Inference for Continuous Semimartingales Observed at High Frequency *

Per A. Mykland
The University of Chicago

Lan Zhang
University of Illinois at Chicago

First version: September 9, 2007.
This version: February 12, 2009.

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 derived and documented. Several examples of estimation are provided: powers of volatility, leverage effect, and integrated betas. The first order approximations based on local constancy can be over the period of one observation, or over blocks of successive observations. It has the advantage of gaining 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, Dan Christina Wang, Ting Zhang, and the Editor 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 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, 2003), Barndorff-Nielsen and Shephard (2001, 2002), and Dacorogna, Gençay, Müller, Olsen, and Pictet (2001). Earlier results in probability theory go back to Jacod (1994) and Jacod and Protter (1998). Our own work in this area goes back to Zhang (2001) and Mykland and Zhang (2006). Further references are given below in the Introduction, and in Section 2.5.

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 quantities 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 $t_i,\ 0 = t_0 < t_1 < ... < t_n = T$, are observation times, it is assumed that one can validly make one period approximations of the form

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

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

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

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 invoke approximations (1)-(2)? 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. Once parametric techniques have been used locally in each blocks, estimators of integrated quantities may then be obtained by aggregating local estimators. Any error incurred from this analysis can be corrected directly in the final asymptotic distribution, using adjustments that we provide.

The advantages to thinking parametrically is threefold, 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. A review is given in Fan, Farmen, and Gijbels (1998), and this paper should be consulted for further references. See also Chen and Spokoiny (2007) and Cizek, Härdle, and Spokoiny (2007) for recent papers in this area involving volatility.

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 section (Section 4) 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, 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 different kind of blocking, *pre-averaging*, is used by Podolskij and Vetter (2009) and Jacod, Li, Mykland, Podolskij, and Vetter (2009) 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 rise to contiguity in the absence of noise, but we conjecture that contiguity results can be found under 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. A follow-up discussion on estimation with moving windows, and on how to use this technology for asynchronous observations, can be found in Mykland and Zhang (2009).

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.

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 3-4, and Appendix B-C. 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 $X_t = (X_t^{(1)}, ..., X_t^{(p)})^T$ is called an Itô process provided it satisfies*

$$dX_t = \mu_t dt + \sigma_t dW_t, X_0 = x_0, \tag{3}$$

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 probability distribution will be called P.

If we set

$$\zeta_t = \sigma_t \sigma_t^T, \tag{4}$$

(where "T" in this case means transpose) then the (matrix) integrated covariance process is given as

$$\langle X, X \rangle_t = \int_0^t \zeta_u du. \tag{5}$$

The process (5) is also known as the quadratic covariation of X. We shall sometimes use "integrated volatility" as shorthand in the scalar ($p = 1$) case.

We shall suppose that the process X_t is observed at times $0 = t_0 < t_1 < ... < t_n = T$. Thus, for the moment, we assume synchronous observation of all the p components of the vector X_t. We explain in Mykland and Zhang (2009) 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 sometimes be suppressed). The grids $\mathcal{G}_n = \{0 = t_{n,0} < t_{n,1} < ... < t_{n,n} = T\}$ will not be assumed to be nested when n varies. We then do asymptotics as $n \to \infty$. The basic assumption is that*

$$\max_{1 \le i \le n} |t_{n,j} - t_{n,j-1}| = o(1). \tag{6}$$

We also suppose that the observation times $t_{n,j}$ are nonrandom, but they are allowed to be irregularly spaced. 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 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^*

$$dX_t = \sigma_t dW_t^*, X_0 = x_0, \tag{7}$$

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

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

with

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

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,

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

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 $n \to \infty$ 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, $EI_A g(Z_n) \to EI_A g(Z)$ as $n \to \infty$. The same definition applies to triangular arrays.*

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. $EI_A g(Z_n) = E^* \frac{dP}{dP^*} I_A g(Z_n) \to E^* \frac{dP}{dP^*} I_A g(Z) = EI_A g(Z)$ by uniform integrability of $\frac{dP}{dP^*} I_A g(Z_n)$, and since $\frac{dP}{dP^*}$ is $\mathcal{F}_T$-measurable. ∎

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 5 below. One should always keep in mind that drift most likely is present, and may affect inference. □

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:

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

where $\Delta X_{t_{n,j+1}} = X_{t_{n,j+1}} - X_{t_{n,j}}$, and similarly for $\Delta \breve{W}_{t_{n,j+1}}$ 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 $\breve{W}_t$ 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 $\Delta \breve{W}_{t_{n,j+1}}$ are independent for different j (for fixed n), and that $\Delta \breve{W}_{t_{n,j+1}}$ has conditional distribution $N(0, I\Delta t_{n,j+1})$. 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).*

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

for $j = 0, ..., n$. Here, the quantity $\langle \sigma, W \rangle'_t$ is a three $(p \times p \times p)$ dimensional object (tensor) consisting of elements $\langle \sigma^{(r_1,r_2)}, W^{(r_3)} \rangle'_t$ ($r_1 = 1, ..., p, r_2 = 1, ..., p, r_3 = 1, ..., p$), 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 $U_{t_{n,\iota}}$, $\iota = 0, ..., j$.

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 $(\mathcal{X}_{n,j})_{0 \le j \le n}$, 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) For $j \geq 0$, the conditional P_n^*-distribution of $U^{(1)}_{t_{n,j+1}}$ given $U_0, ..., U_{t_{n,j}}$ is given by (11); and
(iii) For $j \geq 0$, the conditional P_n^*-distribution of $U^{(2)}_{t_{n,j+1}}$ given $U_0, ..., U_{t_{n,j}}, U^{(1)}_{t_{n,j+1}}$ 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

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

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

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

and

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

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 $U_{t_{n,\iota}}$, $\iota = 0, ..., j$. Finally, set

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

In the univariate case, we have the representations

$$k_t = 3\frac{1}{\sigma_t^2} \langle \sigma^2, W \rangle_t' = 6\frac{1}{\sigma_t} \langle \sigma, W \rangle_t' = 6 \langle \log \sigma, W \rangle_t' \tag{17}$$

and

$$\Gamma_0 = \frac{1}{24} \int_0^T k_t^2 dt. \tag{18}$$

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 $U_{t_{n,j}}$, $j = 0, ..., n$. Furthermore, let $(dP^*/dP_n^*)(U_{t_{n,0}}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}})$ be the likelihood ratio (Radon-Nikodym derivative) on $\mathcal{X}_{n,n}$. Then,*

$$\frac{dP^*}{dP_n^*}(U_{t_{n,0}}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}}) \xrightarrow{\mathcal{L}} \exp\{\Gamma_0^{1/2} N(0,1) - \frac{1}{2}\Gamma_0\} \tag{19}$$

stably in law, under P_n^*, as $n \to \infty$. $N(0,1)$ is independent of $\mathcal{F}_T$.

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 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^*). □

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 4.2, where $\langle \sigma, W\rangle_t'$ can be nonzero due to the nonlinear relationship between two securities. □

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

Following (11), write

$$\Delta \breve{W}_{t_{n,j+1}} = \sigma_{t_{n,j}}^{-1} \Delta X_{t_{n,j+1}}. \tag{20}$$

Under the approximating measure P_n^*, $\Delta \breve{W}_{t_{n,j+1}}$ 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

$$M_n^{(0)} = \frac{1}{12} \sum_{j=0}^{n-1} (\Delta t_{n,j+1})^{1/2} \sum_{r_1,r_2,r_3=1}^{p} k_{t_{n,j}}^{(r_1,r_2,r_3)} h_{r_1 r_2 r_3}(\Delta \breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}) \tag{21}$$

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

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

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). □

2.5 Some initial examples

The following is meant for illustration only. The in-depth applications are in Section 4. 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 a scaled version of $\sum_{j=1}^n|\Delta X_{t_{n,j}}|^r$. A general theory for this is given in Barndorff-Nielsen and Shephard (2004b) and Jacod (1994, 2008). For the important cases $r=2$ and $r=4$, see also Barndorff-Nielsen and Shephard (2002), Jacod and Protter (1998), Mykland and Zhang (2006), Zhang (2001), and other work by the same authors.

To reanalyze this estimator with the technology of this paper, 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

$$\begin{aligned}
E_n^*(|\Delta X_{t_{n,j+1}}|^r\mid\mathcal{X}_{n,j})&=|\sigma_{t_{n,j}}|^r E|N(0,1)|^r\Delta t_{n,j+1}^{r/2}\\
\mathrm{Var}_n^*(|\Delta X_{t_{n,j+1}}|^r\mid\mathcal{X}_{n,j})&=|\sigma_{t_{n,j}}|^{2r}\mathrm{Var}(|N(0,1)|^r)\Delta t_{n,j+1}^{r}\text{ and}\\
\mathrm{Cov}_n^*(|\Delta X_{t_{n,j+1}}|^r,\Delta\check{W}_{t_{n,j+1}}\mid\mathcal{X}_{n,j})&=0.
\end{aligned} \tag{23}$$

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

$$\hat{\theta}_n = \frac{1}{E|N(0,1)|^r} \sum_{j=0}^{n-1} \Delta t_{n,j+1}^{1-\frac{r}{2}} |\Delta X_{t_{n,j+1}}|^r. \tag{24}$$

Absolute normal moments can be expressed analytically as in (56) in Section 4.1 below.

From (23), 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 $\frac{\mathrm{Var}(|N(0,1)|^r)}{(E|N(0,1)|^r)^2} \sum_{j=0}^{n-1} |\sigma_{t_{n,j}}|^{2r} \Delta t_{n,j+1}^2$. 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

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} Z \times \left(\frac{\mathrm{Var}(|N(0,1)|^r)}{(E|N(0,1)|^r)^2} T \int_0^T \sigma_t^{2r} dH(t) \right)^{1/2} \tag{25}$$

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

$$H(t) = \lim_{n\to\infty} \frac{n}{T} \sum_{t_{n,j+1} \le t} (t_{n,j+1} - t_{n,j})^2, \tag{26}$$

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 $H(t) = t$.

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 (25) 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) 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^*. □

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

$$Z_n = \frac{n}{T} \sum_{j=0}^{n-1} (\Delta X_{t_{n,j+1}})^3. \tag{27}$$

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).

We shall see in Section 4.3 that quantities similar to (27) 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 (27) 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.

By the same reasoning as in Example 1,

$$\begin{aligned} E_n^*(\Delta X_{t_{n,j+1}}^3 \mid \mathcal{X}_{n,j}) &= 0 \\ \mathrm{Var}_n^*(\Delta X_{t_{n,j+1}}^3 \mid \mathcal{X}_{n,j}) &= \sigma_{t_{n,j}}^6 \mathrm{Var}(N(0,1)^3)\Delta t^3 = 15\sigma_{t_{n,j}}^6 \Delta t^3 \text{ and} \\ \mathrm{Cov}_n^*(\Delta X_{t_{n,j+1}}^3, \Delta \breve{W}_{t_{n,j+1}} \mid \mathcal{X}_{n,j}) &= \sigma_{t_{n,j}}^3 \mathrm{Cov}(N(0,1)^3, N(0,1))\Delta t^2 = 3\sigma_{t_{n,j}}^3 \Delta t^2. \end{aligned} \tag{28}$$

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

$$\begin{aligned} b &= 3\int_0^T \sigma_t^3 dW_t^* \text{ and} \\ a^2 &= 6\int_0^T \sigma_t^6 dt. \end{aligned} \tag{29}$$

Remark 5. (Sample of calculation). To see in more detail how (29) 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 (28), and by interchanging limits and quadratic variation (Chapter VI.6 (p. 376-388) in Jacod and Shiryaev (2003), cf Remark 4 above), we get

$$\begin{aligned} \langle V, V\rangle_t &= 15\int_0^t \sigma_u^6 du \text{ and} \\ \langle V, W^*\rangle_t &= 3\int_0^t \sigma_u^6 du. \end{aligned} \tag{30}$$

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 (30),

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

In particular, $g_t^2 = 6\sigma_t^6$. This yields (29).

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^*. □

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,

$$\begin{aligned}\mathrm{Cov}_n^*(\Delta X_{t_{n,j+1}}^3, h_3(\Delta \breve{W}_{t_{n,j+1}}/\Delta t^{1/2}) \mid \mathcal{X}_{n,j})\Delta t^{1/2} &= \sigma_{t_{n,j}}^3 \mathrm{Cov}(N(0,1)^3, h_3(N(0,1)))\Delta t^2 \\ &= 6\sigma_{t_{n,j}}^3 \Delta t^2 \end{aligned} \tag{31}$$

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

$$\begin{aligned}A_{12} &= \frac{1}{2}\int_0^T k_t \sigma_t^3 dt \\ &= \frac{3}{2}\langle \sigma^2, X\rangle_T,\end{aligned} \tag{32}$$

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 (29), while

$$b' = 3\int_0^T \sigma_t^3 dW_t^* + \frac{3}{2}\langle \sigma^2, X\rangle_T. \tag{33}$$

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

$$b' = 3\int_0^T \sigma_t^3 (dW_t + \sigma_t^{-1}\mu_t dt) + \frac{3}{2}\langle \sigma^2, X\rangle_T. \tag{34}$$

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 happens again in Example 5 in Section 4.3. □

A direct derivation of this same limit is given in Example 6 of Kinnebrock and Podolskij (2008). 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

$$\mathcal{H}_n = \{0 = \tau_{n,0} < \tau_{n,1} < ... < \tau_{n,K_n} = T\} \subseteq \mathcal{G}_n \tag{35}$$

If we set

$$\begin{aligned} M_{n,i} &= \#\{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]\} \\ &= \text{ number of intervals } (t_{n,j-1}, t_{n,j}] \text{ in } (\tau_{n,i-1}, \tau_{n,i}] \end{aligned} \tag{36}$$

we shall suppose that

$$\max_i M_{n,i} = O(1) \text{ as } n \to \infty, \tag{37}$$

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

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

$$\begin{aligned} &X_0 = x_0 \\ &\text{for each } i = 1, K_n : \\ &\Delta X_{t_{n,j+1}} = \sigma_{\tau_{n,i-1}} \Delta W^Q_{t_{n,j+1}} \text{ for } t_{n,j+1} \in (\tau_{n,i-1}, \tau_{n,i}]. \end{aligned} \tag{38}$$

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) For $j \geq 0$, the conditional Q_n-distribution of $U^{(1)}_{t_{n,j+1}}$ given $U_0, ..., U_{t_{n,j}}$ is given by (38), where $\Delta W^Q_{t_{n,j+1}}$ is conditionally normal with mean zero and variance $\Delta t_{n,j+1}$; and
(iii) For $j \geq 0$, the conditional Q_n-distribution of $U^{(2)}_{t_{n,j+1}}$ given $U_0, ..., U_{t_{n,j}}, U^{(1)}_{t_{n,j+1}}$ 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

$$\ell(\Delta x; \zeta) = -\frac{1}{2} \log \det(\zeta) - \frac{1}{2} \Delta x^T \zeta^{-1} \Delta x. \tag{39}$$

We then obtain directly that

Proposition 2. *The likelihood ratio between Q_n and P^*_n is given by*

$$\begin{aligned} &\log \frac{dQ_n}{dP^*_n}(U_{t_{n,0}}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}}) \\ &\quad = \sum_i \sum_{\tau_{n,i-1} \leq t_{n,j} < \tau_{n,i}} \left\{ \ell(\Delta X_{t_{n,j+1}}; \zeta_{\tau_{n,i-1}} \Delta t_{n,j+1}) - \ell(\Delta X_{t_{n,j+1}}; \zeta_{t_{n,j}} \Delta t_{n,j+1}) \right\} \end{aligned} \tag{40}$$

Definition 6. *To measure the extent to which we hold the volatility constant, we define the following "Asymptotic Decoupling Delay" (ADD) by*

$$K(t) = \lim_{n\to\infty} \sum_i \sum_{t_{n,j}\in(\tau_{n,i-1},\tau_{n,i})\cap[0,t]} (t_{n,j} - \tau_{n,i-1}), \tag{41}$$

provided the limit exists.

From (6) and (37), 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

$$K(t) = \frac{1}{2}(M-1)t. \tag{42}$$

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 (41)) exists. Set*

$$Z_n^{(1)} = \frac{1}{2}\sum_i \sum_{t_{n,j}\in[\tau_{n,i-1},\tau_{n,i})} \left(\Delta X_{t_{n,j+1}}^T (\zeta_{t_{n,j}}^{-1} - \zeta_{\tau_{n,i-1}}^{-1})\Delta X_{t_{n,j+1}} \Delta t_{n,j+1}^{-1}\right) \tag{43}$$

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

$$\Gamma_1 = \frac{1}{2}\int_0^T \mathrm{tr}(\zeta_t^{-2}\langle\zeta,\zeta\rangle_t')dK(t), \tag{44}$$

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

$$\log\frac{dQ_n}{dP_n^*} = M_n^{(1)} - \frac{1}{2}\Gamma_1 + o_p(1). \tag{45}$$

Furthermore, if $M_n^{(0)}$ is as defined in (21), then the pair $(M_n^{(0)}, M_n^{(1)})$ converges stably under P_n^* 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.

EXAMPLE 4. For a scalar process on the form $dX_t = \mu_t dt + \sigma_t dW_t$, and with equidistant observations of X, Γ_1 in (44) can be written

$$\Gamma_1 = \frac{M-1}{4} \int_0^T \sigma_t^{-4} \langle \sigma^2, \sigma^2 \rangle_t' dt. \tag{46}$$

From (17)-(18), $\Gamma_0 = \frac{3}{8} \int_0^T \sigma_t^{-6} (\langle \sigma^2, X \rangle_t')^2 dt$. Thus, Γ_0 is related to the leverage effect, while Γ_1 is related to the volatility of volatility. 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

$$\Gamma_0 = \frac{3}{8}(\rho\gamma)^2 \int_0^T \sigma_t^{-2} dt \quad \text{and} \quad \Gamma_1 = \frac{1}{4}\gamma^2 (M-1) \int_0^T \sigma_t^{-2} dt. \tag{47}$$

□

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

$$\log \frac{dP}{dQ_n} = \log \frac{dP}{dP^*} + \log \frac{dP^*}{dP_n^*} + \log \frac{dP_n^*}{dQ_n} \ . \tag{48}$$

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 (21), with $\Delta W_{t_{n,j+1}}^Q$ replacing $\Delta \breve{W}_{t_{n,j+1}}$. (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, (45) should be replaced by

$$\log \frac{dP_n^*}{dQ_n} = M_n^{(1,Q)} - \frac{1}{2}\Gamma_1 + o_p(1). \tag{49}$$

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,

$$M_n^{(0,Q)} = \frac{1}{12} \sum_{j=0}^{n-1} (\Delta t_{n,j+1})^{1/2} \sum_{r_1,r_2,r_3=1}^{p} k_{t_{n,j}}^{(r_1,r_2,r_3)} h_{r_1 r_2 r_3}(\Delta W_{t_{n,j+1}}^Q / (\Delta t_{n,j+1})^{1/2}), \text{ and}$$
$$M_n^{(1,Q)} = -\frac{1}{2} \sum_i \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} \operatorname{tr}\left(\sigma_{\tau_{n,i-1}}^T (\zeta_{t_{n,j}}^{-1} - \zeta_{\tau_{n,i-1}}^{-1}) \sigma_{\tau_{n,i-1}} h_{\cdot\cdot}(\Delta W_{t_{n,j+1}}^Q / (\Delta t_{n,j+1})^{1/2}) \right). \tag{50}$$

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

$$M_n^{(G,Q)} = \sum_{j=0}^{n-1} (\Delta t_{n,j+1})^{1/2} \left(\sigma_{\tau_{n,i-1}}^{-1} \mu_{\tau_{n,i-1}} h.(\Delta W_{t_{n,j+1}}^Q / (\Delta t_{n,j+1})^{1/2}) \right) \tag{51}$$

("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

$$\Gamma_G = \int_0^T \mu_t^T (\sigma_t^T \sigma_t)^{-1} \mu_t dt. \tag{52}$$

We therefore get the following summary of our results:

$$\begin{aligned} \log \frac{dP}{dP^*} &= M_n^{(G,Q)} - \frac{1}{2}\Gamma_G + o_p(1) \\ \log \frac{dP^*}{dP_n^*} &= M_n^{(0,Q)} - \frac{1}{2}\Gamma_0 + o_p(1) \\ \log \frac{dP_n^*}{dQ_n} &= M_n^{(1,Q)} - \frac{1}{2}\Gamma_1 + o_p(1). \end{aligned} \tag{53}$$

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. The term $M_n^{(G,Q)}$ 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 5 in Section 4.3 below. □

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.

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). The question of irregular sampling is discussed in Mykland and Zhang (2009). 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

$$\hat{\sigma}^2_{\tau_{n,i}} = \frac{1}{\Delta t_n (M_n - 1)} \sum_{t_{n,j} \in (\tau_{n,i}, \tau_{n,i+1}]} (\Delta X_{t_{n,j}} - \overline{\Delta X}_{\tau_{n,i}})^2 \text{ and}$$
$$\overline{\Delta X}_{\tau_{n,i}} = \frac{1}{M_n} \sum_{t_{n,j} \in (\tau_{n,i}, \tau_{n,i+1}]} \Delta X_{t_{n,j}} = \frac{1}{M_n}(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \tag{54}$$

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 $N(0, \sigma^2_{\tau_{n,i}} \Delta t_n)$ 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

$$\widehat{|\sigma_{\tau_{n,i}}|^r} = c^{-1}_{M-1,r} (\hat{\sigma}^2_{\tau_{n,i}})^{r/2} \tag{55}$$

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

$$c_{M,r} = E\left((\chi^2_M / M)^{r/2}\right)$$
$$= \left(\frac{2}{M}\right)^{r/2} \frac{\Gamma(\frac{r+M}{2})}{\Gamma(\frac{M}{2})} \tag{56}$$

where χ^2_M 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

$$\hat{\theta}_n = (M\Delta t)\sum_i \widehat{|\sigma_{\tau_{n,i}}|^r} \tag{57}$$

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,

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0,1)\left(TM\left(\frac{c_{M-1,2r}}{c^2_{M-1,r}} - 1\right)\int_0^T \sigma_t^{2r}dt\right)^{1/2} \tag{58}$$

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

$$\begin{aligned}\operatorname{Var}((M\Delta t)\sum_i (\widehat{|\sigma_{\tau_{n,i}}|^r}) \mid \mathcal{Y}_{n,i}) &= \sigma^{2r}_{\tau_{n,i}}(M\Delta t)^2 c^{-2}_{M-1,r}\operatorname{Var}\left((\chi^2_{M-1}/(M-1))^{r/2}\right) \\ &= \sigma^{2r}_{\tau_{n,i}}(M\Delta t)\frac{TM}{n}\left(\frac{c_{M-1,2r}}{c^2_{M-1,r}} - 1\right). \end{aligned} \tag{59}$$

Remark 8. (Not taking out the mean). One can replace $\hat{\sigma}^2_{\tau_{n,i}}$ by

$$\tilde{\sigma}^2_{\tau_{n,i}} = \frac{1}{\Delta t_n M_n}\sum_{t_{n,j}\in(\tau_{n,i},\tau_{n,i+1}]}(\Delta X_{t_{n,j}})^2, \tag{60}$$

and take

$$\widetilde{|\sigma_{\tau_{n,i}}|^r} = c^{-1}_{M,r}(\tilde{\sigma}^2_{\tau_{n,i}})^{r/2} \tag{61}$$

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

$$TM\left(\frac{c_{M,2r}}{c^2_{M,r}} - 1\right)\int_0^T \sigma_t^{2r}dt. \tag{62}$$

□

4.1.2 Asymptotic Efficiency

We note that for large M,

$$\text{asymptotic variance of } n^{1/2}(\hat{\theta}_n - \theta) \downarrow T\frac{r^2}{2}\int_0^T \sigma_t^{2r}dt. \tag{63}$$

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 (63), the earlier estimator has asymptotic relative efficiency (ARE)

$$ARE(\text{estimator from Example 1}) = \frac{\text{asymptotic variance in (63)}}{\text{asymptotic variance of estimator from Example 1}}$$
$$= \frac{r^2}{2}\left(\frac{c_{1,2r}}{c_{1,r}^2}-1\right)^{-1} \tag{64}$$

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 (57).

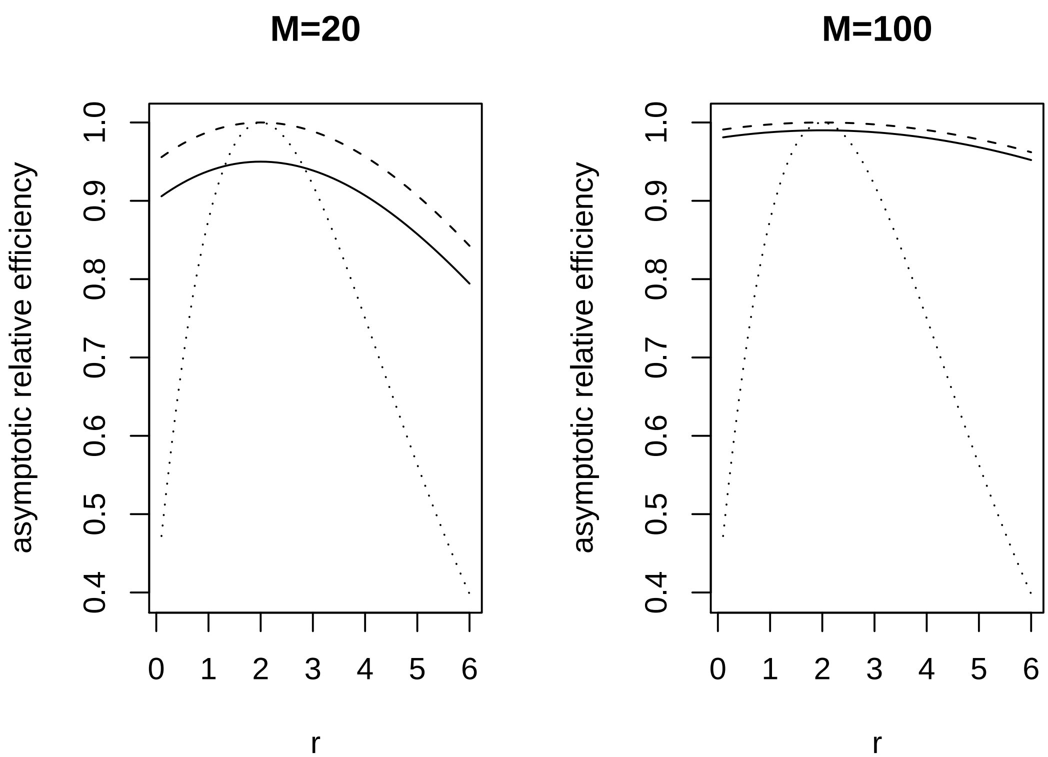

Figure 1. Asymptotic relative efficiency (ARE) of three estimators of $\theta=\int_0^T|\sigma|_t^r dt$, as a function of r. The dotted curve corresponds to the traditional estimator, which is proportional to $\sum_{j=1}^n|\Delta X_{t_{n,j}}|^r$. The solid and dashed lines are the ARE's of the block based 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).

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. □

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 (63) is only approximately attained. It would be desirable to build a theory where $M \to \infty$ as $n \to \infty$. Such a theory would presumably be able to pick up any biases due to the blocking. □

4.2 Integrated betas

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

$$dY_t = \sum_{i=1}^{p} \beta_t^{(k)} dX_t^{(k)} + dZ_t, \text{ with } \langle X^{(k)}, Z \rangle_t = 0 \text{ for all } t \text{ and } k. \tag{65}$$

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 (65) becomes, for the observables,

$$\Delta Y_{t_{n,j}} = \sum_{k=1}^{p} \beta_{\tau_{n,i-1}}^{(k)} \Delta X_{t_{n,j}}^{(k)} + \Delta Z_{t_{n,j}}. \tag{66}$$

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

$$\hat{\theta}_n^{(k)} = \sum_i \hat{\beta}_{\tau_{n,i-1}}^{(k)} M \Delta t. \tag{67}$$

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

$$n(M\Delta t)^2 \sum_i \mathrm{Cov}_{Q_n}(\hat{\beta}_{\tau_{n,i-1}} - \beta_{\tau_{n,i-1}} \mid \mathcal{Y}_{n,i-1}). \tag{68}$$

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

$$\mathrm{Cov}_{Q_n}(\hat{\beta}_{\tau_{n,i-1}} - \beta_{\tau_{n,i-1}} \mid \mathcal{Y}'_{n,i-1}) = \mathrm{Var}_{Q_n}(\Delta Z_{t_{n,j}} \mid \mathcal{Y}'_{n,i-1}) \times (\Delta X^T \Delta X)^{-1} \tag{69}$$

where, with some abuse of notation, ΔX is the matrix of $\Delta X^{(k)}_{t_{n,j}}$, where $k = 1, ..., p$, 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)

$$E_{Q_n}((\Delta X^T \Delta X)^{-1} \mid \mathcal{Y}_{n,i-1}) = (\langle X, X\rangle'_{\tau_{n,i-1}})^{-1} \Delta t_n^{-1}/(M - p - 1). \tag{70}$$

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

$$\mathrm{Cov}_{Q_n}(\hat{\beta}_{\tau_{n,i-1}} - \beta_{\tau_{n,i-1}} \mid \mathcal{Y}_{n,i-1}) = \langle Z, Z\rangle'_{\tau_{n,i-1}} (\langle X, X\rangle'_{\tau_{n,i-1}})^{-1}/(M - p - 1). \tag{71}$$

It follows that the limit of (68) is

$$\frac{MT}{M - p - 1} \int_0^T \langle Z, Z\rangle'_t (\langle X, X\rangle'_t)^{-1} dt. \tag{72}$$

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 (72), 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

$$\widetilde{\langle \sigma^2, X\rangle}_T = \sum_i (\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \tag{73}$$

where $\hat{\sigma}^2_{\tau_{n,i}}$ and $\overline{\Delta X}_{\tau_{n,i}}$ are given above in (54). It turns out, however, that this estimator is asymptotically biased, as follows:

PROPOSITION 3. *Let $M \geq 2$. In the equally spaced case, under both P^* and P, and as $n \to \infty$,*

$$\widetilde{\langle \sigma^2, X \rangle}_T \xrightarrow{\mathcal{L}} \frac{1}{2}\langle \sigma^2, X \rangle + N(0,1) \times \left(\frac{4}{M-1} \int_0^T \sigma_t^6 dt \right)^{1/2} \tag{74}$$

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

The derivation of this result, along with that of the result in Example 5 below, is given in Appendix C. This appendix gives what we think is a typical way of showing results based on the general theory of Sections 2-3.

Accordingly, we define an asymptotically unbiased estimator of leverage effect by

$$\widehat{\langle \sigma^2, X \rangle}_T = 2 \sum_i (\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \tag{75}$$

In other words, $\widehat{\langle \sigma^2, X \rangle}_T = 2\widetilde{\langle \sigma^2, X \rangle}_T$. Following Proposition 3,

$$\widehat{\langle \sigma^2, X \rangle}_T - \langle \sigma^2, X \rangle \xrightarrow{\mathcal{L}} c_M^{1/2} N(0,1) \tag{76}$$

stably under P^* and P, where

$$c_M = \frac{16}{M-1} \int_0^T \sigma_t^6 dt. \tag{77}$$

It is important to note that the bias in $\widetilde{\langle \sigma^2, X \rangle}_T$ comes from error induced by both the one period and multi period discretizations (the adjustment from P^* to P_n^*, then to Q_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 $n \to \infty$. The presumed optimal convergence rate of $\widehat{\langle \sigma^2, X \rangle}_T - \langle \sigma^2, X \rangle_T$ is $O_p(n^{-1/4})$, in analogy with the results in Zhang (2006). This makes sense because there is an inherited noisy measure $\hat{\sigma}_t^2$ of σ_t^2 in the definition the estimator $\widehat{\langle \sigma^2, X \rangle}_T$, see (75). 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 $M \to \infty$ with n, but this is beyond the scope of this paper. □

EXAMPLE 5. (THE RÔLE OF μ: THE EFFECT OF NOT REMOVING THE MEAN FROM THE ESTIMATE OF σ^2). In the development above, the drift μ did not surface. This 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.

Suppose that one wishes to use the estimator (75), but replacing $\hat{\sigma}^2_{\tau_{n,i}}$ by the estimator $\tilde{\sigma}^2_{\tau_{n,i}}$ from (60). An estimator analogous to $\widehat{\langle \sigma^2, X \rangle}_T$ is then

$$\widehat{\langle \sigma^2, X \rangle}_T^{\text{with mean}} = 2 \sum_i (\tilde{\sigma}^2_{\tau_{n,i+1}} - \tilde{\sigma}^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \tag{78}$$

We show in Section C.2 that, for $M \geq 2$,

$$\widehat{\langle \sigma^2, X\rangle}_T^{\text{with mean}} \xrightarrow{\mathcal{L}} \frac{M-2}{M}\langle \sigma^2, X\rangle_T - \frac{4}{M}\int_0^T \sigma_t^3(dW_t + \sigma_t^{-1}\mu_t dt) + N(0,1)\left(16\frac{M+1}{M^2}\int_0^T \sigma_t^6 dt\right)^{1/2}. \tag{79}$$

Hence, with this estimator, μ does show up in asymptotic expressions. The estimation of leverage effect is therefore a case where it is important to remove the mean in each block. □

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 $E_{Q_n}(V^2 \mid \mathcal{Y}_{n,i-1}) = \sigma^2_{\tau_{n,i-1}} EU^2_{(k)}$, 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 (65), but now the purpose is to estimate $\langle Z, Z\rangle_T$, *i.e.*, the residual quadratic variation of Y after regressing on X. Blocking can here be used in much the same way as in Section 4.2.

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

$$\theta = \sum_i \theta_{n,i} + O_p(n^{-1}) \tag{80}$$

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

$$\hat{\theta}_n = \sum_i \hat{\theta}_{n,i}, \tag{81}$$

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,

$$\begin{aligned}
&\sum_i \mathrm{Var}_n^Q(Z_{n,i} \mid \mathcal{Y}_{n,i}) \to \int_0^T f_t^2 dt \\
&\sum_i \mathrm{Cov}_n^Q(Z_{n,i}, W^Q_{\tau_{n,i}} - W^Q_{\tau_{n,i-1}} \mid \mathcal{Y}_{n,i}) \to \int_0^T g_t dt
\end{aligned} \tag{82}$$

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

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

and

$$A_{13} = -\frac{1}{2} \lim_{n\to\infty} \sum_i \mathrm{Cov}_n^Q\left(Z_{n,i}, \sum_{t_{n,j}\in(\tau_{n,i-1},\tau_{n,i}]} \left(\sigma^2_{\tau_{n,i-1}}(\zeta^{-1}_{t_{n,j}} - \zeta^{-1}_{\tau_{n,i-1}}) h_2(\Delta W^Q_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2})\right) \mid \mathcal{Y}_{n,i}\right). \tag{83}$$

We finally obtain

Theorem 5. *(Summary of method in the scalar case). In the setting described, and subject to regularity conditions,*

$$n^\alpha(\hat{\theta}_n - \theta_n) \xrightarrow{\mathcal{L}} b + A_{12} + A_{13} + N(0,1)\left(\int_0^T (f_t^2 - g_t^2) dt\right)^{1/2} \tag{84}$$

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

$$b = \int_0^T g_t dW_t^* = \int_0^T g_t(dW_t + \sigma_t^{-1}\mu_t dt). \tag{85}$$

5 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 in Section 4.

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 asymptotic likelihood ratios can be 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

$$\log \frac{dQ_\infty}{dP} \;=\; \log \frac{dQ_\infty}{dP^*_\infty} \;+\; \log \frac{dP^*_\infty}{dP^*} \;+\; \log \frac{dP^*}{dP}. \tag{86}$$

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.

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 asymptotic limits, denoted by subscript "∞". This connects to the statistical concept of contiguity, cf. Remark 2.

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 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.

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 $M \to \infty$ 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, IL 60637, U.S.A.; mykland@galton.uchicago.edu; http://galton.uchicago.edu/∼mykland.

and

Department of Finance, The University Illinois at Chicago, Chicago, IL 60607, U.S.A.; lanzhang@uic.edu; http://tigger.uic.edu/∼lanzhang/.

REFERENCES

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., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (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.

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., M. Farmen, and I. Gijbels (1998): "Local Maximum Likelihood Estimation and Inference," *Journal of the Royal Statistcal Society ser. B*, 60, 591–608.

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

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.

Jacod, J. (1994): "Limit of Random Measures Associated with the Increments of a Brownian Semimartingale," Tech. rep., Université de Paris VI.

——— (2008): "Asymptotic Properties of Realized Power Variations and Related Functionals of Semimartingales," *Stochastic Processes and Applications*, 118, 517–559.

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Applications, forthcoming.*

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.

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

Kinnebrock, S. and M. Podolskij (2008): "A Note on the Central Limit Theorem for Bipower Variation of General Functions," *Stochastic Processes and their Applications*, 118, 1056–1070.

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.

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.

Mykland, P. A. and L. Zhang (2006): "ANOVA for Diffusions and Itô Processes," *Annals of Statistics*, 34, 1931–1963.

——— (2009): "The Econometrics of High Frequency Data," (submitted to *Statistical Methods for Stochastic Differential Equations*, M. Kessler, A. Lindner, and M. Sørensen, eds., Chapman and Hall/CRC Press).

Podolskij, M. and M. Vetter (2009): "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli, forthcoming.*

Renault, E. and B. J. Werker (2006): "Causality effects in return volatility measures with random times," *Journal of Econometrics* (forthcoming).

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. 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 Statistican Association*, 82, 559–567.

Weisberg, S. (1985): *Applied Linear Regression*, New York: Wiley, second ed.

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.

——— (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 avoid having superscript "$*$" everywhere, use the notation P for P^*, until the end of the proof of Theorem 1 only, and without loss of generality. This is only a matter of notation. One understands the differential $\sigma_t dW_t$ to be a p-dimensional vector with r_1'th component $\sum_{r_2=1}^{p} \sigma_t^{(r_1,r_2)} dW_t^{(r_2)}$. To study the properties of this approximation, consider the following "strong approximation". Set

$$d\sigma_t = \tilde{\sigma}_t dt + f_t dW_t + g_t dB_t \tag{A.1}$$

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

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

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

$$f_t^{(r_1,r_2,r_3)} = \langle \sigma^{(r_1,r_2)}, W^{(r_3)} \rangle_t'. \tag{A.3}$$

To proceed with the proof, some further notation. Define

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

$\breve{\sigma}_t^{(r_1,r_2)}$ and $\breve{f}_t(r_1, r_2, r_3)$ 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 $d\breve{\zeta}_t = \sigma_t^{-1} d\sigma_t + (\sigma_t^{-1} d\sigma_t)^T + dt$ terms. Hence

$$\langle \breve{\zeta}^{(r_1,r_2)}, W^{(r_3)} \rangle_t' = \breve{f}_t^{(r_1,r_2,r_3)} + \breve{f}_t^{(r_2,r_1,r_3)}. \tag{A.5}$$

Also

$$k_t^{(r_1,r_2,r_3)} = \langle \breve{\zeta}^{(r_1,r_2)}, W^{(r_3)} \rangle_t'[3] = \breve{f}_t^{(r_1,r_2,r_3)}[6] \tag{A.6}$$

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

Proof of Theorem 1. Note that, from (20) and (A.2)

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

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

$$
\begin{aligned}
&\mathrm{cum}_3(\Delta \breve{W}^{(r_1)}_{t_{n,j+1}}, \Delta W^{(r_2)}_{t_{n,j+1}}, \Delta W^{(r_3)}_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}) \\
&\quad \doteq \mathrm{cum}(\sum_{s_2,s_3} \breve{f}^{(r_1,s_2,s_3)}_{t_{n,j}} \int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{t} dW^{(s_3)}_u \right) dW^{(s_2)}_t, \Delta W^{(r_2)}_{t_{n,j+1}}, \Delta W^{(r_3)}_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}) \\
&\quad = \sum_{s_2,s_3} \breve{f}^{(r_1,s_2,s_3)}_{t_{n,j}} \mathrm{cum}(\int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{t} dW^{(s_3)}_u \right) dW^{(s_2)}_t, \Delta W^{(r_2)}_{t_{n,j+1}}, \Delta W^{(r_3)}_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}}) \\
&\quad = \sum_{s_2,s_3} \breve{f}^{(r_1,s_2,s_3)}_{t_{n,j}} \mathrm{Cov}(\int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{t} dW^{(s_3)}_u \right) dt \delta^{s_2,r_2}, \Delta W^{(r_3)}_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}})[2] \\
&\quad = \sum_{s_3} \breve{f}^{(r_1,r_2,s_3)}_{t_{n,j}} \mathrm{Cov}(\int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{t} dW^{*(s_3)}_u \right) dt, \Delta W^{(r_3)}_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}})[2] \\
&\quad = \int_{t_{n,j}}^{t_{n,j+1}} dt \sum_{s_3} \breve{f}^{(r_1,r_2,s_3)}_{t_{n,j}} \mathrm{Cov}(\int_{t_{n,j}}^{t} dW^{*(s_3)}_u, \Delta W^{(r_3)}_{t_{n,j+1}} | \mathcal{F}_{t_{n,j}})[2] \\
&\quad = \int_{t_{n,j}}^{t_{n,j+1}} dt \sum_{s_3} \breve{f}^{(r_1,r_2,s_3)}_{t_{n,j}} (t - t_{n,j}) \delta_{s_3,r_3}[2] \\
&\quad = \frac{1}{2} \Delta t^2_{n,j+1} \breve{f}^{(r_1,r_2,r_3)}_{t_{n,j}} \; [2],
\end{aligned} \tag{A.8}
$$

where "[2]" represents the swapping of r_2 and r_3 (see p. 29-30 of McCullagh (1987) for a discussion of the notation). In the third transition, we have used the third Bartlett type identity for martingales. Hence

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

by symmetry. Set $\kappa^{r_1,r_2,r_3} = \mathrm{cum}_3(\Delta \breve{W}^{(r_1)}_{t_{n,j+1}} / \Delta t^{1/2}_{n,j+1}, \Delta \breve{W}^{(r_2)}_{t_{n,j+1}} / \Delta t^{1/2}_{n,j+1}, \Delta \breve{W}^{(r_3)}_{t_{n,j+1}} / \Delta t^{1/2}_{n,j+1} | \mathcal{F}_{t_{n,j}})$, and similarly for other cumulants. From (15) and (A.9),

$$
\kappa^{r_1,r_2,r_3} = \frac{1}{2} \Delta t^{1/2}_{n,j+1} k^{(r_1,r_2,r_3)}_{t_{n,j}} + O_p(\Delta t). \tag{A.10}
$$

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

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

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

$$\kappa^{r_1,r_2} = \delta^{r_1,r_2}_{t_{n,j}} + O_p(\Delta t). \tag{A.12}$$

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

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 $\Delta\breve{W}_{t_{n,j+1}}/\Delta t^{1/2}_{n,j+1}$ given $\mathcal{F}_{t_{n,j}}$, on the log scale as

$$\begin{aligned}\log f_{n,j+1}(x) = \log\phi(x;\delta^{r_1,r_2}) &+ \frac{1}{3!}\kappa^{r_1,r_2,r_3}h_{r_1r_2r_3}(x)\\ &+ \frac{1}{2}(\kappa^{r_1,r_2} - \lambda^{r_1,r_2})h_{r_1r_2}(x) + \frac{1}{4!}\kappa^{r_1,r_2,r_3,r_4}h_{r_1r_2r_3r_4}(x)\\ &+ \kappa^{r_1,r_2,r_3}\kappa^{r_4,r_5,r_6}h_{r_1r_2r_3r_4r_5r_6}(x)\frac{[10]}{6!}\\ &- \frac{1}{72}(\kappa^{r_1,r_2,r_3}h_{r_1r_2r_3}(x))^2 + O_p(\Delta t^{3/2})\end{aligned} \tag{A.13}$$

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 $h_{r_1r_2r_3} = h_{r_1}h_{r_2}h_{r_3} - h_{r_1}\delta_{r_2,r_3}[3]$, with $h_{r_1} = \delta_{r_1,r_2}x^{r_2}$. Observe that

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

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

$$h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}) = Z_{r_1}Z_{r_2}Z_{r_3} - Z_{r_1}\delta_{r_2,r_3}[3] \tag{A.15}$$

Under the approximating measure, therefore, $E_n(h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2})|\mathcal{F}_{t_{n,j}}) = 0$, while

$$\mathrm{Cov}_n(h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}), h_{r_4r_5r_6}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2})|\mathcal{F}_{t_{n,j}}) = \delta_{r_1,r_4}\delta_{r_2,r_5}\delta_{r_3,r_6}[6] \tag{A.16}$$

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

$$\begin{aligned}&\mathrm{Var}_n(\frac{1}{3!}\kappa^{r_1,r_2,r_3}h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2})|\mathcal{F}_{t_{n,j}})\\ &\quad= \frac{1}{36}\kappa^{r_1,r_2,r_3}\kappa^{r_4,r_5,r_6}\mathrm{Cov}_n(h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}), h_{r_4r_5r_6}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2})|\mathcal{F}_{t_{n,j}})\\ &\quad= \frac{1}{6}\kappa^{r_1,r_2,r_3}\kappa^{r_4,r_5,r_6}\delta_{r_1,r_4}\delta_{r_2,r_5}\delta_{r_3,r_6}\\ &\quad= \Delta t_{n,j+1}\frac{1}{24}k^{r_1,r_2,r_3}_{t_{n,j}}k^{r_4,r_5,r_6}_{t_{n,j}}\delta_{r_1,r_4}\delta_{r_2,r_5}\delta_{r_3,r_6} + O_p(\Delta t^{3/2})\end{aligned} \tag{A.17}$$

by symmetry of the κ's. Thus

$$\sum_{t_{n,j+1}\le t} \mathrm{Var}_n(\frac{1}{3!}\kappa^{r_1,r_2,r_3} h_{r_1 r_2 r_3}(\Delta \breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2})|\mathcal{F}_{t_{n,j}}) \xrightarrow{p} \int_0^t \frac{1}{24} k_u^{r_1,r_2,r_3} k_u^{r_4,r_5,r_6} \delta_{r_1,r_4}\delta_{r_2,r_5}\delta_{r_3,r_6} du \tag{A.18}$$

under P_n^*, still using the summation convention. Note that (A.18), with $t = T$, is the same as Γ_0 in (16). By the same methods, and since Hermite polynomials of different orders are orthogonal under the approximating measure,

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

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

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

converges stably in law to a normal distribution with random variance Γ_0. (Note that $\breve{M}_n^{(0)} = M_n^{(0)} + O_p(\Delta t^{1/2})$ from (21), and that we are still using the summation convention). We now observe that, in the notation of (A.13),

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

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

Remark 12. 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

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

where $M_n^{(0)}$ is as defined in equation (21). 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,

$$
\begin{aligned}
E^* g(Z_n) &= E_n^* g(Z_n) \exp\{M_n^{(0)} - \frac{1}{2}\Gamma_0\}(1 + o(1)) \\
&\to E g(Z) \exp\{M - \frac{1}{2}\Gamma_0\} \\
&= E_\infty^* g(b_1 + c_1\Gamma_0^{1/2} V_1 + c_2 V_2) \exp\{\Gamma_0^{1/2} V_1 - \frac{1}{2}\Gamma_0\} \\
&= \int_{-\infty}^{\infty} E_\infty^* g(b_1 + c_1\Gamma_0^{1/2} v + c_2 V_2) \exp\{\Gamma_0^{1/2} v - \frac{1}{2}\Gamma_0\}(2\pi)^{-1/2} \exp\{-\frac{1}{2}v^2\} dv \\
&= \int_{-\infty}^{\infty} E_\infty^* g(b_1 + c_1\Gamma_0^{1/2}(u + \Gamma_0^{1/2}) + c_2 V_2)(2\pi)^{-1/2} \exp\{-\frac{1}{2}u^2\} du \quad (u = v - \Gamma_0^{1/2}) \\
&= E_\infty^* g(Z + c_1\Gamma_0) \qquad \text{(A.23)}
\end{aligned}
$$

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

B Proof of Theorem 3

Let $Z_n^{(1)}$ be given by (43). Set

$$\Delta Z_{n,t_{n,j+1}}^{(1)} = \frac{1}{2}\Delta X_{t_{n,j+1}}^T (\zeta_{t_{n,j}}^{-1} - \zeta_{\tau_{n,i-1}}^{-1}) \Delta X_{t_{n,j+1}} \Delta t_{n,j+1}^{-1} \qquad \text{(B.24)}$$

and note that $Z_n^{(1)} = \sum_j \Delta Z_{n,t_{n,j+1}}^{(1)}$. Set $A_j = \zeta_{t_{n,j}}^{1/2} \zeta_{\tau_{n,i-1}}^{-1} \zeta_{t_{n,j}}^{1/2} - I$.

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

$$E_{P_n^*}(\Delta Z_{n,t_{n,j+1}}^{(1)} | \mathcal{X}_{n,t_{n,j}}) = -\frac{1}{2}\mathrm{tr}(A_j) \qquad \text{(B.25)}$$

and

$$\text{conditional variance of } \Delta Z_{n,t_{n,j+1}}^{(1)} = \frac{1}{2}\mathrm{tr}(A_j^2) \qquad \text{(B.26)}$$

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

$$M_n^{(1)} = Z_n^{(1)} + (1/2)\sum_j \mathrm{tr}(A_j). \qquad \text{(B.27)}$$

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

$$\langle M_n^{(1)}, M_n^{(1)} \rangle^{\mathcal{G}} = \frac{1}{2}\sum_j \mathrm{tr}(A_j^2). \qquad \text{(B.28)}$$

Now note that, by analogy to the development in Zhang (2001), Mykland and Zhang (2006), Zhang, Mykland, and Aït-Sahalia (2005), and Zhang (2006),

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

where K is the ADD given by equation (41).

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

$$\sup_j \mathrm{tr}(A_j^2) \to 0 \text{ as } n \to \infty. \tag{B.30}$$

Since also,

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

it follows that

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

Now 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 Aït-Sahalia (2005). The result is thus proved. ∎

C Proofs concerning the Leverage Effect (Section 4.3)

C.1 Proof of Proposition 3

We here show how to arrive at the final result in Proposition 3. This serves as a fairly extensive illustration of how to apply the theory development in the earlier sections.

By rearranging terms, write

$$\begin{aligned}\widetilde{\langle\sigma^2, X\rangle}_T = &\sum_i (\sigma^2_{\tau_{n,i+1}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \\ &+ \sum_i (\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}) \\ &- \sum_i (\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) + O_p(n^{-1}),\end{aligned} \tag{C.33}$$

where the $O_p(n^{-1})$ term comes from edge effects. Note that by conditional Gaussianity, both the two last sums in (C.33) are Q_n-martingales with respect to the sigma-fields $\mathcal{Y}_{n,i}$. They are also orthogonal, in the sense that

$$\mathrm{Cov}^Q_n((\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}), (\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \mid \mathcal{Y}_{n,i}) = 0 \tag{C.34}$$

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

$$\begin{aligned}&\mathrm{Var}^Q_n((\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}) \mid \mathcal{Y}_{n,i}) \\ &= \sigma^4_{\tau_{n,i}}(M-1)^{-2}(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})^2\mathrm{Var}(\chi^2_{M-1}) \\ &= 2\sigma^4_{\tau_{n,i}}(M-1)^{-1}(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})^2\end{aligned} \tag{C.35}$$

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

$$\frac{2}{M-1}\int_0^T \sigma^6_t dt. \tag{C.36}$$

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 $\sum_i (\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})$ converges stably under P to a normal distribution with mean zero and variance (C.36).

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

$$\begin{aligned}&\mathrm{Var}^Q_n((\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \mid \mathcal{Y}_{n,i}) \\ &= \sigma^6_{\tau_{n,i}}(M\Delta t)\mathrm{Var}((\frac{\chi^2_{M-1}}{M-1} - 1)N(0,1)) \\ &= \frac{2}{M-1}\sigma^6_{\tau_{n,i}}(M\Delta t).\end{aligned} \tag{C.37}$$

Hence the asymptotic quadratic variation is

$$\frac{2}{M-1}\int_0^T \sigma_t^6 dt. \tag{C.38}$$

The sum is asymptotically uncorrelated with W^Q, since

$$\begin{aligned}
&\mathrm{Cov}_n^Q((\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), W_{\tau_{n,i+1}} - W_{\tau_{n,i}} \mid \mathcal{Y}_{n,i}) \\
&= \sigma^3_{\tau_{n,i}}(M\Delta t)\mathrm{Cov}((\frac{\chi^2_{M-1}}{M-1} - 1)N(0,1), N(0,1)) \\
&= 0.
\end{aligned} \tag{C.39}$$

Overall, under Q_n, we have the stable convergence

$$\sum_i (\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \xrightarrow{\mathcal{L}} N(0,1)\left(\frac{2}{M-1}\int_0^T \sigma_t^6 dt\right)^{1/2} \tag{C.40}$$

There is, however, covariation between this sum and $M_n^{(0,Q)}$. It is shown below in Remark 13 (see equation (C.47)) 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). Similarly, there is covariation with $M_n^{(1,Q)}$, and one can show that $A_{13} = \frac{M-3}{2M}\langle\sigma^2, X\rangle_T$. Thus, by Theorem 4, under P^*, we have (stably)

$$\sum_i (\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \xrightarrow{\mathcal{L}} \frac{1}{2}\langle\sigma^2, X\rangle_T + N(0,1)\left(\frac{2}{M-1}\int_0^T \sigma_t^6 dt\right)^{1/2}. \tag{C.41}$$

Because of the orthogonality (C.34), and since $\sum_i(\sigma^2_{\tau_{n,i+1}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) - \langle\sigma^2, X\rangle_T = O_p(n^{-1/2})$ by Proposition 1 of Mykland and Zhang (2006), it follows that $\widetilde{\langle\sigma^2, X\rangle}_T - \frac{1}{2}\langle\sigma^2, X\rangle$ converges stably (under P^*) to a normal distribution with mean as in equation (C.41), and variance contributed by the second and third terms on the right hand side of (C.33). We have thus shown Proposition 3.

Remark 13. (Sample of calculation). To see how the reasoning works in the case of covariations, consider the case of covariation between $\sum_i(\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})$ and $M_n^{(0,Q)}$. We proceed as follows.

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

$$G_{r,i} = \sum_{t_{n,j}\in(\tau_{n,i},\tau_{n,i+1}]} h_r(\Delta W^Q_{t_{n,j}}/\Delta t^{1/2}), \tag{C.42}$$

note that

$$\begin{aligned}
X_{\tau_{n,i+1}} - X_{\tau_{n,i}} &= \sigma_{\tau_{n,i}}\Delta t^{1/2}G_{1,i} \text{ and} \\
\hat\sigma^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}} &= \frac{\sigma^2_{\tau_{n,i}}}{M-1}\left(G_{2,i} - \frac{1}{M}G^2_{1,i} + 1\right)
\end{aligned} \tag{C.43}$$

At the same time,

$$M_n^{(0,Q)} = \frac{1}{12}(\Delta t)^{1/2} \sum_i k_{\tau_{n,i}} G_{3,i} + o_p(1) \tag{C.44}$$

The covariance for each i-increment becomes

$$\begin{aligned}
&\mathrm{Cov}_n^Q((\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \frac{1}{12}(\Delta t)^{1/2} k_{\tau_{n,i}} G_{3,i} \mid \mathcal{Y}_{n,i}) \\
&\quad = \frac{1}{12}\Delta t \frac{k_{\tau_{n,i}}\sigma^3_{\tau_{n,i}}}{M-1} \mathrm{Cov}_n^Q((G_{2,i} - \frac{1}{M}G^2_{1,i} + 1)G_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\
&\quad = \frac{1}{2}(M\Delta t)\frac{k_{\tau_{n,i}}\sigma^3_{\tau_{n,i}}}{M}
\end{aligned} \tag{C.45}$$

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

$$\begin{aligned}
&\mathrm{Cov}_n^Q((G_{2,i} - \frac{1}{M}G^2_{1,i} + 1)G_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\
&\quad = \mathrm{cum}^Q_{3,n}(G_{1,i}, G_{2,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) - \frac{1}{M}\mathrm{cum}^Q_{4,n}(G_{1,i}, G_{1,i}, G_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\
&\quad = M\mathrm{cum}_3(h_1(N(0,1)), h_2(N(0,1), h_3(N(0,1))) \\
&\quad - \mathrm{cum}_4(h_1(N(0,1)), h_1(N(0,1), h_1(N(0,1)), h_3(N(0,1))) \\
&\quad = 6(M-1).
\end{aligned} \tag{C.46}$$

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

$$\begin{aligned}
A_{12} &= \frac{1}{2M}\int_0^T k_t \sigma_t^3 dt \\
&= \frac{3}{2M}\langle \sigma^2, X\rangle_T
\end{aligned} \tag{C.47}$$

as in (32). □

C.2 Proof for Example 5

In analogy with (73), define

$$\widetilde{\langle \sigma^2, X\rangle}_T^{\text{with mean}} = \sum_i (\tilde{\sigma}^2_{\tau_{n,i+1}} - \tilde{\sigma}^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \tag{C.48}$$

We have the representation

$$\tilde{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}} = \frac{\sigma^2_{\tau_{n,i}}}{M} G_{2,i}. \tag{C.49}$$

We now consider the terms analogous to those in (C.33). The analysis of $\sum_i(\tilde{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})$ is unaffected by this change, except that (C.36) is replaced by $\frac{2}{M}\int_0^T \sigma_t^6 dt$. However, this is not true for the term $\sum_i(\tilde{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})$, which we analyze in the following.

Observe that $\tilde{\sigma}^2_{\tau_{n,i}} = \frac{M-1}{M}\hat{\sigma}^2_{\tau_{n,i}} + (\overline{\Delta X}_{\tau_{n,i}})^2/\Delta t$. Hence,

$$\sum_i (\tilde{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) = \frac{M-1}{M}\sum_i (\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})$$
$$+ \frac{1}{M}\frac{K_n}{T}\sum_i (X_{\tau_{n,i+1}} - X_{\tau_{n,i}})^3 - \frac{1}{M}\int_0^T \sigma_t^2 dX_t + o_p(1), \tag{C.50}$$

where $K_n = n/M$ and the $o_p(1)$ term comes (only) from the approximation of $-\frac{1}{M}\sum_i \sigma^2_{\tau_{n,i}}(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})$ by $-\frac{1}{M}\int_0^T \sigma_t^2 dX_t$. It is easy to see that the first two terms on the right hand side of (C.50) have zero Q_n-covariation, and hence, asymptotically, zero P^*-covariation (Remark 4 in Section 2.4). Since we are thus in a position to easily aggregate the normal parts of the limiting distributions, we obtain the limit of the first term from (C.41), and the limit of the second term from Example 3 in Section 2.5. Hence, stably under P^*, with U_1 and U_2 as independent standard normal,

$$\sum_i (\tilde{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \xrightarrow{\mathcal{L}} \frac{M-1}{M}\left(\frac{1}{2}\langle\sigma^2, X\rangle_T + U_1\left(\frac{2}{M-1}\int_0^T \sigma_t^6 dt\right)^{1/2}\right)$$
$$+ \frac{1}{M}\left(3\int_0^T \sigma_t^3 dW_t^* + \frac{3}{2}\langle\sigma^2, X\rangle_T + U_2\left(6\int_0^T \sigma_t^6 dt\right)^{1/2}\right) - \frac{1}{M}\int_0^T \sigma_t^2 dX_t$$
$$= \frac{2}{M}\int_0^T \sigma_t^3 dW_t^* + \frac{M+2}{2M}\langle\sigma^2, X\rangle_T + N(0,1)\left(\frac{2M+4}{M^2}\int_0^T \sigma_t^6 dt\right)^{1/2}. \tag{C.51}$$

Since the terms in (C.50) have zero Q_n-covariation with $\sum_i (\tilde{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})$, the result in Example 5 follows. ∎