Inference for Continuous Semimartingales Observed at High Frequency: A General Approach *

Per A. Mykland
The University of Chicago

Lan Zhang
University of Illinois at Chicago

First version: September 9, 2007.

This version: August 22, 2008.

Abstract

The econometric literature of high frequency data often relies on moment estimators which are derived from assuming local constancy of volatility and related quantities. We here study this local-constancy approximation as a general approach to estimation in such data. We show that the technique yields asymptotic properties (consistency, normality) that are correct subject to an ex post adjustment involving asymptotic likelihood ratios. These adjustments are given. Several examples of estimation are provided: powers of volatility, leverage effect, integrated betas, bipower, and covariance under asynchronous observation. The first order approximations in this study can be over the period of one observation, or over blocks of successive observations. The advantage of blocking is a gain in transparency in defining and analyzing estimators. The theory relies heavily on the interplay between stable convergence and measure change, and on asymptotic expansions for martingales.

KEYWORDS: consistency, cumulants, contiguity, continuity, discrete observation, efficiency, equivalent martingale measure, Itô process, leverage effect, likelihood inference, realized beta, realized volatility, stable convergence.

JEL CODES: C02; C13; C14; C15; C22

*We are grateful to Oliver Linton, Nour Meddahi, Eric Renault, Neil Shephard, and two referees, for helpful comments and suggestions. Financial support from the National Science Foundation under grants DMS 06-04758 and SES 06-31605 is also gratefully acknowledged.

1 Introduction

An important development in econometrics and statistics is the invention of estimation of financial volatility on the basis of high frequency data. The earliest results in this direction were in probability theory (Jacod (1994), Jacod and Protter (1998)). The econometric literature first focused on instantaneous volatility (Foster and Nelson (1996), Comte and Renault (1998)). The econometrics of integrated volatility was pioneered in Andersen, Bollerslev, Diebold, and Labys (2001) and Barndorff-Nielsen and Shephard (2001, 2002). Further literature includes, in particular, Andersen, Bollerslev, Diebold, and Labys (2000, 2003), Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006), Dacorogna, Gençay, Müller, Olsen, and Pictet (2001), Goncalves and Meddahi (2005), Jacod (2006), Kalnina and Linton (2007), Kinnebrock and Podolskij (2007), Mykland and Zhang (2006), and Zhang (2001).

The quantities that can be estimated from high frequency data are not confined to volatility. Problems that are attached to the estimation of covariations between two processes are discussed in, for example, Barndorff-Nielsen and Shephard (2004a), Hayashi and Yoshida (2005) and Zhang (2005). There is a literature on power variations and bi- and multi-power estimation (see Examples 1-2 in Section 2.5 for references). There is an analysis of variance/variation (ANOVA) based on high frequency observations (see Section 4.4.2). We shall see in this paper that one can also estimate such things as integrated betas, and the leverage effect.

The literature on high frequency data often relies on moment estimators derived from assuming local constancy of volatility and related quantities. To be specific, if t_i , $0 < t_1 < \dots < t_n = T$, are observation times, it is assumed that one can validly make one period approximations of the form

$$\int_{t_i}^{t_{i+1}} f_s dW_s \approx f_{t_i}(W_{t_{i+1}} - W_{t_i}), \quad (1)$$

where $\{W_t\}$ is a standard Brownian motion. The cited work on mixed normal distributions uses similar approximations to study stochastic variances. In the case of volatility, one can under weak regularity conditions make the approximation

$$\sum_i \left(\int_{t_i}^{t_{i+1}} \sigma_t dW_t \right)^2 - \int_0^T \sigma_t^2 dt \approx \sum_i \sigma_{t_i}^2 (W_{t_{i+1}} - W_{t_i})^2 - \sum_i \sigma_{t_i}^2 (t_{i+1} - t_i) \quad (2)$$

without affecting asymptotic properties (the error in (2) is of order $o_p(n^{-1/2})$). Thus the asymptotic distribution of realized volatility (sums of squared returns) can be inferred from discrete time martingale central limit theorems. In the special case where the σ_t^2 process is independent of W_t , one can even talk about unbiasedness of the estimator.

This raises two questions. First of all, (i) can one always do this? Or, does the approximation in formula (1) only work for a handful of cases such as volatility? Also, (ii) if one can pretend that volatility characteristics are constant from t_{i-1} to t_i , then can one also pretend constancy over

successive blocks of M ($M > 1$) observations, from, say t_{i-M} to t_i ? If this were true, a whole arsenal of additional statistical techniques would become available.

This paper will show that, subject to some adjustments, the answer to both these questions is yes. There are two main gains from this. One is easy derivation of asymptotic results. The other is to give a framework for how to set up inference procedures, as follows. If σ_t is treated as constant over a block of M observations, then the returns (the first differences of the observations) are simply Gaussian, and one can therefore think “parametrically” when setting up and analyzing estimators.

More specifically, in answer to (i), we show in Section 2 that first order approximations (like (1)-(2)) are “almost” valid for purposes of inference in high frequency data. The error is controllable by a general procedure, as follows. Let P be the true probability distribution governing the asset price process. The first order approximation (1) is (in a suitable sense) exact under a distribution P_n^* . We show that the likelihood ratio dP_n^*/dP has an asymptotic (nondegenerate) limit, which we can call dP_∞^*/dP . One can therefore proceed as follows:

- (1) Find the asymptotic distribution of estimators under P_n^* (i.e., pretending that the first order approximation (1) is valid); and then
- (2) Correct for the error in the final asymptotic distribution by multiplying by dP/dP_∞^* .

In other words, this is a generic device for correcting asymptotic distributions ex post when first order approximations are used in the analysis. The relationship between P and P_n^* is one of *contiguity*, cf. Remark 2 in Section 2.3 below.

The paper then turns to question (ii). We show in Section 3 that by holding volatility characteristics constant over successive blocks of M ($M > 1$) observations, from, say, t_{i-M} to t_i , one implicitly behaves as if the probability distribution were Q_n , where dQ_n/dP has an asymptotic (nondegenerate) limit, which we can call dQ_∞/dP . The conclusion is therefore much the same as under one period discretization (1).

The implication is that parametric techniques can be used locally over such blocks. Estimators of integrated quantities may then be obtained by aggregating local estimators. There are three advantages to blocking, as illustrated by examples in Section 4:

- Efficiency. In the case of quantities like $\int_0^T |\sigma|_t^r dt$, there can be substantial reduction in asymptotic variance (see Section 4.1).
- Transparency. Section 4.2 shows that the analysis of integrated betas reduces to ordinary least squares regression. Similar considerations apply to the examples (realized quantiles, ANOVA) in Section 4.4.
- Definition of new estimators. In the case of the leverage effect, blocking is a *sine qua non*, as will be clear from Sections 2.5 and 4.3.

Local parametric inference appears to have been introduced by Tibshirani and Hastie (1987),

and there is an extensive literature on the subject, such as Fan (1993), Fan and Gijbels (1996), Hjort and Jones (1996), Loader (1996), Fan, Gijbels, and King (1997), Kauermann and Opsomer (2003), Chen and Spokoiny (2007) and Cizek, Härdle, and Spokoiny (2007). A review is given in Fan, Farnen, and Gijbels (1998), and this paper should be consulted for further references.

Our current paper establishes, therefore, the connection of high-frequency-data inference to local parametric inference. We make this link with the help of contiguity. It will take time and further research to harvest the existing knowledge in the area of local likelihood for use in high frequency semimartingale inference. In fact, the estimators discussed in the applications sections (Sections 4-5) are rather obvious once a local likelihood perspective has been adapted; they are more of a beginning than an end. For example, local adaptation is not considered.

We emphasize that the main outcome of the paper is to provide direction on how to create estimators, and an easy way to analyze them. It is, however, perfectly possible to derive asymptotic results for such estimators by other existing methods of proof, as used in many of the papers cited above. In fact, direct proof will permit the most careful study of the precise conditions needed for consistency and mixed asymptotic normality for any given procedure.

A related but different form of diffusion block approximation was discussed in Mykland (2006), in the context of (conditionally) nonrandom volatility. Like Mykland (2006), we here use contiguity arguments to assess the approximation. The two approximations, however, are not the same. For example, the approximation in Mykland (2006) does not give any room for the estimation of leverage effect. Also, it does not guarantee that the stated properties of estimators remain valid when leverage effect is present. On the other hand, Mykland (2006) detailed the methodology and applications in ANOVA and Hayashi-Yoshida-type covariance estimation, which are not our current focus.

To further understand the difference between the two approaches, consider the distinction between (i) one step approximation, and (ii) approximation with block sizes larger than one. In the case of Mykland (2006), problem (i) does not occur: the approximating measure is exact for the observed data. Hence, the development in the current Section 2 is entirely orthogonal to the other paper, while there is some conceptual overlap in Section 3.1. It should also be noted that the current paper does not supersede Mykland (2006). The approximation in Mykland (2006) is easier to use. Also, since the one step discretization is exact in the other paper, it permits certain results (such as the maximum likelihood property of the Hayashi and Yoshida (2005) estimator) that can not be harvested from our current development.

The idea of blocking has also been exploited by Christensen, Oomen, and Podolskij (2008) to define volatility estimators based on local quantiles, citing Mykland (2006), but also developing theory in the case of leverage effect. See also Section 4.4.1.

A different kind of blocking, *pre-averaging*, is used by Podolskij and Vetter (2006) and Jacod, Li, Mykland, Podolskij, and Vetter (2008) in the context of inference in the presence of microstructure

noise. In these papers, the (latent) semimartingale is itself given a locally constant approximation. This approximation would not give contiguity in the absence of noise, but we conjecture that contiguity results can be found for common types of microstructure.

In the current paper, we do not deal with microstructure. This would be a study in itself, and is deferred to a later paper. This reflects our belief that one needs to understand the no-microstructure case before studying microstructure. Also, the relevance of the no-microstructure case is increased with the pre-averaging approach discussed in the previous paragraph.

The plan for the paper is that Section 2 discusses measure changes in detail, and their relationship to high frequency inference. It then analyzes the one period ($M = 1$) discretization. Section 3 discusses longer block sizes ($M > 1$). Major applications are given in Section 4, with a summary of the methodology (for the scalar case) in Section 4.5. Section 5 discusses how to handle more complex data structures or inference schemes: Section 5.1 looks at estimation with moving windows, and Section 5.2 at inference for a vector process under asynchronous observation. We conclude in Section 6, including a technical summary. Proofs are in the Appendix.

A reader's guide: We emphasize that the two approximations (to block size $M = 1$, and then from $M = 1$ to $M > 1$) are quite different in their methodologies. If you are only interested in the one period approximation, the material to read is Section 2 and Appendix A. (Though consequences for estimation of leverage effect is discussed in Section 4.3). The block ($M > 1$) approximation is mainly described in Sections 3-5, and Appendix B. An alternative way of reading the paper is to head for Section 4.5 first; this section should in any case be consulted early and kept in mind while reading the rest of the paper.

2 Approximate Systems.

We here discuss the discretization to block size $M = 1$. As a preliminary, we define some notation, and discuss measure change and stable convergence. This section can be read independently of the rest of the paper.

2.1 Data Generating Mechanism

In general, we shall work with a broad class of continuous semimartingales, namely *Itô processes*.

DEFINITION 1. A p -variate process $X_t = (X_t^{(1)}, \dots, X_t^{(p)})^T$ is called an *Itô process* provided it satisfies

$$dX_t = \mu_t dt + \sigma_t dW_t, X_0 = x_0, \quad (3)$$

where μ_t and σ_t are adapted locally bounded random processes, of dimension p and $p \times p$ respectively, and W_t is a p -dimensional Brownian motion. The underlying filtration will be called $(\mathcal{F}_t)$. The prob-

ability distribution will be called P . One understands the differential $\sigma_t dW_t$ to be a p -dimensional vector with r_1 'th component $\sum_{r_2=1}^p \sigma_t^{(r_1, r_2)} dW_t^{(r_2)}$.

If we set

$$\zeta_t = \sigma_t \sigma_t^T, \quad (4)$$

(where “ T ” in this case means transpose) then the (matrix) integrated volatility process is given as

$$\langle X, X \rangle_t = \int_0^t \zeta_u du. \quad (5)$$

A more precise name for (5) would perhaps be the integrated covariance matrix, but we shall use integrated volatility as shorthand.

We shall suppose that the process X_t is observed at times $0 = t_0 < t_1 < \dots < t_n = T$. Thus, for the moment, we assume synchronous observation of all the p components of the vector X_t . We explain in Section 5.2 how the results encompass the asynchronous case.

ASSUMPTION 1. (*Sampling times*). In asymptotic analysis, we suppose that $t_j = t_{n,j}$ (the additional subscript will normally be suppressed). The grids $\mathcal{G}_n = \{0 = t_{n,0} < t_{n,1} < \dots < t_{n,n} = T\}$ will not be assumed to be nested when n varies. We then do asymptotics as $n \rightarrow \infty$. The basic assumption is that

$$\max_{1 \leq i \leq n} |t_{n,i} - t_{n,i-1}| = o(1). \quad (6)$$

We also suppose that the observation times $t_{n,j}$ are nonrandom, but they can be irregular. By conditioning, this means that we include the case of random times independent of the X_t process.

We thus preclude dependence between the observation times and the process. Such dependence does appear to exist in some cases, cf. Renault and Werker (2006), and we hope to return to this question in a later paper.

2.2 A simplifying strategy for inference

When carrying out inference for observations in a fixed time interval $[0, T]$, the process μ_t cannot be consistently estimated. This follows from Girsanov's Theorem (see, for example, Chapter 5.5 of Karatzas and Shreve (1991)). For most purposes, μ_t simply drops out of the calculations, and is only a nuisance parameter. It is also a nuisance in that it complicates calculations substantially.

To deal with this most effectively, we shall borrow an idea from asset pricing theory, and consider a probability distribution P^* which is a measure theoretically equivalent to P , and under which X_t is a (local) martingale (Ross (1976), Harrison and Kreps (1979), Harrison and Pliska (1981), see also Duffie (1996)). Specifically, under P^*

$$dX_t = \sigma_t dW_t^*, X_0 = x_0, \quad (7)$$

where W_t^* is a P^* -Brownian motion. Following Girsanov's Theorem

$$\log \frac{dP^*}{dP} = - \int_0^T \sigma_t^{-1} \mu_t dW_t - \frac{1}{2} \int_0^T \mu_t^T (\sigma_t^T \sigma_t)^{-1} \mu_t dt, \quad (8)$$

with

$$dW_t^* = dW_t + \sigma_t^{-1} \mu_t dt. \quad (9)$$

Our plan is now the following: carry out the analysis under P^* , and adjust results back to P using the likelihood ratio (Radon-Nikodym derivative) dP^*/dP . Specifically suppose that θ is a quantity to be estimated (such as $\int_0^T \sigma_t^2 dt$, $\int_0^T \sigma_t^4 dt$, or the leverage effect). An estimator $\hat{\theta}_n$ is then found with the help of P^* , and an asymptotic result is established whereby, say,

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(b, a^2), \quad (10)$$

under P^* . It then follows directly from the measure theoretic equivalence that $n^{1/2}(\hat{\theta}_n - \theta)$ also converges in law under P . *In particular, consistency and rate of convergence is unaffected by the change of measure.* We emphasize that this is due to the finite (fixed) time horizon T .

The asymptotic law may be different under P^* and P . While the normal distribution remains, the distributions of b and a^2 (if random) may change. The main concept is *stable convergence*.

DEFINITION 2. *Suppose that all relevant processes (X_t , σ_t , etc) are adapted to filtration $(\mathcal{F}_t)$. Let Z_n be a sequence of $\mathcal{F}_T$ -measurable random variables, We say that Z_n converges stably in law to Z as $n \rightarrow \infty$ if Z is measurable with respect to an extension of $\mathcal{F}_T$ so that for all $A \in \mathcal{F}_T$ and for all bounded continuous g , $E I_{Ag}(Z_n) \rightarrow E I_{Ag}(Z)$ as $n \rightarrow \infty$.*

In the context of (10), $Z_n = n^{1/2}(\hat{\theta}_n - \theta)$ and $Z = N(b, a^2)$. For further discussion of stable convergence, see Rényi (1963), Aldous and Eagleson (1978), Chapter 3 (p. 56) of Hall and Heyde (1980), Rootzén (1980) and Section 2 (p. 169-170) of Jacod and Protter (1998).

With this tool in hand, assume that the convergence in (10) is stable. Then the same convergence holds under P . The technical result is as follows.

PROPOSITION 1. *Suppose that Z_n is a sequence of random variables which converges stably to $N(b, a^2)$ under P^* . By this we mean that $N(b, a^2) = b + aN(0, 1)$, where $N(0, 1)$ is a standard normal variable independent of $\mathcal{F}_T$, also a and b are $\mathcal{F}_T$ measurable. Then Z_n converges stably in law to $b + aN(0, 1)$ under P , where $N(0, 1)$ remains independent of $\mathcal{F}_T$ under P .*

PROOF OF PROPOSITION. $E I_{Ag}(Z_n) = E^* \frac{dP}{dP^*} I_{Ag}(Z_n) \rightarrow E^* \frac{dP}{dP^*} I_{Ag}(Z) = E I_{Ag}(Z)$ by uniform integrability of $\frac{dP}{dP^*} I_{Ag}(Z_n)$. $\blacksquare$

Proposition 1 substantially simplifies calculations and results. In fact, the same strategy will be helpful for the localization results that come next in the paper. It will turn out that the relationship between the localized and continuous process can also be characterized by absolute continuity and likelihood ratios.

REMARK 1. It should be noted that after adjusting back from P^* to P , the process μ_t may show up in expressions for asymptotic distributions. For instances of this, see Examples 3 and 6 below. One should always keep in mind that drift most likely is present, and may affect inference. $\square$

In order to use the measure change (8) in the subsequent development, we impose the following condition.

ASSUMPTION 2. (*Structure of the instantaneous volatility*). We assume that the matrix process σ_t is itself an Itô processes, and that if $\lambda_t^{(p)}$ is the smallest eigenvalue of σ_t , then $\inf_t \lambda_t^{(p)} > 0$ a.s.

2.3 Main result concerning one period discretization

Our main result in this section is that for the purposes of high frequency inference one can replace the system (7) by the following approximation:

$$P_n^* : \quad \Delta X_{t_{n,j+1}} = \sigma_{t_{n,j}} \Delta \check{W}_{t_{n,j+1}} \text{ for } j = 0, \dots, n-1; X_0 = x_0, \quad (11)$$

where $\Delta X_{t_{n,j+1}} = X_{t_{n,j+1}} - X_{t_{n,j}}$, and similarly for $\Delta \check{W}_{t_{n,j+1}}$ and $\Delta t_{n,j+1}$. One can view (11) as holding σ_t constant for one period, from $t_{n,j}$ to $t_{n,j+1}$. We call this a one period discretization (or localization). We are not taking a position on what the $\check{W}_t$ process looks like in continuous time, or even on whether it exists for other t than the sampling times $t_{n,j}$. The only assumption is that the random variables $\Delta \check{W}_{t_{n,j+1}}$ are independent for different j (for fixed n), and that $\Delta \check{W}_{t_{n,j+1}}$ has distribution $N(0, \Delta t_{n,j+1})$. Note that we here follow the convention from options pricing theory, whereby, when the measure changes, the process (X_t) doesn't change, while the driving Brownian motion changes.

To formally describe the nature of our approximations we go through two definitions:

DEFINITION 3. (*Specification of the time discrete process subject to measure change*).

$$\begin{aligned} U_{t_{n,j}}^{(1)} &= X_{t_{n,j}} \\ U_{t_{n,j}}^{(2)} &= (\sigma_{t_{n,j}}, \langle \sigma, W \rangle'_{t_{n,j}}, \langle \sigma, \sigma \rangle'_{t_{n,j}}) \\ U_{t_{n,j}} &= (U_{t_{n,j}}^{(1)}, U_{t_{n,j}}^{(2)}). \end{aligned} \quad (12)$$

for $j = 0, \dots, n$. Here, the quantity $\langle \sigma, W \rangle'_t$ is a three $(p \times p \times p)$ dimensional object (tensor) consisting of elements $\langle \sigma^{(r_1, r_2)}, W^{(r_3)} \rangle'_t$ ($r_1 = 1, \dots, p, r_2 = 1, \dots, p, r_3 = 1, \dots, p$), where prime denotes differentiation with respect to time. Similarly, $\langle \sigma, \sigma \rangle'_t$ is a four dimensional tensor with elements of the form $\langle \sigma^{(r_1, r_2)}, \sigma^{(r_3, r_4)} \rangle'_t$. Finally, denote by $\mathcal{X}_{n,j}$ the σ -field generated by $U_{t_{n,\iota}}$, $\iota = 0, \dots, j$.

We note here that $\langle \sigma, W \rangle'_t$ and $\langle \sigma, \sigma \rangle'_t$ are the usual continuous time quadratic variations, but they are only observed at the times $t_{n,j}$. Through $U_{t_{n,j}}^{(2)}$, however, we do incorporate information

about the continuous time system into discrete time observations: the σ_t process, the leverage effect (via the tensor $\langle \sigma, W \rangle'_t$), and the volatility of volatility (via $\langle \sigma, \sigma \rangle'_t$).

For each n , the approximate probability P_n^* will live on the filtration $(\mathcal{X}_{n,j})_{0 \leq j \leq n}$, as follows:

DEFINITION 4. (*Specification of the first order approximation*). Define the probability P_n^* recursively by:

- (i) U_0 has same distribution under P_n^* as under P^* ;
- (ii) The conditional P_n^* -distribution of $U_{t_{n,j+1}}^{(1)}$ given $U_0, \dots, U_{t_{n,j}}$ is given by (11); and
- (iii) The conditional P_n^* -distribution of $U_{t_{n,j+1}}^{(2)}$ given $U_0, \dots, U_{t_{n,j}}, U_{t_{n,j+1}}^{(1)}$ is the same as under P^* .

To the extent that conditional densities are defined, one can describe the relationship between P^* and P_n^* as

$$f(U_{t_{n,1}}, \dots, U_{t_{n,j}}, \dots, U_{t_{n,n}} | U_0) = \underbrace{\prod_{j=1}^n f(U_{t_{n,j}}^{(1)} | U_0, \dots, U_{t_{n,j-1}})}_{\text{altered from } P^* \text{ to } P_n^*} \underbrace{\prod_{j=1}^n f(U_{t_{n,j}}^{(2)} | U_0, \dots, U_{t_{n,j-1}}, U_{t_{n,j}}^{(1)})}_{\text{unchanged from } P^* \text{ to } P_n^*} \quad (13)$$

where $f(y|x)$ is the density of the regular conditional distribution of y given x with respect to a reference (say, Lebesgue) measure.

To state the main theorem, define

$$d\check{\zeta}_t = \sigma_t^{-1} d\zeta_t (\sigma_t^T)^{-1} \quad (14)$$

and

$$k_t^{(r_1, r_2, r_3)} = \langle \check{\zeta}^{(r_1, r_2)}, W^{(r_3)} \rangle'_t [3] \quad (15)$$

where the “[3]” means that the right hand side of (15) is a sum over three terms where r_3 can change position with either r_1 or r_2 : $\langle \check{\zeta}^{(r_1, r_2)}, W^{(r_3)} \rangle'_t [3] = \langle \check{\zeta}^{(r_1, r_2)}, W^{(r_3)} \rangle'_t + \langle \check{\zeta}^{(r_1, r_3)}, W^{(r_1)} \rangle'_t + \langle \check{\zeta}^{(r_3, r_2)}, W^{(r_1)} \rangle'_t$ (note that $\langle \check{\zeta}^{(r_1, r_2)}, W^{(r_3)} \rangle'_t$ is symmetric in its two first arguments). For further discussion of this notation, see Chapter 2.3 (p. 29-30) of McCullagh (1987). Note that $k_{t_{n,j}}^{(r_1, r_2, r_3)}$ is measurable with respect to the σ -field $\mathcal{X}_{n,j}$ generated by $U_{t_{n,\iota}}, \iota = 0, \dots, j$. Finally, set

$$\Gamma_0 = \frac{1}{24} \int_0^T \sum_{r_1, r_2, r_3=1}^p (k_t^{(r_1, r_2, r_3)})^2 dt. \quad (16)$$

In the univariate case, we have the representations

$$k_t = 3 \frac{1}{\sigma_t^2} \langle \sigma^2, W \rangle'_t = 6 \frac{1}{\sigma_t} \langle \sigma, W \rangle'_t = 6 \langle \log \sigma, W \rangle'_t$$

and

$$\Gamma_0 = \frac{1}{24} \int_0^T k_t^2 dt.$$

We now state the main result for one period discretization.

THEOREM 1. P^* and P_n^* are mutually absolutely continuous on the σ -field $\mathcal{X}_{n,n}$ generated by $U_{t_{n,j}}$, $j = 0, \dots, n$. Furthermore, let $(dP^*/dP_n^*)(U_{t_0}, \dots, U_{t_{n,j}}, \dots, U_{t_{n,n}})$ be the likelihood ratio (Radon-Nikodym derivative) on $\mathcal{X}_{n,n}$. Then,

$$\frac{dP^*}{dP_n^*}(U_{t_0}, \dots, U_{t_{n,j}}, \dots, U_{t_{n,n}}) \xrightarrow{\mathcal{L}} \exp\{\Gamma_0^{1/2}N(0,1) - \frac{1}{2}\Gamma_0\} \quad (17)$$

stably in law, under P_n^* , as $n \rightarrow \infty$.

Based on Theorem 1, one can (for a fixed time period) carry out inference under the model (11), and asymptotic results will transfer back to the continuous model (7) by absolute continuity. This is much the same strategy as the one to eliminate the drift described in Section 2.2. The main difference is that we use an asymptotic version of absolute continuity. This concept is known as *contiguity*, and is well known in classical statistical literature (see Remark 2 below). We state the following result, in analogy with Proposition 1. A sequence Z_n is called *tight* if every subsequence if every subsequence has a further subsequence which converges in law (see Chapter VI of Jacod and Shiryaev (2003)). Tightness is the compactness concept which goes along with convergence in law.

COROLLARY 1. Suppose that Z_n (say, $n^{1/2}(\hat{\theta}_n - \theta)$) is tight in the sense of stable convergence under P_n^* . The same statement then holds under P^* and P . The converse is also true.

In particular, if an estimator is consistent under P_n^* , it is also consistent under P^* (and P).

Unlike the situation in Section 2.2, the stable convergence in Corollary 1 does not assure that $n^{1/2}(\hat{\theta}_n - \theta)$ is asymptotically independent of the normal distribution $N(0,1)$ in Theorem 1. It only assures independence from $\mathcal{F}_T$ -measurable quantities. The asymptotic law of $n^{1/2}(\hat{\theta}_n - \theta)$ may, therefore, require an adjustment from P_n^* to P^* .

REMARK 2. Theorem 1 says that P^* and the approximation P_n^* are *contiguous* in the sense of Hájek and Sidak (1967) (Chapter IV), LeCam (1986), LeCam and Yang (1986), and Jacod and Shiryaev (2003) (Chapter IV). This follows from Theorem 1 since dP^*/dP_n^* is uniformly integrable under P_n^* (since the sequence dP_n^*/dP^* is nonnegative, also the limit integrates to one under P^*). $\square$

REMARK 3. A nonzero $\langle \sigma, W \rangle'_t$ can occur in other cases than what is usually termed “leverage effect”. An important instance of this occurs in Section 4.4.2, where $\langle \sigma, W \rangle'_t$ can be nonzero due to the nonlinear relationship between an option Y and the underlying security X . $\square$

2.4 Adjusting for the Change from P^* to P_n^*

Following (11), write

$$\Delta \check{W}_{t_{n,j+1}} = \sigma_{t_{n,j}}^{-1} \Delta X_{t_{n,j+1}}. \quad (18)$$

Under the approximating measure P_n^* , $\Delta\check{W}_{t_{n,j+1}}$ has distribution $N(0, I\Delta t_{n,j+1})$ and is independent of the past.

Define the third order Hermite polynomials by $h_{r_1 r_2 r_3}(x) = x^{r_1} x^{r_2} x^{r_3} - x^{r_1} \delta^{r_2, r_3} [3]$, where, again, “[3]” represents the sum over all three possible terms for this form, and $\delta^{r_2, r_3} = 1$ if $r_2 = r_3$, and zero otherwise. In the univariate case, $h_{111}(x) = x^3 - 3x$. Set

$$M_n^{(0)} = \frac{1}{12} \sum_{j=0}^{n-1} (\Delta t_{n,j+1})^{1/2} \sum_{r_1, r_2, r_3=1}^p k_{t_{n,j}}^{(r_1, r_2, r_3)} h_{r_1 r_2 r_3}(\Delta\check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}) \quad (19)$$

Note that $k_{t_{n,j}}^{(r_1, r_2, r_3)}$ is $\mathcal{X}_{n,j}$ -measurable. The adjustment result is now as follows:

THEOREM 2. *Assume the setup in Theorem 1. Suppose that under P_n^* , $(Z_n, M_n^{(0)})$ converges stably to a bivariate distribution $b + aN(0, I)$, where $N(0, I)$ is a bivariate standard normal vector independent of $\mathcal{F}_T$ and where the vector $b = (b_1, b_2)^T$ and the symmetric 2×2 matrix a are $\mathcal{F}_T$ measurable. Set $A = aa^T$. It is then the case that Z_n converges stably under P^* to $b_1 + A_{12} + (A_{11})^{1/2}N(0, 1)$, where $N(0, 1)$ is independent of $\mathcal{F}_T$.*

Note that under the conditions of Theorem 1, $M_n^{(0)}$ converges stably under P_n^* to a (mixed) normal distribution with mean zero and (random, but $\mathcal{F}_T$ -measurable) variance Γ_0 (so $b_2 = 0$ and $A_{22} = \Gamma_0$). Thus, when adjusting from P_n^* to P^* , the asymptotic variance of Z_n is unchanged, while the asymptotic bias may change.

REMARK 4. The logic behind this result is as follows. On the one hand, the asymptotic variance remains unchanged in Theorem 2 as a special case of a stochastic process property (the preservation of quadratic variation under limit operations). We refer to the discussion in Chapter VI.6 (p. 376-388) in Jacod and Shiryaev (2003) for a general treatment.

On the other hand, it follows from the proof of Theorem 1 that

$$\log \frac{dP^*}{dP_n^*} = M_n^{(0)} - \frac{1}{2}\Gamma_0 + o_p(1). \quad (20)$$

Thus, to the extent that the random variables Z_n are correlated with $M_n^{(0)}$, their asymptotic mean will change from P_n^* to P^* . This change of mean is precisely the value A_{12} , which is the asymptotic covariance of Z_n and $M_n^{(0)}$. This is a standard phenomenon in situations of contiguity, cf. Hájek and Sidak (1967). $\square$

2.5 Some initial examples

The following is meant for illustration only. The in-depth applications are in Sections 4-5. We here only consider one dimensional systems ($p = 1$).

EXAMPLE 1. (Integral of absolute powers of ΔX). For $r > 0$, it is customary to estimate $\int_0^T |\sigma_t|^r dt$ by $\sum_{j=1}^n |\Delta X_{t_{n,j}}|^r$. A general theory for this is given in Jacod (1994, 2006) and Barndorff-Nielsen and Shephard (2004b), see also Andersen and Bollerslev (1998) for an early use of such estimators. For the important cases $r = 2$ and $r = 4$, see also Jacod and Protter (1998), Zhang (2001), Barndorff-Nielsen and Shephard (2002), Mykland and Zhang (2006), and other work by the same authors.

We here see how the same estimators can be analyzed using the technology of this paper. Further developments concerning the estimation of $\int_0^T |\sigma_t|^r dt$ are discussed in Sections 4.1 and 5.1.2. However, there is some added conceptual understanding to the present analysis.

To do the analysis, note that under P_n^* , the law of $|\Delta X_{t_{n,j+1}}|^r$ given $\mathcal{X}_{n,j}$ is $|\sigma_{t_{n,j}} N(0, 1)|^r \Delta t_{n,j+1}^{r/2}$, whereby

$$\begin{aligned} E_n^*(|\Delta X_{t_{n,j+1}}|^r \mid \mathcal{X}_{n,j}) &= |\sigma_{t_{n,j}}|^r E|N(0, 1)|^r \Delta t_{n,j+1}^{r/2} \\ \text{Var}_n^*(|\Delta X_{t_{n,j+1}}|^r \mid \mathcal{X}_{n,j}) &= |\sigma_{t_{n,j}}|^{2r} \text{Var}(|N(0, 1)|^r) \Delta t_{n,j+1}^r \text{ and} \\ \text{Cov}_n^*(|\Delta X_{t_{n,j+1}}|^r, \Delta \check{W}_{t_{n,j+1}} \mid \mathcal{X}_{n,j}) &= 0. \end{aligned} \quad (21)$$

Thus, the natural estimator of $\theta = \int_0^T |\sigma_t|^r dt$ becomes

$$\hat{\theta}_n = \frac{1}{E|N(0, 1)|^r} \sum_{j=0}^{n-1} \Delta t_{n,j+1}^{1-\frac{r}{2}} |\Delta X_{t_{n,j+1}}|^r. \quad (22)$$

From (21), it follows that $\hat{\theta}_n - \sum_{j=0}^{n-1} |\sigma_{t_{n,j}}|^r \Delta t_{n,j+1}$ is the end point of a martingale orthogonal to W , and with discrete time quadratic variation $\frac{\text{Var}(|N(0, 1)|^r)}{(E|N(0, 1)|^r)^2} \sum_{j=0}^{n-1} |\sigma_{t_{n,j}}|^{2r} \Delta t_{n,j+1}^2$. By the usual martingale central limit considerations (Jacod and Shiryaev (2003)), and since $\theta - \sum_{j=0}^{n-1} |\sigma_{t_{n,j}}|^r \Delta t_{n,j+1} = O_p(n^{-1})$, it follows that

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} Z \times \left(\frac{\text{Var}(|N(0, 1)|^r)}{(E|N(0, 1)|^r)^2} T \int_0^T \sigma_t^{2r} dH(t) \right)^{1/2} \quad (23)$$

stably in law under P_n^* , where Z is a standard normal random variable. Here, $H(t)$ is the “Asymptotic Quadratic Variation of Time” (AQVT), given by

$$H(t) = \lim_{n \rightarrow \infty} \frac{n}{T} \sum_{t_{n,j+1} \leq t} (t_{n,j+1} - t_{n,j})^2, \quad (24)$$

provided that the limit exists. For further references on this quantity, see (Zhang (2001, 2006), and Mykland and Zhang (2006)).

Note that in the case of *equally spaced observations*, $\hat{\theta}_n$ is proportional to $\sum_{j=1}^n |\Delta X_{t_{n,j}}|^r$, and also $H(t) = t$. Finally, recall that

$$E|N(0, 1)|^r = 2^{\frac{r}{2}} \pi^{-\frac{1}{2}} \Gamma\left(\frac{r+1}{2}\right) \quad (25)$$

where Γ is the Gamma-function. (And similarly for variances).

To get from the convergence under P_n^* to convergence under P^* , we note that $|N(0,1)|^r$ is uncorrelated with $N(0,1)$ and $N(0,1)^3$. We therefore obtain from Theorems 1 and 2 that the stable convergence in (23) holds under P^* . The same is true under the true probability P by Proposition 1.

□

EXAMPLE 2. (Bi- and multipower estimators.) The same considerations as in Example 1 apply to bi- and multipower estimators (see, in particular, Barndorff-Nielsen and Shephard (2004b), Woerner (2004) and Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006)). The derivations are much the same. In particular, no adjustment is needed from P_n^* to P^* . We continue the discussion of bipower in Section 5.1.1.

□

EXAMPLE 3. (Sum of third moments). We here consider quantities of the form

$$Z_n = \frac{n}{T} \sum_{j=0}^{n-1} (\Delta X_{t_{n,j+1}})^3. \quad (26)$$

To avoid clutter, we shall look at the equally spaced case only ($\Delta t_{n,j+1} = \Delta t = T/n$ for all j, n).

Our study of the quantity Z_n started out as the consideration of a strange example, but has gradually grown into something more important. We shall see in Section 4.3 that similar quantities can be parlayed into estimators of the leverage effect. For now, we just show what the simplest calculation will bring. An important issue, which sets (26) apart from most other cases is that there is a need for an adjustment from P_n^* to P^* , and also from P^* to P . This is what will, in turn, yield interesting results in Section 4.3.

By the same reasoning as in Example 1,

$$\begin{aligned} E_n^*(\Delta X_{t_{n,j+1}}^3 \mid \mathcal{X}_{n,j}) &= 0 \\ \text{Var}_n^*(\Delta X_{t_{n,j+1}}^3 \mid \mathcal{X}_{n,j}) &= \sigma_{t_{n,j}}^6 \text{Var}(N(0,1)^3) \Delta t^3 = 15 \sigma_{t_{n,j}}^6 \Delta t^3 \text{ and} \\ \text{Cov}_n^*(\Delta X_{t_{n,j+1}}^3, \Delta \check{W}_{t_{n,j+1}} \mid \mathcal{X}_{n,j}) &= \sigma_{t_{n,j}}^3 \text{Cov}(N(0,1)^3, N(0,1)) \Delta t^2 = 3 \sigma_{t_{n,j}}^3 \Delta t^2. \end{aligned} \quad (27)$$

Thus, Z_n is the end point of a P_n^* martingale, and, $Z_n \xrightarrow{\mathcal{L}} N(b, a^2)$ stably under P_n^* , where

$$\begin{aligned} b &= 3 \int_0^T \sigma_t^3 dW_t^* \text{ and} \\ a^2 &= 6 \int_0^T \sigma_t^6 dt. \end{aligned} \quad (28)$$

REMARK 5. (Sample of calculation). To see in more detail how (28) comes about, let $V_t^{(n)}$ be the P^* martingale for which $V_T^{(n)} = Z_n$. Let (X_t, V_t) be the process corresponding to the limiting

distribution of $(X_t, V_t^{(n)})$ under P_n^* . (The prelimiting process is only defined on the grid points $t_{n,i}$). From the two last equations in (27), and by interchanging limits and quadratic variation (Chapter VI.6 (p. 376-388) in Jacod and Shiryaev (2003), cf Remark 4 above), we get

$$\begin{aligned}\langle V, V \rangle_t &= 15 \int_0^t \sigma_u^6 du \text{ and} \\ \langle V, W^* \rangle_t &= 3 \int_0^t \sigma_u^6 du.\end{aligned}\tag{29}$$

Now consider the representation

$$dV_t = f_t dW_t^* + g_t dB_t$$

where B_t is a Brownian motion independent of $\mathcal{F}_T$ (this is by Lévy's Theorem; see, for example, Theorem II.4.4 (p. 102) of Jacod and Shiryaev (2003), or Theorem 3.16 (p. 157) of Karatzas and Shreve (1991)). From (29),

$$\begin{aligned}f_t^2 dt + g_t^2 dt &= 15\sigma_t^6 dt \text{ and} \\ f_t dt &= 3\sigma_t^6 dt.\end{aligned}$$

In particular, $g_t^2 = 6\sigma_t^6$. This yields (28).

What happens here is that the full quadratic variation of V_t splits in a bias and a variance term. This is due to the non-zero covariation of V and W^* . $\square$

In this example, $b \neq 0$. Even more interestingly, the distributional result needs to be adjusted from P_n^* to P^* . To see this, denote $h_3(x) = x^3 - 3x$ (the third Hermite polynomial in the scalar case). Then,

$$\begin{aligned}\text{Cov}_n^*(\Delta X_{t_{n,j+1}}^3, h_3(\Delta \check{W}_{t_{n,j+1}}/\Delta t^{1/2}) \mid \mathcal{X}_{n,j})\Delta t^{1/2} &= \sigma_{t_{n,j}}^3 \text{Cov}(N(0,1)^3, h_3(N(0,1)))\Delta t^2 \\ &= 6\sigma_{t_{n,j}}^3 \Delta t^2\end{aligned}\tag{30}$$

Thus, if $M_n^{(0)}$ is as given in Section 2.4, it follows that $(Z_n, M_n^{(0)})$ converge jointly, and stably, under P_n^* to a normal distribution, where the asymptotic covariance is

$$\begin{aligned}A_{12} &= \frac{1}{2} \int_0^T k_t \sigma_t^3 dt \\ &= \frac{3}{2} \langle \sigma^2, X \rangle_T,\end{aligned}\tag{31}$$

since $k_t \sigma_t^3 dt = 3\sigma_t^{-2} \langle \zeta, W \rangle_t' \sigma_t^3 dt = 3d\langle \zeta, X \rangle_t = 3d\langle \sigma^2, X \rangle_t$. Thus, by Theorem 2, under P^* , $Z_n \xrightarrow{\mathcal{L}} N(b', a^2)$ stably, where a^2 is as in (28), while

$$b' = 3 \int_0^T \sigma_t^3 dW_t^* + \frac{3}{2} \langle \sigma^2, X \rangle_T.\tag{32}$$

We thus have a limit which relates to the leverage effect, which is interesting, but unfortunately obscured by the rest of b' , and by the random term with variance a^2 .

There is finally a need to adjust from P^* to P . From (9), we have $dW_t^* = dW_t + \sigma_t^{-1}\mu_t dt$, it follows that

$$b' = 3 \int_0^T \sigma_t^3 (dW_t + \sigma_t^{-1}\mu_t dt) + \frac{3}{2} \langle \sigma^2, X \rangle_T. \quad (33)$$

Thus, b' is unchanged from P^* to P , but it has different distributional properties. In particular, μ_t now appears in the expression. This is unusual in the high frequency context.

It seems to be a general phenomenon that if there is random bias under P^* , then μ will occur in the expression for bias under P . This occurs again in Example 6 in Section 4.3.2. $\square$

A direct derivation of this same limit is given in Example 6 of Kinnebrock and Podolskij (2007). In their notation, $\sigma_t' dt = 2\sigma^{-2} d\langle \sigma^2, X \rangle_t$.

3 Holding σ constant over longer time periods

3.1 Setup

We have shown in the above that it is asymptotically valid to consider systems where σ is constant from one time point to the next. We shall in the following show that it is also possible to consider approximate systems where σ is constant over longer time periods.

We suppose that there are K_n intervals of constancy, on the form $(\tau_{n,i-1}, \tau_{n,i}]$, where

$$\mathcal{H}_n = \{0 = \tau_{n,0} < \tau_{n,1} < \dots < \tau_{n,K_n} = T\} \subseteq \mathcal{G}_n \quad (34)$$

If we set

$$\begin{aligned} M_{n,i} &= \#\{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]\} \\ &= \text{number of intervals } (t_{n,j-1}, t_{n,j}] \text{ in } (\tau_{n,i-1}, \tau_{n,i}] \end{aligned} \quad (35)$$

we shall suppose that

$$\max_i M_{n,i} = O(1) \text{ as } n \rightarrow \infty, \quad (36)$$

from which it follows that K_n is of exact order $O(n)$.

We now define the approximate measure, called Q_n , given by

$$\begin{aligned} X_0 &= x_0 \\ \text{for each } i &= 1, K_n : \\ \Delta X_{t_{j+1}} &= \sigma_{\tau_{n,i-1}} \Delta W_{t_{j+1}}^Q \text{ for } t_{n,j+1} \in (\tau_{n,i-1}, \tau_{n,i}]. \end{aligned} \quad (37)$$

To implement this, we use a variation over Definition 4. Formally, we define the approximation as follows.

DEFINITION 5. (*Block approximation*). Define the probability Q_n recursively by:

- (i) U_0 has same distribution under Q_n as under P^* ;
- (ii) The conditional Q_n -distribution of $U_{t_{n,j+1}}^{(1)}$ given $U_0, \dots, U_{t_{n,j}}$ is given by (37); and
- (iii) The conditional P_n^* -distribution of $U_{t_{n,j+1}}^{(2)}$ given $U_0, \dots, U_{t_{n,j}}, U_{t_{n,j+1}}^{(1)}$ is the same as under P^* .

We can now describe the relationship between Q_n and P_n^* , as follows. Let the Gaussian log likelihood be given by

$$\ell(\Delta x; \zeta) = -\frac{1}{2} \log \det(\zeta) - \frac{1}{2} \Delta x^T \zeta^{-1} \Delta x. \quad (38)$$

We then obtain directly that

PROPOSITION 2. The likelihood ratio between Q_n and P_n^* is given by

$$\begin{aligned} \log \frac{dQ_n}{dP_n^*}(U_{t_0}, \dots, U_{t_{n,j}}, \dots, U_{t_{n,n}}) \\ = \sum_i \sum_{\tau_{i-1} \leq t_j < \tau_i} \{ \ell(\Delta X_{t_{j+1}}; \zeta_{\tau_{n,i-1}} \Delta t_{j+1}) - \ell(\Delta X_{t_{j+1}}; \zeta_{t_{n,j}} \Delta t_{j+1}) \} \end{aligned} \quad (39)$$

DEFINITION 6. To measure the extent to which we hold the volatility constant, we define the following “Asymptotic Decoupling Delay” (ADD) by

$$K(t) = \lim_{n \rightarrow \infty} \sum_i \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}] \cap [0, t]} (t_{n,j} - \tau_{n,i-1}), \quad (40)$$

provided the limit exists.

From (6) and (36), every subsequence has a further subsequence for which $K(\cdot)$ exists (by Helly’s Theorem, see, for example, p. 336 in Billingsley (1995)). Thus one can take the limits to exist without any major loss of generality. Also, when the limit exists, it is Lipschitz continuous.

In the case of equidistant observations and equally sized blocks of M observations, the ADD takes the form

$$K(t) = \frac{1}{2}(M-1)t. \quad (41)$$

3.2 Main Contiguity Theorem for the Block Approximation

We obtain the following main result, which is proved in Appendix B.

THEOREM 3. (*Contiguity of P_n^* and Q_n*). Suppose that Assumptions 1-2 are satisfied. Assume that the Asymptotic Decoupling Delay (K , equation (40)) exists. Set

$$Z_n^{(1)} = \frac{1}{2} \sum_i \sum_{t_{n,j} \in [\tau_{n,i-1}, \tau_{n,i})} \left(\Delta X_{t_{n,j+1}}^T (\zeta_{t_{n,j}}^{-1} - \zeta_{\tau_{n,i-1}}^{-1}) \Delta X_{t_{n,j+1}} \Delta t_{n,j+1}^{-1} \right) \quad (42)$$

and let $M_n^{(1)}$ be the end point of the P_n^* -martingale part of $Z_n^{(1)}$ (see (B.25) and (B.27) in Appendix B for the explicit formula). Define

$$\Gamma_1 = \frac{1}{2} \int_0^T \text{tr}(\zeta_t^{-2} \langle \zeta, \zeta \rangle'_t) dK(t), \quad (43)$$

where “tr” denotes the trace of the matrix. Then, as $n \rightarrow \infty$, $M_n^{(1)}$ converges stably in law under P^* to a normal distribution with mean zero and variance Γ_1 . Also, under P^* ,

$$\log \frac{dQ_n}{dP_n^*} = M_n^{(1)} - \frac{1}{2} \Gamma_1 + o_p(1). \quad (44)$$

Furthermore, if $M_n^{(0)}$ is as defined in (19), then the pair $(M_n^{(0)}, M_n^{(1)})$ converges stably to $(\Gamma_0^{1/2} V_0, \Gamma_1^{1/2} V_1)$, where V_0 and V_1 are iid $N(0,1)$, and independent $\mathcal{F}_T$.

The theorem says that P_n^* and the approximation Q_n are contiguous, cf. Remark 2 in Section 2.3. By the earlier Theorem 1, it follows that Q_n and P^* (and P) are contiguous. In particular, as before, if an estimator is consistent under Q_n , it is also consistent under P^* and P . Rates of convergence (typically $n^{1/2}$) are also preserved, but the asymptotic distribution may change.

REMARK 6. (Which probability?) We have now done several approximations. The true probability is P , and we are proposing to behave as if it is Q_n . We thus have the following alterations of probability

$$\log \frac{dP}{dQ_n} = \log \frac{dP}{dP^*} + \log \frac{dP^*}{dP_n^*} + \log \frac{dP_n^*}{dQ_n}. \quad (45)$$

To make matters slightly more transparent, we have stated Theorem 3 under the same probability (P_n^*) as Theorems 1 and 2. Since computations would normally be made under Q_n , however, we note that Theorem 2 applies equally if one replaces P_n^* by Q_n , and $M_n^{(0)}$ by $M_n^{(0,Q)}$, given as in (19), with $\Delta W_{t_n, j+1}^Q$ replacing $\Delta \check{W}_{t_n, j+1}$. (Since $M_n^{(0,Q)} = M_n^{(0)} + o_p(1)$). Similarly, if one lets $M_n^{(1,Q)}$ be endpoint of the Q_n -martingale part of $-Z_n^{(1)}$, one gets the same stable convergence under Q^n . Obviously, (44) should be replaced by

$$\log \frac{dP_n^*}{dQ_n} = M_n^{(1,Q)} - \frac{1}{2} \Gamma_1 + o_p(1). \quad (46)$$

and $M_n^{(1,Q)} = -M_n^{(1)} + \Gamma_1 + o_p(1)$. □

3.3 Measure change and Hermite polynomials

The three measure changes in Remark 6 turn out to all have a representation in terms of Hermite polynomials.

Recall that the standardized Hermite polynomials are given by $h_{r_1}(x) = x^{r_1}$, $h_{r_1 r_2}(x) = x^{r_1} x^{r_2} - \delta^{r_1, r_2}$, and $h_{r_1 r_2 r_3}(x) = x^{r_1} x^{r_2} x^{r_3} - x^{r_1} \delta^{r_2, r_3} [3]$, where, again, “[3]” represents the sum over all three

possible combinations, and $\delta^{r_2, r_3} = 1$ if $r_2 = r_3$, and zero otherwise. In the scalar case, $h_1(x) = x$, $h_{11}(x) = x^2 - 1$, and $h_{111}(x) = x^3 - 3x$. From Remark 6,

$$\begin{aligned} M_n^{(0,Q)} &= \frac{1}{12} \sum_{j=0}^{n-1} (\Delta t_{n,j+1})^{1/2} \sum_{r_1, r_2, r_3=1}^p k_{t_{n,j}}^{(r_1, r_2, r_3)} h_{r_1 r_2 r_3}(\Delta W_{t_{n,j+1}}^Q / (\Delta t_{n,j+1})^{1/2}), \text{ and} \\ M_n^{(1,Q)} &= -\frac{1}{2} \sum_i \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} \text{tr} \left(\sigma_{\tau_{n,i-1}}^T (\zeta_{t_{n,j}}^{-1} - \zeta_{\tau_{n,i-1}}^{-1}) \sigma_{\tau_{n,i-1}} h_{..}(\Delta W_{t_{n,j+1}}^Q / (\Delta t_{n,j+1})^{1/2}) \right). \end{aligned} \quad (47)$$

Similarly, define a discretized version of $M^{(G)} = \int_0^T \sigma_t^{-1} \mu_t dW_t^*$ by

$$M_n^{(G,Q)} = \sum_{j=0}^{n-1} (\Delta t_{n,j+1})^{1/2} \left(\sigma_{\tau_{n,i-1}}^{-1} \mu_{\tau_{n,i-1}} h_{..}(\Delta W_{t_{n,j+1}}^Q / (\Delta t_{n,j+1})^{1/2}) \right) \quad (48)$$

(“G” is for Girsanov; $h_{..}$ is the vector of first order Hermite polynomials, similarly $h_{..}$ is the matrix of second order such polynomials). We also set

$$\Gamma_G = \int_0^T \mu_t^T (\sigma_t^T \sigma_t)^{-1} \mu_t dt. \quad (49)$$

We therefore get the following summary of our results:

$$\begin{aligned} \log \frac{dP}{dP^*} &= M_n^{(G,Q)} - \frac{1}{2} \Gamma_G + o_p(1) \\ \log \frac{dP^*}{dP_n^*} &= M_n^{(0,Q)} - \frac{1}{2} \Gamma_0 + o_p(1) \\ \log \frac{dP_n^*}{dQ_n} &= M_n^{(1,Q)} - \frac{1}{2} \Gamma_1 + o_p(1). \end{aligned} \quad (50)$$

Furthermore, by the Hermite polynomial property, we obtain that these three martingales have, by construction, zero predictable covariation (under Q_n). In particular, the triplet $(M_n^{(G,Q)}, M_n^{(0)}, M_n^{(1)})$ converges stably to $(M^{(G)}, \Gamma_0^{1/2} V_0, \Gamma_1^{1/2} V_1)$, where V_0 and V_1 are iid $N(0,1)$, and independent $\mathcal{F}_T$.

REMARK 7. Note that the $M_n^{(G,Q)}$ is in many ways different from $M_n^{(0)}$ and $M_n^{(1)}$. The convergence of the former is in probability, while the latter converge only in law. Thus, for example, the property discussed in Remark 4 (see also Theorem 4 in the next section) does not apply to $M_n^{(G,Q)}$. If Z_n and $M_n^{(G,Q)}$ have joint covariation, this yields a smaller asymptotic variance for Z_n , but also bias. For instances of this, see Example 3 in Section 2.5, and also Example 6 in Section 4.3.2 below. $\square$

3.4 Adjusting for the Change from P^* to Q_n

The adjustment result is now similar to that of Section 2.4

THEOREM 4. *Assume the setup in Theorems 1-3. Suppose that, under Q_n , $(Z_n, M_n^{(0)}, M_n^{(1)})$ converges stably to a trivariate distribution $b + aN(0, I)$, where $N(0, I)$ is a trivariate vector independent of $\mathcal{F}_T$, where the vector $b = (b_1, b_2, b_3)^T$ and the symmetric 3×3 matrix a are $\mathcal{F}_T$ measurable. Set $A = aa^T$. Then Z_n converges stably under P^* to $b_1 + A_{12} + A_{13} + (A_{11})^{1/2}N(0, 1)$, where $N(0, 1)$ is independent of $\mathcal{F}_T$.*

Recall that $b_2 = b_3 = A_{23} = 0$, $A_{22} = \Gamma_0$, and $A_{33} = \Gamma_1$. The proof is the same as for Theorem 2.

Theorem 4 states that when adjusting from Q_n to P^* , the asymptotic variance of Z_n is unchanged, while the asymptotic bias may change.

3.5 Irregularly spaced data: A second block approximation.

The approximation in this section reduces the problem (in each block) to a case of independent (but not identically distributed) increments. Can we do better than this, and go to iid observations? We here give a criterion for this to be the case.

Consider yet another approximate probability measure R_n :

$$\begin{aligned} X_0 &= x_0 \\ \text{for each } i &= 1, K_n : \\ \Delta X_{t_{j+1}} &= \sigma_{\tau_{n,i-1}} \left(\frac{\Delta \tau_i}{\Delta t_{j+1} M_i} \right)^{1/2} \Delta W_{t_{j+1}}^* \text{ for } t_{n,j+1} \in (\tau_{n,i-1}, \tau_{n,i}] \end{aligned} \quad (51)$$

Formally, we define the approximation as follows.

DEFINITION 7. R_n is defined as Q_n in Definition 5, but with (51) replacing (37).

The crucial fact will be that under R_n , the observables $\Delta X_{t_{j+1}}$ are conditionally iid $N(0, \zeta_{\tau_{n,i-1}} \Delta \tau_i / M_i)$ for $t_{n,j+1} \in (\tau_{n,i-1}, \tau_{n,i}]$

We can now describe the relationship between R_n and P_n^* . In analogy with the previous proposition, we obtain that

PROPOSITION 3.

$$\begin{aligned} &\log \frac{dR_n}{dP_n^*}(U_{t_0}, \dots, U_{t_{n,j}}, \dots, U_{t_{n,n}}) \\ &= \sum_i \sum_{\tau_{i-1} \leq t_j < \tau_i} \{ \ell(\Delta X_{t_{j+1}}; \zeta_{\tau_{n,i-1}} \Delta \tau_i / M_i) - \ell(\Delta X_{t_{j+1}}; \zeta_{t_{n,j}} \Delta t_{j+1}) \} \end{aligned} \quad (52)$$

The contiguity question is then addressed as follows. Recall that

$$\log \frac{dR_n}{dP_n^*} = \log \frac{dR_n}{dQ_n} + \log \frac{dQ_n}{dP_n^*}. \quad (53)$$

Define

$$B_{n,j} = \left(\Delta t_{n,j+1} \left(\frac{\Delta \tau_{n,i}}{M_i} \right)^{-1} - 1 \right) \quad (54)$$

THEOREM 5. (*Asymptotic relationship between P_n^* , Q_n and R_n*). Assume the conditions of Theorem 3. Assume that the following limits exist:

$$\Gamma_2 = \frac{p}{2} \lim_{n \rightarrow \infty} \sum_j B_{n,j}^2 \text{ and } \Gamma_3 = \frac{p}{2} \lim_{n \rightarrow \infty} \sum_j \log(1 + B_{n,j}). \quad (55)$$

Let $Z_n^{(1)}$ and $M_n^{(1)}$ be as in Theorem 3, set

$$Z_n^{(2)} = \frac{1}{2} \sum_i \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}] } \Delta X_{t_{n,j}}^T (\zeta_{\tau_{n,i-1}}^{-1}) \Delta X_{t_{n,j}} \left(\Delta t_{n,j+1}^{-1} - \left(\frac{\Delta \tau_{n,i}}{M_i} \right)^{-1} \right), \quad (56)$$

and let $M_n^{(2)}$ be the end point of the martingale part of $Z_n^{(2)}$ (see (B.25) and (B.27) in Appendix B for the explicit formula). Then, as $n \rightarrow \infty$, $(M_n^{(1)}, M_n^{(2)})$ converges stably in law under P^* to a normal distribution with mean zero and diagonal variance matrix with diagonal elements Γ_1 and Γ_2 . Also, under P^* ,

$$\log \frac{dR_n}{dQ_n} = M_n^{(2)} + \Gamma_3 + o_p(1). \quad (57)$$

The theorem can be viewed from the angle of contiguity:

COROLLARY 2. Under the assumptions of Theorem 5, the following statements are equivalent, as $n \rightarrow \infty$:

- (i) R_n is contiguous to P_n^* .
- (ii) R_n is contiguous to Q_n .
- (iii) The following relationship holds:

$$\Gamma_3 = -\frac{1}{2} \Gamma_2. \quad (58)$$

As we shall see, the requirement (58) is a substantial restriction. Corollary 2 says that unlike the case of Q_n , inference under R_n may not give rise to desired results. Part of the probability mass under Q_n (and hence P^*) is not preserved under R_n .

To understand the requirement (58), note that

$$\frac{p}{2} \sum_j \log(1 + B_{n,j}) = -\frac{p}{4} \sum_j B_{n,j}^2 + \frac{p}{6} \sum_j B_{n,j}^3 - \dots \quad (59)$$

since $\sum_j B_{n,j} = 0$. Hence, (58) will, for example, be satisfied if $\max_j |B_{n,j}| \rightarrow 0$ as $n \rightarrow \infty$. One such example is

$$t_{n,j} = f(j/n) \text{ and } f \text{ is continuously differentiable.} \quad (60)$$

However, (60) will not hold in more general settings, as we shall see from the following examples.

EXAMPLE 4. (POISSON SAMPLING.) Suppose that the sampling time points follow a Poisson process with parameter λ . If one conditions on the number of sampling points n , these points behave like the order statistics of n uniformly distributed random variables (see, for example, Chapter 2.3 in Ross (1996)). Consider the case where $M_i = M$ for all but (possibly) the last interval in $\mathcal{H}_n$. In this case, K_n is the smallest integer larger than or equal to n/M . Let Y_i be the M -tuple $(B_j, \tau_{i-1} \leq t_j < \tau_i)$.

We now obtain, by passing between the conditional and unconditional, that $Y_1, \dots, Y_{K_n-1}$ are iid, and the distribution can be described by

$$Y_1 = M(U_{(1)}, U_{(2)} - U_{(1)}, \dots, U_{(M-1)} - U_{(M-2)}, 1 - U_{(M-1)}) - 1, \quad (61)$$

where $U_{(1)}, \dots, U_{(M-1)}$ is the order statistic of $M - 1$ independent uniform random variables on $(0, 1)$. It follows that

$$\begin{aligned} \sum_j B_{n,j}^2 &= \frac{n}{M} (M^2 E U_{(1)}^2 - 1) + o_p(n) \\ \sum_j \log(1 + B_{n,j}) &= \frac{n}{M} E \log(M U_{(1)}) + o_p(n) \end{aligned} \quad (62)$$

since $E U_{(1)}^2 = 2/(M+1)(M+2)$. Hence, both Γ_2 and Γ_3 are infinite. The contiguity between R_n and the other probabilities fails. On the other hand all our assumptions up to Section 3 are satisfied, and so P , P^* , P_n^* and Q_n are all contiguous. The AQVT (equation (24)) is given by $H(t) = 2t$. Also, if the block size is constant (size M), the ADD is $K(t) = (M-1)t$. $\square$

EXAMPLE 5. (SYSTEMATIC IRREGULARITY.) Let ϵ be a small positive number, and let $\Delta t_{n,j} = (1+\epsilon)T/n$ for odd j and $\Delta t_{n,j} = (1-\epsilon)T/n$ for even j (with $\Delta t_{n,j} = T/n$ for odd n). Again, all our assumptions up to Section 3 are satisfied. The AQVT is given by $H(t) = t(1+\epsilon^2)$. If we suppose that all $M_i = 2$, the ADD becomes $K(t) = t$. On the other hand, $B_{n,j} = \pm\epsilon$, so that, again, both Γ_2 and Γ_3 are infinite. The contiguity between R_n and the other probabilities thus fails in the same radical fashion as in the case of Poisson sampling. $\square$

4 First applications

We here discuss various applications of our theory. For simplicity, assume in following that sampling is equispaced (so $\Delta t_{n,j} = \Delta t_n = T/n$ for all j). We return to the question of irregular sampling in the context of asynchronicity, in Section 5.2. Except in Sections 4.2 and 4.4.2, we also take (X_t)

to be a scalar process. We take the block size M to be independent of i (except possibly for the first and last block, and this does not matter for asymptotics).

Define

$$\begin{aligned}\hat{\sigma}_{\tau_{n,i}}^2 &= \frac{1}{\Delta t_n(M_n - 1)} \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}] } (\Delta X_{t_{n,j+1}} - \overline{\Delta X}_{\tau_{n,i}})^2 \text{ and} \\ \overline{\Delta X}_{\tau_{n,i}} &= \frac{1}{M_n} \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}] } \Delta X_{t_{n,j+1}} = \frac{1}{M_n} (X_{\tau_{n,i}} - X_{\tau_{n,i-1}})\end{aligned}\quad (63)$$

To analyze estimators, denote by $\mathcal{Y}_{n,i}$ the information at time $\tau_{n,i}$. Note that $\mathcal{Y}_{n,i} = \mathcal{X}_{n,j}$, where j is such that $t_{n,j} = \tau_{n,i}$.

4.1 Estimation of integrals of $|\sigma_t|^r$

We return to the question of estimating

$$\theta = \int_0^T |\sigma_t|^r dt.$$

We shall not use estimators of the form $\sum_{j=1}^n |\Delta X_{t_{n,j}}|^r$, as in Example 1. We show how to get more efficient estimators by using the block approximation.

4.1.1 Analysis

We observe that under Q_n , the $\Delta X_{t_{n,j+1}}$ are iid $N(0, \sigma_{\tau_{n,i}}^2 \Delta t_n)$ within each block. From the theory of unbiased minimum variance (UMVU) estimation (see, for example, Lehmann (1983)), the optimal estimator of $|\sigma_{\tau_{n,i}}|^r$ is

$$\widehat{|\sigma_{\tau_{n,i}}|^r} = c_{M-1,r}^{-1} (\hat{\sigma}_{\tau_{n,i}}^2)^{r/2} \quad (64)$$

This also follows from sufficiency considerations. Here, $c_{M,r}$ is the normalizing constant which gives unbiasedness, namely

$$\begin{aligned}c_{M,r} &= E \left((\chi_M^2/M)^{r/2} \right) \\ &= \left(\frac{2}{M} \right)^{r/2} \frac{\Gamma(\frac{r+M}{2})}{\Gamma(\frac{M}{2})}\end{aligned}\quad (65)$$

where χ_M^2 has the standard χ^2 distribution with M degrees of freedom, and Γ is the Gamma function.

Our estimator of θ (which is blockwise UMVU under Q_n) therefore becomes

$$\hat{\theta}_n = (M \Delta t) \sum_i \widehat{|\sigma_{\tau_{n,i}}|^r} \quad (66)$$

It is easy to see that $\hat{\theta}_n$ asymptotically has no covariation with any of the Hermite polynomials in Section 3.3, and so, by standard arguments,

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0, 1) \left(TM \left(\frac{c_{M-1,2r}}{c_{M-1,r}^2} - 1 \right) \int_0^T \sigma_t^{2r} dt \right)^{1/2} \quad (67)$$

stably in law, under P (and P^* , P_n^* , and Q_n). This is because, under Q_n ,

$$\begin{aligned} \text{Var}((M\Delta t) \sum_i (|\widehat{\sigma_{\tau_{n,i}}}|^r) \mid \mathcal{Y}_{n,i}) &= \sigma_{\tau_{n,i}}^{2r} (M\Delta t)^2 c_{M-1,r}^{-2} \text{Var} \left((\chi_{M-1}^2 / (M-1))^{r/2} \right) \\ &= \sigma_{\tau_{n,i}}^{2r} (M\Delta t) \frac{TM}{n} \left(\frac{c_{M-1,2r}}{c_{M-1,r}^2} - 1 \right). \end{aligned} \quad (68)$$

REMARK 8. (NOT TAKING OUT THE MEAN). One can replace $\hat{\sigma}_{\tau_{n,i}}^2$ by

$$\tilde{\sigma}_{\tau_{n,i}}^2 = \frac{1}{\Delta t_n M_n} \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}] } (\Delta X_{t_{n,j+1}})^2, \quad (69)$$

and take

$$\widetilde{|\sigma_{\tau_{n,i}}|^r} = c_{M,r}^{-1} (\tilde{\sigma}_{\tau_{n,i}}^2)^{r/2} \quad (70)$$

and define $\tilde{\theta}_n$ accordingly. The above analysis goes through. The (random) asymptotic variance becomes

$$TM \left(\frac{c_{M,2r}}{c_{M,r}^2} - 1 \right) \int_0^T \sigma_t^{2r} dt. \quad (71)$$

□

4.1.2 Asymptotic Efficiency

We note that for large M ,

$$\text{asymptotic variance of } n^{1/2}(\hat{\theta}_n - \theta) \downarrow T \frac{r^2}{2} \int_0^T \sigma_t^{2r} dt. \quad (72)$$

This is also the minimal asymptotic variance of the parametric MLE when σ^2 is constant. Thus, by choosing M largeish, say $M = 20$, one can get close to parametric efficiency (see Figure 1).

To see the gain from the procedure, compare to the asymptotic variance of the estimator in Example 1, which can be written as $T \left(\frac{c_{1,2r}}{c_{1,r}^2} - 1 \right) \int_0^T \sigma_t^{2r} dt$. Compared to the variance in (72), the earlier estimator has asymptotic relative efficiency (ARE)

$$\begin{aligned} ARE(\text{estimator from Example 1}) &= \frac{\text{asymptotic variance in (72)}}{\text{asymptotic variance of estimator from Example 1}} \\ &= \frac{r^2}{2} \left(\frac{c_{1,2r}}{c_{1,r}^2} - 1 \right)^{-1} \end{aligned} \quad (73)$$

Note that except for $r = 2$, $ARE < 1$. Figure 1 gives a plot of the ARE as a function of r . As one can see, there can be substantial gain from using the proposed estimator (66).

REMARK 9. In terms of asymptotic distribution, there is further gain in using the estimator from Remark 8. Specifically, $ARE_M(\tilde{\theta})/ARE_M(\hat{\theta}) = M/(M - 1)$. This is borne out by Figure 1. However, it is likely that the drift μ , as well as the block size M , would show up in a higher order bias calculation. This would make $\tilde{\sigma}$ less attractive. In connection with estimating the leverage effect, it is crucial to use $\hat{\sigma}$ rather than $\tilde{\sigma}$, cf. Section 4.3.2. $\square$

REMARK 10. We emphasize again that M has to be fixed in the present calculation, so that the ideal asymptotic variance on the right hand side of (72) is only approximately attained. It would be desirable to build a theory where $M \rightarrow \infty$ as $n \rightarrow \infty$. Such a theory would presumably be able to pick up any biases due to the blocking. $\square$

Figure 1. ASYMPTOTIC RELATIVE EFFICIENCY (ARE) OF THREE ESTIMATORS OF $\theta = \int_0^T |\sigma|_t^r dt$, AS A FUNCTION OF r . THE DOTTED CURVE CORRESPONDS TO THE TRADITIONAL ESTIMATOR, WHICH IS PROPORTIONAL TO $\sum_{j=1}^n |\Delta X_{t_{n,j}}|^r$. THE SOLID AND DASHED LINES ARE THE ARE'S OF THE BLOCK BASED ESTIMATORS USING, RESPECTIVELY, $\hat{\sigma}$ (SOLID) AND $\tilde{\sigma}$ (DASHED). BLOCK SIZES $M = 20$ AND $M = 100$ ARE GIVEN. THE IDEAL VALUE IS $ARE = 1$. BLOCKING IS SEEN

TO IMPROVE EFFICIENCY, ESPECIALLY AWAY FROM $r = 2$. THERE IS SOME COST TO REMOVING THE MEAN IN EACH BLOCK (THE DIFFERENCE BETWEEN THE DASHED AND THE SOLID CURVE).

4.2 Integrated betas

Consider processes $X_t^{(1)}, \dots, X_t^{(p)}$ and Y_t which are observed synchronously at times $0 = t_{n,0} < t_{n,1} < \dots < t_{n,n_1} = T$. Suppose that these processes are related by

$$dY_t = \sum_{k=1}^p \beta_t^{(k)} dX_t^{(k)} + dZ_t, \text{ with } \langle X^{(k)}, Z \rangle_t = 0 \text{ for all } t \text{ and } k. \quad (74)$$

We consider the question of estimating $\theta^{(k)} = \int_0^T \beta_t^{(k)} dt$. This estimation problem is conceptually closely related to the realized regressions studied in Barndorff-Nielsen and Shephard (2004a) and Dovonon, Goncalves, and Meddahi (2008). The ANOVA in Mykland and Zhang (2006) is concerned with the residuals in this same model.

Under the approximation Q_n , in each block $\tau_{n,i-1} < t_{n,j} \leq \tau_{n,i}$ the regression (74) becomes, for the observables,

$$\Delta Y_{t_{n,j}} = \sum_{k=1}^p \beta_{\tau_{n,i-1}}^{(k)} \Delta X_{t_{n,j}}^{(k)} + \Delta Z_{t_{n,j}}. \quad (75)$$

It is therefore natural to take the estimator $(\hat{\beta}_{\tau_{n,i-1}}^{(1)}, \dots, \hat{\beta}_{\tau_{n,i-1}}^{(p)})$ of $(\beta_{\tau_{n,i-1}}^{(1)}, \dots, \beta_{\tau_{n,i-1}}^{(p)})$ to be the regular least squares estimator (without intercept) based on the observables $(\Delta X_{t_{n,j}}^{(1)}, \dots, \Delta X_{t_{n,j}}^{(p)}, \Delta Y_{t_{n,j}})$ inside the block. The overall estimate of the vector of θ 's is then

$$\hat{\theta}_n^{(k)} = \sum_i \hat{\beta}_{\tau_{n,i-1}}^{(k)} M \Delta t. \quad (76)$$

From the unbiasedness of linear regression, we inherit that $n^{1/2}(\hat{\theta}_n - \theta)$ is the end point of an $(\mathcal{Y}_{n,i}, Q_n)$ martingale, with discrete time quadratic covariation matrix

$$n(M \Delta t)^2 \sum_i \text{Cov}_{Q_n}(\hat{\beta}_{\tau_{n,i-1}} - \beta_{\tau_{n,i-1}} \mid \mathcal{Y}_{n,i-1}). \quad (77)$$

To see how the martingale property follows, let $\mathcal{Y}'_{n,i-1}$ be the smallest sigma-field containing $\mathcal{Y}_{n,i-1}$ and $\sigma(\Delta X_{t_{n,j}}, \tau_{n,i-1} < t_{n,j} \leq \tau_{n,i})$. The precise implication of the classical unbiasedness is that $E_{Q_n}(\hat{\beta}_{\tau_{n,i-1}} - \beta_{\tau_{n,i-1}} \mid \mathcal{Y}'_{n,i-1}) = 0$, whence the stated martingale property follows by the law of iterated expectations (or tower property).

To compute (77), note that from standard regression theory (see, *e.g.*, p. 44 in Weisberg (1985))

$$\text{Cov}_{Q_n}(\hat{\beta}_{\tau_{n,i-1}} - \beta_{\tau_{n,i-1}} \mid \mathcal{Y}'_{n,i-1}) = \text{Var}_{Q_n}(\Delta Z_{t_{n,j}} \mid \mathcal{Y}'_{n,i-1}) \times (\Delta X^T \Delta X)^{-1} \quad (78)$$

where, with some abuse of notation, ΔX is the matrix of $\Delta X_{t_{n,j}}^{(k)}$, where $k = 1, \dots, p$, and the $t_{n,j}$ are in block number i . Now observe that under Q_n , the conditional distribution of ΔX given $\mathcal{Y}_{n,i-1}$ is

that of M independent rows, each row being a p -variate normal distribution with mean zero and covariance matrix $\langle X, X \rangle'_{\tau_{n,i-1}} \Delta t_n$. (Recall that “prime” here denotes differentiation w.r.t. time t). Hence, $\Delta X^T \Delta X$ has Wishart distribution with scale matrix $\langle X, X \rangle'_{\tau_{n,i-1}} \Delta t_n$, and M degrees of freedom. (We refer to p. 66 of Mardia, Kent, and Bibby (1979) for the definition of the Wishart distribution). It follows that (*ibid.*, p. 85)

$$E_{Q_n}((\Delta X^T \Delta X)^{-1} \mid \mathcal{Y}_{n,i-1}) = (\langle X, X \rangle'_{\tau_{n,i-1}})^{-1} \Delta t_n^{-1} / (M - p - 1). \quad (79)$$

Since $\text{Var}_{Q_n}(\Delta Z_{t_{n,j}} \mid \mathcal{Y}'_{n,i-1}) = \langle Z, Z \rangle'_{\tau_{n,i-1}} \Delta t_n$, we finally get that

$$\text{Cov}_{Q_n}(\hat{\beta}_{\tau_{n,i-1}} - \beta_{\tau_{n,i-1}} \mid \mathcal{Y}_{n,i-1}) = \langle Z, Z \rangle'_{\tau_{n,i-1}} (\langle X, X \rangle'_{\tau_{n,i-1}})^{-1} / (M - p - 1). \quad (80)$$

It follows that the limit of (77) is

$$\frac{MT}{M - p - 1} \int_0^T \langle Z, Z \rangle'_t (\langle X, X \rangle'_t)^{-1} dt. \quad (81)$$

For the same reasons as in Sections 2.5 and 4.1 it then follows that $n^{1/2}(\hat{\theta}_n - \theta)$ converges stably to a multivariate mixed normal distribution, with mean zero and covariance matrix given by (81), under all of Q_n , P_n^* , P^* , and P .

4.3 Estimation of Leverage Effect

We here seek to estimate $\langle \sigma^2, X \rangle_T$. We have seen in Example 3 that this quantity can appear in asymptotic distributions, and we shall here see how the sum of third powers can be refined into an estimate of this quantity.

The natural estimator would be

$$\widetilde{\langle \sigma^2, X \rangle}_T = \sum_i (\hat{\sigma}_{\tau_{n,i+1}}^2 - \hat{\sigma}_{\tau_{n,i}}^2) (X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \quad (82)$$

where $\hat{\sigma}_{\tau_{n,i}}^2$ and $\overline{\Delta X}_{\tau_{n,i}}$ are given above in (63). Section 4.3.1 shows that this estimator is asymptotically biased. Accordingly, we define an asymptotically unbiased estimator of leverage effect by

$$\widehat{\langle \sigma^2, X \rangle}_T = \frac{2M}{2M + 3} \sum_i (\hat{\sigma}_{\tau_{n,i+1}}^2 - \hat{\sigma}_{\tau_{n,i}}^2) (X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \quad (83)$$

In other words, $\widehat{\langle \sigma^2, X \rangle}_T = \frac{2M}{2M+3} \widetilde{\langle \sigma^2, X \rangle}_T$. Following Proposition 4 in Section 4.3.1,

$$\widehat{\langle \sigma^2, X \rangle}_T - \langle \sigma^2, X \rangle \xrightarrow{\mathcal{L}} c_M^{1/2} N(0, 1) \quad (84)$$

stably under P^* and P , where

$$c_M = \frac{4}{M - 1} \left(\frac{2M}{2M + 3} \right)^2 \int_0^T \sigma_t^6 dt. \quad (85)$$

It is important to note that the bias comes from error induced by the one period discretization (the adjustment from P^* to P_n^*). Thus, this is an instance where naïve discretization does not work.

For fixed M , the estimator $\widehat{\langle \sigma^2, X \rangle}_T$ is not consistent. By choosing large M , however, one can make the error as small as one wishes.

REMARK 11. It is conjectured that there is an optimal rate of $M = O(n^{1/2})$ as $n \rightarrow \infty$. The presumed optimal convergence rate is $n^{1/4}$, in analogy with the results in Zhang (2006). This makes sense because $\hat{\sigma}_t^2$ is a noisy measurement of σ_t^2 . The problem of estimating $\langle \sigma^2, X \rangle_T$ is therefore similar to estimating volatility in the presence of microstructure noise.

It would clearly be desirable to have a theory for the case where M depends on n , but this is beyond the scope of this paper. $\square$

4.3.1 Analysis

We here show how to arrive at the final result (84). This serves as a fairly extensive illustration of how to apply the theory development in the earlier sections.

By rearranging terms, write

$$\begin{aligned} \widetilde{\langle \sigma^2, X \rangle}_T &= \sum_i (\sigma_{\tau_{n,i+1}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \\ &\quad + \sum_i (\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}) \\ &\quad - \sum_i (\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) + O_p(n^{-1}), \end{aligned} \quad (86)$$

where the $O_p(n^{-1})$ term comes from edge effects. Note that by conditional Gaussianity, both the two last sums in (86) are Q_n -martingales with respect to the sigma-fields $\mathcal{Y}_{n,i}$. They are also orthogonal, in the sense that

$$\text{Cov}_n^Q((\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}), (\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \mid \mathcal{Y}_{n,i}) = 0 \quad (87)$$

Under Q_n and conditionally on the information up to time $\tau_{n,i-1}$, $\hat{\sigma}_{\tau_{n,i}}^2 = \sigma_{\tau_{n,i}}^2 \chi_{M-1}^2 / (M-1)$ and $\overline{\Delta X}_{\tau_{n,i}} = \sigma_{\tau_{n,i}}(\Delta t/M)^{1/2} N(0,1)$, where χ_{M-1}^2 and $N(0,1)$ are independent. It follows that

$$\begin{aligned} \text{Var}_n^Q((\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}) \mid \mathcal{Y}_{n,i}) &= \sigma_{\tau_{n,i}}^4 (M-1)^{-2} (X_{\tau_{n,i}} - X_{\tau_{n,i-1}})^2 \text{Var}(\chi_{M-1}^2) \\ &= 2\sigma_{\tau_{n,i}}^4 (M-1)^{-1} (X_{\tau_{n,i}} - X_{\tau_{n,i-1}})^2 \end{aligned} \quad (88)$$

Hence, under Q_n , the quadratic variation of $\sum_i (\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})$ converges to

$$\frac{2}{M-1} \int_0^T \sigma_t^6 dt. \quad (89)$$

At the same time, it is easy to see that this sum has asymptotically zero covariation with the increments of $M_n^{(0,Q)}$ and $M_n^{(1,Q)}$, and also with W^Q . Hence $\sum_i (\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})$ converges stably under P to a normal distribution with mean zero and variance (89).

The situation with the other sum $\sum_i (\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})$ is more complicated. First of all,

$$\begin{aligned} & \text{Var}_n^Q((\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \mid \mathcal{Y}_{n,i}) \\ &= \sigma_{\tau_{n,i}}^6 (M\Delta t) \text{Var}((\frac{\chi_{M-1}^2}{M-1} - 1)N(0,1)) \\ &= \frac{2}{M-1} \sigma_{\tau_{n,i}}^6 (M\Delta t). \end{aligned} \quad (90)$$

Hence the asymptotic quadratic variation is

$$\frac{2}{M-1} \int_0^T \sigma_t^6 dt. \quad (91)$$

The sum is asymptotically uncorrelated with $M_n^{(1,Q)}$ and W^Q . To see the latter, note that

$$\begin{aligned} & \text{Cov}_n^Q((\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), W_{\tau_{n,i+1}} - W_{\tau_{n,i}} \mid \mathcal{Y}_{n,i}) \\ &= \sigma_{\tau_{n,i}}^3 (M\Delta t) \text{Cov}((\frac{\chi_{M-1}^2}{M-1} - 1)N(0,1), N(0,1)) \\ &= 0. \end{aligned} \quad (92)$$

Overall, under both Q_n and P_n^* ,

$$\sum_i (\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \xrightarrow{\mathcal{L}} N(0,1) \left(\frac{2}{M-1} \int_0^T \sigma_t^6 dt \right)^{1/2}, \quad (93)$$

stably.

There is, however, covariation between this sum and $M_n^{(0,Q)}$. It is shown below in Remark 12 (see equation (101)) that $A_{12} = \frac{3}{2M} \langle \sigma^2, X \rangle_T$, where A_{12} has the same meaning as in Theorems 2 and 4 (in Sections 2.4 and 3.4, respectively). Thus, by these theorems, under P^* , we have (stably)

$$\sum_i (\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \xrightarrow{\mathcal{L}} \frac{3}{2M} \langle \sigma^2, X \rangle_T + N(0,1) \left(\frac{2}{M-1} \int_0^T \sigma_t^6 dt \right)^{1/2}. \quad (94)$$

Because of the orthogonality (87), and since $\sum_i (\sigma_{\tau_{n,i+1}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) - \langle \sigma^2, X \rangle_T = O_p(n^{-1/2})$ by Proposition 1 of Mykland and Zhang (2006), it follows that $\widehat{\langle \sigma^2, X \rangle_T} - \langle \sigma^2, X \rangle_T$ converges stably (under P^*) to a normal distribution with mean as in equation (94), and variance contributed by the second and third terms on the right hand side of (86). In other words:

PROPOSITION 4. *In the equally spaced case, under both P^* and P , and as $n \rightarrow \infty$,*

$$\langle \widetilde{\sigma^2}, X \rangle_T \xrightarrow{\mathcal{L}} \left(1 + \frac{3}{2M}\right) \langle \sigma^2, X \rangle + N(0, 1) \times \left(\frac{4}{M-1} \int_0^T \sigma_t^6 dt\right)^{1/2} \quad (95)$$

stably in law, where $N(0, 1)$ is independent of $\mathcal{F}_T$.

This shows the result (84).

REMARK 12. (Sample of calculation). To see how the reasoning works in the case of covariations, consider the case of covariation between $\sum_i (\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})$ and $M_n^{(0,Q)}$. We proceed as follows.

If h_r is the r 'th (scalar) Hermite polynomial, set

$$G_{r,i} = \sum_{t_{n,j} \in (\tau_{n,i}, \tau_{n,i+1}]} h_r(\Delta W_{t_{n,j+1}}^Q / \Delta t^{1/2}), \quad (96)$$

note that

$$\begin{aligned} X_{\tau_{n,i+1}} - X_{\tau_{n,i}} &= \sigma_{\tau_{n,i}} \Delta t^{1/2} G_{1,i} \text{ and} \\ \hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2 &= \frac{\sigma_{\tau_{n,i}}^2}{M-1} \left(G_{2,i} - \frac{1}{M} G_{1,i}^2 + 1 \right) \end{aligned} \quad (97)$$

At the same time,

$$M_n^{(0,Q)} = \frac{1}{12} (\Delta t)^{1/2} \sum_i k_{\tau_{n,i}} G_{3,i} + o_p(1) \quad (98)$$

The covariance for each i -increment becomes

$$\begin{aligned} &\text{Cov}_n^Q((\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \frac{1}{12} (\Delta t)^{1/2} k_{\tau_{n,i}} G_{3,i} \mid \mathcal{Y}_{n,i}) \\ &= \frac{1}{12} \Delta t \frac{k_{\tau_{n,i}} \sigma_{\tau_{n,i}}^3}{M-1} \text{Cov}_n^Q((G_{2,i} - \frac{1}{M} G_{1,i}^2 + 1) G_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\ &= \frac{1}{2} (M \Delta t) \frac{k_{\tau_{n,i}} \sigma_{\tau_{n,i}}^3}{M} \end{aligned} \quad (99)$$

since, by orthogonality of the Hermite polynomials, and by normality,

$$\begin{aligned} &\text{Cov}_n^Q((G_{2,i} - \frac{1}{M} G_{1,i}^2 + 1) G_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\ &= \text{cum}_{3,n}^Q(G_{1,i}, G_{2,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) - \frac{1}{M} \text{cum}_{4,n}^Q(G_{1,i}, G_{1,i}, G_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\ &= M \text{cum}_3(h_1(N(0, 1)), h_2(N(0, 1)), h_3(N(0, 1))) \\ &\quad - \text{cum}_4(h_1(N(0, 1)), h_1(N(0, 1)), h_1(N(0, 1)), h_3(N(0, 1))) \\ &= 6(M-1). \end{aligned} \quad (100)$$

The covariation with $M_n^{(0,Q)}$ therefore converges to

$$\begin{aligned} A_{12} &= \frac{1}{2M} \int_0^T k_t \sigma_t^3 dt \\ &= \frac{3}{2M} \langle \sigma^2, X \rangle_T \end{aligned} \quad (101)$$

as in (31). $\square$

4.3.2 The rôle of μ

In the development above, the drift μ did not surface. The following example gives evidence that the drift can matter. We shall see that if one does not take out the drift when estimating σ^2 , μ can appear in the asymptotic bias.

EXAMPLE 6. (NOT REMOVING THE MEAN FROM THE ESTIMATE OF σ^2). Suppose that one wishes to use the estimator (82), but replacing $\hat{\sigma}_{\tau_{n,i}}^2$ by the estimator $\tilde{\sigma}_{\tau_{n,i}}^2$ from (69). An estimator analogous to $\widetilde{\langle \sigma^2, X \rangle}_T$ is then

$$\widetilde{\langle \sigma^2, X \rangle}_T^{\text{with mean}} = \sum_i (\tilde{\sigma}_{\tau_{n,i+1}}^2 - \tilde{\sigma}_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \quad (102)$$

We have the representation

$$\tilde{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2 = \frac{\sigma_{\tau_{n,i}}^2}{M} G_{2,i}. \quad (103)$$

We now consider the terms in (86). First note that $\sum_i (\tilde{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})$ is unaffected by this change. However, this is not true for the term $\sum_i (\tilde{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})$, which we analyze in the following.

Since $\text{cum}_3(h_1(N(0,1)), h_2(N(0,1)), h_3(N(0,1))) = 6$, it now follows as in (99) that

$$\begin{aligned} \text{Cov}_n^Q((\tilde{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \frac{1}{12} \Delta t^{1/2} k_{\tau_{n,i}} G_{3,i} \mid \mathcal{Y}_{n,i}) &= \Delta t \frac{k_{\tau_{n,i}} \sigma_{\tau_{n,i}}^3}{2M} \\ &= \Delta t \frac{3 \langle \sigma^2, X \rangle'_{\tau_{n,i}}}{2M}. \end{aligned} \quad (104)$$

More importantly,

$$\begin{aligned}
& \text{Cov}_n^Q((\tilde{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), W_{\tau_{n,i+1}}^Q - W_{\tau_{n,i}}^Q \mid \mathcal{Y}_{n,i}) \\
&= \text{Cov}_n^Q\left(\frac{\sigma_{\tau_{n,i}}^2}{M} G_{2,i} \sigma_{\tau_{n,i}} \Delta t^{1/2} G_{1,i}, \Delta t^{1/2} G_{1,i} \mid \mathcal{Y}_{n,i}\right) \\
&= \Delta t \frac{\sigma_{\tau_{n,i}}^3}{M} \text{Cov}_n^Q(G_{2,i} G_{1,i}, G_{1,i} \mid \mathcal{Y}_{n,i}) \\
&= \Delta t \sigma_{\tau_{n,i}}^3 \text{Cov}(h_2(N(0,1))h_1(N(0,1)), h_1(N(0,1))) \\
&= 2\Delta t \sigma_{\tau_{n,i}}^3.
\end{aligned} \tag{105}$$

Finally,

$$\begin{aligned}
& \text{Var}_n^Q((\tilde{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \mid \mathcal{Y}_{n,i}) \\
&= \Delta t \frac{\sigma_{\tau_{n,i}}^6}{M^2} \text{Var}(G_{2,i} G_{1,i} \mid \mathcal{Y}_{n,i}) \\
&= (2 + 6M^{-1}) \Delta t \sigma_{\tau_{n,i}}^6
\end{aligned} \tag{106}$$

Since the term $\sum_i (\tilde{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2)(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})$ remains a martingale, we therefor obtain that under Q_n ,

$$\widetilde{\langle \sigma^2, X \rangle}_T^{\text{with mean}} \xrightarrow{\mathcal{L}} \langle \sigma^2, X \rangle_T + \frac{2}{M} \int_0^T \sigma^3 d\check{W} + N(0,1) \left(2 \frac{M+1}{M} \int_0^T \sigma^6 dt \right)^{1/2}. \tag{107}$$

stably in law. From (104), we obtain from Theorem 4 that under P^*

$$\frac{2M}{2M+3} \widetilde{\langle \sigma^2, X \rangle}_T^{\text{with mean}} \xrightarrow{\mathcal{L}} \langle \sigma^2, X \rangle_T + \frac{4}{2M+3} \int_0^T \sigma^3 dW^* + N(0,1) \left(8 \frac{M(M+1)}{(2M+3)^2} \int_0^T \sigma^6 dt \right)^{1/2} \tag{108}$$

We now, therefore, have an asymptotic bias. If one replaces P^* by P , we get that the bias becomes

$$\frac{4}{2M+3} \int_0^T \sigma^3 (dW + \sigma_t^{-1} \mu_t dt). \tag{109}$$

□

4.4 Other examples

We here summarize two additional examples of application that have been studied more carefully elsewhere.

4.4.1 Realized quantile-based estimation of integrated volatility

This methodology has been studied in a recent paper by Christensen, Oomen, and Podolskij (2008). In the case of fixed block size and no micro-structure, their results (Theorem 1-2) can be deduced

from Theorem 1-3 of this paper. The key observation is that if V is the k 'th quantile among $\Delta X_{t_{n,j}}$, with $\tau_{n,i-1} < t_{n,j} \leq \tau_{n,i}$, then $E_{Q_n}(V^2 \mid \mathcal{Y}_{n,i-1}) = \sigma_{\tau_{n,i-1}}^2 EU_{(k)}^2$, where $U_{(k)}$ is the k 'th quantile of M iid standard normal random variables. Blockwise L-statistics can be constructed similarly.

We emphasize that the paper by Christensen, Oomen, and Podolskij (2008) goes much further in developing the quantile-based estimation technology, including increasing block size and allowing for micro-structure.

4.4.2 ANOVA (Analysis of variance/variation)

A related problem to the one discussed above in Section 4.2 is that of analysis of variance/variation (Zhang (2001) and Mykland and Zhang (2006)). We are again in the situation of the regression (74), but now the purpose is to estimate $\langle Z, Z \rangle_T$, *i.e.*, the residual quadratic variation of Y after regressing on X . The use of blocking to study this problem is discussed in the conditionally Gaussian case in Mykland (2006). It is straightforward to see that the conditions of Theorem 1-3 can be verified using similar techniques to those used in the earlier paper. The conditional convergence of Theorem 7 of that paper becomes stable convergence in the more general setting of this current article. The asymptotic variance remains the same.

4.5 Abstract summary of applications

We here summarize the procedure which is implemented in the applications section above. We remain in the scalar case.

In the type of problems we have considered, the parameter θ to be estimated can be written as

$$\theta = \sum_i \theta_{n,i} + O_p(n^{-1}) \quad (110)$$

where, under the approximating measure, $\theta_{n,i}$ is approximately an integral from $\tau_{n,i-1}$ to $\tau_{n,i}$. Estimators are of the form

$$\hat{\theta}_n = \sum_i \hat{\theta}_{n,i}, \quad (111)$$

where $\hat{\theta}_{n,i}$ uses M or (in the case of the leverage effect) $2M$ increments. If one sets $Z_{n,i} = n^\alpha(\hat{\theta}_{n,i} - \theta_{n,i})$, we need that $Z_{n,i}$ is a martingale under Q_n . α can be 0, 1/2 or any other number smaller than 1. We then show in each individual case that, in probability,

$$\begin{aligned} \sum_i \text{Var}_n^Q(Z_{n,i} \mid \mathcal{Y}_{n,i}) &\rightarrow \int_0^T f_t^2 dt \\ \sum_i \text{Cov}_n^Q(Z_{n,i}, W_{\tau_{n,i+1}}^Q - W_{\tau_{n,i}}^Q \mid \mathcal{Y}_{n,i}) &\rightarrow \int_0^T g_t dt \end{aligned} \quad (112)$$

for some functions (processes) f_t and g_t . We also find the following limits in probability:

$$A_{12} = \frac{1}{12} \lim_{n \rightarrow \infty} \sum_i \text{Cov}_n^Q \left(Z_{n,i}, \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} (\Delta t_{n,j+1})^{1/2} k_{t_{n,j}} h_3(\Delta W_{t_{n,j+1}}^Q / (\Delta t_{n,j+1})^{1/2}) \mid \mathcal{Y}_{n,i} \right)$$

and

$$A_{13} = -\frac{1}{2} \lim_{n \rightarrow \infty} \sum_i \text{Cov}_n^Q \left(Z_{n,i}, \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} \left(\sigma_{\tau_{n,i-1}}^2 (\zeta_{t_{n,j}}^{-1} - \zeta_{\tau_{n,i-1}}^{-1}) h_2(\Delta W_{t_{n,j+1}}^Q / (\Delta t_{n,j+1})^{1/2}) \right) \mid \mathcal{Y}_{n,i} \right). \quad (113)$$

We finally obtain

THEOREM 6. *(Summary of method in the scalar case). In the setting described, and subject to regularity conditions,*

$$n^\alpha (\hat{\theta}_n - \theta_n) \xrightarrow{\mathcal{L}} b + A_{12} + A_{13} + N(0, 1) \left(\int_0^T (f_t^2 - g_t^2) dt \right)^{1/2} \quad (114)$$

stably in law under P^* and P , with $N(0, 1)$ independent of $\mathcal{F}_T$. b is given by

$$b = \int_0^T g_t dW_t^* = \int_0^T g_t (dW_t + \sigma_t^{-1} \mu_t dt). \quad (115)$$

5 Dealing with more complex cases

5.1 Moving windows

The paper so far has considered chopping n data up into non-overlapping windows of size M each. We here show by example that the methodology can be adapted to the moving window case. We consider two examples: bipower, and the estimation of $\theta = \int_0^T |\sigma_t|^p dt$, as in Section 4.1. It should be noted that the moving window is close to the concept of a moving kernel, and this may be a promising avenue of further investigation. See, in particular, Linton (2007).

5.1.1 Bipower

We assume equidistant observations as in Section 4. The bipower estimator of $\theta = \int_0^T \sigma_t^2 dt$ (Barndorff-Nielsen and Shephard (2004b)) is given by

$$\hat{\theta} = \frac{\pi}{2} \sum_{i=1}^{n-1} |\Delta X_{t_{n,i}}| |\Delta X_{t_{n,i+1}}|. \quad (116)$$

The asymptotic of this estimator is well known from Barndorff-Nielsen and Shephard (2004b, 2006), and Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006)), so we are not here about to say anything new. However, the estimator (116) gives rise to a nice illustration of how to adapt the current technology to moving windows, and for this reason we present this example before approaching the more complicated discussion in Section 5.1.2.

The first thing to observe is that bipower is naturally based on blocks, of size two. Thus, if one sums over every second term in (116)), the estimator can be trivially analyzed using the theorems in this paper. From sufficiency considerations, however, this would be less efficient than the usual bipower estimator.

To cope with this problem, we use two block sizes, 2 and $M \gg 2$. Decompose

$$\begin{aligned}\hat{\theta}_n &= \hat{\theta}_n^{(1)} + \hat{\theta}_n^{(2)}, \text{ where} \\ \hat{\theta}_n^{(1)} &= \frac{\pi}{2} \sum_k \sum_{i=kM+1}^{(k+1)M-1} |\Delta X_{t_n,i}| |\Delta X_{t_n,i+1}| \text{ and} \\ \hat{\theta}_n^{(2)} &= \frac{\pi}{2} \sum_k |\Delta X_{t_n,kM}| |\Delta X_{t_n,kM+1}|.\end{aligned}$$

For the first term, note that since $E|N(0,1)| = (2/\pi)^{1/2}$,

$$n^{1/2} \sum_k \left(\frac{\pi}{2} \sum_{i=kM+1}^{(k+1)M-1} |\Delta X_{t_n,i}| |\Delta X_{t_n,i+1}| - \frac{M-1}{M} \sigma_{\tau_n,k}^2 \right)$$

is a sum of $(\mathcal{Y}_{n,k}, Q_n)$ martingale increments with quadratic variation

$$\begin{aligned}n \sum_k \text{Var}_{Q_n} \left(\frac{\pi}{2} \sum_{i=kM+1}^{(k+1)M-1} |\Delta X_{t_n,i}| |\Delta X_{t_n,i+1}| - \frac{M-1}{M} \sigma_{\tau_n,k}^2 \mid \mathcal{Y}_{n,k} \right) \\ = n \sum_k \sigma_{\tau_n,k}^4 \left(\frac{\pi}{2} \right)^2 \Delta t_n^2 \text{Var} \left(\sum_{i=1}^{M-1} |U_i| |U_{i+1}| \right) \\ \rightarrow \frac{T}{M} \left(\frac{\pi}{2} \right)^2 \text{Var} \left(\sum_{i=1}^{M-1} |U_i| |U_{i+1}| \right) \int_0^T \sigma_t^4 dt,\end{aligned} \tag{117}$$

in probability as $n \rightarrow \infty$, where the U_i are iid $N(0,1)$. By the same arguments as before, we thus get that $n^{1/2}(\hat{\theta}_n^{(1)} - \theta(M-1)/M)$ converges stably under Q_n to a mixed normal distribution with mean zero and variance (117). It is easy to see that there is no adjustment to P_n^* , P^* , or P .

Similar arguments, using a first block of size one, and then blocks of size M , yield that $n^{1/2}(\hat{\theta}_n^{(2)} - \theta/M)$ converges stably, to a variance which is of $o_p(1)$ as $M \rightarrow \infty$ (M is sent to infinity *after* n .) Standard weak convergence arguments involving tightness and subsequences of subsequences thus

yield that $n^{1/2}(\hat{\theta}_n^{(1)} - \theta)$ converges stably under P to a mixed normal limit with mean zero and a random variance which is the limit of (117) as $M \rightarrow \infty$, in other words,

$$T \left(\frac{\pi}{2} \right)^2 \frac{\pi^2 + 4\pi - 12}{\pi^2} \int_0^T \sigma_t^4 dt, \quad (118)$$

since $\text{Var}(|U_1||U_2|) + 2\text{Cov}(|U_1||U_2|, |U_2||U_3|) = (\pi^2 + 4\pi - 12)/\pi^2$.

As mentioned above, this result is well known from the papers cited. Most cases of overlapping blocks, however, can be handled in similar fashion. We now approach a more difficult situation of this sort.

5.1.2 Return to the estimation of integrals of $|\sigma_t|^r$

We again consider the estimation of $\theta = \int_0^T |\sigma_t|^r dt$, as in Section 4.1 (and we also use the notation from there).

We still use block length M , and we use for simplicity the estimator $\tilde{\sigma}^2$ from Remark 8. To be precise, $\tilde{\sigma}_{t_{n,i}}^2$ is now the estimate of $\sigma_{t_{n,i}}^2$ based on $\Delta X_{t_{n,j+1}}$, $j = i, \dots, i + M - 1$. The moving window estimate of θ is now

$$\tilde{\theta}_n^{MW} = (\Delta t) \sum_{i=0}^{n-M} \widetilde{|\sigma_{t_{n,i}}|^r}.$$

It is easy to see that

$$\tilde{\theta}_n^{MW} = \frac{1}{M} \sum_{m=1}^M \tilde{\theta}_{n,m} + O_p(n^{-1}),$$

where $\tilde{\theta}_{n,m}$ is the non-overlapping block estimator, with block number one starting at $t_{n,m}$. In view of this representation, it is once again clear from sufficiency considerations that the moving window estimator will have an asymptotic variance which is smaller (or, at least, no larger) than the estimator based on non-overlapping blocks. We now carry out the precise asymptotic analysis.

To analyze this estimator, let $\mathcal{M} > M$, and let $A_n = \{i = 0, \dots, n - M : [t_{n,i}, t_{n,i+M}] \subseteq [k\mathcal{M}, (k+1)\mathcal{M}] \text{ for some } k\}$, with $B_n = \{0, \dots, n - M\} - A_n$. Write

$$\begin{aligned} n^{1/2}(\tilde{\theta}_n^{MW} - \theta) &= n^{1/2} \Delta t \sum_k \sum_{i: [t_{n,i}, t_{n,i+M}] \subseteq [k\mathcal{M}, (k+1)\mathcal{M}]} (\widetilde{|\sigma_{t_{n,i}}|^r} - |\sigma_{t_{k\mathcal{M}}}|^r) \\ &\quad + n^{1/2} \Delta t \sum_{i \in B_n} (\widetilde{|\sigma_{t_{n,i}}|^r} - |\sigma_{t_{n,i}}|^r) + O_p(n^{-1/2}). \end{aligned} \quad (119)$$

Now apply Theorems 2 and 3 *with block size $\mathcal{M}$* to the first term in (119). Under this block approximation, the inner sum in the first term is based on conditionally i.i.d. observations, in fact,

for $[t_{n,i}, t_{n,i+M}] \subseteq [k\mathcal{M}/n, (k+1)\mathcal{M}/n]$, $\tilde{\sigma}_{t_{n,i}}^2 = \sigma_{k\mathcal{M}/n}^2 S_i$, in law, where

$$S_i = M^{-1} \sum_{j=i}^{i+M-1} U_j^2, \quad U_0, U_1, U_2, \dots \text{ iid standard normal.} \quad (120)$$

As in Section 4.1, there is no adjustment due to the asymptotic likelihood ratios, and so the first term in (119) converges stably to a mixed normal with random variance as the limit of $n\Delta t_n^2 \sum_k |\sigma|_{k\mathcal{M}/n}^r \text{Var} \left(c_{M,r}^{-1} \sum_{i=0}^{M-M} S_i^{r/2} \right)$, which is

$$T c_{M,r}^{-2} \frac{1}{\mathcal{M}} \text{Var} \left(\sum_{i=0}^{M-M} S_i^{r/2} \right) \int_0^T |\sigma|_t^r dt. \quad (121)$$

Similarly, one can apply the same technique to the second term in (119), but now with the k 'th block ($k \geq 2$) starting at $k\mathcal{M} - M$. This analysis yields that the second term is also asymptotically mixed normal, but with a variance what is of order $o_p(1)$ as $\mathcal{M} \rightarrow \infty$. (In other words, once again, first send n to infinity, and then, afterwards, do the same to $\mathcal{M}$). This yields that, overall, and in the sense of stable convergence,

$$n^{1/2}(\tilde{\theta}_n^{MW} - \theta) \xrightarrow{\mathcal{L}} N(0, 1) \times \left(c_{M,r}^{-2} \alpha_{M,r} T \int_0^T |\sigma|_t^r dt \right)^{1/2}, \quad (122)$$

where, from (121), $\alpha_{M,r} = \lim_{\mathcal{M} \rightarrow \infty} \text{Var} \left(\sum_{i=0}^{M-M} S_i^{r/2} \right) / \mathcal{M}$, i.e.,

$$\alpha_{M,r} = \text{Var}(S_0^{r/2}) + 2 \sum_{i=1}^{M-1} \text{Cov}(S_0^{r/2}, S_i^{r/2}),$$

where the S_i are given in (120).

5.2 Asynchronous data

For purposes of analysis, asynchronous data does not pose any conceptual difficulty when applying the results from Sections 2-3.2. One includes all observation times when computing the likelihood ratios in the contiguity theorems. It does not matter that some components of the vector are not observed at all these times. In a sense, they are just treated as missing data. Just as in the case of irregular times for scalar processes, this does not necessarily mean that it is straightforward to write down sensible estimators.

For example, consider a bivariate process $(X_t^{(1)}, X_t^{(2)})$. If process $(X_t^{(r)})$ is observed at times :

$$\mathcal{G}_n^{(r)} = \{0 \leq t_{n,0}^{(r)} < t_{n,1}^{(r)} < \dots < t_{n,n_1}^{(r)} \leq T\}, \quad (123)$$

one would normally use the grid $\mathcal{G}_n = \mathcal{G}_n^{(1)} \cup \mathcal{G}_n^{(2)} \cup \{0, T\}$ to compute the quantities in Theorems 1-3.

To focus the mind with an example, consider the estimation of covariation under asynchronous data. It is shown in Mykland (2006) that the Hayashi-Yoshida estimator (Hayashi and Yoshida (2005)) can be seen as a nonparametric maximum likelihood estimator (MLE). We shall here see that blocking induces an additional class of local likelihood based MLEs. The difference between the former and the latter depends on the continuity assumptions made on the volatility process, and is a little like the difference between the Kaplan-Meier (Kaplan and Meier (1958)) and Nelson-Aalen (Nelson (1969), Aalen (1976, 1978)) estimators in survival analysis. (Note that the variance estimate for the Haysahi-Yoshida estimator from Section 5.3 of Mykland (2006) obviously also remains valid in the setting of this paper).

For simplicity, work with a bivariate process, and let the grid $\mathcal{G}_n$ be given by (123). For now, let the block dividers τ be any subset of $\mathcal{G}_n$. Under the approximate measure Q_n , note that for

$$\tau_{n,i-1} \leq t_{n,j-1}^{(1)} < t_{n,j}^{(1)} \leq \tau_{n,i} \text{ and } \tau_{n,i-1} \leq t_{k-1}^{(2)} < t_k^{(2)} \leq \tau_{n,i} \quad (124)$$

the set of returns $X_{t_{n,j}^{(1)}}^{(1)} - X_{t_{n,j-1}^{(1)}}^{(1)}$ and $X_{t_{n,k}^{(2)}}^{(2)} - X_{t_{n,k-1}^{(2)}}^{(2)}$ are conditionally jointly normal with mean zero and covariances

$$\text{Cov}_{Q_n}(X_{t_{n,j}^{(r)}}^{(r)} - X_{t_{n,j-1}^{(r)}}^{(r)}, X_{t_{n,k}^{(s)}}^{(s)} - X_{t_{n,k-1}^{(s)}}^{(s)} \mid \mathcal{F}_{\tau_{n,i-1}}) = (\zeta_{\tau_{n,i-1}})_{r,s} d\{(t_{n,j-1}^{(r)}, t_{n,j}^{(r)}) \cap (t_{n,k-1}^{(s)}, t_{n,k}^{(s)})\} \quad (125)$$

where d is length (Lebesgue measure). Set $\kappa_{r,s;j,k} = \zeta d\{(t_{n,j-1}^{(r)}, t_{n,j}^{(r)}) \cap (t_{n,k-1}^{(s)}, t_{n,k}^{(s)})\}$. The Q_n log likelihood ratio based on observations fully in block $(\tau_{n,i-1}, \tau_{n,i}]$ is therefore given as

$$\ell(\zeta) = -\frac{1}{2} \ln \det(\kappa) - \frac{1}{2} \sum_{r,s;j,k} \kappa^{r,s;j,k} (X_{t_{n,j}^{(r)}}^{(r)} - X_{t_{n,j-1}^{(r)}}^{(r)}) (X_{t_{n,k}^{(s)}}^{(s)} - X_{t_{n,k-1}^{(s)}}^{(s)}) - \frac{N_i}{2} \ln(2\pi), \quad (126)$$

where $\kappa^{r,s;j,k}$ are the elements of the matrix inverse of $(\kappa_{r,s;j,k})$, and N_i is a measure of block sample size. The sum in (j, k) is over all intersections $(t_{n,j-1}^{(r)}, t_{n,j}^{(r)}) \cap (t_{n,k-1}^{(s)}, t_{n,k}^{(s)})$ with positive length satisfying (124). Call the number of such terms

$$m_{n,i}^{(r,s)} = \# \text{ nonempty intersections } (t_{n,j-1}^{(r)}, t_{n,j}^{(r)}) \cap (t_{n,k-1}^{(s)}, t_{n,k}^{(s)}) \text{ satisfying (124)}. \quad (127)$$

The “parameter” ζ corresponds to $\zeta_{\tau_{n,i-1}}$. The block MLE is thus given as

$$\hat{\zeta}_{\tau_{n,i-1}}^{(r,s)} = \frac{1}{m_{n,i}^{(r,s)}} \sum_{j,k} \frac{(X_{t_{n,j}^{(r)}}^{(r)} - X_{t_{n,j-1}^{(r)}}^{(r)}) (X_{t_{n,k}^{(s)}}^{(s)} - X_{t_{n,k-1}^{(s)}}^{(s)})}{d\{(t_{n,j-1}^{(r)}, t_{n,j}^{(r)}) \cap (t_{n,k-1}^{(s)}, t_{n,k}^{(s)})\}} \quad (128)$$

where the sum is over j, k satisfying (124) for which the denominator in the summand is nonzero. The overall estimate of covariation is thus

$$\langle \widehat{X^{(r)}}, \widehat{X^{(s)}} \rangle_T = \sum_i \hat{\zeta}_{\tau_{n,i-1}}^{(r,s)} (\tau_{n,i} - \tau_{n,i-1}). \quad (129)$$

We suppose, of course, that each block is large enough for $m_{n,i}^{(r,s)}$ to be always greater than zero.

Under Q_n , $E_{Q_n}(\hat{\zeta}_{\tau_{n,i-1}}|\mathcal{F}_{\tau_{n,i-1}}) = \zeta_{\tau_{n,i-1}}$, and

$$\begin{aligned} \text{Var}_{Q_n}(\hat{\zeta}_{\tau_{n,i-1}}^{(r,s)}|\mathcal{F}_{\tau_{n,i-1}}) &= \left(\frac{1}{m_{n,i}^{(r,s)}}\right)^2 \left(\zeta_{\tau_{n,i-1}}^{(r,r)} \zeta_{\tau_{n,i-1}}^{(s,s)} \sum_{j,k} \frac{(t_{n,j}^{(r)} - t_{n,j-1}^{(r)})(t_{n,k}^{(s)} - t_{n,k-1}^{(s)})}{d\{(t_{n,j-1}^{(r)}, t_{n,j}^{(r)}) \cap (t_{n,k-1}^{(s)}, t_{n,k}^{(s)})\}}^2 \right. \\ &\quad \left. + (\zeta_{\tau_{n,i-1}}^{(r,s)})^2 \sum_{j_1, j_2, k_1, k_2} \frac{d\{(t_{n,j_1-1}^{(r)}, t_{n,j_1}^{(r)}) \cap (t_{n,k_2-1}^{(s)}, t_{n,k_2}^{(s)})\} d\{(t_{n,j_2-1}^{(r)}, t_{n,j_2}^{(r)}) \cap (t_{n,k_1-1}^{(s)}, t_{n,k_1}^{(s)})\}}}{d\{(t_{n,j_1-1}^{(r)}, t_{n,j_1}^{(r)}) \cap (t_{n,k_1-1}^{(s)}, t_{n,k_1}^{(s)})\} d\{(t_{n,j_2-1}^{(r)}, t_{n,j_2}^{(r)}) \cap (t_{n,k_2-1}^{(s)}, t_{n,k_2}^{(s)})\}} \right), \end{aligned} \quad (130)$$

The first sum is over the same (j, k) as in (128), and the second sum is over all j_1, j_2, k_1, k_2 satisfying (124), again for which the denominator in the summand is nonzero.

It is therefore easy to see that subject to conditions on the observation times $t_{n,i}^{(r)}$ and $t_{n,i}^{(s)}$, $n^{1/2}(\widehat{\langle X^{(r)}, X^{(s)} \rangle}_T - \langle X^{(r)}, X^{(s)} \rangle_T)$ converges stably (under Q_n), to a mixed normal distribution with variance as the limit of

$$n \sum_i \text{Var}_{Q_n}(\hat{\zeta}_{\tau_{n,i-1}}^{(r,s)}|\mathcal{F}_{\tau_{n,i-1}})(\tau_{n,i} - \tau_{n,i-1})^2. \quad (131)$$

It is straightforward to see that there is no adjustment from Q_n to P^* . A formal asymptotic analysis would be tedious, and has therefore been omitted. In any case, to estimate the asymptotic variance, one would use (130)-(131), with $\hat{\zeta}_{\tau_{n,i-1}}$ replacing $\zeta_{\tau_{n,i-1}}$ in (130).

REMARK 13. An important difference from the Hayashi-Yoshida estimator is that (128) depends on the observation times. This is in many instances undesirable, and the choice of estimator will depend on the degree to which these times are trusted. The Hayashi-Yoshida estimator is also aesthetically more pleasing. We note, however, that from likelihood considerations, the estimator (128) will have an asymptotic variance which, as the block size tends to infinity, converges to a limit which corresponds to the efficient minimum for constant volatility matrix.

This phenomenon can be best illustrated for a scalar process (so there is no asynchronicity). In this case, our estimator (128) of $\langle X, X \rangle_T$ becomes (for block size M fixed)

$$\widehat{\langle X, X \rangle}_T = \sum_i (\tau_{n,i} - \tau_{n,i-1}) \frac{1}{M} \sum_{j: \tau_{n,i-1} < t_{n,j} \leq \tau_{n,i}} \frac{\Delta X_{t_{n,j}}^2}{\Delta t_{n,j}}. \quad (132)$$

It is easy to see, by the methods in this paper, or directly, that for this estimator, the asymptotic variance is $2T \int_0^T \sigma_t^4 dt$, while for the standard realized volatility, the corresponding expression is $2T \int_0^T \sigma_t^4 H'(t) dt$, where $H(t)$ is the asymptotic quadratic variation of time (24). It is always the case that $H'(t) \geq 1$, and when observations are sufficiently irregular (under, say, Poisson sampling, cf. Example 4) the inequality is strict. Thus, (132) is more efficient than regular realized volatility, but since the times can in many cases not be trusted, the realized volatility remains a main tool for estimating volatility. $\square$

6 Conclusion

The main finding of the paper is that one can in broad generality use first order approximations when defining and analyzing estimators. Such approximations require an ex post adjustment involving asymptotic likelihood ratios, and these are given. Several examples are provided (Sections 4-5).

The theory relies heavily on the interplay between stable convergence and measure change, and on asymptotic expansions for martingales. We here give a technical summary of the findings.

The paper deals with two forms of discretization: to block size $M = 1$, and then to block size $M > 1$. Each of these has to be adjusted for by using an asymptotic measure change. Accordingly, the likelihood ratios are called dP_∞^*/dP and dQ_∞/dP . There is similarity here to the measure change dP^*/dP used in option pricing theory, where P^* is an equivalent martingale measure (a probability distribution under which the drift of an underlying process has been removed; for our purposes, discounting is not an issue); for more discussion and references, see Section 2.2. In fact, for the reasons given in that section, we can for simplicity assume that the probabilities P_n^* and Q_n also are such that the (observed discrete time) process has no drift.

It is useful to write the likelihood ratio decomposition

$$\log \frac{dQ_\infty}{dP} = \log \frac{dQ_\infty}{dP_\infty^*} + \log \frac{dP_\infty^*}{dP^*} + \log \frac{dP^*}{dP}. \quad (133)$$

We saw in Section 3.3 that these three likelihood ratios are of similar form, and can be represented in terms of Hermite polynomials of the increments of the observed process. The connections are summarized in Table 1.

The three approximations all lead to adjustments that are absolutely continuous. This fact means that for estimators, consistency and rate of convergence are unaffected by the approximation. It turned out that asymptotic variances are similarly unaffected (Remark 4 in Section 2.4). Asymptotic distributions can be changed through their means only (Sections 2.4, 3.4). We emphasize that this is not the same as introducing inconsistency.

type of approximation	compensating likelihood ratio (LR)	size of LR is related to	order of relevant Hermite polynomial
one period discretization ($M = 1$)	dP_∞^*/dP^*	leverage effect	3
multi period discretization (block $M > 1$)	dQ_∞/dP_∞^*	volatility of volatility	2
removal of drift	dP^*/dP	mean	1

TABLE 1. Measure changes (likelihood ratios) tied to three procedures modifying properties of the observed process. P is the true probability distribution, P^* is the equivalent martingale measure

(as in option pricing theory). P_n^* is the probability for which (1) is exact, and Q_n is the probability for which one can use $\int_{t_i-M}^{t_i} f_s dW_s \approx f_{t_i-M}(W_{t_i} - W_{t_i-M})$. The two measure changes dP_n^*/dP^* and dQ_n/dP_n^* have an asymptotic limit. This connects to the statistical concept of contiguity, cf. Remark 2.

As an illustration for Table 1, consider a scalar process on the form $dX_t = \mu_t dt + \sigma_t dW_t$, and with equidistant observations of X . The terms in (133) can be written

$$\log \frac{dP_\infty^*}{dP^*} = Z_1 \eta_1 - \frac{1}{2} \eta_1^2 \quad \text{and} \quad \log \frac{dQ_\infty}{dP_\infty^*} = Z_2 \eta_2 - \frac{1}{2} \eta_2^2, \quad (134)$$

where Z_1 and Z_2 are independent standard normal (independent of the data), and

$$\eta_1^2 = \underbrace{\frac{3}{8} \int_0^T \sigma_t^{-6} (\langle \sigma^2, X \rangle'_t)^2 dt}_{\text{related to leverage effect}} \quad \text{and} \quad \eta_2^2 = \underbrace{\frac{M-1}{4} \int_0^T \sigma_t^{-4} \langle \sigma^2, \sigma^2 \rangle'_t dt}_{\text{related to volatility of volatility}} \quad (135)$$

The angle brackets $\langle \cdot, \cdot \rangle$ refers to covariation (sum of instantaneous covariance between processes, see Chapter I.4e (p. 51-58), of Jacod and Shiryaev (2003), see also (5) in Section 2.1). As is well known from stochastic calculus, $\log \frac{dP^*}{dP} = - \int_0^T \frac{\mu_t}{\sigma_t} dW_t - \frac{1}{2} \int_0^T \frac{\mu_t^2}{\sigma_t^2} dt$.

In the case of a Heston (1993) model, where $d\sigma_t^2 = \kappa(\alpha - \sigma_t^2)dt + \gamma\sigma_t dB_t$, and B is a Brownian motion correlated with W , $d\langle B, W \rangle_t = \rho dt$, one obtains

$$\eta_1^2 = \frac{3}{8}(\rho\gamma)^2 \int_0^T \sigma_t^{-2} dt \quad \text{and} \quad \eta_2^2 = \frac{1}{4}\gamma^2(M-1) \int_0^T \sigma_t^{-2} dt. \quad (136)$$

A number of unsolved questions remain. The approach provides is a tool for *analyzing* estimators, and it does not always give guidance as to how to define estimators in the first place. Also, the theory requires block sizes (M) to stay bounded as the number of observations increases. It would be desirable to have a theory where $M \rightarrow \infty$ with n . This is not possible with the likelihood ratios we consider, but may be available in other settings, such as with microstructure noise. Causality effects from observation times to the process, such as in Renault and Werker (2006), would also need an extended theory.

Department of Statistics, The University of Chicago, Chicago, IL60637-1514, U.S.A.;
mykland@galton.uchicago.edu; <http://galton.uchicago.edu/~mykland>.

and

Department of Finance, The University Illinois at Chicago, Chicago, IL60607, U.S.A.;
lanzhang@uic.edu; <http://tigger.uic.edu/~lanzhang/>.

REFERENCES

- AALLEN, O. (1976): "Nonparametric inference in connection with multiple decrement models," *Scandinavian Journal of Statistics*, 3, 15–27.
- (1978): "Nonparametric inference for a family of counting processes," *Annals of Statistics*, 6, 701–726.
- AÏT-SAHALIA, Y. (2002): "Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223–262.
- ALDOUS, D. J. AND G. K. EAGLESON (1978): "On Mixing and Stability of Limit Theorems," *Annals of Probability*, 6, 325–331.
- ANDERSEN, T. G. AND T. BOLLERSLEV (1998): "Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies," *Journal of Finance*, 53, 219–265.
- ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2000): "Great realizations," *Risk*, 13, 105–108.
- (2001): "The Distribution of Exchange Rate Realized Volatility," *Journal of the American Statistical Association*, 96, 42–55.
- (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.
- BARNDORFF-NIELSEN, O., S. GRAVERSEN, J. JACOD, M. PODOLSKIJ, AND N. SHEPHARD (2006): "A central limit theorem for realised bipower variations of continuous semimartingales," in *From Stochastic Calculus to Mathematical Finance, The Shiryaev Festschrift*, ed. by Y. Kabanov, R. Liptser, and J. Stoyanov, Berlin: Springer Verlag, 33–69.
- BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2001): "Non-Gaussian Ornstein-Uhlenbeck-Based Models And Some Of Their Uses In Financial Economics," *Journal of the Royal Statistical Society, B*, 63, 167–241.
- (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, B*, 64, 253–280.
- (2004a): "Econometric analysis of realised covariation: high frequency based covariance, regression and correlation in financial economics," *Econometrica*, 72, 885–925.
- (2004b): "Power and bipower variation with stochastic volatility and jumps (with discussion)," *Journal of Financial Econometrics*, 2, 1–48.
- (2006): "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of Financial Econometrics*, 4, 1–30.
- BILLINGSLEY, P. (1995): *Probability and Measure*, New York: Wiley, third ed.
- CHEN, Y. AND V. SPOKOINY (2007): "Robust Risk Management. Accounting for Nonstationarity and Heavy Tails," Tech. rep., Weierstrass Institute, Berlin.

- CHRISTENSEN, K., R. OOMEN, AND M. PODOLSKIJ (2008): “Realised Quantile-Based Estimation of the Integrated Variance,” Tech. rep.
- CIZEK, P., W. HÄRDLE, AND V. SPOKOINY (2007): “Adaptive pointwise estimation in time-inhomogeneous time series models,” Tech. rep., Weierstrass Institute, Berlin.
- COMTE, F. AND E. RENAULT (1998): “Long memory in continuous-time stochastic volatility models,” *Mathematical Finance*, 8, 291–323.
- DACOROGNA, M. M., R. GENÇAY, U. MÜLLER, R. B. OLSEN, AND O. V. PICTET (2001): *An Introduction to High-Frequency Finance*, San Diego: Academic Press.
- DOVONON, P., S. GONCALVES, AND N. MEDDAHI (2008): “Bootstrapping realized multivariate volatility measures,” Tech. rep.
- DUFFIE, D. (1996): *Dynamic Asset Pricing Theory*, Princeton, N.J.: Princeton University Press.
- FAN, J. (1993): “Local Linear Regression Smoothers and Their Minimax Efficiencies,” *Annals of Statistics*, 21, 196–216.
- FAN, J., M. FARMEN, AND I. GIJBELS (1998): “Local Maximum Likelihood Estimation and Inference,” *Journal of the Royal Statistical Society ser. B*, 60, 591–608.
- FAN, J. AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*, London, U.K.: Chapman and Hall.
- FAN, J., I. GIJBELS, AND M. KING (1997): “Local Likelihood and Local Partial Likelihood in Hazard Regression,” *Annals of Statistics*, 25, 1661–1690.
- FOSTER, D. AND D. NELSON (1996): “Continuous record asymptotics for rolling sample variance estimators,” *Econometrica*, 64, 139–174.
- GONCALVES, S. AND N. MEDDAHI (2005): “Bootstrapping realized volatility,” Tech. rep., Université de Montréal.
- HÁJEK, J. AND Z. SIDAK (1967): *Theory of Rank Tests*, New York: Academic Press.
- HALL, P. AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*, Boston: Academic Press.
- HARRISON, M. AND D. KREPS (1979): “Martingales and Arbitrage in Multiperiod Securities Markets,” *Journal of Economic Theory*, 20, 381–408.
- HARRISON, M. AND S. PLISKA (1981): “Martingales and Stochastic Integrals in the Theory of Continuous Trading,” *Stochastic Processes and Their Applications*, 11, 215–260.
- HAYASHI, T. AND N. YOSHIDA (2005): “On Covariance Estimation of Non-synchronously Observed Diffusion Processes,” *Bernoulli*, 11, 359–379.
- HESTON, S. (1993): “A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options,” *Review of Financial Studies*, 6, 327–343.

- HJORT, N. L. AND M. JONES (1996): “Locally parametric nonparametric density estimation,” *The Annals of Statistics*, 24, 1619–1647.
- JACOD, J. (1994): “Limit of Random Measures Associated with the Increments of a Brownian Semimartingale,” Tech. rep., Université de Paris VI.
- (2006): “Asymptotic Properties of Realized Power Variations and related Functionals of Semimartingales,” *Stochastic Processes and Applications*, (to appear).
- JACOD, J., Y. LI, P. A. MYKLAND, M. PODOLSKIJ, AND M. VETTER (2008): “Microstructure Noise in the Continuous Case: The Pre-Averaging Approach,” Tech. rep., University of Chicago.
- JACOD, J. AND P. PROTTER (1998): “Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations,” *Annals of Probability*, 26, 267–307.
- JACOD, J. AND A. N. SHIRYAEV (2003): *Limit Theorems for Stochastic Processes*, New York: Springer-Verlag, second ed.
- KALNINA, I. AND O. LINTON (2007): “Inference about Realized Volatility using Infill Subsampling,” Tech. rep., London School of Economics.
- KAPLAN, E. AND P. MEIER (1958): “Nonparametric Estimation from Incomplete Observations,” *Journal of the American Statistical Association*, 53, 457–481.
- KARATZAS, I. AND S. E. SHREVE (1991): *Brownian Motion and Stochastic Calculus*, New York: Springer-Verlag.
- KAUERMANN, G. AND J. OPSOMER (2003): “Local Likelihood Estimation in Generalized Additive Models,” *Scandinavian Journal of Statistics*, 30, 317–337.
- KINNEBROCK, S. AND M. PODOLSKIJ (2007): “A Note on the Central Limit Theorem for Bipower Variation of General Functions,” *Stochastic Processes and their Applications*, forthcoming.
- LECAM, L. (1986): *Asymptotic Methods in Statistical Decision Theory*, New York: Springer-Verlag.
- LECAM, L. AND G. YANG (1986): *Asymptotics in Statistics: Some Basic Concepts*, New York: Springer-Verlag, second ed.
- LEHMANN, E. (1983): *Theory of Point Estimation*, New York: John Wiley and Sons, second ed.
- LINTON, O. (2007): “Notes about Gaussian calculus,” Presented in Imperial College Financial Econometrics Conference.
- LOADER, C. R. (1996): “Local likelihood density estimation,” *The Annals of Statistics*, 24, 1602–1618.
- MARDIA, K. V., J. KENT, AND J. BIBBY (1979): *Multivariate Analysis*, London: Academic Press.
- MCCULLAGH, P. (1987): *Tensor Methods in Statistics*, London, U.K.: Chapman and Hall.
- MYKLAND, P. A. (1993): “Asymptotic Expansions for Martingales,” *Annals of Probability*, 21, 800–818.

- (1995a): “Embedding and Asymptotic Expansions for Martingales,” *Probability Theory and Related Fields*, 103, 475–492.
- (1995b): “Martingale Expansions and Second Order Inference,” *Annals of Statistics*, 23, 707–731.
- (2006): “A Gaussian Calculus for Inference from High Frequency Data,” Tech. rep., University of Chicago.
- MYKLAND, P. A. AND L. ZHANG (2006): “ANOVA for Diffusions and Itô Processes,” *Annals of Statistics*, 34, 1931–1963.
- NELSON, W. (1969): “Hazard plotting for incomplete failure data,” *J. Qual. Tech.*, 1, 27–52.
- PODOLSKIJ, M. AND M. VETTER (2006): “Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps,” Tech. rep., Ruhr-Universität Bochum.
- RENAULT, E. AND B. J. WERKER (2006): “Causality effects in return volatility measures with random times,” Tech. rep., Tilburg University.
- RÉNYI, A. (1963): “On Stable Sequences of Events,” *Sankyā Series A*, 25, 293–302.
- ROOTZÉN, H. (1980): “Limit Distributions for the Error in Approximations of Stochastic Integrals,” *Annals of Probability*, 8, 241–251.
- ROSS, S. (1996): *Stochastic Processes*, New York, N.Y.: Wiley, second ed.
- ROSS, S. M. (1976): “The arbitrage theory of capital asset pricing,” *Journal of Economic Theory*, 13, 341–360.
- TIBSHIRANI, R. AND T. HASTIE (1987): “Local Likelihood Estimation,” *Journal of the American Statistician Association*, 82, 559–567.
- WEISBERG, S. (1985): *Applied Linear Regression*, New York: Wiley, second ed.
- WOERNER, J. H. (2004): “Power and multipower variation: Inference for high frequency data,” Tech. rep., University of Göttingen.
- ZHANG, L. (2001): “From Martingales to ANOVA: Implied and Realized Volatility,” Ph.D. thesis, The University of Chicago, Department of Statistics.
- (2005): “Estimating Covariation: Epps Effect and Microstructure Noise,” Tech. rep., University of Illinois at Chicago (working paper).
- (2006): “Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach,” *Bernoulli*, 12, 1019–1043.
- ZHANG, L., P. A. MYKLAND, AND Y. AÏT-SAHALIA (2005): “A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data,” *Journal of the American Statistical Association*, 100, 1394–1411.

APPENDIX: PROOFS

A Proofs of Theorems 1 and 2

To understand the properties of this approximation, consider the following “strong approximation”. Set

$$d\sigma_t = \tilde{\sigma}_t dt + f_t dW_t + g_t dB_t \quad (\text{A.1})$$

where f_t is a tensor and $g_t dB_t$ is a matrix, with B a Brownian motion independent of W (g and B can be tensor processes). For example, component (r_1, r_2) of the matrix $f_t dW_t$ is $\sum_{r_3=1}^p f_t^{(r_1, r_2, r_3)} dW^{(r_3)}$. Note that σ_t is an Itô process by Assumption 2. Then

$$\begin{aligned} \Delta X_{t_{n,j+1}} &= \sigma_{t_{n,j}} \Delta W_{t_{n,j+1}} + \int_{t_{n,j}}^{t_{n,j+1}} (\sigma_u - \sigma_{t_{n,j}}) dW_u \\ &= \sigma_{t_{n,j}} \Delta W_{t_{n,j+1}} + f_{t_{n,j}} \int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^u dW_u \right) dW_t \\ &\quad + dB\text{-term} + \text{higher order terms} . \end{aligned} \quad (\text{A.2})$$

It will turn out that the two first terms on the right hand side will matter in our approximation. Note first that by taking quadratic covariations, one obtains

$$f_t^{(r_1, r_2, r_3)} = \langle \sigma^{(r_1, r_2)}, W^{(r_3)} \rangle'_t. \quad (\text{A.3})$$

To proceed with the proof, some further notation. Define

$$d\check{\sigma}_t = \sigma_t^{-1} d\sigma_t \text{ and } \check{f}_t^{(r_1, r_2, r_3)} = \langle \check{\sigma}^{(r_1, r_2)}, W^{(r_3)} \rangle'_t = \sum_{r_4=1}^p (\sigma_t^{-1})^{(r_1, r_4)} f_t^{(r_1, r_2, r_3)} \quad (\text{A.4})$$

$\check{\sigma}_t^{(r_1, r_2)}$ and $\check{f}_t^{(r_1, r_2, r_3)}$ are not symmetric in (r_1, r_2) . However, since $d\zeta_t = d(\sigma_t \sigma_t^T) = \sigma_t d\sigma_t + (\sigma_t d\sigma_t)^T + dt$ terms, we obtain from (14) that $d\check{\zeta}_t = \sigma_t^{-1} d\sigma_t + (\sigma_t^{-1} d\sigma_t)^T + dt$ terms. Hence

$$\langle \check{\zeta}^{(r_1, r_2)}, W^{(r_3)} \rangle'_t = \check{f}_t^{(r_1, r_2, r_3)} + \check{f}_t^{(r_2, r_1, r_3)}. \quad (\text{A.5})$$

Also

$$k_t^{(r_1, r_2, r_3)} = \langle \check{\zeta}^{(r_1, r_2)}, W^{(r_3)} \rangle'_t[3] = \check{f}_t^{(r_1, r_2, r_3)}[6] \quad (\text{A.6})$$

Finally, we let $\Delta t = T/n$ (the average $\Delta t_{n,j+1}$).

PROOF OF THEOREM 1. Note that, from (18) and (A.2)

$$\begin{aligned} \Delta \check{W}_{t_{n,j+1}} &= \Delta W_{t_{n,j+1}} + \check{f}_{t_{n,j}} \int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^u dW_u \right) dW_t \\ &\quad + dB\text{-term} + \text{higher order terms} . \end{aligned} \quad (\text{A.7})$$

In the representation (A.7), we obtain, up to $O_p(\Delta t^{5/2})$, .

$$\begin{aligned}
& \text{cum}_3(\Delta \check{W}_{t_{n,j+1}}^{(r_1)}, \Delta W_{t_{n,j+1}}^{(r_2)}, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}}) \\
& \doteq \text{cum}(\sum_{s_2, s_3} \check{f}_{t_{n,j}}^{(r_1, s_2, s_3)} \int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^u dW_u^{(s_3)} \right) dW_t^{(s_2)}, \Delta W_{t_{n,j+1}}^{(r_2)}, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}}) \\
& = \sum_{s_2, s_3} \check{f}_{t_{n,j}}^{(r_1, s_2, s_3)} \text{cum}(\int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^u dW_u^{(s_3)} \right) dW_t^{(s_2)}, \Delta W_{t_{n,j+1}}^{(r_2)}, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}}) \\
& = \sum_{s_2, s_3} \check{f}_{t_{n,j}}^{(r_1, s_2, s_3)} \text{Cov}(\int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^u dW_u^{(s_3)} \right) dt \delta^{s_2, r_2}, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}}) [2] \\
& = \sum_{s_3} \check{f}_{t_{n,j}}^{(r_1, r_2, s_3)} \text{Cov}(\int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^u dW_u^{*(s_3)} \right) dt, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}}) [2] \\
& = \int_{t_{n,j}}^{t_{n,j+1}} dt \sum_{s_3} \check{f}_{t_{n,j}}^{(r_1, r_2, s_3)} \text{Cov}(\int_{t_{n,j}}^u dW_u^{*(s_3)}, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}}) [2] \\
& = \int_{t_{n,j}}^{t_{n,j+1}} dt \sum_{s_3} \check{f}_{t_{n,j}}^{(r_1, r_2, s_3)} (u - t_{n,j}) \delta s_3, r_3 [2] \\
& = \frac{1}{2} \Delta t_{n,j+1}^2 \check{f}_{t_{n,j}}^{(r_1, r_2, r_3)} [2], \tag{A.8}
\end{aligned}$$

where “[2]” represents the swapping of r_2 and r_3 (see p. 29-30 of McCullagh (1987) for a discussion of the notation). Hence

$$\begin{aligned}
& \text{cum}_3(\Delta \check{W}_{t_{n,j+1}}^{(r_1)}, \Delta \check{W}_{t_{n,j+1}}^{(r_2)}, \Delta \check{W}_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}}) \\
& = \frac{1}{2} \Delta t_{n,j+1}^2 \check{f}_{t_{n,j}}^{(r_1, r_2, r_3)} [6] + O_p(\Delta t^{5/2}) \\
& = \frac{1}{2} \Delta t_{n,j+1}^2 \langle \check{\zeta}^{(r_1, r_2)}, W^{(r_3)} \rangle'_{t_{n,j}} [3] + O_p(\Delta t^{5/2}) \tag{A.9}
\end{aligned}$$

by symmetry, Hence, from (15),

$$\begin{aligned}
\kappa^{r_1, r_2, r_3} & = \frac{1}{2} \Delta t_{n,j+1}^{1/2} \check{f}_{t_{n,j}}^{(r_1, r_2, r_3)} [3] + O_p(\Delta t) \\
& = \frac{1}{2} \Delta t_{n,j+1}^{1/2} k_{t_{n,j}}^{(r_1, r_2, r_3)} + O_p(\Delta t). \tag{A.10}
\end{aligned}$$

At the same time ($d\zeta = \tilde{\zeta} dt + d$ martingale),

$$\begin{aligned}
\text{Cov}(\Delta X_{t_{n,j+1}}^{(r_1)}, \Delta X_{t_{n,j+1}}^{(r_2)} | \mathcal{F}_{t_{n,j}}) & = \Delta t_{n,j+1} \zeta_{t_{n,j}}^{(r_1, r_2)} + E(\int_{t_{n,j}}^{t_{n,j+1}} (\zeta_u^{(r_1, r_2)} - \zeta_{t_{n,j}}^{(r_1, r_2)}) du | \mathcal{F}_{t_{n,j}}) \\
& = \Delta t_{n,j+1} \zeta_{t_{n,j}}^{(r_1, r_2)} + E(\int_{t_{n,j}}^{t_{n,j+1}} du \int_{t_{n,j}}^u \tilde{\zeta}_v^{(r_1, r_2)} dv | \mathcal{F}_{t_{n,j}}) \\
& = \Delta t_{n,j+1} \zeta_{t_{n,j}}^{(r_1, r_2)} + \frac{1}{2} \Delta t_{n,j+1}^2 \tilde{\zeta}_{t_{n,j}}^{(r_1, r_2)} + O_p(\Delta t^3), \tag{A.11}
\end{aligned}$$

so that $\text{Cov}(\Delta\check{W}_{t_{n,j+1}}^{(r_1)}, \Delta\check{W}_{t_{n,j+1}}^{(r_2)} | \mathcal{F}_{t_{n,j}}) = \Delta t_{n,j+1} \delta^{r_1, r_2} + O_p(\Delta t^2)$, and

$$\kappa^{r_1, r_2} = \delta_{t_{n,j}}^{r_1, r_2} + O_p(\Delta t). \quad (\text{A.12})$$

Since X is a martingale, we also have $\kappa^r = E(\Delta\check{W}_{t_{n,j+1}}^{(r)} | \mathcal{F}_{t_{n,j}}) = 0$.

In the notation of Chapter 5 of McCullagh (1987), we take $\lambda^{r_1, r_2} = \delta^{r_1, r_2}$, and let the other λ 's be zero. From now on, we also use the summation convention. By the development in Chapter 5.2.2 of this work, we obtain the Edgeworth expansion for the density $f_{n,j+1}$ of $\Delta X_{t_{n,j+1}}$ given $\mathcal{F}_{t_{n,j}}$, on the log scale as

$$\begin{aligned} \log f_{n,j+1}(x) &= \log \phi(x; \delta^{r_1, r_2}) + \frac{1}{3!} \kappa^{r_1, r_2, r_3} h_{r_1 r_2 r_3}(x) \\ &\quad + \frac{1}{2} (\kappa^{r_1, r_2} - \lambda^{r_1, r_2}) h_{r_1 r_2}(x) + \frac{1}{4!} \kappa^{r_1, r_2, r_3, r_4} h_{r_1 r_2 r_3 r_4}(x) \\ &\quad + \kappa^{r_1, r_2, r_3} \kappa^{r_4, r_5, r_6} h_{r_1 r_2 r_3 r_4 r_5 r_6}(x) \frac{[10]}{6!} \\ &\quad + \frac{1}{72} (\kappa^{r_1, r_2, r_3} h_{r_1 r_2 r_3}(x))^2 + O_p(\Delta t^{3/2}) \end{aligned} \quad (\text{A.13})$$

where we for simplicity have used the summation convention. Note that the three last lines contain terms of order $O_p(\Delta t)$ (or smaller).

We note, following formula (5.7) (p. 149) in McCullagh (1987), that $h_{r_1 r_2 r_3} = h_{r_1} h_{r_2} h_{r_3} - h_{r_1} \delta_{r_2, r_3} [3]$, with $h_{r_1} = \delta_{r_1, r_2} x^{r_2}$. Observe that

$$Z_{r_1} = h_{r_1} (\Delta\check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}) = \delta_{r_1, r_2} \Delta\check{W}_{t_{n,j+1}}^{r_2} / (\Delta t_{n,j+1})^{1/2}. \quad (\text{A.14})$$

Under the approximating measure, therefore, the vector consisting of elements Z_{r_1} is conditionally normally distributed with mean zero and covariance matrix δ_{r_1, r_2} .

It follows that

$$h_{r_1 r_2 r_3} (\Delta\check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}) = Z_{r_1} Z_{r_2} Z_{r_3} - Z_{r_1} \delta_{r_2, r_3} [3] \quad (\text{A.15})$$

Under the approximating measure, therefore, $E_n(h_{r_1 r_2 r_3} (\Delta\check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}) | \mathcal{F}_{t_{n,j}}) = 0$, while

$$\text{Cov}_n(h_{r_1 r_2 r_3} (\Delta\check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}), h_{r_4 r_5 r_6} (\Delta\check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}) | \mathcal{F}_{t_{n,j}}) = \delta_{r_1, r_4} \delta_{r_2, r_5} \delta_{r_3, r_6} [6] \quad (\text{A.16})$$

where the “[6]” refers to all six combinations where each δ has one index from $\{r_1, r_2, r_3\}$ and one from $\{r_4, r_5, r_6\}$. It follows that

$$\begin{aligned} \text{Var}_n &\left(\frac{1}{3!} \kappa^{r_1, r_2, r_3} h_{r_1 r_2 r_3} (\Delta\check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}) | \mathcal{F}_{t_{n,j}} \right) \\ &= \frac{1}{36} \kappa^{r_1, r_2, r_3} \kappa^{r_4, r_5, r_6} \text{Cov}_n(h_{r_1 r_2 r_3} (\Delta\check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}), h_{r_4 r_5 r_6} (\Delta\check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}) | \mathcal{F}_{t_{n,j}}) \\ &= \frac{1}{6} \kappa^{r_1, r_2, r_3} \kappa^{r_4, r_5, r_6} \delta_{r_1, r_4} \delta_{r_2, r_5} \delta_{r_3, r_6} \\ &= \Delta t_{n,j+1} \frac{1}{24} k_{t_{n,j}}^{r_1, r_2, r_3} k_{t_{n,j}}^{r_4, r_5, r_6} \delta_{r_1, r_4} \delta_{r_2, r_5} \delta_{r_3, r_6} + O_p(\Delta t^{3/2}) \end{aligned} \quad (\text{A.17})$$

by symmetry of the κ 's. Thus

$$\sum_{t_{n,j+1} \leq t} \text{Var}_n \left(\frac{1}{3!} \kappa^{r_1, r_2, r_3} h_{r_1 r_2 r_3} (\Delta \check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}) | \mathcal{F}_{t_{n,j}} \right) \xrightarrow{p} \int_0^t \frac{1}{24} k_t^{r_1, r_2, r_3} k_t^{r_4, r_5, r_6} \delta_{r_1, r_4} \delta_{r_2, r_5} \delta_{r_3, r_6} dt \quad (\text{A.18})$$

under P_n^* . By the same methods, and since Hermite polynomials of different orders are otherogonal under the approximating measure,

$$\sum_{t_{n,j+1} \leq t} \text{Cov}_n (h_{r_1 r_2 r_3} (\Delta \check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}), h_{r_4} (\Delta X_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2} | \mathcal{F}_{t_{n,j}}) \xrightarrow{p} 0. \quad (\text{A.19})$$

By the methods of Jacod and Shiryaev (2003), it follows that

$$\check{M}_n^{(0)} = \sum_{j=0}^n \frac{1}{3!} \kappa^{r_1, r_2, r_3} h_{r_1 r_2 r_3} (\Delta \check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}) \quad (\text{A.20})$$

converges stably in law to a normal distribution with random variance Γ_0 . (Note that $\check{M}_n^{(0)} = M_n^{(0)} + O_p(\Delta t^{1/2})$ from (19)). We now note that, in the notation of (A.13),

$$\log \frac{dP^*}{dP_n^*} = \sum_{j=0}^{n-1} (f_{n,j+1} - \log \phi) (\Delta \check{W}_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}). \quad (\text{A.21})$$

By the same reasoning as above, the terms other than $\check{M}_n^{(0)}$ and its discrete time quadratic variation (A.18), go away. Thus $\log \frac{dP^*}{dP_n^*} = \check{M}_n^{(0)} - \frac{1}{2} \Gamma_0 + o_p(1)$, and the result follows. ■

REMARK 14. The proof of Theorem 1 uses the Edgeworth expansion (A.13). The proof of the broad availability of such expansions in the martingale case goes back to Mykland (1993, 1995b,a), which uses a test function topology. The formal existence of Edgeworth expansions in our current case is proved by iterating the expansion (A.2) as many times as necessary, and bounding the remainder. In the diffusion case, similar arguments have been used in the estimation and computation theory in Aït-Sahalia (2002). □

PROOF OF THEOREM 2. It follows from the development in the proof of Theorem 1 that

$$\log \frac{dP^*}{dP_n^*} = M_n^{(0)} - \frac{1}{2} \Gamma_0 + o_p(1) \quad (\text{A.22})$$

where $M_n^{(0)}$ is as defined in equation (19). Write that, under P_n^* , $(Z_n, M_n^{(0)}) \xrightarrow{\mathcal{L}} (Z, M)$, with $M = \Gamma_0^{1/2} V_1$, and $Z = b_1 + c_1 M + c_2 V_2$, where V_1 and V_2 are independent and standard normal (independent of $\mathcal{F}_T$). Denote the distribution of (Z, M) as P_∞^* to avoid confusion.

It follows that, for bounded and continuous g , and by uniform integrability,

$$\begin{aligned}
E^*g(Z_n) &= E_n^*g(Z_n) \exp\{M_n^{(0)} - \frac{1}{2}\Gamma_0\}(1 + o(1)) \\
&\rightarrow Eg(Z) \exp\{M - \frac{1}{2}\Gamma_0\} \\
&= E_\infty^*g(b_1 + c_1\Gamma_0^{1/2}V_1 + c_2V_2) \exp\{\Gamma_0^{1/2}V_1 - \frac{1}{2}\Gamma_0\} \\
&= \int_{-\infty}^{\infty} E_\infty^*g(b_1 + c_1\Gamma_0^{1/2}v + c_2V_2) \exp\{\Gamma_0^{1/2}v - \frac{1}{2}\Gamma_0\} (2\pi)^{-1/2} \exp\{-\frac{1}{2}v^2\} dv \\
&= \int_{-\infty}^{\infty} E_\infty^*g(b_1 + c_1\Gamma_0^{1/2}(u + \Gamma_0^{1/2}) + c_2V_2) (2\pi)^{-1/2} \exp\{-\frac{1}{2}u^2\} du \quad (u = v - \Gamma_0^{1/2}) \\
&= E_\infty^*g(Z + c_1\Gamma_0)
\end{aligned} \tag{A.23}$$

The result then follows since $c_1\Gamma_0 = A_{12}$. ■

B Proof of Theorems 3 and 5

COMMON INITIAL DEVELOPMENT.

Let $Z_n^{(1)}$ and $Z_n^{(2)}$ be given by (42) and (56), respectively. Set

$$\begin{aligned}
\Delta Z_{n,t_{n,j+1}}^{(1)} &= \frac{1}{2} \Delta X_{t_{n,j+1}}^T (\zeta_{t_{n,j}}^{-1} - \zeta_{\tau_{n,i-1}}^{-1}) \Delta X_{t_{n,j+1}} \Delta t_{n,j+1}^{-1} \\
\Delta Z_{n,t_{n,j+1}}^{(2)} &= \frac{1}{2} \Delta X_{t_{n,j+1}}^T (\zeta_{\tau_{n,i-1}}^{-1}) \Delta X_{t_{n,j+1}} \left(\Delta t_{n,j+1}^{-1} - \left(\frac{\Delta \tau_{n,i}}{M_i} \right)^{-1} \right)
\end{aligned} \tag{B.24}$$

and note that $Z_n^{(v)} = \sum_j \Delta Z_{n,t_{n,j+1}}^{(v)}$ for $v = 1, 2$. Set $A_j = \zeta_{t_{n,j}}^{1/2} \zeta_{\tau_{n,i-1}}^{-1} \zeta_{t_{n,j}}^{1/2} - I$ and $B_j = \left(\Delta t_{n,j+1} \left(\frac{\Delta \tau_{n,i}}{M_i} \right)^{-1} - 1 \right)$ (the latter is a scalar).

Set $C_j = \zeta_{t_{n,j}}^{1/2} \zeta_{\tau_{n,i-1}}^{-1} \zeta_{t_{n,j}}^{1/2} \left(\Delta t_{n,j+1} \left(\frac{\Delta \tau_{n,i}}{M_i} \right)^{-1} - 1 \right) = (I + A_j)B_j$.

Since $\Delta X_{t_{n,j}}$ is conditionally Gaussian, we obtain (under P_n^*)

$$\begin{aligned}
E_{P_n^*}(\Delta Z_{n,t_{n,j+1}}^{(1)} | \mathcal{X}_{n,t_{n,j}}) &= -\frac{1}{2} \text{tr}(A_j) \\
E_{P_n^*}(\Delta Z_{n,t_{n,j+1}}^{(2)} | \mathcal{X}_{n,t_{n,j}}) &= -\frac{1}{2} \text{tr}(C_j) = -\frac{1}{2} (p + \text{tr}(A_j)) B_j
\end{aligned} \tag{B.25}$$

and

$$\begin{aligned}
&\text{conditional covariance of} \\
&\Delta Z_{n,t_{n,j+1}}^{(1)} \text{ and } \Delta Z_{n,t_{n,j+1}}^{(2)} = \frac{1}{2} \begin{pmatrix} \text{tr}(A_j^2) & \text{tr}(A_j C_j) \\ \text{tr}(A_j C_j) & \text{tr}(C_j^2) \end{pmatrix}
\end{aligned} \tag{B.26}$$

Finally, let $M_n^{(v)}$ be the (end point of the) martingale part (under P^*) of $Z_n^{(v)}$ ($v = 1, 2$), so that

$$M_n^{(1)} = Z^{(1)} + (1/2) \sum_j \text{tr}(A_j) \text{ and } M_n^{(2)} = Z^{(2)} + (1/2) \sum_j \text{tr}(C_j). \quad (\text{B.27})$$

If $\langle \cdot, \cdot \rangle^{\mathcal{G}}$ represents discrete time predictable quadratic variation on the grid $\mathcal{G}$, then equation (B.26) yields

$$\begin{pmatrix} \langle M_n^{(1)}, M_n^{(1)} \rangle^{\mathcal{G}} & \langle M_n^{(1)}, M_n^{(2)} \rangle^{\mathcal{G}} \\ \langle M_n^{(1)}, M_n^{(2)} \rangle^{\mathcal{G}} & \langle M_n^{(2)}, M_n^{(2)} \rangle^{\mathcal{G}} \end{pmatrix} = \frac{1}{2} \sum_j \begin{pmatrix} \text{tr}(A_j^2) & \text{tr}(A_j C_j) \\ \text{tr}(A_j C_j) & \text{tr}(C_j^2) \end{pmatrix}. \quad (\text{B.28})$$

PROOF OF THEOREM 3. First note that, by analogy to the development in Zhang (2001), Mykland and Zhang (2006), Zhang, Mykland, and Aït-Sahalia (2005), and Zhang (2006),

$$\begin{aligned} \langle M_n^{(1)}, M_n^{(1)} \rangle^{\mathcal{G}} &= \frac{1}{2} \sum_j \text{tr}(\zeta_{\tau_{n,i-1}}^{-2} (\zeta_{t_{n,j}} - \zeta_{\tau_{n,i-1}})^2) \\ &= \frac{1}{2} \sum_j \text{tr}(\zeta_{\tau_{n,i-1}}^{-2} (\langle \zeta, \zeta \rangle_{t_{n,j}} - \langle \zeta, \zeta \rangle_{\tau_{n,i-1}})) + o_p(1) \\ &= \frac{1}{2} \sum_j \text{tr}(\zeta_{\tau_{n,i-1}}^{-2} \langle \zeta, \zeta \rangle'_{\tau_{n,i-1}}) (t_{n,j} - \tau_{n,i-1}) + o_p(1) \\ &= \frac{1}{2} \int_0^T \text{tr}(\zeta_t^{-2} \langle \zeta, \zeta \rangle'_t) dK(t) + o_p(1) \\ &= \Gamma_1 + o_p(1), \end{aligned} \quad (\text{B.29})$$

where K is the ADD given by equation (40).

At this point, note that Assumption 2 entails, in view of Lemma 2 in Mykland and Zhang (2006), that

$$\sup_j \text{tr}(A_j^2) \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (\text{B.30})$$

Since also,

$$\text{for } r > 2, \quad |\text{tr}(A_j^r)| \leq \text{tr}(A_j^2)^{r/2}, \quad (\text{B.31})$$

it follows that

$$\begin{aligned}
\log \frac{dQ_n}{dP_n^*} &= Z_n^{(1)} + \frac{1}{2} \sum_i \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} (\log \det \zeta_{t_{n,j}} - \log \det \zeta_{\tau_{n,i-1}}) \\
&= Z_n^{(1)} + \frac{1}{2} \sum_j \log \det(I + A_j) \\
&= Z_n^{(1)} + \frac{1}{2} \sum_j (\text{tr}(A_j) - \text{tr}(A_j^2)/2 + \text{tr}(A_j^3)/3 + \dots) \\
&= M_n^{(1)} - \frac{1}{4} \sum_j \text{tr}(A_j^2) + \frac{1}{6} \sum_j \text{tr}(A_j^3) + \dots \\
&= M_n^{(1)} - \frac{1}{2} \langle M_n^{(1)}, M_n^{(1)} \rangle^{\mathcal{G}} + o_p(1)
\end{aligned} \tag{B.32}$$

At this point, let $\langle M_n^{(1)}, M_n^{(1)} \rangle$ be the quadratic variation of the continuous martingale that coincides at points $t_{n,j}$ with the discrete time martingale leading up to the end point $M_n^{(1)}$. By a standard quarticity argument (as in the proof of Remark 2 in Mykland and Zhang (2006)), (B.29)-(B.31) and the conditional normality of $\Delta Z_{n,t_{n,j+1}}^{(1)}$ yield that $\langle M_n^{(1)}, M_n^{(1)} \rangle = \langle M_n^{(1)}, M_n^{(1)} \rangle^{\mathcal{G}} + o_p(1)$. The stable convergence to a normal distribution with variance Γ_1 then follows by the same methods as in Zhang, Mykland, and Ait-Sahalia (2005). The result is thus proved.

PROOF OF THEOREM 5. The joint stable convergence of $M_n^{(1)}$ and $M_n^{(1)}$ follows analogously to the proof of Theorem 3 by referring to (B.28) and observing that, by (B.29)-(B.31),

$$\begin{aligned}
\sum_j \text{tr}(C_j^2) &= \sum_j \text{tr}(I_p) B_j^2 + \sum_j \text{tr}(A_j) B_j^2 + \sum_j \text{tr}(A_j^2) B_j^2 \\
&= p \sum_j B_j^2 + o_p(1) \\
&= 2\Gamma_2 + o_p(1),
\end{aligned} \tag{B.33}$$

and

$$\begin{aligned}
\sum_j \text{tr}(A_j C_j) &= \sum_j \text{tr}(A_j) B_j + \sum_j \text{tr}(A_j^2) B_j \\
&= \sum_j \text{tr}(A_j) B_j + o_p(1) \\
&= o_p(1)
\end{aligned} \tag{B.34}$$

where the last transition follows by Assumption 2. Meanwhile, since

$$\log \frac{dR_n}{dP_n^*} = \log \frac{dR_n}{dQ_n} + \log \frac{dQ_n}{dP_n^*}, \tag{B.35}$$

we obtain similarly that

$$\begin{aligned}\log \frac{dR_n}{dQ_n} &= Z_n^{(2)} + \frac{p}{2} \sum_j \log(1 + B_j) \\ &= M_n^{(2)} + \Gamma_3 + o_p(1),\end{aligned}\tag{B.36}$$

whence the Theorem is proved. ■