g,h)\in B_n} \kappa_{k,i;l,h}\kappa_{k,g;l,j},
\end{aligned} \tag{14}$$

$$\begin{aligned}
B_n = \{(i,j,g,h) :\ & i \neq g, h \neq j, (t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j})\\
& \neq \emptyset, (t_{k,g-1}, t_{g,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset,\\
& (t_{k,i-1}, t_{k,i}) \cap (t_{l,h-1}, t_{l,h}) \neq \emptyset \text{ and } (t_{k,g-1}, t_{k,g}) \cap (t_{l,h-1}, t_{l,h}) \neq \emptyset\}.
\end{aligned} \tag{15}$$

The expressions Eqs. 12 and 14, however, do not show how to estimate $\mathrm{Var}_\omega\left(\widehat{\langle X^{(k)}, X^{(l)}\rangle_T}\right)$. For example,

$$E_\omega(\hat{\kappa}_{k,i;k,i}\hat{\kappa}_{l,j;l,j}) = E_\omega((\hat{\kappa}_{k,i;l,j})^2) = \kappa_{k,i;k,i}\kappa_{l,j;l,j} + 2(\kappa_{k,i;l,j})^2, \tag{16}$$

so that even if $t_{k,i} = t_{l,i}$ (which is to say that there is no asynchronicity, so that the second term in Eq. 14 vanishes), there is no directly obtainable estimate of the quantity in Eq. 14. We shall now see how this type of issue can be remedied. We return to the problem of the Hayashi–Yoshida estimator in Sect. 5.2.

4 Locally constant volatility

The basic problem in estimating the quantity Eq. 14 is that there are not enough observations per parameter. So long as there are a few observations per parameter, one can use standard theory of sample cumulants (see, for example, Chapter 4 of McCullagh 1987) to find unbiased estimators. After that, at least under Assumption 2, consistency of the variance estimates will take care of itself. There is some overhead in setting up the approximation (Sect. 4.1), but the investment will pay off when we get to Sect. 4.2 and beyond.

4.1 Setup

We shall in the following show that it is valid, at least asymptotically, to consider approximate systems of the type (Eq. 1) were we take $\mu_t = 0$, and in addition suppose that there are times $0 = \tau_{N,0} < \tau_{N,1} < \cdots < \tau_{N,v_N}$ so that $\mathrm{d}\langle X^{(k)}, X^{(l)}\rangle_t$ is constant on $(\tau_{N,\iota-1}, \tau_{N,\iota}]$ for each ι. The requirement for this is that

$$\max_\iota \#\{t_{N,k,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}]\} = O(1) \quad \text{as } N \to \infty, \tag{17}$$

and

$$v_N = O(N). \tag{18}$$

Assumption 3 (*Structure of the quadratic variation process*) Set $\zeta_t^{(k,l)} = \mathrm{d}\langle X^{(k)}, X^{(l)}\rangle_t / \mathrm{d}t$ (this process exists under the model (Eq. 1) and the Kunita–Watanabe inequality). We assume that the matrix ζ_t is an Itô process. We shall further assume that the drift part of each ζ_t is absolutely continuous with locally bounded derivative, and that each $\langle \zeta^{(k,l)}, \zeta^{(m,n)}\rangle_t$ is continuously differentiable. We finally assume that ζ_t is locally bounded, and that if $\lambda_t^{(p)}$ is the smallest eigenvalue of ζ_t, then $\inf_t \lambda_t^{(p)} > 0$ a.s.

We are proposing to hold the characteristics of the process constant in small time periods. We first define our time periods, which have to be joint for all p coordinates of the process X.

Definition 1 (*A reference set of time points*) We shall let $\mathcal{G}_N$ be the ordered set which contains (but can be bigger that) all points $t_{N,k,j}$ and the $\tau_{N,\iota}$. Represent $\mathcal{G}_N = \{0 = \theta_{N,0} < \theta_{N,1} < \cdots < \theta_{N,w_N} = T\}$. We suppose for simplicity that there is a number M so that $\tau_{N,\iota} = \theta_{N,M\iota}$. This is without loss of generality in view of assumption Eq. 17, since we can add point to $\mathcal{G}_N$ until this is true (one can state results also without this assumption, but they look unnecessarily dreadful). Note that $w_N = M v_N$.

The purpose of the grid $\mathcal{G}_N$ is to have a maximal set of time points at which we have to worry about the difference between the true probability P and the approximate probability distribution we are about to construct. The grid $\mathcal{G}_N$ includes all observations times (for all components), and all the break points $\tau_{N,\iota}$ on which we hall base estimators. By constructing below the likelihood ratio between the two probabilities based on the values of the process at the times in $\mathcal{G}_N$, we shall capture the effect of the approximation to the extent required by our observations and our modeling. Thus, while the role of the τ's is to construct an approximation, and resulting estimators, the purpose of the θ's is to facilitate analysis.

We shall need the quadratic variation of the above set of time points: the "Asymptotic Quadratic Variation of Time" ("AQVT") $H(t)$ is defined by

$$H(t) = \lim_{N\to\infty} \frac{w_N}{T} \sum_{\theta_{N,j+1}\le t} (\theta_{N,j} - \theta_{N,j-1})^2, \tag{19}$$

provided the limit exists. From Eq. 3,

$$\max_{1\le i\le w_N} |\theta_{N,i} - \theta_{N,i-1}| = O(N^{-1}), \tag{20}$$

whence every subsequence has a further subsequence for which H exists. Also, when the limit exists, it is Lipschitz continuous. We shall be using the following assumption.

Assumption 4 (*Structure of the AQVT*) Assume that the AQVT H exists, and that $H'(t)^{-1}$ is integrable. Further suppose that there is a transformation $G : [0, T] \to [0, T]$ so that $\sum_i (G(\theta_{N,i+1}) - G(\theta_{N,i}) - (T/N))^2 = o(n^{-1})$. Note that in this case, $G'(t) = H'(t)^{-1}$.

The assumption is more restrictive than the one made in Mykland and Zhang (2006), but is still quite broad. An easy extension can make G random, covering, in particular, the construction in Sect. 5.3 (pp. 1505–1507) of Barndorff-Nielsen et al. (2008).

We finally define the covariance matrix on the grid $\mathcal{G}_N$:

$$\nu_{k,l,i} = \mathrm{Cov}_\omega\left(X^{(k)}_{\theta_{N,i}} - X^{(k)}_{\theta_{N,i-1}}, X^{(l)}_{\theta_{N,i}} - X^{(l)}_{\theta_{N,i-1}}\right). \tag{21}$$

To hold the characteristics of the process constant over the interval $(\tau_{N,\iota-1}, \tau_{N,\iota}]$, we set

$$\bar{\nu}_{k,l,\iota} = \frac{1}{M} \sum_{\theta_{N,j}\in(\tau_{N,\iota-1},\tau_{N,\iota}]} \nu_{k,l,j}. \tag{22}$$

The idea is now to approximate our actual probability P by one which we shall call P^N, for which $\left(X^{(k)}_{\theta_{N,j}} - X^{(k)}_{\theta_{N,j-1}}, k = 1, \ldots, p\right)$ are iid random vectors with mean zero and covariances $\bar{\nu}_{k,l,\iota}$ as $\theta_{N,j}$ ranges over $(\tau_{N,\iota-1}, \tau_{N,\iota}]$ (for fixed (N, ι)). In general, of course, there are missing values from these observations, but this is not always a problem, as we shall see.

It will be useful at this point to use matrix notation. Write ν_j for the $p \times p$ matrix $(\nu_{k,l,j})$ and similarly with $\bar{\nu}_\iota$ (the $p \times p$ matrix $(\bar{\nu}_{k,l,\iota})$) and ΔX_{θ_j} (the $p \times 1$ vector $X^{(k)}_{\theta_j} - X^{(k)}_{\theta_{j-1}}$). The log likelihood ratio between the two measures (on the σ-field generated by the processes $\langle X^{(k)}, X^{(l)}\rangle_t$ and the random variables $X^{(k)}_{\theta_j}$) becomes

$$\log \frac{\mathrm{d}P}{\mathrm{d}P^N} = \frac{1}{2} \sum_\iota \sum_{\theta_{N,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}]} \Bigg(\log\det(\bar{\nu}_\iota) - \log\det(\nu_j) - \Delta X^*_{\theta_j} (\nu_j^{-1} - \bar{\nu}_\iota^{-1}) \Delta X_{\theta_j} \Bigg). \tag{23}$$

Note that

$$\log \frac{\mathrm{d}P}{\mathrm{d}P^N} = \log \frac{\mathrm{d}P_\omega}{\mathrm{d}P^N_\omega}. \tag{24}$$

4.2 Main contiguity theorem

We obtain the following main result, which is proved in Sect. 6.

Theorem 1 (Contiguity of P^N and P) *Suppose that Assumptions 1–4 are satisfied. Set*

$$Z_N = -\frac{1}{2} \sum_\iota \sum_{\theta_{N,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}]} \left(\Delta X^*_{\theta_j} \left(\nu_j^{-1} - \bar{\nu}_\iota^{-1} \right) \Delta X_{\theta_j} \right). \tag{25}$$

Let $\gamma_t = \int_0^t \zeta_t^{-1/2}(\,d\zeta_t)\zeta_t^{-1/2}$, where $\zeta_t^{-1/2}$ us the symmetric square root of ζ_t^{-1}. Define

$$\Gamma_{ZZ} = \frac{1}{12}(M-1) \int_0^T H'(t)^2 \sum_{k=1}^p \langle \gamma^{(k,k)}, \gamma^{(k,k)} \rangle'_t \mathrm{d}t, \tag{26}$$

Then (for almost all ω), as $N \to \infty$, Z_N converges in law under P^N_ω to a normal distribution with mean $-\Gamma_{ZZ}$ and variance Γ_{ZZ}. Also, under P^N,

$$\log \frac{\mathrm{d}P}{\mathrm{d}P^N} = Z_N + \frac{1}{2}\Gamma_{ZZ} + o_p(1). \tag{27}$$

The theorem says that P_ω and the approximation P_ω^N are *contiguous* in the sense of, for example Chapter IV Hájek and Sidak (1967) and Chapter VI of Jacod and Shiryaev (2003). This is because it follows from the theorem that $\mathrm{d}P_\omega/\mathrm{d}P_\omega^N$ is uniformly integrable under P_ω^N.

In particular, if an estimator is consistent under P^N, it is also consistent under P. In other words, one can, for purposes of consistency, assume that $(X^{(k)}_{\theta_{N,j}} - X^{(k)}_{\theta_{N,j-1}}, k = 1, \ldots, p)$ are iid random vectors with mean zero and covariances $\bar{\upsilon}_{k,l,\iota}$ as $\theta_{N,j}$ ranges over $(\tau_{N,\iota-1}, \tau_{N,\iota}]$ (for fixed (N, ι)).

Rates of convergence (typically $n^{1/2}$) are also preserved, but the asymptotic distribution may be biased. One fairly general result is as follows.

Theorem 2 *Assume the conditions of Theorem 1. Let $\xi_{N,\iota}$ be a function of $(\xi_t, \tau_{N,\iota-1} \le t \le \tau_{N,\iota})$. Let $\hat{\xi}_{N,\iota}$ be an estimator based on $(X^{(k)}_{t_{k,j}} - X^{(k)}_{t_{k,j-1}}, \tau_{N,\iota-1} \le t_{k,j-1}$ and $t_{k,j} \le \tau_{N,\iota}, k = 1, \ldots, p)$ (that is to say, that $\hat{\xi}_{N,\iota}$ is based on the actually observable increments in the time interval $[\tau_{N,\iota-1}, \tau_{N,\iota}]$). Set $\xi_N = \sum_\iota \xi_{N,\iota}$ (this quantity would normally be almost independent of N) and $\hat{\xi}_N = \sum_\iota \hat{\xi}_{N,\iota}$. Assume that $E^N_\omega(\hat{\xi}_{N,\iota}) = \xi_{N,\iota}$ (P^N_ω-unbiasedness), and assume the existence of the limits*

$$\Gamma_{\xi\xi} = \lim_{N\to\infty} \mathrm{Var}^N_\omega(N^{1/2}(\hat{\xi}_N - \xi_N)) \quad \text{and} \quad \Gamma_{\xi Z} = \lim_{N\to\infty} \mathrm{Cov}^N_\omega(N^{1/2}(\hat{\xi}_N - \xi_N), Z_N). \tag{28}$$

Also assume the Lindeberg Condition: for every $\epsilon > 0$,

$$\lim_{N\to\infty} E^N_\omega \sum_\iota g(N^{1/2}(\hat{\xi}_{N,\iota} - \xi_{N,\iota})) = 0, \tag{29}$$

where $g(x) = x^2 I_{\{|x|>\epsilon\}}$. Then under P_ω,

$$N^{1/2}(\hat{\xi}_N - \xi_N) \to N(\Gamma_{\xi Z}, \Gamma_{\xi\xi}) \tag{30}$$

in law, for almost every ω.

The result follows directly from Theorem 1 in view of Lindeberg's Central Limit Theorem (see, for example, Theorem 27.2 (pp. 359–360) of Billingsley 1995) and LeCam's Third Lemma (see the lemma on p. 208 in Hájek and Sidak 1967). Note that by conditional independence,

$$\Gamma_{\xi\xi} = \lim_{N\to\infty} N \sum_{\iota=1}^{\upsilon_N} \mathrm{Var}_{N\omega}(\hat{\xi}_{N,\iota}), \tag{31}$$

and similarly for $\Gamma_{\xi Z}$.

In other words, the price for using P^N rather than P is to incur a bias $N^{-1/2}\Gamma_{\xi Z} + o(N^{-1/2})$. In our examples, it will be the case that $\Gamma_{\xi Z} = 0$ (note that there is also a possibility that ξ_N is not exactly the quantity desired in estimation. This difference is typically negligible, as we shall see in our examples).

The main obstacle to using Theorem 2 may be to calculate the Γ's, but we shall see in the next Section how this can be implemented.

Remark 2 (*Role of the Itô assumption*) It should be clear from the above that it would be a radical oversimplification to assume that the process ζ_t is, say, continuously differentiable. If one does that, then $\Gamma_{ZZ} = \Gamma_{\xi Z} = 0$, and some of the predictive value of the above theorems would be lost.

Finally, it is worth asserting the contiguity even in the case where the $\tau_{N,\iota}$ are less regular than assumed in Theorem 1.

Theorem 3 (Contiguity of P^N and P in irregular cases) *Suppose that Assumptions 1–4 are satisfied. Set* $M_{N,\iota} = \#\{\theta_{N,k,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}]\}$, *and suppose that* $\max_\iota M_\iota = O(1)$. *Replace* M *by* M_ι *in Eq. 22. Then* dP^N/dP *is uniformly integrable. In particular, any sequence which converges in probability under* P^N *also converges in probability under* P.

4.3 When the drift μ is not zero

Assumption 5 (*A conditionally Gaussian system, with drift*) We modify Assumption 2 to let μ_t be nonzero. We let Q be a probability distribution on the form Eq. 1, where μ_t is locally bounded. We now take μ_t and quadratic variations and covariations $\langle X^{(k)}, X^{(l)}\rangle_t$, to be $\mathcal{F}_0$-measurable, so the system remains conditionally Gaussian. The process μ_t and quadratic variations and covariations $\langle X^{(k)}, X^{(l)}\rangle_t$ must also be adapted to a filtration for which the $W_t^{(k)}$ are Brownian motions.

We first consider consistency. Under Assumptions 3 and 5, Girsanov's Theorem (see, for example, Chapter 5.5 of Karatzas and Shreve 1991) yields that there is a probability measure P satisfying Assumption 2 so that for almost all ω

$$Q_\omega \text{ and } P_\omega \text{ are mutually absolutely continuous.} \tag{32}$$

This means that consistency holds under Q_ω if and only if it holds under P_ω. Thus, for instance, the final sentence in Theorem 3 remains true also under Q_ω (one still verifies the conditions under P and P^N).

To discuss this issue on the level of asymptotic distributions, we need the concept of stable convergence, which we here adapt to contiguous sequences

Definition 2 Let P be a probability measure on a σ-field $\mathcal{X}$, and let P^N be a sequence of probabilities equivalent and contiguous to P, so that for all sets $A \in \mathcal{X}$, $P^N(A) \to P^\infty(A)$. If Z_N is a sequence of $\mathcal{X}$-measurable random variables, then Z_N converges stably in law to Z as $N \to \infty$ under probability measures P^N if there is an extension of $\mathcal{X}$ so that, for all $A \in \mathcal{X}$ and for all bounded continuous g, $E^N I_A g(Z_n) \to E^\infty I_A g(Z)$ as $N \to \infty$.

For further discussion of stable convergence, see Rényi (1963), Aldous and Eagleson (1978), Chapter 3 (p. 56) of Hall and Heyde (1980), Rootzén (1980), Sect. 2 (pp. 169–170) of Jacod and Protter (1998), and Chapter IX.6–7 (pp. 575–591) of Jacod and Shiryaev (2003). It is a useful device in operationalizing asymptotic conditionality, and we have earlier used it in Mykland and Zhang (2006), Zhang et al. (2005), and other papers. In the current context, it will permit us to deal with the drift μ.

First, we generalize the above Theorems 1–2. Let $\mathcal{X}$ be the σ-field generated by the process X_t, $0 \leq t \leq T$ (for example, one can consider the Borel σ-field on the space of p-dimensional càdlàg functions, as in Chapter VI of Jacod and Shiryaev 2003). On can consider P_ω as a measure on $\mathcal{X}$, while P itself is a measure on the product σ-field $\mathcal{F}_0 \times \mathcal{X}$.

Using the methodology from the cited papers. it is easy to see from our proofs of the theorems in this paper that the convergence is, in fact, stable, under a side condition:

$$\lim_{N\to\infty} \mathrm{Cov}_\omega^N(N^{1/2}(\hat{\xi}_N - \xi_N), W_t^{(k)}) = 0 \text{ for all } t \in [0, T] \text{ and for } k = 1, \ldots, p. \tag{33}$$

Specifically, we obtain

Theorem 4 *The convergence in Theorem 1 is stable under P_ω^N, for (almost) every ω. Similarly, assume the conditions of Theorem 2, and also Eq. 33. Then the convergence in this theorem is stable under P_ω^N, for (almost) every ω.*

Proof of Theorem 1 The two parts of the proof are similar since the main fact we use about Z_N is that

$$\lim_{N\to\infty} \mathrm{Cov}_\omega^N(Z_N, W_t^{(k)}) = 0 \text{ for all } t \in [0, T] \text{ and for } k = 1, \ldots, p. \tag{34}$$

To see the first part of the theorem, consider first the reference filtration $\mathcal{F}_t' = \mathcal{F}_0 \vee \sigma\{W_s^{(k)}, 0 \leq s \leq t, \text{ all } k\}$. Let Z_t^N be the interpolated partial sum (for $t_j \leq t$) in Eq. 25. The interpolation is analogous to that used in the proof of Proposition 2 (p. 1952) of Mykland and Zhang (2006). From the proof of Theorem 1, one obtains that the conditions of Theorem IX.7.3 (p. 584) of Jacod and Shiryaev (2003) are satisfied, where their martingale Z is our (p-dimensional) martingale W. In their condition 7.4, $G_t = 0$ identically, from Eq. 34. Their condition 7.5 is satisfied since there are (in their notation) no martingales N. This is by the Martingale Representation Theorem (see, e.g., Theorem III.4.33 (p. 189) of Jacod and Shiryaev 2003) by definition the filtration $(\mathcal{F}_t')$. This shows the stable convergence with respect to the filtration $(\mathcal{F}_t')$, and the result for filtration $(\mathcal{F}_t)$ follows by conditional independence. The second part of Theorem 4 is shown similarly. □

From Theorem 4 it follows directly that

Theorem 5 *Assume the conditions of Theorem 2, but replace Assumption 2 by Assumption 5. Also assume Eq. 33.*

Then the (stable) convergence in law (Eq. 30) holds also under Q_ω, for almost every ω.

The proof here is similar to that of Proposition 1 in Mykland and Zhang (2009).

Remark 3 In the more general case where the limit in Eq. 33 can be nonzero, the stable convergence in Theorem 2 will be to a normal limit with smaller variance but extra bias, subject to suitable regularity conditions. We have not further discussed this case since such estimators are clearly inefficient.

5 Some applications of the theory

We here present two applications of the theory, as proof of principle. A number of other applications are presented in other papers, see, in particular, Sects. 4.1–4.2 of Mykland and Zhang (2009), and Sects. 6.2–6.4 of Mykland and Zhang (2010). The methodology is also cited as the conceptual background for the quantile based estimation of Christensen et al. (2008). In many cases, one has a choice of whether to invoke the theory from this paper or the one from Mykland and Zhang (2009). The latter is more general, but harder to apply, and with less transparent conditions. We note that a main purpose of the theory is to gain efficiency, see, in particular, the example in Sect. 4.1 of Mykland and Zhang (2009) (where one could equally well have used the technology of the current paper).

5.1 ANOVA with multiple regression and finite smoothing

We here revisit the problem from Zhang (2001) and Mykland and Zhang (2006). There are processes $X_t^{(1)}, \ldots, X_t^{(p)}$ and Y_t which are observed synchronously at times $0 = t_{N,0} < t_{N,1} < \cdots < t_{N,n_1} = T$ (the asynchronous problem is also interesting, but beyond what we are planning to do in this paper). In this treatment, we shall take the times to be equidistant, so $\Delta t = T/n_1$, but this can of course be generalized. The two processes are related by

$$dY_t = \sum_{i=1}^{p} f_s^{(i)}\, dX_s^{(i)} + dZ_t, \quad \text{with } \langle X^{(i)}, Z\rangle_t = 0 \text{ for all } t \text{ and } i. \tag{35}$$

The problem is now to estimate $\langle Z, Z\rangle_T$, that is to say the residual quadratic variation of Y after regressing on X. As documented in Zhang (2001) and Mykland and Zhang (2006), this is useful for statistical and trading purposes.

We take the θ_i's to be identical with the t_i's. We assume that $M > p$. For simplicity of argument, we also assume that n_1 is a multiple of M (this does not really affect the conclusions).

We now describe the system under P^N. In each of υ_N intervals $[\tau_{\iota-1}, \tau_\iota]$ we get M iid observations of $(\Delta X_{t_j}, \Delta Y_{t_j})$, which are normal with mean zero and covariance matrix $\bar{\nu}_\iota$. For $t_j \in (\tau_{\iota-1}, \tau_\iota]$, in obvious notation,

$$\bar{\nu}_{ZZ,\iota} = E_\omega^N \Delta Z_{t_j}^2 = \bar{\nu}_{YY,\iota} - \bar{\nu}_{XY,\iota}^* \bar{\nu}_{XX,\iota}^{-1} \bar{\nu}_{XY,\iota} \tag{36}$$

The strategy is now as follows: in each time interval, regress ΔY_{t_j} on ΔX_{t_j} linearly, and without intercept. Call the residuals $\widehat{\Delta Z}_{t_j}$. Set

$$\hat{\xi}_{\iota,N} = \frac{M}{M-p} \sum_{t_j \in (\tau_{\iota-1},\tau_\iota]} \widehat{\Delta Z}^2_{t_j}$$
$$\text{and} \quad \xi_{\iota,N} = E^N_\omega \left(\sum_{t_j \in (\tau_{\iota-1},\tau_\iota]} \Delta Z^2_{t_j} \right) \tag{37}$$

Standard regression theory (see, for example, Weisberg 1985) yields that, because of the Gaussianity

$$E^N_\omega \left(\hat{\xi}_{\iota,N} \mid \Delta X_{t_j}, \text{ all } j = 1, \ldots, n_1 \right) = \xi_{\iota,N}, \tag{38}$$

and, in fact, more generally,

$$(M-p)\frac{\hat{\xi}_{\iota,N}}{\xi_{\iota,N}}, \iota = 1, \ldots, v_N \text{ are iid } \chi^2_{M-p} \text{ under } P^N_\omega \text{ given all } \Delta X_{t_j}, j=1, \ldots, n_1. \tag{39}$$

A natural estimator for $\langle Z, Z\rangle_T$ is therefore

$$\widehat{\langle Z, Z\rangle}_T = \frac{M}{M-p} \sum_{j=1}^{n_1} \widehat{\Delta Z}^2_{t_j} = \sum_{\iota=1}^{v_N} \hat{\xi}_{\iota,N}. \tag{40}$$

From Eqs. 36 and 38, and by Assumption 3,

$$\begin{aligned}
E^N_\omega(\widehat{\langle Z, Z\rangle}_T) &= E^N_\omega \left(\sum_{j=1}^{n_1} \Delta Z^2_{t_j} \right) \\
&= \int_0^T \zeta^{(YY)}_t \mathrm{d}t - \sum_{\iota=1}^{v_N} \left(\int_{\tau_{\iota-1}}^{\tau_\iota} \zeta^{(XY)}_t \mathrm{d}t \right)^* \left(\int_{\tau_{\iota-1}}^{\tau_\iota} \zeta^{(XX)}_t \mathrm{d}t \right)^{-1} \left(\int_{\tau_{\iota-1}}^{\tau_\iota} \zeta^{(XY)}_t \mathrm{d}t \right) \\
&= \int_0^T \zeta^{(YY)}_t \mathrm{d}t - \frac{T}{v_N} \sum_{\iota=1}^{v_N} \left(\zeta^{(XY)}_{\tau_{\iota-1}} + \frac{v_N}{T} \int_{\tau_{\iota-1}}^{\tau_\iota} (\tau_\iota - t)\, \mathrm{d}\zeta^{(XY)}_t \right)^* \\
&\quad \times \left(\left(\zeta^{(XX)}_{\tau_{\iota-1}} \right)^{-1} - \frac{v_N}{T} \left(\zeta^{(XY)}_{\tau_{\iota-1}} \right)^* \int_{\tau_{\iota-1}}^{\tau_\iota} (\tau_\iota - t)\, \mathrm{d}\zeta^{(XX)}_t \zeta^{(XY)}_{\tau_{\iota-1}} \right) \\
&\quad \times \left(\zeta^{(XY)}_{\tau_{\iota-1}} + \frac{v_N}{T} \int_{\tau_{\iota-1}}^{\tau_\iota} (\tau_\iota - t)\, \mathrm{d}\zeta^{(XY)}_t \right) + o_p(N^{-1/2}) \\
&= \langle Z, Z\rangle_T + o_p(N^{-1/2}).
\end{aligned} \tag{41}$$

Similarly, from Eqs. 38–39,

$$\mathrm{Var}_{\omega}^{N}(\widehat{\langle Z, Z\rangle_T}) = \frac{2}{M-p}\sum_{\iota=1}^{v_N}\xi_{\iota,N}^2$$
$$= n_1^{-1}\, 2\frac{M}{M-p}\int_0^T \left(\langle Z, Z\rangle_t'\right)^2 \mathrm{d}t + o_p\left(n_1^{-1}\right) \tag{42}$$

In seeking to apply Theorem 2, let $\mathcal{X}$ be the $p\times M$-matrix whose jth row is made up of $\left(\Delta X^{(1)}_{t_{M\iota+j}}, \ldots, \Delta X^{(p)}_{t_{M\iota+j}}\right)$. Set $H = \mathcal{X}(\mathcal{X}^*\mathcal{X})^{-1}\mathcal{X}^*$, and note that this is the standard "hat matrix" for the regression in time period # ι (see Chapter 5.1 of Weisberg 1985 for this and for the following manipulations). Call the jth diagonal element of H as $h_{M\iota+j}$. Finally set $\Delta Z = (\Delta Z_{t_{M\iota+1}}, \ldots, \Delta Z_{t_{M(\iota+1)}})^*$. Note that our regression means that $\widehat{\Delta Z} = (I-H)\Delta Z$, and so $\hat{\xi}_{N,i} = (M/(M-p))\Delta Z^*(I-H)\Delta Z$.

To match notation with Theorem 2, finally let $X^{(p+1)} = Y$. Obtain from Eqs. 38–39 and the normality of the observations that

$$\mathrm{Cov}_{\omega}^{N}\left(\sum_{t_{N,j}\in(\tau_{N,\iota-1},\tau_{N,\iota}]}\Delta X^*_{t_j}\left(\nu_j^{-1}-\bar{\nu}_\iota^{-1}\right)\Delta X_{t_j}, \hat{\xi}_{N,\iota}\mid\mathcal{X}\right)$$
$$= \mathrm{Cov}_{\omega}^{N}\left(\sum_{t_{N,j}\in(\tau_{N,\iota-1},\tau_{N,\iota}]}\Delta Z_{t_j}\left(\nu_j^{-1}-\bar{\nu}_\iota^{-1}\right)^{(YY)}\Delta Z_{t_j}, \hat{\xi}_{N,\iota}\mid\mathcal{X}\right)$$
$$= \frac{M}{M-p}\sum_{t_{N,j}\in(\tau_{N,\iota-1},\tau_{N,\iota}]}\mathrm{Cov}_{\omega}^{N}\left(\left(\nu_j^{-1}-\bar{\nu}_\iota^{-1}\right)^{(YY)}\Delta Z^2_{t_j}, (1-h_j)\Delta Z^2_{t_j}\right)$$
$$= 2\frac{M}{M-p}\bar{\nu}^2_{ZZ,\iota}\sum_{t_{N,j}\in(\tau_{N,\iota-1},\tau_{N,\iota}]}\left(\nu_j^{-1}-\bar{\nu}_\iota^{-1}\right)^{(YY)}(1-h_j). \tag{43}$$

At this point, note that by symmetry, $E_{\omega}^{N}(1-h_j)$ is independent of j, and so $E_{\omega}^{N}(1-h_j) = E_{\omega}^{N}\mathrm{tr}(I-H)/M = (M-p)/M$. Thus, from Eqs. 38 and 43,

$$\mathrm{Cov}_{\omega}^{N}\left(\sum_{t_{N,j}\in(\tau_{N,\iota-1},\tau_{N,\iota}]}\Delta X^*_{t_j}\left(\nu_j^{-1}-\bar{\nu}_\iota^{-1}\right)\Delta X_{t_j}, \hat{\xi}_{N,\iota}\right)$$
$$= 2\bar{\nu}^2_{ZZ,\iota}\sum_{t_{N,j}\in(\tau_{N,\iota-1},\tau_{N,\iota}]}\left(\nu_j^{-1}-\bar{\nu}_\iota^{-1}\right)^{(YY)}. \tag{44}$$

Hence, in the notation of Theorem 2, we obtain that

$$\mathrm{Cov}_{\omega}^{N}(\hat{\xi}_N, Z_N) = 2\sum_{\iota}\bar{\nu}^2_{ZZ,\iota}\sum_{t_{N,j}\in(\tau_{N,\iota-1},\tau_{N,\iota}]}\left(\nu_j^{-1}-\bar{\nu}_\iota^{-1}\right)^{(YY)}$$
$$= o_p(N^{-1/2}). \tag{45}$$

Finally, the Lindeberg condition in Theorem 2 is satisfied from Eq. 39. We have therefore shown the following corollary to this theorem:

Theorem 6 (Multiple ANOVA in the Gaussian case) *Let* $\widehat{\langle Z, Z\rangle}_T$ *be as defined in Eq.* 40. *Under the conditions of Theorem* 1,

$$n_1^{1/2}(\widehat{\langle Z, Z\rangle}_T - \langle Z, Z\rangle_T) \tag{46}$$

converges in law under P_ω *to a normal distribution with mean zero and variance*

$$2\frac{M}{M-p}\int_0^T (\langle Z, Z\rangle_t')^2 \mathrm{d}t. \tag{47}$$

Compared to the results of Zhang (2001) and Mykland and Zhang (2006), the difference in method is that M is here finite and fixed, while in the earlier paper, $M \to \infty$ with n_1. In terms of results, there is here no asymptotic bias (whereas this is present, though correctable from the data, in the earlier work). On the other hand, the current estimator is not quite efficient, as the asymptotic variance in Zhang (2001) and Mykland and Zhang (2006) is

$$2\int_0^T (\langle Z, Z\rangle_t')^2 \mathrm{d}t \tag{48}$$

(as in the single series case in Jacod and Protter 1998 and Barndorff-Nielsen and Shephard 2002). Of course, the expression in Eq. 47 converges to that of Eq. 48 as $M \to \infty$. Comparing the two sets of results, one is lead to conjecture that there is a bias-variance tradeoff where M should go very slowly to infinity n_1, but exploring that is beyond the scope of this paper.

5.2 Estimating the variability in the Hayashi–Yoshida estimator

We now return to Example 2 from Sect. 3. We consider the variance Eq. 12. Fix k and l ($k \neq l$), and recall that

$$\begin{aligned}\mathrm{Var}_\omega\left(\widehat{\left\langle X^{(k)}, X^{(l)}\right\rangle}_T\right) &= \sum_{i,j:(t_{k,i-1},t_{k,i})\cap(t_{l,j-1},t_{l,j})\neq\emptyset} (\kappa_{k,i;k,i}\kappa_{l,j;l,j} + (\kappa_{k,i;l,j})^2) \\ &\quad + \sum_{(i,j,g,h)\in B_n} \kappa_{k,i;l,h}\kappa_{k,g;l,j},\end{aligned} \tag{49}$$

where B_n is given by Eq. 15. We are seeking to consistently estimate this variance.

We here provide a moment based estimate in the style of Barndorff-Nielsen and Shephard (2002). In this section, we only consider consistency of the estimate, though, of course, the same principles as before can be used to find asymptotic normality.

To follow up on Eq. 16, note this equality also holds in the asynchronous case. Also note that

$$E_\omega(\hat{\kappa}_{k,i;l,h}\hat{\kappa}_{k,g;l,j}) = \kappa_{k,i;l,h}\kappa_{k,g;l,j} + \kappa_{k,i;l,j}\kappa_{k,g;l,h} + \kappa_{k,i;k,g}\kappa_{l,h;l,j} \tag{50}$$

Thus, if we set

$$\begin{aligned}\hat{V}_1 &= \sum_{i,j:(t_{k,i-1},t_{k,i})\cap(t_{l,j-1},t_{l,j})\neq\emptyset} \hat{\kappa}_{k,i;k,i}\hat{\kappa}_{l,j;l,j} \\ \hat{V}_2 &= \sum_{(i,j,g,h)\in B_n} \hat{\kappa}_{k,i;l,h}\hat{\kappa}_{k,g;l,j},\end{aligned} \tag{51}$$

we obtain

$$\begin{aligned}V_1 &= E_\omega \hat{V}_1 = \sum_{i,j:(t_{k,i-1},t_{k,i})\cap(t_{l,j-1},t_{l,j})\neq\emptyset} (\kappa_{k,i;k,i}\kappa_{l,j;l,j} + 2(\kappa_{k,i;l,j})^2) \\ V_2 &= E_\omega \hat{V}_2 = 2\sum_{(i,j,g,h)\in B_n} \kappa_{k,i;l,h}\kappa_{k,g;l,j}.\end{aligned} \tag{52}$$

Obviously, $\hat{V}_1$ and $\hat{V}_2$ are consistent estimators of their expectations (note that the precise technical statement here is that $N(\hat{V}_i - V_i) \to 0$ in probability as $N \to \infty$).

Since $\mathrm{Var}_\omega\left(\widehat{\langle X^{(k)}, X^{(l)}\rangle}_T\right)$ is not a linear combination of V_1 and V_2, we need to find a third estimator to make up the difference. There would seem to be many ways of accomplishing this, but the following struck us as appealing.

Note that it is sufficient to obtain a consistent estimate for

$$V_3 = \sum_{i,j:(t_{k,i-1},t_{k,i})\cap(t_{l,j-1},t_{l,j})\neq\emptyset} (\kappa_{k,i;l,j})^2 \tag{53}$$

We shall use the theory from the previous section and find a consistent estimator under a suitable P^N.

Assume that a grid $\mathcal{G}_N = \{\theta_{N,j}\}$ is given as in the previous section. Define as follows:

$$\begin{aligned}D_{k,\iota} &= \{i : (t_{k,i-1}, t_{k,i}) \subseteq (\tau_{\iota-1}, \tau_\iota)\} \\ D_{k,l,\iota} &= \{(i,j) : (t_{k,i-1}, t_{k,i}) \cup (t_{l,j-1}, t_{l,j}) \subseteq (\tau_{\iota-1}, \tau_\iota) \\ &\quad \text{and } (t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset\} \\ M_{k,i} &= \text{No. of intervals } (\theta_{N,h-1}, \theta_{N,h}) \subseteq (t_{k,i-1}, t_{k,i})) \\ M_{k,i;l,j} &= \text{No. of intervals } (\theta_{N,h-1}, \theta_{N,h}) \subseteq (t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \\ m_{k,\iota} &= \sum_{i\in D_{k,\iota}} M_{k,i} \\ m_{k,l,\iota} &= \sum_{(i,j)\in D_{k,l,\iota}} M_{k,i;l,j}.\end{aligned} \tag{54}$$

Now note that under the approximate measure P^N

$$
\begin{aligned}
E^N_\omega & \sum_{(i,j)\in(D_{k,\iota}\times D_{l,\iota}-D_{k,l,\iota})} \hat{\kappa}_{k,i;k,i}\hat{\kappa}_{l,j;l,j} \\
&= \bar{v}^2_{k,l,\iota} \sum_{(i,j)\in(D_{k,\iota}\times D_{l,\iota}-D_{k,l,\iota})} M(k,i;k,i)M(l,j;l,j) \\
&= \bar{v}^2_{k,l,\iota}(m_{k,\iota}m_{l,\iota}-m_{k,l,\iota}). \qquad (55)
\end{aligned}
$$

On the other hand, set

$$
\begin{aligned}
\bar{D}_{k,l,\iota} &= \{(i,j) : t_{k,i-1}, t_{l,j-1} \in [\tau_{\iota-1}, \tau_\iota) \text{ and } (t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset\} \\
\bar{m}^{(2)}_{k,l,\iota} &= \sum_{(i,j)\in\bar{D}_{k,l,\iota}} M^2_{k,i;l,j}. \qquad (56)
\end{aligned}
$$

The difference between $D_{k,l,\iota}$ and $\bar{D}_{k,l,\iota}$ is that (up to a negligible set) the union of intervals $(t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j})$, $(i,j) \in \bar{D}_{k,l,\iota}$ covers $[0,T]$, while this is normally not the case for $D_{k,l,\iota}$. Note that $D_{k,l,\iota} \subseteq \bar{D}_{k,l,\iota}$. By continuity (Assumption 3),

$$
V_3 = \sum_\iota \bar{v}^2_{k,l,\iota}\bar{m}^{(2)}_{k,l,\iota} + o(N^{-1}). \qquad (57)
$$

It therefore follows from the theory in the previous section that if we set

$$
\hat{V}_3 = \sum_\iota \frac{\bar{m}^{(2)}_{k,l,\iota}}{m_{k,\iota}m_{l,\iota}-m_{k,l,\iota}} \sum_{(i,j)\in(D_{k,\iota}\times D_{l,\iota}-D_{k,l,\iota})} \hat{\kappa}_{k,i;k,i}\hat{\kappa}_{l,j;l,j}, \qquad (58)
$$

then $\hat{V}_3 - V_3 = o_P(N^{-1})$.

Take the grid $\mathcal{G}_N = \{\theta_{N,j}\}$ to consist of all points of the form $t_{k,i}$ or $t_{l,j}$. Thus $w_N \le n_k + n_l$. Fix M as a sufficiently large integer (more about the exact condition below). We let τ_1 be one of the first M points of $\mathcal{G}_N$, and then let the following τ_ι be every Mth point in $\mathcal{G}_N$ (we can without impacting the asymptotics ignore the first and last interval, which may not have M points).

One can therefore, finally, use the estimate

$$
\widehat{\mathrm{Var}_\omega}\left(\widehat{\langle X^{(k)}, X^{(l)}\rangle}_T\right) = \hat{V}_1 + \frac{1}{2}\hat{V}_2 - \hat{V}_3. \qquad (59)
$$

As an example of grid, one can let $\mathcal{G}_N = \{\theta_{N,j}\}$ consist of all points of the form $t_{k,i}$ or $t_{l,j}$. Thus $w_N \le n_k + n_l$. Fix M as a sufficiently large integer. We let τ_1 be one of the first M points of $\mathcal{G}_N$, and then let the following τ_ι be every Mth point in $\mathcal{G}_N$ (we can without impacting the asymptotics ignore the first and last interval, which may not have M points). In line with general principles of sufficiency, one should, of course, average $\hat{V}_3$ over the M choices of initial point $\tau_{N,1}$. This does not, of course, affect consistency.

6 Proof of Theorem 1

Proof of Theorem 1 Set $U_j = \bar{\nu}_\iota^{1/2} \nu_j^{-1} \bar{\nu}_\iota^{1/2} - I$, where $\bar{\nu}_\iota^{1/2}$ is the symmetric square root of $\bar{\nu}_\iota$. By standard Gaussian arguments,

$$\begin{aligned} E^N_\omega(\Delta X^*_{\theta_j}\left(\nu_j^{-1} - \bar{\nu}_\iota^{-1}\right)\Delta X_{\theta_j}) &= \mathrm{tr}(U_j) \\ \text{and} \quad \mathrm{Var}^N_\omega(\Delta X^*_{\theta_j}\left(\nu_j^{-1} - \bar{\nu}_\iota^{-1}\right)\Delta X_{\theta_j}) &= 2\mathrm{tr}\left(U_j^2\right). \end{aligned} \tag{60}$$

Set $V_j = \nu_j - \bar{\nu}_\iota$, and note that,

$$\mathrm{tr}\left(U_j^r\right) = \mathrm{tr}\left(\left(-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^r\right) + r\,\mathrm{tr}\left(\left(-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^{r+1}\right) + \cdots. \tag{61}$$

For $r \geq 2$,

$$\left|\mathrm{tr}\left(\left(\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^r\right)\right| \leq \left|\mathrm{tr}\left(\left(\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^2\right)\right|^{r/2} \tag{62}$$

(this is true for any symmetric matrix). Now note that

$$\mathrm{tr}\left(\left(-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^2\right) = \sum_{r,s}\left(\left(-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)_{rs}\right)^2,$$

(where A_{rs} denotes component (r, s) of matrix A). Using Lemma 2 (given below), and since $\bar{\nu}_\iota$ sufficiently approximated by $\zeta_{\tau_\iota}(\tau_{\iota+1} - \tau_\iota)/M$ (by continuity), we obtain that, under P^N,

$$\sum_j \mathrm{tr}\left(\left(-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^2\right) = 2\Gamma_{ZZ} + o_p(1). \tag{63}$$

Meanwhile, from Eq. 62, we obtain that for $r > 2$,

$$\begin{aligned} \left|\sum_j \mathrm{tr}\left(\left(-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^r\right)\right| \leq \sup_j \left|\mathrm{tr}\left(\left(-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^2\right)\right|^{(r-2)/2} \\ \times \sum_j \mathrm{tr}\left(\left(-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^2\right). \end{aligned}$$

Thus Eq. 61 yields that

$$\mathrm{tr}\left(U_j^2\right) = \mathrm{tr}\left(\left(-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^2\right) \times \left(1 + f\left(\sup_j \left|\mathrm{tr}\left(\left(-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2}\right)^2\right)\right|\right)\right)$$

where $f(x) = 2x^{1/2} + 3x + \cdots = o(1)$ as $x \to 0$. Since $\sup_j |\mathrm{tr}((-\bar{\nu}_\iota^{-1/2} V_j \bar{\nu}_\iota^{-1/2})^2)| = o_p(1)$, we obtain from Eq. 63 that

$$\sum_j \mathrm{tr}\left(U_j^2\right) = 2\Gamma_{ZZ} + o_p(1). \tag{64}$$

Also, since

$$\sum_{\theta_{N,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}]} \mathrm{tr}\left(\bar{\nu}_\iota^{-1} V_j\right) = 0 \tag{65}$$

by construction, we obtain

$$\begin{aligned} \sum_\iota \sum_{\theta_{N,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}]} \mathrm{tr}(U_j) &= \sum_j \mathrm{tr}\left(U_j^2\right) + o_p(1) \\ &= 2\Gamma_{ZZ} + o_p(1) \end{aligned} \tag{66}$$

By the Lindeberg's Central Limit Theorem (see, for example, Theorem 27.2 (pp. 359–360) of Billingsley 1995), the result for Z_N follows (the Lindeberg condition is similarly checked, using normality).

$$\begin{aligned} &\frac{1}{2} \sum_\iota \sum_{\theta_{N,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}]} \left(\log\det(\bar{\nu}_\iota) - \log\det(\nu_j)\right) \\ &= \frac{1}{2} \sum_j \log\det(I + U_j) \\ &= \frac{1}{2} \sum_j \left(\mathrm{tr}(U_j) - \frac{1}{2}\mathrm{tr}\left(U_j^2\right)\right) + o_p(1) \\ &= \frac{1}{2}\Gamma_{ZZ} + o_p(1). \end{aligned} \tag{67}$$

This shows the Theorem. □

It remains to prove

Lemma 2 *Assume the conditions of Theorem 1. Let f_t be a continuous process adapted to the filtration $\sigma\left(\langle X^{(k)}, X^{(l)}\rangle_s,\ 0 \le s \le t\right)$. Set $\zeta_t^{(k,l)} = \langle X^{(k)}, X^{(l)}\rangle_t'$. Then*

$$\begin{aligned} &\sum_\iota \sum_{\theta_{N,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}]} f_{\tau_{N,\iota-1}} (\nu_{k,l,j} - \bar{\nu}_{k,l,\iota})(\nu_{m,n,j} - \bar{\nu}_{m,n,\iota}) \\ &= \frac{1}{6} \frac{M-1}{M^2} T^2 \upsilon_N^{-2} \int_0^T f_t H'(t)^2 \mathrm{d}\langle \zeta^{(k,l)}, \zeta^{(m,n)}\rangle_t + o_p(N^{-2}). \end{aligned} \tag{68}$$

Proof of Lemma 2 Without loss of generality, by Girsanov's Theorem, we assume that $\zeta_t^{(k,l)}$ is a martingale.

First, suppose that

$$\sum_{j=1}^{w_N}\left(\theta_{N,j}-\theta_{N,j-1}-\frac{T}{w_N}\right)^2=o(N^{-1}). \tag{69}$$

Write

$$\begin{aligned}
\nu_{k,l,j}-\bar{\nu}_{k,l,\iota} &= \int_{\theta_{N,j-1}}^{\theta_{N,j}} \zeta_u^{(k,l)}\mathrm{d}u-\frac{1}{M}\int_{\tau_{N,\iota-1}}^{\tau_{N,\iota}} \zeta_u^{(k,l)}\mathrm{d}u\\
&= \int_{\theta_{N,j-1}}^{\theta_{N,j}} \left(\zeta_u^{(k,l)}-\zeta_{\theta_{N,j-1}}^{(k,l)}\right)\mathrm{d}u+(\theta_{N,j}-\theta_{N,j-1})\left(\zeta_{\theta_{N,j-1}}^{(k,l)}-\zeta_{\tau_{N,\iota-1}}^{(k,l)}\right)\\
&\quad -\frac{1}{M}\int_{\tau_{N,\iota-1}}^{\tau_{N,\iota}}\left(\zeta_u^{(k,l)}-\zeta_{\tau_{N,\iota-1}}^{(k,l)}\right)\mathrm{d}u\\
&\quad +\left((\theta_{N,j}-\theta_{N,j-1})-\frac{\tau_{N,\iota}-\tau_{N,\iota-1}}{M}\right)\zeta_{\tau_{N,\iota-1}}^{(k,l)}\\
&= \int_{\theta_{N,j-1}}^{\theta_{N,j}} (\theta_{N,j}-u)\,\mathrm{d}\zeta_u^{(k,l)}+(\theta_{N,j}-\theta_{N,j-1})\left(\zeta_{\theta_{N,j-1}}^{(k,l)}-\zeta_{\tau_{N,\iota-1}}^{(k,l)}\right)\\
&\quad -\frac{1}{M}\int_{\tau_{N,\iota-1}}^{\tau_{N,\iota}}(\tau_{N,\iota}-u)\,\mathrm{d}\zeta_u^{(k,l)}\\
&\quad +\left((\theta_{N,j}-\theta_{N,j-1})-\frac{\tau_{N,\iota}-\tau_{N,\iota-1}}{M}\right)\zeta_{\tau_{N,\iota-1}}^{(k,l)}\\
&= \int_{\tau_{N,\iota-1}}^{\tau_{N,\iota}} g_{N,j}(u)\,\mathrm{d}\zeta_u^{(k,l)}+\left((\theta_{N,j}-\theta_{N,j-1})-\frac{\tau_{N,\iota}-\tau_{N,\iota-1}}{M}\right)\zeta_{\tau_{N,\iota-1}}^{(k,l)}
\end{aligned} \tag{70}$$

where

$$\begin{aligned}
g_{N,j}(u) &= I_{\{\theta_{N,j-1}<u\le\theta_{N,j}\}}(\theta_{N,j}-u)+I_{\{\tau_{N,\iota-1}<u\le\theta_{N,j-1}\}}(\theta_{N,j}-\theta_{N,j-1})\\
&\quad -\frac{1}{M}I_{\{\tau_{N,\iota-1}<u\le\tau_{N,\iota}\}}(\tau_{N,\iota}-u).
\end{aligned}$$

The last term in Eq. 70 is ignorable in view of assumption Eq. 69. Thus,

$$\sum_{\iota} \sum_{\theta_{N,j}\in(\tau_{N,\iota-1},\tau_{N,\iota}]} f_{\tau_{N,\iota-1}}(\nu_{k,l,j}-\bar{\nu}_{k,l,\iota})(\nu_{m,n,j}-\bar{\nu}_{m,n,\iota})$$
$$= \sum_{\iota} f_{\tau_{N,\iota-1}} \int_{\tau_{N,\iota-1}}^{\tau_{N,\iota}} \sum_{\theta_{N,j}\in(\tau_{N,\iota-1},\tau_{N,\iota}]} g_{N,j}(u)^2 \, \mathrm{d}\left\langle \zeta^{(k,l)}, \zeta^{(m,n)} \right\rangle_u + o_p(N^{-2})$$
$$= c_M M^{-2} T^2 \upsilon_N^{-2} \sum_{\iota} f_{\tau_{N,\iota-1}} \left\langle \zeta^{(k,l)}, \zeta^{(m,n)} \right\rangle'_{\tau_{N,\iota-1}} (\tau_{N,\iota}-\tau_{N,\iota-1}) + o_p(N^{-2})$$
$$= c_M M^{-2} T^2 \upsilon_N^{-2} \int_0^T f_t \, \mathrm{d}\left\langle \zeta^{(k,l)}, \zeta^{(m,n)} \right\rangle_t + o_p(N^{-2}) \tag{71}$$

where

$$c_M = \left(\frac{2}{3}\frac{1}{M^3}\sum_{j=1}^{M-1} j^3 + \frac{1}{M^3}\sum_{j=1}^{M-1} j^2 - \frac{1}{6}\left(\frac{M-1}{M}\right)^2 \right) = \frac{1}{6}(M-1).$$

This shows the result under assumption Eq. 69 (note that under this assumption, $H'(t) = 1$ identically).

For the more general case, set

$$G(t) = \int_0^t \frac{1}{H'(s)} \, \mathrm{d}s. \tag{72}$$

Define $\tilde{\theta}_{N,j} = G(\theta_{N,j})$ $\tilde{\tau}_{N,j} = G(\tau_{N,j})$, $\tilde{T} = G(T)$, $\tilde{X}^{(k)}_{G(t)} = X^{(k)}_t$, and $\tilde{f}^{(k)}_{G(t)} = f^{(k)}_t$. If $\tilde{\zeta}^{(k,l)}_u = \langle \tilde{X}^{(k)}, \tilde{X}^{(l)} \rangle'_u$, then $\tilde{\zeta}^{(k,l)}_{G(t)} = \zeta^{(k,l)}_t H'(t)$. Since,

$$\int_0^{\tilde{T}} \tilde{f}_u \, \mathrm{d}\left\langle \tilde{\zeta}^{(k,l)}, \tilde{\zeta}^{(m,n)} \right\rangle_t = \int_0^T f_t H'(t)^2 \, \mathrm{d}\left\langle \zeta^{(k,l)} \zeta^{(m,n)} \right\rangle_t, \tag{73}$$

the result of the lemma follows. □

7 Conclusion

We have sought to demonstrate that the conditional Gaussian assumption is useful in finding estimators for high frequency financial data. We also show general asymptotic theorems for this case. In application, we develop new estimators for quarticity, for residual variance in ANOVA, and for the variance of an estimator of covariation. A particular feature of the methodology is that classical techniques for normal data can be used to propose and analyze estimators. We conjecture that the machinery can be used for a whole range of other problems with similar data.

This paper can be seen as a companion paper to Mykland and Zhang (2009). The setting in the current paper is more restrictive, but it is also possible to obtain sharper results with the Gaussian calculus.

Acknowledgments The author would like to thank Mark Podolskij, Eric Renault, and anonymous referees on an earlier version, who helped greatly to improve the paper. The support of the National Science Foundation under grants DMS 06-04758 and SES 06-31605 is gratefully acknowledged.

References

Aldous, D.J., Eagleson, G.K.: On mixing and stability of limit theorems. Ann Probab **6**, 325–331 (1978)

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: Great realizations. Risk **13**, 105–108 (2000)

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: The distribution of realized exchange rate volatility. J Am Stat Assoc **96**, 42–55 (2001)

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P.: Modeling and forecasting realized volatility. Econometrica **71**, 579–625 (2003)

Barndorff-Nielsen, O.E., Shephard, N.: Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J R Stat Soc B **64**, 253–280 (2002)

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

Billingsley, P.: Probability and Measure. 3rd edn. Wiley, New York (1995)

Chernov, M., Ghysels, E.: A study towards a unified approach to the joint estimation of objective and risk neutral measures for the purpose of options valuation. J Financ Econ **57**, 407–458 (2000)

Christensen, K., Oomen, R., Podolskij, M.: Realised quantile-based estimation of the integrated variance. Tech Rep (2008)

Comte, F., Renault, E.: Long memory in continuous-time stochastic volatility models. Math Financ **8**, 291–323 (1998)

Dacorogna, M.M., Gençay, R., Müller, U., Olsen, R.B., Pictet, O.V.: An Introduction to High-Frequency Finance. Academic Press, San Diego (2001)

Foster, D., Nelson, D.: Continuous record asymptotics for rolling sample variance estimators. Econometrica **64**, 139–174 (1996)

Gallant, A.R., Hsu, C.T., Tauchen, G.T.: Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. Rev Econ Stat **81**, 617–631 (1999)

Gonçalves, S., Meddahi, N.: Bootstrapping realized volatility. Econometrica **77**, 283–306 (2009)

Hájek, J., Sidak, Z.: Theory of Rank Tests. Academic Press, New York (1967)

Hall, P., Heyde, C.C.: Martingale Limit Theory and Its Application. Academic Press, Boston (1980)

Hayashi, T., Yoshida, N.: On covariance estimation of non-synchronously observed diffusion processes. Bernoulli **11**, 359–379 (2005)

Jacod, J., Protter, P.: Asymptotic error distributions for the euler method for stochastic differential equations. Ann Probab **26**, 267–307 (1998)

Jacod, J., Shiryaev, A.N.: Limit Theorems for Stochastic Processes. 2nd edn. Sringer, New York (2003)

Karatzas, I., Shreve, S.E.: Brownian Motion and Stochastic Calculus. Springer, New York (1991)

Mardia, K.V., Kent, J., Bibby, J.: Multivariate Analysis. Academic Press, London (1979)

McCullagh, P.: Tensor Methods in Statistics. Chapman and Hall, London (1987)

Mykland, P.A., Zhang, L.: ANOVA for diffusions and Itô processes. Ann Stat **34**, 1931–1963 (2006)

Mykland, P.A., Zhang, L.: Inference for volatility type objects and implications for hedging. Stat Its Interface **1**, 255–278 (2008)

Mykland, P.A., Zhang, L.: Inference for continuous semimartingales observed at high frequency. Econometrica **77**, 1403–1455 (2009)

Mykland, P.A., Zhang, L.: The econometrics of high frequency data. In: Kessler, M., Lindner, A., Sørensen, M. (eds.) Statistical methods for stochastic differential equations. Chapman and Hall/CRC Press (2010)

Protter, P.: Stochastic Integration and Differential Equations: A New Approach. 2nd edn. Springer, New York (2004)

Rényi, A.: On stable sequences of events. Sankyā Ser A **25**, 293–302 (1963)

Rootzén, H.: Limit distributions for the error in approximations of stochastic integrals. Ann Probab **8**, 241–251 (1980)

Weisberg, S.: Applied Linear Regression. 2nd edn. Wiley, New York (1985)

Zhang, L.: From martingales to ANOVA: Implied and realized volatility. PhD thesis, The University of Chicago, Department of Statistics (2001)

Zhang, L.: Estimating covariation: Epps effect and microstructure noise. J Econ (forthcoming) (2010)

Zhang, L., Mykland, P.A., Aït-Sahalia, Y.: A tale of two time scales: determining integrated volatility with noisy high-frequency data. J Am Stat Assoc **100**, 1394–1411 (2005)