The University of Chicago
Department of Statistics
TECHNICAL REPORT SERIES

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

Per A. Mykland
The University of Chicago

Lan Zhang
University of Illinois at Chicago

TECHNICAL REPORT NO. 576

Departments of Statistics
The University of Chicago
Chicago, Illinois 60637

September 9, 2007

* We are grateful to Oliver Linton for his comments and suggestions. Financial support from the National Science Foundation under grants DMS 06-04758 and SES 06-31605 is also gratefully acknowledged.

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

Per A. Mykland
The University of Chicago

Lan Zhang
University of Illinois at Chicago

This version: September 9, 2007.

Abstract

The econometric literature of high frequency data usually relies on moment estimators which are derived from assuming local constancy of volatility and related quantities. We here show that this first order approximation is not always valid if used naïvely. We find that such approximations require an ex post adjustment involving asymptotic likelihood ratios. These are given. Several examples (powers of volatility, leverage effect, ANOVA) are provided. The first order approximations in this study can be over the period of one observation, or over blocks of successive observations. The theory relies heavily on the interplay between stable convergence and measure change, and on asymptotic expansions for martingales.

Practically, the procedure permits (1) the definition of estimators of hard to reach quantities, such as the leverage effect, (2) the improvement in efficiency in classical estimators, and (3) easy analysis. More conceptually, we show that the approximation induces a measure change similar to that occurring in options pricing theory. In particular, localization over one observation induces a measure change related to the leverage effect, while localization over a block of observations creates an effect that connects to the volatility of volatility. Another conceptual gain is the relationship to Hermite polynomials. The three measure changes mentioned relate, respectively, to the first, third, and second such polynomial.

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

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

*We are grateful to Oliver Linton for his 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

1.1 Setting and Overview

Recent years have seen an explosion of literature in the area of estimating volatility on the basis of high frequency data. The concepts go back to stochastic calculus, see, for example, Karatzas and Shreve (1991) (Section 1.5), Jacod and Shiryaev (2003) (Theorem I.4.47 on page 52), and Protter (2004) (Theorem II-22 on page 66). An early econometric discussion of this relationship can be found in Andersen, Bollerslev, Diebold, and Labys (2000). Recent work both from the probabilistic and econometric side show that the approximation error has a mixed normal distribution. References include Jacod (1994, 2006), Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002), Zhang (2001), Mykland and Zhang (2006), and Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006). Further econometric literature includes, in particular, Gallant, Hsu, and Tauchen (1999), Chernov and Ghysels (2000), Engle (2000), Andersen, Bollerslev, Diebold, and Labys (2001, 2003), Dacorogna, Gençay, Müller, Olsen, and Pictet (2001), Goncalves and Meddahi (2005) and Kalnina and Linton (2007). Estimating instantaneous volatility at each point in time goes back to Foster and Nelson (1996), see also Mykland and Zhang (2001), but this has not caught on quite as much in the econometric application. There is also an emerging literature on micro structure, but we are not planning to address this question in any depth here.

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 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 4.2). Also, the separation of jumps and continuous evolution is a highly discussed topic, cf. Section 8.1. We shall see in this paper that one can also estimate such things as the leverage effect.

The literature of high frequency data relies on moment estimators derived from assuming local constancy of volatility and related quantities. To be specific, if $t_i,\ 0 < t_1 < ... < t_n = T$, 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 effectively 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 several questions. First of all, *(i)* can one always do this? Or, does the approximation in formula (1) only work for a handful of cases such as volatility? Is there any case where the first order approximation fails to give the (asymptotically) correct answer in high frequency inference? Second, *(ii)* there is a mysterious absence of leverage effect from existing results for high frequency data. The *leverage effect* here refers to the dependence between a process and its own volatility (often in the form of covariance or correlation), and is an important ingredient in many models, such as the one by Heston (1993). From the existing literature, however, one might be tempted to think that one can simply assume the absence of leverage effect, and that asymptotic results will remain valid even when this assumption fails. Can this be true? Finally, *(iii)* 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 the answer to this were yes, a whole arsenal of additional statistical techniques would become available.

The purpose of this paper is to provide answers to these questions. In answer to *(i)*, we show in Section 2 that first order approximations (like (1)-(2)) are not always valid for purposes of inference in high frequency data. The error, however, is controllable by a general procedure which we shall provide.

Specifically, let P be the true probability distribution governing the asset price process. The first order approximation (1) is (in a suitable sense) exact under a distribution P_n^*. We show that the likelihood ratio dP_n^*/dP has an asymptotic (nondegenerate) limit, which we can call dP_∞^*/dP. One can therefore proceed as follows:
(1) Find the asymptotic distribution of estimators under P_n^* (*i.e.*, pretending that the first order approximation (1) is valid); and then
(2) Correct for the error in the final asymptotic distribution by multiplying by dP/dP_∞^*.

In other words, we provide a generic device for correcting asymptotic distributions ex post when first order approximations are used in the analysis.

In the process, we also give the answer to question *(ii)* above. Specifically, dP_∞^*/dP is closely related to the size of the leverage effect.

Our result is more of an opportunity than a crisis. The usual asymptotic results concerning power variations (as estimators of $\int_0^T |\sigma|_t^p dt$) are unaffected by the adjustment. The new result does permit, however, the estimation of more elusive quantites such as the leverage effect (see Section 4.3; otherwise the development described above is in Section 2).

The paper then turns to question *(iii)*. We show in Section 3 that by holding volatility characteristics constant over successive blocks of M ($M > 1$) observations, from, say, t_{i-M} to t_i, one implicitly behaves as if the probability distribution were Q_n, where dQ_n/dP has an asymptotic

(nondegenerate) limit, which we can call dQ_∞/dP. The conclusion is therefore much the same as under one period discretization (1). In this case, the size of the additional adjustment turns out to related to the volatility of volatility

The implication is that parametric techniques can be used locally over such blocks. Estimators of integrated quantities may then be obtained by aggregating local estimators. The advantage of such an approach is to gain efficiency. In the case of quantities like $\int_0^T |\sigma|_t^p dt$, there can be substantial gain (see Section 4.1). In the case of the leverage effect, such blocking is a *sine qua non*, as will be clear from Sections 2.5 and 4.3.

Local parametric inference appears to have been introduced by Tibshirani and Hastie (1987), and there is an extensive literature on the subject, such as Fan (1993), Fan and Gijbels (1996), Hjort and Jones (1996), Loader (1996), Fan, Gijbels, and King (1997), Kauermann and Opsomer (2003), Chen and Spokoiny (2007) and Cizek, Härdle, and Spokoiny (2007). A review is given in Fan, Farmen, and Gijbels (1998), and this paper should be consulted for further references. A related but different form of diffusion block approximation was also discussed in Mykland (2006), in the context of (conditionally) nonrandom volatility.

This paper establishes 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.

The plan for the paper is that Section 1.2 provides a technical preview about various measure changes in our inference. 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 ($M > 1$). The major applications are given in Section 4, with a summary of the methodology in Section 5. The later sections are concerned with irregular and asynchronous data (Section 6), and also with overlapping blocks (moving windows, Section 7). Section 8 discusses the extension to jumps and microstructure. Section 9 discusses the use of Milstein type schemes for the one period approximation. We conclude in Section 10. Proofs are in the Appendix.

A reader's guide: We emphasize that the two approximations (to block size $M = 1$, and then from $M = 1$ to $M > 1$) are quite different in their methodologies. If you are only interested in the one period approximation, the material to read is Sections 2 and 9, 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 Section 3-5, and Appendix B.

1.2 Technical Preview

The theory in this paper deals with two forms of discretization: to block size $M = 1$, and then to block size $M > 1$. Each of these has to be adjusted for by using an asymptotic measure change. Accordingly, the likelihood ratios are called dP^*_∞/dP and dQ_∞/dP. There is similarity here to the measure change dP^*/dP used in option pricing theory, where P^* is an equivalent martingale measure (a probability distribution under which the drift of an underlying process has been removed; for our purposes, discounting is not an issue); for more discussion and references, see Remark 2 in Section 2.2. In fact, for the reasons given in that section, we shall 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

$$\log \frac{dQ_\infty}{dP} \;=\; \log \frac{dQ_\infty}{dP^*_\infty} \;+\; \log \frac{dP^*_\infty}{dP^*} \;+\; \log \frac{dP^*}{dP}. \tag{3}$$

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

The three approximations all lead to adjustments that are absolutely continuous. This fact means that for estimators, consistency and rate of convergence are unaffected by the the approximation. It will also turn out that asymptotic variances are similarly unaffected (Remark 8 in Section 2.4). Asymptotic distributions can be changed through their means only (Sections 2.4, 3.4). We emphasize that this is not the same as introducing inconsistency.

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

TABLE 1. Measure changes (likelihood ratios) tied to three procedures modifying properties of the observed process. P is the true probability distribution, P^* is the equivalent martingale measure (as in option pricing theory). P^*_n is the probability for which (1) is exact, and Q_n is the probability for which one can use $\int_{t_{i-M}}^{t_i} f_s dW_s \approx f_{t_{i-M}}(W_{t_i} - W_{t_{i-M}})$. The two measure changes dP^*_n/dP^* and dQ_n/dP^*_n have an asymptotic limit. This connects to the statistical concept of contiguity, cf. Remark 5.

As an illustration for Table 1, consider a scalar process on the form $dX_t = \mu_t dt + \sigma_t dW_t$, and with equidistant observations of X. The terms in (3) can be written

$$\log \frac{dQ_\infty}{dP^*_\infty} = Z_1\eta_1 - \frac{1}{2}\eta_1^2 \quad \text{and} \quad \log \frac{dP^*_\infty}{dP^*} = Z_2\eta_2 - \frac{1}{2}\eta_2^2, \tag{4}$$

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

$$\eta_1^2 = \underbrace{\frac{3}{8}\int_0^T \sigma_t^{-6}(\langle \sigma^2, X\rangle'_t)^2 dt}_{\text{related to leverage effect}} \quad \text{and } \eta_2^2 = \underbrace{\frac{M-1}{4}\int_0^T \sigma_t^{-4}\langle \sigma^2, \sigma^2\rangle'_t dt}_{\text{related to volatility of volatility}} \tag{5}$$

The angle brackets $\langle \cdot, \cdot \rangle$ refers to covariation (sum of instantaneous covariance between processes, see Chapter I.4e (p. 51-58), of Jacod and Shiryaev (2003), see also (9) in the upcoming Section 2.1). As is well known from stochastic calculus, $\log \frac{dP^*}{dP} = -\int_0^T \frac{\mu_t}{\sigma_t} dW_t - \frac{1}{2}\int_0^T \frac{\mu_t^2}{\sigma_t^2} dt$.

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

$$\eta_1^2 = \frac{3}{8}(\rho\gamma)^2 \int_0^T \sigma_t^{-2} dt \quad \text{and} \quad \eta_2^2 = \frac{1}{4}\gamma^2(M-1)\int_0^T \sigma_t^{-2} dt. \tag{6}$$

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. Along with Section 9 and Appendix A, this section can be read independently of the rest of the paper.

2.1 Data Generating Mechanism

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

$$dX_t = \mu_t dt + \sigma_t dW_t, X_0 = x_0, \tag{7}$$

where μ_t and σ_t are adapted 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

$$\zeta_t = \sigma_t^T \sigma_t, \tag{8}$$

then the integrated volatility process is given as

$$\langle X, X\rangle_t = \int_0^t \zeta_u du. \tag{9}$$

Note that 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)), one can take σ_t to be the symmetric square root of ζ_t. One understands the differential $\sigma_t dW_t$ to be a p-dimensional vector with r_1'th component $\sum_{r_2=1}^{p} \sigma_t^{(r_1,r_2)} dW_t^{(r_2)}$.

DEFINITION 1. *A process on the form (7) will be called an Itô process. We always assume (as part of the definition) that both μ_t and ζ_t are locally bounded.*

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 Section 6.3 how the results encompass the asynchronous case.

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

$$\max_{1 \leq i \leq n} |t_{n,j} - t_{n,j-1}| = o(1). \tag{10}$$

We also suppose that the observation times $t_{n,j}$ are nonrandom, but irregular.

2.2 A simplifying strategy for inference

We bring out the heavy machinery right away. The advantage is that this substantially cleans up the further presentation. The general principles laid out here will also be an ingredient in the presentation of the later discretization results.

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. Specifically, under P^*

$$dX_t = \sigma_t dW_t^*, X_0 = x_0, \tag{11}$$

where W_t^* is a P^*-Brownian motion.

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,

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(b, a^2), \tag{12}$$

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 $n \to \infty$.*

In the context of (12), $Z_n = n^{1/2}(\hat{\theta}_n - \theta)$ 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 (12) 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 $\frac{dP}{dP^*} I_A g(Z_n)$. ∎

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

REMARK 2. As noted, our device is comparable to the use of equivalent martingale measures in options pricing theory (Ross (1976), Harrison and Kreps (1979), Harrison and Pliska (1981), see also Duffie (1996)) in that it affords a convenient probability distribution with which to make computations. In our econometric case, one can always take the drift to be zero, while in the options pricing case, this can only be done for discounted securities prices. In both cases, however, the computational purpose is to get rid of a nuisance "dt term".

The idea of combining stable convergence with measure change appears to go back to Rootzén (1980). We have earlier used the idea in Zhang, Mykland, and Aït-Sahalia (2005), cf the second paragraph of the proof of Theorem 2 in that paper (p. 1410). The current description is meant to be a more complete description of the same procedure. □

REMARK 3. Note that following Girsanov's Theorem

$$\frac{dP^*}{dP} = \exp\left\{-\int_0^T \sigma_t^{-1}\mu_t dW_t - \frac{1}{2}\int_0^T \mu_t^T(\sigma_t^T\sigma_t)^{-1}\mu_t dt\right\}, \tag{13}$$

with

$$dW_t^* = dW_t + \sigma_t^{-1}\mu_t dt. \tag{14}$$

□

In order to carry out the measure change, we make the following assumption.

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 (11) by the following approximation:

$$P_n^*: \quad \Delta X_{t_{n,j+1}} = \sigma_{t_{n,j}}\Delta\breve{W}_{t_{n,j+1}} \text{ for } j = 0, ..., n-1; X_0 = x_0, \tag{15}$$

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 (15) 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 distribution $N(0, \Delta t_{n,j+1})$. Note that we here follow the convention from options pricing theory, whereby, when the measure changes, the process (X_t) doesn't change, while the driving Brownian motion changes.

To describe the nature of our approximations, focus on the random variables:

$$\begin{aligned} U_{t_{n,j}}^{(1)} &= X_{t_{n,j}} \\ U_{t_{n,j}}^{(2)} &= (\sigma_{t_{n,j}}, \langle\sigma, W\rangle'_{t_{n,j}}, \langle\sigma, \sigma\rangle'_{t_{n,j}}) \\ U_{t_{n,j}} &= (U_{t_{n,j}}^{(1)}, U_{t_{n,j}}^{(2)}). \end{aligned} \tag{16}$$

for $j = 0, ..., n$.

DEFINITION 3. *(First order approximation). Define the probability P_n^* recursively by:*
(i) U_0 has same distribution under P_n^* as under P^*;
(ii) The conditional P_n^*-distribution of $U^{(1)}_{t_{n,j+1}}$ given $U_0, ..., U_{t_{n,j}}$ is given by (15); and
(iii) 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^*.

Also, denote by $\mathcal{X}_{n,j}$ the σ-field generated by $U_{t_{n,\iota}}$, $\iota = 0, ..., j$.

Here $\langle \sigma, W \rangle'_t$ 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 $\langle \sigma^{(r_1,r_2)}, \sigma^{(r_3,r4)} \rangle'_t$.

REMARK 4. The quantity $U^{(2)}_{t_{n,j+1}}$ thus contains information about the behavior of the σ_t process, and also (via the tensor $\langle \sigma, W \rangle'_t$) leverage effect. □

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

$$f(U_{t_{n,1}}..., U_{t_{n,j}}, ..., U_{t_{n,n}}|U_0) = \underbrace{\prod_{j=1}^{n} f(U^{(1)}_{t_{n,j}}|U_0, ..., U_{t_{n,j-1}})}_{\text{altered from } P^* \text{ to } P_n^*} \underbrace{\prod_{j=1}^{n} f(U^{(2)}_{t_{n,j}}|U_0, ..., U_{t_{n,j-1}}, U^{(1)}_{t_{n,j}})}_{\text{unchanged from } P^* \text{ to } P_n^*} \tag{17}$$

To state the main theorem, define

$$d\check{\zeta}_t = \sigma_t^{-1} d\zeta_t \sigma_t^{-1} \tag{18}$$

and

$$k_t^{(r_1,r_2,r_3)} = \langle \check{\zeta}^{(r_1,r_2)}, W^{(r_3)} \rangle'_t[3] \tag{19}$$

where the "[3]" means that the right hand side of (19) 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 $k^{(r_1,r_2,r_3)}_{t_{n,j}}$ is measurable with respect to the σ-field $\mathcal{X}_{n,j}$ generated by $U_{t_{n,\iota}}$, $\iota = 0, ..., j$. Finally, set

$$\Gamma_0 = \frac{1}{24} \int_0^T \sum_{r_1,r_2,r_3=1}^{p} (k_t^{(r_1,r_2,r_3)})^2 dt. \tag{20}$$

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 $(dP^*/dP_n^*)(U_{t_0}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}})$ be the likelihood ratio (Radon-Nikodym derivative) on $\mathcal{X}_{n,n}$. Then,*

$$\frac{dP^*}{dP_n^*}(U_{t_0}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}}) \xrightarrow{\mathcal{L}} \exp\{\Gamma_0^{1/2} N(0,1) - \frac{1}{2}\Gamma_0\} \tag{21}$$

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

Based on Theorem 1, one can (for a fixed time period) carry out inference under the model (15), and asymptotic results will transfer back to the continuous model (11) 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 5 below). We state the following result, in analogy with Proposition 1.

COROLLARY 1. *Suppose that Z_n (say, $n^{1/2}(\hat{\theta}_n - \theta)$) converges stably in law 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 5. 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 6. 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 an option Y and the underlying security X. □

REMARK 7. The approximation (15) goes along with the high frequency setting, where T is fixed and the data is sampled more finely with n. It is in this setting that (15) yields estimators that are consistent and have the right order of convergence. For a discussion (in the parametric framework) of such approximation in the low frequency setting where $T \to \infty$, we refer to the discussion in Sections 4.2-4.3 (p. 2201-2206) of Aït-Sahalia and Mykland (2003). For intermediate types of data frequency, there may be some uncertainty about what form of asymptotic regime is most suitable, and a combined regime may be appropriate. See, for example, Kessler (1997) and Sørensen (2007). □

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

Following (15), write

$$\Delta \breve{W}_{t_{n,j+1}} = \sigma_{t_{n,j}}^{-1} \Delta X_{t_{n,j+1}}. \tag{22}$$

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 $h_{r_1r_2r_3}(x) = x^{r_1}x^{r_2}x^{r_3} - x^{r_1}\delta^{r_2,r_3}[3]$, 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. Set

$$M_n^{(0)} = \frac{1}{12}\sum_{j=0}^{n-1}(\Delta t_{n,j+1})^{1/2}\sum_{r_1,r_2,r_3=1}^{p} k_{t_{n,j}}^{(r_1,r_2,r_3)} h_{r_1r_2r_3}(\Delta \breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}) \tag{23}$$

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 $A = a^2$. 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 8. The fact that the asymptotic variance remains unchanged in Theorem 2 is 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. □

2.5 Some initial examples

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

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

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

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

whereby

$$\begin{aligned} E_n^*(|\Delta X_{t_{n,j+1}}|^p \mid \mathcal{X}_{n,j}) &= |\sigma_{t_{n,j}}|^p E|N(0,1)|^p \Delta t_{n,j+1}^{p/2} \\ \mathrm{Var}_n^*(|\Delta X_{t_{n,j+1}}|^p \mid \mathcal{X}_{n,j}) &= |\sigma_{t_{n,j}}|^{2p} \mathrm{Var}(|N(0,1)|^p) \Delta t_{n,j+1}^p \text{ and} \\ \mathrm{Cov}_n^*(|\Delta X_{t_{n,j+1}}|^p, \Delta \breve{W}_{t_{n,j+1}} \mid \mathcal{X}_{n,j}) &= 0. \end{aligned} \tag{24}$$

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

$$\hat{\theta}_n = \frac{1}{E|N(0,1)|^p} \sum_{j=0}^{n-1} \Delta t_{n,j+1}^{1-\frac{p}{2}} |\Delta X_{t_{n,j+1}}|^p. \tag{25}$$

From (24), it follows that $\hat{\theta}_n - \sum_{j=0}^{n-1} |\sigma_{t_{n,j}}|^p \Delta t_{n,j+1}$ is the end point of a martingale orthogonal to W, and with discrete time quadratic variation $\frac{\mathrm{Var}(|N(0,1)|^p)}{(E|N(0,1)|^p)^2} \sum_{j=0}^{n-1} |\sigma_{t_{n,j}}|^{2p} \Delta t_{n,j+1}^2$. By the usual martingale central limit considerations (Jacod and Shiryaev (2003)), and since $\theta - \sum_{j=0}^{n-1} |\sigma_{t_{n,j}}|^p \Delta t_{n,j+1} = O_p(n^{-1})$, it follows that

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} Z \times \left(\frac{\mathrm{Var}(|N(0,1)|^p)}{(E|N(0,1)|^p)^2} T \int_0^T \sigma_t^{2p} dH(t) \right)^{1/2} \tag{26}$$

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

$$H(t) = \lim_{n\to\infty} \frac{n}{T} \sum_{t_{n,j+1} \le t} (t_{n,j+1} - t_{n,j})^2, \tag{27}$$

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 $\sum_{j=1}^n |\Delta X_{t_{n,j}}|^p$, and also $H(t) = t$. Finally, recall that

$$E|N(0,1)|^r = 2^{\frac{r}{2}} \pi^{-\frac{1}{2}} \Gamma(\frac{r+1}{2}) \tag{28}$$

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

To get from the convergence under P_n^* to convergence under P^*, we note that $|N(0,1)|^p$ is uncorrelated with $N(0,1)$ and $N(0,1)^3$. We therefore obtain from Theorems 1 and 2 that the stable convergence in (26) 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 (2004), Woerner (2004) and Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006)). The derivations are much the same. In particular, no adjustment is needed from P_n^* to P^*. □

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

$$Z_n = \frac{n}{T}\sum_{j=0}^{n-1}(\Delta X_{t_{n,j+1}})^3. \tag{29}$$

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

Our study of the quantity Z_n started out as the consideration of a strange example, but has gradually grown into something more important. We shall see in Section 4.3 that similar quantities can be parlayed into estimators of the leverage effect. For now, we just show what the simplest calculation will bring. An important issue, which sets (29) apart from most other cases is that there is a need for an adjustment from P_n^* to P^*, and also from P^* to P. This is what will, in turn, yield interesting results in Section 4.3.

By the same reasoning as in Example 1,

$$\begin{aligned} E_n^*(\Delta X^3_{t_{n,j+1}} \mid \mathcal{X}_{n,j}) &= 0 \\ \mathrm{Var}_n^*(\Delta X^3_{t_{n,j+1}} \mid \mathcal{X}_{n,j}) &= \sigma^6_{t_{n,j}}\mathrm{Var}(N(0,1)^3)\Delta t^3 = 15\sigma^6_{t_{n,j}}\Delta t^3 \text{ and} \\ \mathrm{Cov}_n^*(\Delta X^3_{t_{n,j+1}}, \Delta \breve{W}_{t_{n,j+1}} \mid \mathcal{X}_{n,j}) &= \sigma^3_{t_{n,j}}\mathrm{Cov}(N(0,1)^3, N(0,1))\Delta t^2 = 3\sigma^3_{t_{n,j}}\Delta t^2. \end{aligned} \tag{30}$$

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

$$\begin{aligned} b &= 3\int_0^T \sigma_t^3 dW_t^* \text{ and} \\ a^2 &= 6\int_0^T \sigma_t^6 dt. \end{aligned} \tag{31}$$

Note that 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,

$$\begin{aligned} \mathrm{Cov}_n^*(\Delta X^3_{t_{n,j+1}}, h_3(\Delta \breve{W}_{t_{n,j+1}}/\Delta t^{1/2}) \mid \mathcal{X}_{n,j})\Delta t^{1/2} &= \sigma^3_{t_{n,j}}\mathrm{Cov}(N(0,1)^3, h_3(N(0,1)))\Delta t^2 \\ &= 6\sigma^3_{t_{n,j}}\Delta t^2 \end{aligned} \tag{32}$$

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

$$\begin{aligned} A_{12} &= \frac{1}{2}\int_0^T k_t\sigma_t^3 dt \\ &= \frac{3}{2}\langle \sigma^2, X\rangle_T, \end{aligned} \tag{33}$$

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 (31), while

$$b' = 3\int_0^T \sigma_t^3 dW_t^* + \frac{3}{2}\langle \sigma^2, X\rangle_T. \tag{34}$$

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 (14), we have $dW_t^* = dW_t + \sigma_t^{-1}\mu_t dt$, it follows that

$$b' = 3\int_0^T \sigma_t^3(dW_t + \sigma_t^{-1}\mu_t dt) + \frac{3}{2}\langle\sigma^2, X\rangle_T. \tag{35}$$

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

It seems to be a general phenomenon that if there is random bias under P^*, then μ will occur in the expression for bias under P. This occurs again in Example 4 in Section 4.3.2. □

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

$$\mathcal{H}_n = \{0 = \tau_{n,0} < \tau_{n,1} < ... < \tau_{n,K_n} = T\} \subseteq \mathcal{G}_n \tag{36}$$

If we set

$$\begin{aligned} M_{n,i} &= \#\{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]\} \\ &= \text{ number of intervals } (t_{n,j-1}, t_{n,j}] \text{ in } (\tau_{n,i-1}, \tau_{n,i}] \end{aligned} \tag{37}$$

we shall suppose that

$$\max_i M_{n,i} = O(1) \text{ as } n \to \infty, \tag{38}$$

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

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

$$\begin{aligned} X_0 &= x_0 \\ \text{for each } i &= 1, K_n : \\ \Delta X_{t_{j+1}} &= \sigma_{\tau_{n,i-1}} \Delta W^Q_{t_{j+1}} \text{ for } t_{n,j+1} \in (\tau_{n,i-1}, \tau_{n,i}]. \end{aligned} \tag{39}$$

To implement this, we use a variation over Definition 3. Formally, we define the approximation as follows.

DEFINITION 4. *(Block approximation). Define the probability Q_n recursively by:*
(i) U_0 has same distribution under Q_n as under P^*;
(ii) The conditional Q_n-distribution of $U^{(1)}_{t_{n,j+1}}$ given $U_0, ..., U_{t_{n,j}}$ is given by (39); and
(iii) 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^*.

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

$$\ell(\Delta x; \zeta) = -\frac{1}{2}\log\det(\zeta) - \frac{1}{2}\Delta x^T \zeta^{-1} \Delta x. \tag{40}$$

We then obtain directly that

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

$$\begin{aligned}&\log\frac{dQ_n}{dP^*_n}(U_{t_0}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}}) \\ &\quad = \sum_i \sum_{\tau_{i-1} \le t_j < \tau_i} \left\{ \ell(\Delta X_{t_{j+1}}; \zeta_{\tau_{n,i-1}} \Delta t_{j+1}) - \ell(\Delta X_{t_{j+1}}; \zeta_{t_{n,j}} \Delta t_{j+1}) \right\}\end{aligned} \tag{41}$$

DEFINITION 5. *To measure the extent to which we hold the volatility constant, we define the following "Asymptotic Decoupling Delay" (ADD) by*

$$K(t) = \lim_{n\to\infty} \sum_i \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}] \cap [0,t]} (t_{n,j} - \tau_{n,i-1}), \tag{42}$$

provided the limit exists.

From (10) and (38), 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.

Noted that in the case of equidistant observations and equally sized blocks of M observations, the ADD takes the form

$$K(t) = \frac{1}{2}(M-1)t. \tag{43}$$

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 (42)) exists. Set*

$$Z^{(1)}_n = \frac{1}{2}\sum_i \sum_{t_{n,j} \in [\tau_{n,i-1}, \tau_{n,i})} \left(\Delta X^T_{t_{n,j+1}} (\zeta^{-1}_{t_{n,j}} - \zeta^{-1}_{\tau_{n,i-1}}) \Delta X_{t_{n,j+1}} \Delta t^{-1}_{n,j+1} \right) \tag{44}$$

and let $M_n^{(1)}$ *be the end point of the* P_n^**-martingale part of* $Z_n^{(1)}$ *(see (B.23) and (B.25) in Appendix B for the explicit formula). Define*

$$\Gamma_1 = \frac{1}{2}\int_0^T \mathrm{tr}(\zeta_t^{-2}\langle\zeta,\zeta\rangle_t')dK(t), \tag{45}$$

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

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

Furthermore, if $M_n^{(0)}$ *is as defined in (23), then the pair* $(M_n^{(0)}, M_n^{(1)})$ *converges stably to* $(\Gamma_0^{1/2}V_0, \Gamma_1^{1/2}V_1)$*, where* V_0 *and* V_1 *are iid N(0,1), and independent* $\mathcal{F}_T$*.*

The theorem says that P_n^* and the approximation Q_n are contiguous, cf. Remark 5 in Section 2.3. By the earlier Theorem 1, it follows that Q_n and P^* (and P) are contiguous. In particular, as before, if an estimator is consistent under Q_n, it is also consistent under P^* and P. Rates of convergence (typically $n^{1/2}$) are also preserved, but the asymptotic distribution may change.

Remark 9. (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

$$\log\frac{dP}{dQ_n} = \log\frac{dP}{dP^*} + \log\frac{dP^*}{dP_n^*} + \log\frac{dP_n^*}{dQ_n}\ . \tag{47}$$

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 (23), with $\Delta W^Q_{t_{n,j+1}}$ 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, (46) should be replaced by

$$\log\frac{dP_n^*}{dQ_n} = M_n^{(1,Q)} - \frac{1}{2}\Gamma_1 + o_p(1). \tag{48}$$

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 9 turn out to all have a representation in terms of Hermite polynomials.

Recall that the standardized Hermite polynomials are given by $h_{r_1}(x) = x^{r_1}$, $h_{r_1r_2}(x) = x^{r_1}x^{r_2} - \delta^{r_1,r_2}$, and $h_{r_1r_2r_3}(x) = x^{r_1}x^{r_2}x^{r_3} - x^{r_1}\delta^{r_2,r_3}$[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. From Remark 9,

$$M_n^{(0,Q)} = \frac{1}{12}\sum_{j=0}^{n-1}(\Delta t_{n,j+1})^{1/2}\sum_{r_1,r_2,r_3=1}^{p} k_{t_{n,j}}^{(r_1,r_2,r_3)} h_{r_1r_2r_3}(\Delta W_{t_{n,j+1}}^Q/(\Delta t_{n,j+1})^{1/2}), \text{ and}$$
$$M_n^{(1,Q)} = -\frac{1}{2}\sum_i \sum_{t_{n,j}\in(\tau_{n,i-1},\tau_{n,i}]} \mathrm{tr}\left(\sigma_{\tau_{n,i-1}}(\zeta_{t_{n,j}}^{-1} - \zeta_{\tau_{n,i-1}}^{-1})\sigma_{\tau_{n,i-1}} h_{..}(\Delta W_{t_{n,j+1}}^Q/(\Delta t_{n,j+1})^{1/2})\right). \quad (49)$$

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

$$M_n^{(G,Q)} = \sum_{j=0}^{n-1}(\Delta t_{n,j+1})^{1/2}\left(\sigma_{\tau_{n,i-1}}^{-1}\mu_{\tau_{n,i-1}} h_{.}(\Delta W_{t_{n,j+1}}^Q/(\Delta t_{n,j+1})^{1/2})\right) \quad (50)$$

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

$$\Gamma_G = \int_0^T \mu_t^T(\sigma_t^T\sigma_t)^{-1}\mu_t dt. \quad (51)$$

We therefore get the following summary of our results:

$$\begin{aligned}
\log\frac{dP}{dP^*} &= M_n^{(G,Q)} - \frac{1}{2}\Gamma_G + o_p(1)\\
\log\frac{dP^*}{dP_n^*} &= M_n^{(0,Q)} - \frac{1}{2}\Gamma_0 + o_p(1)\\
\log\frac{dP_n^*}{dQ_n} &= M_n^{(1,Q)} - \frac{1}{2}\Gamma_1 + o_p(1).
\end{aligned} \quad (52)$$

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 10. Note that the $M_n^{(G,Q)}$ is in many ways different from $M_n^{(0)}$ and $M_n^{(1)}$. The convergence of the former is in probability, while the latter converge only in law. Thus, for example, the property discussed in Remark 8 (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 4 in Section 4.3.2 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 $A = a^2$. 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 that when adjusting from Q_n to P^*, the asymptotic variance of Z_n is unchanged, while the asymptotic bias may change.

4 Applications

We here discuss various applications of our theory. For simplicity, assume in following that sampling is equispaced (so $\Delta_{t_{n,j}} = \Delta t_n = T/n$ for all j). We return to the question of irregular sampling in Section 6. Except in Section 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

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

$$\overline{\Delta X}_{\tau_{n,i}} = \frac{1}{M_n} \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} \Delta X_{t_{n,j+1}} = \frac{1}{M_n}(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}) \tag{53}$$

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 $|\sigma_t|^p$

We return to the question of estimating

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

We shall not use estimators of the form $\sum_{j=1}^n |\Delta X_{t_{n,j}}|^p$, 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 $N(0, \sigma^2_{\tau_{n,i}})$ within each block. From the theory of unbiased minimum variance (UMVU) estimation (see, for example, Lehmann (1983)), the optimal estimator of $|\sigma_{\tau_{n,i}}|^p$ is

$$\widehat{|\sigma_{\tau_{n,i}}|^p} = c^{-1}_{M-1,p}(\hat{\sigma}^2_{\tau_{n,i}})^{p/2} \tag{54}$$

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

$$\begin{aligned} c_{M,p} &= E\left((\chi^2_M/M)^{p/2}\right) \\ &= \left(\frac{2}{M}\right)^{p/2} \frac{\Gamma(\frac{p+M}{2})}{\Gamma(\frac{M}{2})} \end{aligned} \tag{55}$$

where χ_M is 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

$$\hat{\theta}_n = (M\Delta t)\sum_i \widehat{|\sigma_{\tau_{n,i}}|^p} \tag{56}$$

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,

$$n^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} N(0,1)\left(TM\left(\frac{c_{M-1,2p}}{c^2_{M-1,p}} - 1\right)\int_0^T \sigma_t^{2p} dt\right)^{1/2} \tag{57}$$

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

$$\begin{aligned} \mathrm{Var}((M\Delta t)\sum_i (\widehat{|\sigma_{\tau_{n,i}}|^p}) \mid \mathcal{Y}_{n,i}) &= \sigma^{2p}_{\tau_{n,i}}(M\Delta t)^2 c^{-2}_{M-1,p}\mathrm{Var}\left((\chi^2_{M-1}/(M-1))^{p/2}\right) \\ &= \sigma^{2p}_{\tau_{n,i}}(M\Delta t)\frac{TM}{n}\left(\frac{c_{M-1,2p}}{c^2_{M-1,p}} - 1\right). \end{aligned} \tag{58}$$

Remark 11. (Not taking out the mean). One can replace $\hat{\sigma}^2_{\tau_{n,i}}$ by

$$\tilde{\sigma}^2_{\tau_{n,i}} = \frac{1}{\Delta t_n M_n}\sum_{t_{n,j}\in(\tau_{n,i-1},\tau_{n,i}]} (\Delta X_{t_{n,j+1}})^2, \tag{59}$$

and take

$$\widetilde{|\sigma_{\tau_{n,i}}|^p} = c^{-1}_{M,p}(\tilde{\sigma}^2_{\tau_{n,i}})^{p/2} \tag{60}$$

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

$$TM\left(\frac{c_{M,2p}}{c^2_{M,p}} - 1\right)\int_0^T \sigma_t^{2p} dt. \tag{61}$$

□

4.1.2 Asymptotic Efficiency

We note that for large M,

$$\text{asymptotic variance of } n^{1/2}(\hat{\theta}_n - \theta) \;\downarrow\; T\frac{p^2}{2}\int_0^T \sigma_t^{2p} dt. \tag{62}$$

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 $T\left(\frac{c_{1,2p}}{c_{1,p}^2} - 1\right)\int_0^T \sigma_t^{2p} dt$. Compared to the variance in (62), the earlier estimator has asymptotic relative efficiency (ARE)

$$\begin{aligned} ARE(\text{estimator from Example 1}) &= \frac{\text{asymptotic variance in (62)}}{\text{asymptotic variance of estimator from Example 1}} \\ &= \frac{p^2}{2}\left(\frac{c_{1,2p}}{c_{1,p}^2} - 1\right)^{-1} \end{aligned} \tag{63}$$

Note that except for $p = 2$, $ARE < 1$. Figure 1 gives a plot of the ARE as a function of p. As one can see, there can be substantial gain from using the proposed estimator (56).

Remark 12. In terms of asymptotic distribution, there is further gain in using the estimator from Remark 11. Specifically, $ARE_M(\tilde{\theta})/ARE_M(\hat{\theta}) = M/(M-1)$. This is borne out by Figure 1. However, it is likely that the drift μ, as well as the block size M, would show up in a higher order bias calculation. This would make $\tilde{\sigma}$ less attractive. In connection with estimating the leverage effect, it is crucial to use $\hat{\sigma}$ rather than $\tilde{\sigma}$, cf. Section 4.3.2. □

Remark 13. We emphasize again that M has to be fixed in the present calculation, so that the ideal asymptotic rate on the right hand side of (62) 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. □

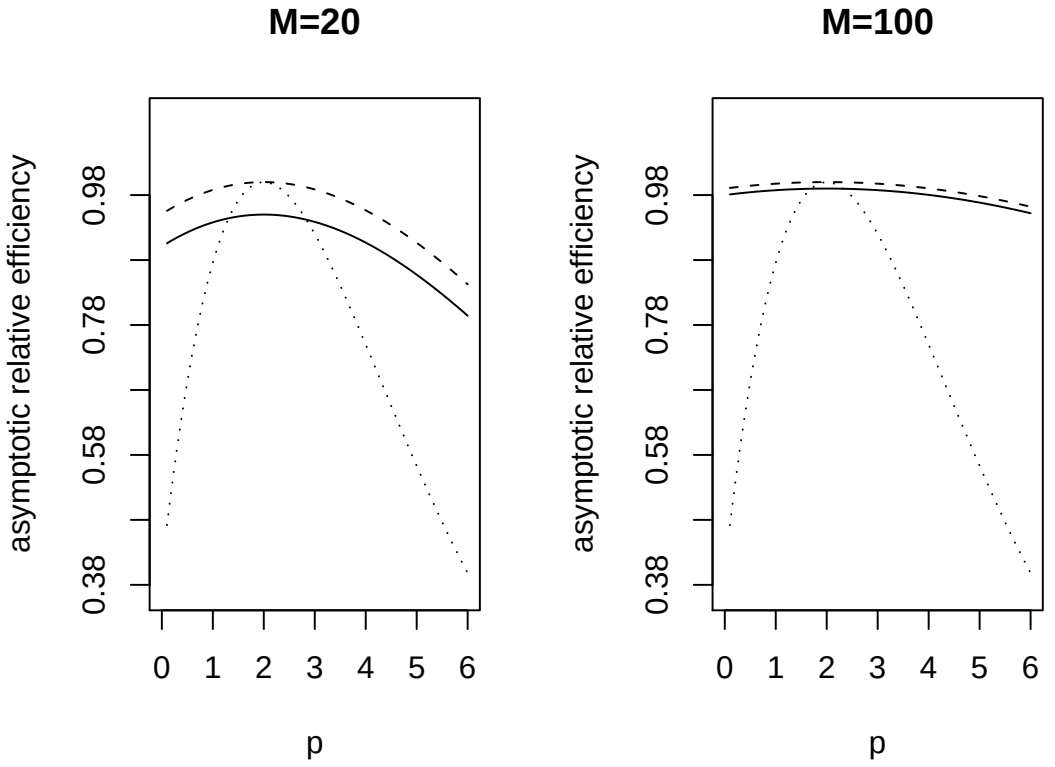

Figure 1. Asymptotic relative efficiency (ARE) of three estimators of $\theta = \int_0^T \sigma_t^p dt$, as a function of p. The dotted curve corresponds to the traditional estimator, which is proportional to $\sum_{j=1}^n |\Delta X_{t_{n,j}}|^p$. The solid and dashed lines are the ARE's of the block 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 $p = 2$. There is some cost to removing the mean in each block (the difference between the dashed and the solid curve).

4.2 ANOVA (Analysis of variance/variation)

We here revisit the problem from Zhang (2001) and Mykland and Zhang (2006). Consider processes $X_t^{(1)}, ..., X_t^{(p)}$ and Y_t which are observed synchronously at times $0 = t_{N,0} < t_{N,1} < ... < t_{N,n_1} = T$. These processes are related by

$$dY_t = \sum_{i=1}^{p} f_s^{(i)} dX_s^{(i)} + dZ_t, \text{ with } \langle X^{(i)}, Z \rangle_t = 0 \text{ for all } t \text{ and } i. \tag{64}$$

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

and trading purposes.

Following the theoretical development in the early sections, one strategy is as follows. Take the block size $M > p$. In each block, regress ΔY_{t_j} on ΔX_{t_j} linearly, and without intercept. Call the residuals $\widehat{\Delta Z}_{t_j}$. A natural estimator for $\langle Z, Z\rangle_T$ is therefore

$$\widehat{\langle Z, Z\rangle_T} = \frac{M}{M-p}\sum_{j=1}^{n_1} \widehat{\Delta Z}_{t_j}^2. \tag{65}$$

This strategy was analyzed in the simpler setting of Mykland (2006). In the more complex setting of this paper, one can see that

$$n^{1/2}(\widehat{\langle Z, Z\rangle_T} - \langle Z, Z\rangle_T) \xrightarrow{\mathcal{L}} N(0,1)\left(2\frac{M}{M-p}\int_0^T (\langle Z, Z\rangle_t')^2 dt\right)^{1/2} \tag{66}$$

stably in law under P.

Remark 14. Compared to the results of Zhang (2001) and Mykland and Zhang (2006), the difference in method is that M is finite and fixed, while in the earlier paper, $M \to \infty$ with n. In terms of results, the estimator in (65) has no asymptotic bias, whereas bias is present, though correctable from the data, in the earlier work. On the other hand, the current estimator is not quite efficient, as the asymptotic variance in Zhang (2001) and Mykland and Zhang (2006) is

$$2\int_0^T (\langle Z, Z\rangle_t')^2 dt. \tag{67}$$

Of course, the variance in (66) converges to that of (67) as $M \to \infty$. □

4.2.1 The case of regression with intercept

The system

$$dY_t = f_s^{(0)} ds + \sum_{i=1}^{p} f_s^{(i)} dX_s^{(i)} + dZ_t, \text{ with } \langle X^{(i)}, Z\rangle_t = 0 \text{ for all } t \text{ and } i. \tag{68}$$

can be treated similarly. In each block, one regresses ΔY_{t_j} on ΔX_{t_j} linearly, but now with intercept. The procedure is otherwise similar. It is easy to see that the asymptotic variance in (66) becomes

$$2\frac{M}{M-(p+1)}\int_0^T (\langle Z, Z\rangle_t')^2 dt. \tag{69}$$

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

$$\widetilde{\langle \sigma^2, X \rangle}_T = \sum_i (\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \tag{70}$$

where $\hat{\sigma}^2_{\tau_{n,i}}$ and $\overline{\Delta X}_{\tau_{n,i}}$ are given above in (53). Section 4.3.1 shows that this estimator is asymptotically biased. Accordingly, we define an asymptotically unbiased estimator of leverage effect by

$$\widehat{\langle \sigma^2, X \rangle}_T = \frac{2M}{2M+3} \sum_i (\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \tag{71}$$

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

$$\widehat{\langle \sigma^2, X \rangle}_T - \langle \sigma^2, X \rangle \xrightarrow{\mathcal{L}} c_M^{1/2} N(0,1) \tag{72}$$

stably under P^* and P, where

$$c_M = \frac{4}{M-1} \left(\frac{2M}{2M+3} \right)^2 \int_0^T \sigma_t^6 dt. \tag{73}$$

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

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

REMARK 15. It is conjectured that there is an optimal rate of $M = O(n^{1/2})$ as $n \to \infty$. The presumed optimal convergence rate is $n^{1/4}$, in analogy with the results in Zhang (2006). This makes sense because $\hat{\sigma}_t^2$ is a noisy measurement of σ_t^2. The problem of estimating $\langle \sigma^2, X \rangle_T$ is therefore similar to estimating volatility in the presence of microstructure noise.

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

4.3.1 Analysis

We here show how to arrive at the final result (72). This serves as a fairly extensive illustration of how to apply the theory development in the earlier sections.

By rearranging terms, write

$$\begin{aligned} \widetilde{\langle \sigma^2, X \rangle}_T = & \sum_i (\sigma^2_{\tau_{n,i+1}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \\ & + \sum_i (\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}) \\ & - \sum_i (\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) + O_p(n^{-1}), \end{aligned} \tag{74}$$

where the $O_p(n^{-1})$ term comes from edge effects. Note that by conditional Gaussianity, both the two last sums in (74) are Q_n-martingales with respect to the sigma-fields $\mathcal{Y}_{n,i}$. They are also orthogonal, in the sense that

$$\mathrm{Cov}_n^Q((\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}), (\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \mid \mathcal{Y}_{n,i}) = 0 \tag{75}$$

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

$$\begin{aligned}
&\mathrm{Var}_n^Q((\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}) \mid \mathcal{Y}_{n,i}) \\
&\quad = \sigma^4_{\tau_{n,i}}(M-1)^{-2}(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})^2\mathrm{Var}(\chi^2_{M-1}) \\
&\quad = 2\sigma^4_{\tau_{n,i}}(M-1)^{-1}(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})^2
\end{aligned} \tag{76}$$

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

$$\frac{2}{M-1}\int_0^T \sigma_t^6 dt. \tag{77}$$

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 (77).

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,

$$\begin{aligned}
&\mathrm{Var}_n^Q((\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \mid \mathcal{Y}_{n,i}) \\
&\quad = \sigma^6_{\tau_{n,i}}(M\Delta t)\mathrm{Var}((\frac{\chi^2_{M-1}}{M-1} - 1)N(0,1)) \\
&\quad = \frac{2}{M-1}\sigma^6_{\tau_{n,i}}(M\Delta t).
\end{aligned} \tag{78}$$

Hence the asymptotic quadratic variation is

$$\frac{2}{M-1}\int_0^T \sigma_t^6 dt. \tag{79}$$

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

$$\begin{aligned}
&\mathrm{Cov}_n^Q((\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), W_{\tau_{n,i+1}} - W_{\tau_{n,i}} \mid \mathcal{Y}_{n,i}) \\
&\quad = \sigma^3_{\tau_{n,i}}(M\Delta t)\mathrm{Cov}((\frac{\chi^2_{M-1}}{M-1} - 1)N(0,1), N(0,1)) \\
&\quad = 0.
\end{aligned} \tag{80}$$

Overall, under both Q_n and P_n^*,

$$\sum_i (\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \xrightarrow{\mathcal{L}} N(0,1)\left(\frac{2}{M-1}\int_0^T \sigma_t^6 dt\right)^{1/2}, \tag{81}$$

stably.

There is, however, covariation between this sum and $M_n^{(0,Q)}$. It is shown below in Remark 16 (see equation (89)) that $A_{12} = \frac{3}{2M}\langle \sigma^2, X\rangle_T$, where A_{12} has the same meaning as in Theorems 2 and 4 (in Sections 2.4 and 3.4, respectively). Thus, by these theorems, under P^*, we have (stably)

$$\sum_i (\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}) \xrightarrow{\mathcal{L}} \frac{3}{2M}\langle \sigma^2, X\rangle_T + N(0,1)\left(\frac{2}{M-1}\int_0^T \sigma_t^6 dt\right)^{1/2}. \tag{82}$$

Because of the orthogonality (75), 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 $\widehat{\langle \sigma^2, X\rangle}_T - \langle \sigma^2, X\rangle$ converges stably (under P^*) to a normal distribution with mean as in equation (82), and variance contributed by the second and third terms on the right hand side of (74). In other words:

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

$$\widetilde{\langle \sigma^2, X\rangle}_T \xrightarrow{\mathcal{L}} \left(1 + \frac{3}{2M}\right)\langle \sigma^2, X\rangle + N(0,1) \times \left(\frac{4}{M-1}\int_0^T \sigma_t^6 dt\right)^{1/2} \tag{83}$$

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

This shows the result (72).

REMARK 16. (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

$$G_{r,i} = \sum_{t_{n,j} \in (\tau_{n,i}, \tau_{n,i+1}]} h_r(\Delta W^Q_{t_{n,j+1}}/\Delta t^{1/2}), \tag{84}$$

note that

$$\begin{aligned} X_{\tau_{n,i+1}} - X_{\tau_{n,i}} &= \sigma_{\tau_{n,i}} \Delta t^{1/2} G_{1,i} \text{ and} \\ \hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}} &= \frac{\sigma^2_{\tau_{n,i}}}{M-1}\left(G_{2,i} - \frac{1}{M}G_{1,i}^2 + 1\right) \end{aligned} \tag{85}$$

At the same time,

$$M_n^{(0,Q)} = \frac{1}{12}(\Delta t)^{1/2} \sum_i k_{\tau_{n,i}} G_{3,i} + o_p(1) \tag{86}$$

The covariance for each i-increment becomes

$$\begin{aligned}
&\mathrm{Cov}_n^Q((\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \frac{1}{12}(\Delta t)^{1/2} k_{\tau_{n,i}} G_{3,i} \mid \mathcal{Y}_{n,i}) \\
&\quad = \frac{1}{12}\Delta t \frac{k_{\tau_{n,i}}\sigma^3_{\tau_{n,i}}}{M-1}\mathrm{Cov}_n^Q((G_{2,i} - \frac{1}{M}G^2_{1,i} + 1)G_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\
&\quad = \frac{1}{2}(M\Delta t)\frac{k_{\tau_{n,i}}\sigma^3_{\tau_{n,i}}}{M}
\end{aligned} \tag{87}$$

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

$$\begin{aligned}
&\mathrm{Cov}_n^Q((G_{2,i} - \frac{1}{M}G^2_{1,i} + 1)G_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\
&\quad = \mathrm{cum}^Q_{3,n}(G_{1,i}, G_{2,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) - \frac{1}{M}\mathrm{cum}^Q_{4,n}(G_{1,i}, G_{1,i}, G_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\
&\quad = M\mathrm{cum}_3(h_1(N(0,1)), h_2(N(0,1), h_3(N(0,1))) \\
&\quad - \mathrm{cum}_4(h_1(N(0,1)), h_1(N(0,1), h_1(N(0,1)), h_3(N(0,1))) \\
&\quad = 6(M-1).
\end{aligned} \tag{88}$$

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

$$\begin{aligned}
A_{12} &= \frac{1}{2M}\int_0^T k_t \sigma_t^3 dt \\
&= \frac{3}{2M}\langle \sigma^2, X\rangle_T
\end{aligned} \tag{89}$$

as in (33). □

4.3.2 The rôle of μ

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

Example 4. (Not removing the mean from the estimate of σ^2). Suppose that one wishes to use the estimator (70), but replacing $\hat{\sigma}^2_{\tau_{n,i}}$ by the estimator $\tilde{\sigma}^2_{\tau_{n,i}}$ from (59). An estimator analogous to $\widetilde{\langle \sigma^2, X\rangle}_T$ is then

$$\widetilde{\langle \sigma^2, X\rangle}_T^{\text{ with mean}} = \sum_i (\tilde{\sigma}^2_{\tau_{n,i+1}} - \tilde{\sigma}^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \tag{90}$$

We have the representation

$$\tilde{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}} = \frac{\sigma^2_{\tau_{n,i}}}{M}G_{2,i}. \tag{91}$$

We now consider the terms in (74). First note that $\sum_i(\tilde{\sigma}^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})(X_{\tau_{n,i}}-X_{\tau_{n,i-1}})$ is unaffected by this change. 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.

Since $\mathrm{cum}_3(h_1(N(0,1)),h_2(N(0,1),h_3(N(0,1)))=6$, it now follows as in (87) that

$$\begin{aligned}\mathrm{Cov}^Q_n((\tilde{\sigma}^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}}-X_{\tau_{n,i}}),\frac{1}{12}\Delta t^{1/2}k_{\tau_{n,i}}G_{3,i}\mid\mathcal{Y}_{n,i})&=\Delta t\frac{k_{\tau_{n,i}}\sigma^3_{\tau_{n,i}}}{2M}\\&=\Delta t\frac{3\langle\sigma^2,X\rangle'_{\tau_{n,i}}}{2M}.\end{aligned}\tag{92}$$

More importantly,

$$\begin{aligned}&\mathrm{Cov}^Q_n((\tilde{\sigma}^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}}-X_{\tau_{n,i}}),W^Q_{\tau_{n,i+1}}-W^Q_{\tau_{n,i}}\mid\mathcal{Y}_{n,i})\\&=\mathrm{Cov}^Q_n(\frac{\sigma^2_{\tau_{n,i}}}{M}G_{2,i}\sigma_{\tau_{n,i}}\Delta t^{1/2}G_{1,i},\Delta t^{1/2}G_{1,i}\mid\mathcal{Y}_{n,i})\\&=\Delta t\frac{\sigma^3_{\tau_{n,i}}}{M}\mathrm{Cov}^Q_n(G_{2,i}G_{1,i},G_{1,i}\mid\mathcal{Y}_{n,i})\\&=\Delta t\sigma^3_{\tau_{n,i}}\mathrm{Cov}(h_2(N(0,1))h_1(N(0,1)),h_1(N(0,1)))\\&=2\Delta t\sigma^3_{\tau_{n,i}}.\end{aligned}\tag{93}$$

Finally,

$$\begin{aligned}&\mathrm{Var}^Q_n((\tilde{\sigma}^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})\mid\mathcal{Y}_{n,i})\\&=\Delta t\frac{\sigma^6_{\tau_{n,i}}}{M^2}\mathrm{Var}(G_{2,i}G_{1,i}\mid\mathcal{Y}_{n,i})\\&=\left(2+6M^{-1}\right)\Delta t\sigma^6_{\tau_{n,i}}\end{aligned}\tag{94}$$

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

$$\widetilde{\langle\sigma^2,X\rangle}_T\overset{\text{with mean}}{\overset{\mathcal{L}}{\to}}\langle\sigma^2,X\rangle_T+\frac{2}{M}\int_0^T\sigma^3d\breve{W}+N(0,1)\left(2\frac{M+1}{M}\int_0^T\sigma^6dt\right)^{1/2}.\tag{95}$$

stably in law. From (92), we obtain from Theorem 4 that under P^*

$$\frac{2M}{2M+3}\widetilde{\langle\sigma^2,X\rangle}_T\overset{\text{with mean}}{\overset{\mathcal{L}}{\to}}\langle\sigma^2,X\rangle_T\;+\;\frac{4}{2M+3}\int_0^T\sigma^3dW^*\;+\;N(0,1)\left(8\frac{M(M+1)}{(2M+3)^2}\int_0^T\sigma^6dt\right)^{1/2}\tag{96}$$

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

$$\frac{4}{2M+3}\int_0^T\sigma^3(dW+\sigma_t^{-1}\mu_tdt).\tag{97}$$

□

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

$$\theta = \sum_i \theta_{n,i} + O_p(n^{-1}) \tag{98}$$

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

$$\hat{\theta}_n = \sum_i \hat{\theta}_{n,i}, \tag{99}$$

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,

$$\begin{aligned} &\sum_i \mathrm{Var}_n^Q(Z_{n,i} \mid \mathcal{Y}_{n,i}) \to \int_0^T f_t^2 dt \\ &\sum_i \mathrm{Cov}_n^Q(Z_{n,i}, W^Q_{\tau_{n,i+1}} - W^Q_{\tau_{n,i}} \mid \mathcal{Y}_{n,i}) \to \int_0^T g_t dt \end{aligned} \tag{100}$$

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

$$A_{13} = -\frac{1}{2} \lim_{n\to\infty} \sum_i \mathrm{Cov}_n^Q \left(Z_{n,i}, \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} \left(\sigma^2_{\tau_{n,i-1}} (\zeta^{-1}_{t_{n,j}} - \zeta^{-1}_{\tau_{n,i-1}}) h_2(\Delta W^Q_{t_{n,j+1}} / (\Delta t_{n,j+1})^{1/2}) \right) \mid \mathcal{Y}_{n,i} \right). \tag{101}$$

We finally obtain

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

$$n^\alpha(\hat{\theta}_n - \theta_n) \xrightarrow{\mathcal{L}} b + A_{12} + A_{13} + N(0,1) \left(\int_0^T (f_t^2 - g_t^2) dt \right)^{1/2} \tag{102}$$

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

$$b = \int_0^T g_t dW_t^* = \int_0^T g_t (dW_t + \sigma_t^{-1} \mu_t dt). \tag{103}$$

6 Irregularly spaced and asynchronous data

The general results stated in the paper cover irregularly spaced data. The theory therefore permits the analysis of estimators in this setting. However, It does not necessarily tell us what is the best estimator in the irregular sampling setting.

To clarify the issue, we show in Section 6.1 that one should not blindly assume that data can be considered as equispaced. We discuss in Section 6.2 some ideas for how one can approach irregularly spaced data.

The same general situation applies to asynchronous data. This is discussed in Section 6.3.

6.1 A second block approximation.

The approximation in Section 3 reduces the problem (in each block) to a case of independent (but not identically distributed) increments. Can we do better than this, and go to iid observations?

To answer this question, consider another approximate probability measure R_n:

$$\begin{aligned} X_0 &= x_0 \\ \text{for each } i &= 1, K_n : \\ \Delta X_{t_{j+1}} &= \sigma_{\tau_{n,i-1}} (\frac{\Delta \tau_i}{\Delta t_{j+1} M_i})^{1/2} \Delta W^*_{t_{j+1}} \text{ for } t_{n,j+1} \in (\tau_{n,i-1}, \tau_{n,i}] \end{aligned} \tag{104}$$

Formally, we define the approximation as follows.

Definition 6. *R_n is defined as Q_n in Definition 4, but with (104) replacing (39).*

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

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

Proposition 4.

$$\begin{aligned} &\log \frac{dR_n}{dP^*_n}(U_{t_0}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}}) \\ &= \sum_i \sum_{\tau_{i-1} \leq t_j < \tau_i} \left\{ \ell(\Delta X_{t_{j+1}}; \zeta_{\tau_{n,i-1}} \Delta \tau_i / M_i) - \ell(\Delta X_{t_{j+1}}; \zeta_{t_{n,j}} \Delta t_{j+1}) \right\} \end{aligned} \tag{105}$$

The contiguity question is then addressed as follows. Recall that

$$\log \frac{dR_n}{dP^*_n} = \log \frac{dR_n}{dQ_n} + \log \frac{dQ_n}{dP^*_n}. \tag{106}$$

Define

$$B_{n,j} = \left(\Delta t_{n,j+1} \left(\frac{\Delta \tau_{n,i}}{M_i} \right)^{-1} - 1 \right) \tag{107}$$

THEOREM 6. *(Asymptotic relationship between P_n^*, Q_n and R_n). Assume the conditions of Theorem 3. Assume that the following limits exist:*

$$\Gamma_2 = \frac{p}{2} \lim_{n\to\infty} \sum_j B_{n,j}^2 \text{ and } \Gamma_3 = \frac{p}{2} \lim_{n\to\infty} \sum_j \log(1 + B_{n,j}). \tag{108}$$

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

$$Z_n^{(2)} = \frac{1}{2} \sum_i \sum_{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]} \Delta X_{t_{n,j}}^T (\zeta_{\tau_{n,i-1}}^{-1}) \Delta X_{t_{n,j}} \left(\Delta t_{n,j+1}^{-1} - \left(\frac{\Delta \tau_{n,i}}{M_i} \right)^{-1} \right), \tag{109}$$

and let $M_n^{(2)}$ be the end point of the martingale part of $Z_n^{(2)}$ (see (B.23) and (B.25) in Appendix B for the explicit formula). Then, as $n \to \infty$, $(M_n^{(1)}, M_n^{(2)})$ converges stably in law under P^* to a normal distribution with mean zero and diagonal variance matrix with diagonal elements Γ_1 and Γ_2. Also, under P^*,

$$\log \frac{dR_n}{dQ_n} = M_n^{(2)} + \Gamma_3 + o_p(1). \tag{110}$$

The theorem can be viewed from the angle of contiguity:

COROLLARY 2. *Under the assumptions of Theorem 6, the following statements are equivalent, as $n \to \infty$:*
(i) R_n is contiguous to P_n^*.
(ii) R_n is contiguous to Q_n.
(iii) The following relationship holds:

$$\Gamma_3 = -\frac{1}{2}\Gamma_2. \tag{111}$$

As we shall see, the requirement (111) is a substantial restriction. Corollary 2 says that unlike the case of Q_n, inference under R_n may not give rise to desired results. Part of the probability mass under Q_n (and hence P^*) is not preserved under R_n.

To understand the requirement (111), note that

$$\frac{p}{2} \sum_j \log(1 + B_{n,j}) = -\frac{p}{4} \sum_j B_{n,j}^2 + \frac{p}{6} \sum_j B_{n,j}^3 - ... \tag{112}$$

since $\sum_j B_{n,j} = 0$. Hence, (111) will, for example, be satisfied if $\max_j |B_{n,j}| \to 0$ as $n \to \infty$. One such example is

$$t_{n,j} = f(j/n) \text{ and } f \text{ is continuously differentiable.} \tag{113}$$

However, (113) will not hold in more general settings, as we shall see from the following examples.

EXAMPLE 5. (POISSON SAMPLING.) Suppose that the sampling time points follow a Poisson process with parameter λ. If one conditions on the number of sampling points n, these points behave like the order statistics of n uniformly distributed random variables (see, for example, Chapter 2.3 in Ross (1996)). Consider the case where $M_i = M$ for all but (possibly) the last interval in $\mathcal{H}_n$. In this case, K_n is the smallest integer larger than or equal to n/M. Let Y_i be the M-tuple $(B_j, \tau_{i-1} \leq t_j < \tau_i)$.

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

$$Y_1 = M(U_{(1)}, U_{(2)} - U_{(1)}, ..., U_{(M-1)} - U_{(M-2)}, 1 - U_{(M-1)}) - 1, \tag{114}$$

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

$$\begin{aligned} \sum_j B_{n,j}^2 &= \frac{n}{M}(M^2 E U_{(1)}^2 - 1) + o_p(n) \\ \sum_j \log(1 + B_{n,j}) &= \frac{n}{M} E \log(M U_{(1)}) + o_p(n) \end{aligned} \tag{115}$$

since $EU_{(1)}^2 = 2/(M+1)(M+2)$. Hence, both Γ_2 and Γ_3 are infinite. The contiguity between R_n and the other probabilities fails. On the other hand all our assumptions up to Section 3 are satisfied, and so P, P^*, P_n^* and Q_n are all contiguous. The AQVT (equation (27)) is given by $H(t) = 2t$. Also, if the block size is constant (size M), the ADD is $K(t) = (M-1)t$. □

EXAMPLE 6. (SYSTEMATIC IRREGULARITY.) Let ϵ be a small positive number, and let $\Delta t_{n,j} = (1+\epsilon)T/n$ for odd j and $\Delta t_{n,j} = (1-\epsilon)T/n$ for even j (with $\Delta t_{n,j} = T/n$ for odd n). Again, all our assumptions up to Section 3 are satisfied. The AQVT is given by $H(t) = t(1+\epsilon^2)$. If we suppose that all $M_i = 2$, the ADD becomes $K(t) = t$. On the other hand, $B_{n,j} = \pm\epsilon$, so that, again, both Γ_2 and Γ_3 are infinite. The contiguity between R_n and the other probabilities thus fails in the same radical fashion as in the case of Poisson sampling. □

6.2 What to do about irregularly spaced data

In view of the above, one cannot blindly assume that data are equispaced. What should one do?

6.2.1 If the sampling times are reliable

In this case, the cure-all is to simply standardize increments to have (within block) equal variance. In other words, for the estimators discussed in Section 4, simply replace $\Delta X_{t_{n,j+1}}$ by $\Delta X_{t_{n,j+1}}(\overline{\Delta t}_{n,i}/\Delta t_{n,j+1})^{1/2}$, where $\overline{\Delta t}_{n,i} = (\tau_{n,i} - \tau_{n,i-1})/M_{n,i}$. Also, normalization by Δt gets replaced (in each block) by normalization by $\overline{\Delta t}_{n,i}$. This reduces all sums of powers to the same form as the Hermite polynomial sums in Section 3.3.

In this approach, results from the equispaced case go through with little modification. For constant block size (in terms of number of observations), the results are the same as in the equispaced case. Thus, the bias and variances given in Section 4 are unchanged.

We note that this approach has been advocated by Jacod (1994).

6.2.2 If the sampling times are unreliable/unavailable

There is no simple solution in this case. We here only give some comments and conjectures. The standard approach would be to write estimators as in the equispaced case, and then try to compensate for the irregularity later. By way of example, consider the estimator of leverage effect discussed in Section 4.3.

In equation (70), write more generally $\widetilde{\langle \sigma^2, X\rangle}_T = \sum_i (\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})$, where

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

$$\overline{\Delta X}_{\tau_{n,i}} = \frac{1}{M_{n,i}} \sum_{t_{n,j}\in(\tau_{n,i-1},\tau_{n,i}]} \Delta X_{t_{n,j+1}} = \frac{1}{M_{n,i}}(X_{\tau_{n,i}} - X_{\tau_{n,i-1}}) \tag{116}$$

It is shown below in Remark 17 that the asymptotic bias (equation (89)) is replaced by

$$A_{12} = \frac{3}{2M}\int \langle \sigma^2, X\rangle'_t dH_{3/2}(t) \tag{117}$$

where

$$H_{\frac{3}{2}}(t) = \lim_{n\to\infty} \sum_{\tau_{n,i+1}\le t} (\tau_{n,i+1} - \tau_{n,i}) \frac{1}{M} \sum_{t_{n,j}\in(\tau_{n,i},\tau_{n,i+1}]} (\Delta t_{j+1}/\overline{\Delta t}_{n,i})^{3/2}. \tag{118}$$

The form of $H_{\frac{3}{2}}(t) = t$ is often simpler in applications. First of all, in the equispaced case, $H_{\frac{3}{2}}(t) = t$. This leads to the corrected estimator $\widehat{\langle \sigma^2, X\rangle}_T$ proposed in Section 4.3. Second, the correction leading to $\widehat{\langle \sigma^2, X\rangle}_T$ is the minimal possible one. To wit, by Jensen's Inequality, $H'_{\frac{3}{2}}(t) \ge 1$ for all t.

In many cases, $H'_{\frac{3}{2}}(t)$ will be a constant. For example, in the case of Poisson sampling,

$$H'_{\frac{3}{2}}(t) \;=\; 3M^{1/2}\frac{2^{2M}(M+1)!M!}{(2M+2)!} \tag{119}$$

(approximately $3\sqrt{\pi}/4 \approx 1.33$ for large M). In the case of constancy, an adjusted estimator might be given by

$$\widehat{\langle \sigma^2, X\rangle}_T = \left(1 + \frac{3}{2M}\hat{H}'_{\frac{3}{2}}\right)^{-1} \widetilde{\langle \sigma^2, X\rangle}_T, \tag{120}$$

where $\hat{H}'_{\frac{3}{2}}$ is a suitably chosen estimator of $H'_{\frac{3}{2}}$. If the estimator $\hat{H}'_{\frac{3}{2}}$ is consistent, then $\widehat{\langle \sigma^2, X\rangle}_T$ will be asymptotically unbiased.

Similarly, if $H'_{\frac{3}{2}}(t)$ is indeed time varying, one may also be able to adjust the sum locally.

There is clearly room for more research on finding good estimators in the unequally spaced case. Our main contribution in this paper is to provide the tool to analyze such estimators, and we do not claim to have final proposals for estimators.

REMARK 17. We here give the calculations for the asymptotic bias of $\widetilde{\langle \sigma^2, X\rangle}_T$, to illustrate how to analyze the unequally spaced case. Set

$$G_{r,i} = \sum_{t_{n,j} \in (\tau_{n,i}, \tau_{n,i+1}]} h_r(\Delta W^Q_{t_{n,j+1}} / \Delta t^{1/2}_{t_{n,j+1}}), \tag{121}$$

and

$$\tilde{G}_{r,i} = \sum_{t_{n,j} \in (\tau_{n,i}, \tau_{n,i+1}]} h_r(\Delta W^Q_{t_{n,j+1}} / \overline{\Delta t}^{1/2}_{n,i}). \tag{122}$$

As in Section 4.3, the bias of $\widetilde{\langle \sigma^2, X\rangle}_T$ is a function of the correlation with $M_n^{(0)}$. In Remark 16, the derivation goes through word for word if $G_{r,i}$ is replaced by $\tilde{G}_{r,i}$ for $r = 1, 2$ (but not $r = 3$). We set $\delta_j = \Delta t_{j+1} / \overline{\Delta t}_{n,i}$, and replace equation (88) by

$$\begin{aligned} &\mathrm{Cov}^Q_n((\tilde{G}_{2,i} - \frac{1}{M}\tilde{G}^2_{1,i} + 1)\tilde{G}_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\ &\quad = 6\frac{M-1}{M} \sum_{t_{n,j} \in (\tau_{n,i}, \tau_{n,i+1}]} \delta_j^{3/2} \end{aligned} \tag{123}$$

Thus, in analogy with (87)

$$\begin{aligned} &\mathrm{Cov}^Q_n((\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}), \frac{1}{12}\overline{\Delta t}^{1/2}_{n,i} k_{\tau_{n,i}} G_{3,i} \mid \mathcal{Y}_{n,i}) \\ &\quad = \frac{1}{12}\overline{\Delta t}_{n,i} \frac{k_{\tau_{n,i}} \sigma^3_{\tau_{n,i}}}{M-1} \mathrm{Cov}^Q_n((G_{2,i} - \frac{1}{M}G^2_{1,i} + 1)G_{1,i}, G_{3,i} \mid \mathcal{Y}_{n,i}) \\ &\quad = \frac{3}{2}(\tau_{n,i+1} - \tau_{n,i}) \frac{\langle \sigma^2, X\rangle'_{\tau_{n,i}}}{M} \frac{1}{M} \sum_{t_{n,j} \in (\tau_{n,i}, \tau_{n,i+1}]} \delta_j^{3/2} \end{aligned} \tag{124}$$

Hence (89) is replaced by

$$A_{12} = \frac{3}{2M} \int \langle \sigma^2, X\rangle'_t dH_{3/2}(t). \tag{125}$$

This shows the asymptotic bias result given above. □

6.3 Asynchronous data

For purposes of analysis, asynchronous data does not pose any difficulty. One includes all observation times when computing the likelihood ratios in the contiguity theorems. It does not matter that some components of the vector are not observed at all these times. (A further elaboration of this principle is discussed in Section 8.2 below). In a sense, they are just treated as missing data. Just as in the case of irregular times for scalar processes, this does not necessarily mean that it is straightforward to write down sensible estimators.

For example, consider a bivariate process $(X_t^{(1)}, X_t^{(2)})$. If process $(X_t^{(r)})$ is observed at times $\mathcal{G}_n^{(r)} = \{0 \leq t_{n,0}^{(r)} < t_{n,1}^{(r)} < ... < t_{n,n_1}^{(r)} \leq T\}$, one would normally use the grid $\mathcal{G}_n = \mathcal{G}_n^{(1)} \cup \mathcal{G}_n^{(2)} \cup \{0, T\}$ to compute the quantities in Theorems 1-3.

7 Moving windows

The paper so far has considered chopping n data up into K non-overlapping windows of size M each. We here give a heuristic account of how the technology can be adapted to the moving window setting. This is given in the style of Section 5. The process (X_t) is assumed to be scalar.

It should be noted that the moving window is close to the concept of a moving kernel, and this may be a promising avenue of further investigation. See, in particular, Linton (2007).

7.1 Description of the result

We still assume that estimators are computed using blocks of size M. To define asymptotic variances, we set up a locally discretized measure as follows. $Q_n^{(\mathcal{M},j)}$ is identical to P_n^* up to observation $X_{t_{n,j}}$. This is followed by one block, of size $2M$, where updating is done as in Definition 4. After time $t_{n,j+\mathcal{M}}$, $Q_n^{(\mathcal{M},j)}$ does not need to be defined.

Suppose the parameter to be estimated is $\theta = \int_0^T \omega_t dt$, and let $\theta_{n,j} = M\Delta t \omega_{t_{n,j-M}}$. Then

$$\theta = \frac{1}{M} \sum_{j=M+1}^{n} \theta_{n,j} + O_p(n^{-1}). \tag{126}$$

Let $\hat{\theta}_{n,j}$ be an estimator of $\theta_{n,j}$, with the property that for any $\mathcal{M} \geq M$,

$$E^{Q_n^{\mathcal{M},j-\mathcal{M}}}(\hat{\theta}_{n,j} - \theta_{n,j} \mid \mathcal{X}_{t_{n,j-\mathcal{M}}}) = 0. \tag{127}$$

We suppose that $\hat{\theta}_{n,j}$ is a function of $\Delta X_{t_{n,j-M+1}}, ..., \Delta X_{t_{n,j}}$ The overall estimator of θ is

$$\hat{\theta}_n = \frac{1}{M} \sum_{j=M+1}^{n} \hat{\theta}_{n,j}. \tag{128}$$

We set $Z_{n,j} = n^\alpha(\hat{\theta}_{n,j} - \theta_{n,j})$, and suppose that the following limits exist in probability:

$$\frac{1}{M}\sum_j \mathrm{Var}^{Q_n^{2M,j-M}}(Z_{n,j} \mid \mathcal{X}_{n,j-M}) + \frac{2}{M^2}\sum_j\sum_{l=1}^{M} \mathrm{Cov}^{Q_n^{2M,j-M}}(Z_{n,j}, Z_{n,j+l} \mid \mathcal{X}_{n,j-M}) \to \int_0^T f_t^2 dt$$
$$\frac{1}{M}\sum_j \mathrm{Cov}^{Q_n^{2M,j-M}}(Z_{n,j}, W^Q_{t_{n,j}} - W^Q_{t_{n,j-M}} \mid \mathcal{X}_{n,j-M}) \to \int_0^T g_t dt \tag{129}$$

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

$$A_{12} = \frac{1}{12M}\lim_{n\to\infty}\sum_j\sum_{l=0}^{M-1} \mathrm{Cov}^{Q_n^{2M,j-M}}\left(Z_{n,j}, (\Delta t_{n,j-l})^{1/2} k_{t_{n,j-M}} h_3(\Delta W^Q_{t_{n,j-l}}/(\Delta t_{n,j-l})^{1/2}) \mid \mathcal{X}_{n,j-M}\right)$$

and

$$A_{13} = -\frac{1}{2M}\lim_{n\to\infty}\sum_j\sum_{l=0}^{M-1} \mathrm{Cov}^{Q_n^{2M,j-M}}\left(Z_{n,j}, \sigma^2_{t_{n,j-M}}(\zeta^{-1}_{t_{n,j-l}} - \zeta^{-1}_{t_{n,j-M}}) h_2(\Delta W^Q_{t_{n,j-l}}/(\Delta t_{n,j-l})^{1/2}) \mid \mathcal{X}_{n,j-M}\right). \tag{130}$$

It is then the case, as in Theorem 5, that subject to regularity conditions,

$$n^\alpha(\hat{\theta}_n - \theta_n) \xrightarrow{\mathcal{L}} b + A_{12} + A_{13} + N(0,1)\left(\int_0^T (f_t^2 - g_t^2)dt\right)^{1/2} \tag{131}$$

stably in law under P^* and P, with $N(0,1)$ independent of $\mathcal{F}_T$. As before, $b = \int_0^T g_t dW_t^* = \int_0^T g_t(dW_t + \sigma_t^{-1}\mu_t dt)$

7.2 The underlying analysis

We emphasize that the following is a quite summary development. Define yet another $Q_n^{\mathcal{M}}$ using larger intervals of constancy $(\tau_{n,i-1}, \tau_{n,i}]$. We assume for simplicity that the first block starts at zero, and that all blocks (except the last one) satisfies.

$$\mathcal{M} = \#\{t_{n,j} \in (\tau_{n,i-1}, \tau_{n,i}]\}, \tag{132}$$

where $\mathcal{M} \geq M$. Write

$$\tilde{Z}_{n,i} = \frac{1}{M}\sum_{\tau_{n,i-1} \leq t_{n,j-M} < t_{n,j} \leq \tau_{n,i}} Z_{n,j}\ , \tag{133}$$

and note that $E_n^Q(\tilde{Z}_{n,i} \mid \mathcal{Y}_{n,i-1}) = 0$. We can now use our theory on the Q_n-martingale sum $\tilde{Z}_n = \sum_i \tilde{Z}_{n,i}$. For large $\mathcal{M}$, the variances and covariances then become as described in the previous section. The remainder $n^\alpha(\hat{\theta}_n - \theta_n) - \tilde{Z}_n$ (which is the sum of the remaining $Z_{n,j}$) is shown (using another blocking, with block size $\mathcal{M}$, but starting the second block at $\mathcal{M} - M$) to be $O_p(1)$, but this term is arbitrarily small by taking $\mathcal{M}$ big enough. This shows the result. (A rigorous proof would, needless to say, take quite long. It is, however, easy to develop using the ideas given).

8 More complicated data generating mechanisms

8.1 Jumps

We only consider the case of finitely many jumps (compound Poisson processes, and similar). The conceptually simplest approach is to remove these jumps using the procedure described in Mancini (2001) and Lee and Mykland (2006). The procedure will detect all intervals $(t_{n,j-1}, t_{n,j}]$ containing jumps, with probability tending to one (exponentially fast) as $n \to \infty$. If one simply removes the detected intervals from the analysis, it is easy to see that our asymptotic results go through unchanged.

The case of infinitely many jumps is more complicated, and beyond the scope of this paper.

Note that there is a range of approaches for estimating the continuous part of volatility in such data. Methods include bi- and multi-power (Barndorff-Nielsen and Shephard (2004), see also Example 2 above). Other devices are considered by Aït-Sahalia (2004), and Aït-Sahalia and Jacod (2007). One can use our method of analysis for all of these approaches.

8.2 Microstructure noise

The presence of noise does not alter the analysis in any major way. Suppose one observes

$$Y_{t_{n,j}} \;=\; X_{t_{n,j}} + \epsilon_{n,j} \tag{134}$$

where the $\epsilon_{n,j}$'s are independent of the (X_t) process. The latter still follows (7). We take the σ-field $\mathcal{X}_{n,n}$ to be generated by $\{X_{t_{n,j}}, \epsilon_{n,j},\ 0 \leq j \leq n\}$. Suppose that P_1 and P_2 are two measures on $\mathcal{X}_{n,n}$ for which: (1) the variables $\{\epsilon_{n,j},\ 0 \leq j \leq n\}$ are independent of $\{X_{n,j},\ 0 \leq j \leq n\}$, and (2) the variables $\{\epsilon_{n,j},\ 0 \leq j \leq n\}$ have the same distribution under P_1 and P_2. Then, from standard results in measure theory,

$$\frac{dP_2}{dP_1}\left((X_{n,j}, \epsilon_{n,j}),\ 0 \leq j \leq n\right) \;=\; \frac{dP_2}{dP_1}\left(X_{n,j},\ 0 \leq j \leq n\right) \tag{135}$$

The results in our theorems are therefore unchanged in the case of microstructure noise (unless one also wants to change the probability distribution of the noise). We note that this remains the case irrespective of the internal dependence structure of the noise.

The key observation which leads to this easy extension is that our results do not require that the observables $Y_{t_{n,j}}$ generate the σ-field $\mathcal{X}_{n,n}$. It is only required that the observables be measurable with respect to this σ-field. The same principle was invoked in Section 6.3.

The extension does not, obviously, solve all problems relating to microstructure noise, since this type of data generating mechanism is best treated with an asymptotics where $M \to \infty$ as $n \to \infty$. This is currently under investigation.

9 Alternative Representations of the One Period Discretization

In the remaining sections, we discuss the more technical aspects of the theory, and give some proofs. This section is concerned with the one period discretization.

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

$$d\sigma_t = \tilde{\sigma}_t dt + f_t dW_t^* + g_t dB_t \tag{136}$$

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

$$\begin{aligned}\Delta X_{t_{n,j+1}} &= \sigma_{t_{n,j}} \Delta W^*_{t_{n,j+1}} + \int_{t_{n,j}}^{t_{n,j+1}} (\sigma_u - \sigma_{t_{n,j}}) dW_u^* \\ &= \sigma_{t_{n,j}} \Delta W_{t_{n,j+1}} + f_{t_{n,j}} \int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{u} dW_u^* \right) dW_t^* \\ &\quad + dB\text{-term} \ + \ \text{higher order terms} \ . \end{aligned} \tag{137}$$

It will turn out that the two first terms on the right hand side will matter in our approximation. We first give this approximation, and then we return to the simpler system (15). Note first that by taking quadratic covariations, one obtains

$$f_t^{(r_1,r_2,r_3)} = \langle \sigma^{(r_1,r_2)}, W^{(r_3)} \rangle'_t. \tag{138}$$

Similarly, a higher order approximation is given (with reference to (16):

Definition 7. *(Second order approximation). Define the probability P_n^{**} recursively by:*
(i) U_0 has same distribution under P_n^* as under P^*;
(ii) The conditional P_n^{**}-distribution of $U^{(1)}_{t_{n,j+1}}$ given $U_0, ..., U_{t_{n,j}}$ is given by

$$\Delta X_{t_{j+1}} = \sigma_{t_{n,j}} \Delta W^*_{t_{j+1}} + f_{t_{n,j}} \int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{u} dW_u^* \right) dW_t^*; \ \textit{and} \tag{139}$$

(iii) 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^*.

We first state the main result, and then comment on it.

Theorem 7. *P^* and P_n^{**} are mutually absolutely continuous on the σ-field $\mathcal{X}_n$ generated by $U_{t_{n,j}}$, $j = 0, ..., n$. Furthermore, let $(dP^*/dP_n^{**})(U_{t_0}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}})$ be the likelihood ratio (Radon-Nikodym derivative) on $\mathcal{X}_n$. Then,*

$$\frac{dP^*}{dP_n^{**}}(U_{t_0}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}}) = 1 + o_p(1) \tag{140}$$

under P_n^{**}, as $n \to \infty$.

Theorem 7 says that one can (for a fixed time period) carry out inference under the model (139), and asymptotic results will transfer to the continuous model (11) by asymptotic absolute continuity (*contiguity*). We state the following result, in analogy with Proposition 1.

COROLLARY 3. *Suppose that* $n^{1/2}(\hat{\theta}_n - \theta)$ *converges stably in law to* $b + aN(0,1)$ *under* P_n^{**}*, where* $N(0,1)$ *is independent of* $\mathcal{F}_T$*. The same statement then holds under* P^* *and* P*. The converse is also true.*

Unlike the situation is Theorems 1-2, there is here no need for adjustment from the approximate probability to P^*.

REMARK 18. (Connection to Milstein schemes) The approximation (137) specializes to a Milstein scheme when (X_t) is a Markov process (cf. Chapter 10.3 (p. 345-351) of Kloeden and Platen (1992)). Theorem 7 says that by using a partial Milstein approximation, econometric properties are unaffected in terms of asymptotic behavior. One can also use the full second order approximation (including the dB terms), but this is not necessary for (140). Presumably, higher order properties, such as Edgeworth expansions, may be affected. □

REMARK 19. (Symmetrized higher order schemes) The approximation (139) involves a double stochastic integral. For our purposes, this can be eliminated as follows. Let P_n^{sym} be as P_n^{**}, but with (139) replaced by

$$\Delta X_{t_{j+1}} = \sigma_{t_{n,j}} \Delta \breve{W}^*_{t_{j+1}} \tag{141}$$

where

$$\Delta \breve{W}^*_{t_{j+1}} = \Delta W^*_{t_{j+1}} + \frac{1}{12} k_{t_{n,j}} (\Delta W^*_{t_{n,j+1}} \Delta W^*_{t_{n,j+1}} - \frac{1}{2} I \Delta t_{n,j+1}). \tag{142}$$

To spell this out in coordinate form, component r_1 of $k_{t_{n,j}}(\Delta W^*_{t_{n,j+1}} \Delta W^*_{t_{n,j+1}} - \frac{1}{2} I \Delta t_{n,j+1})$ is $\sum_{r_2,r_3} k^{(r_1,r_2,r_3)}_{t_{n,j}} (\Delta W^{*,(r_2)}_{t_{n,j+1}} \Delta W^{*,(r_3)}_{t_{n,j+1}} - \frac{1}{2} \delta^{r_2,r_3} \Delta t_{n,j+1})$. It is then shown in Appendix A that

$$\frac{dP^*}{dP_n^{\text{sym}}}(U_{t_0}, ..., U_{t_{n,j}}, ..., U_{t_{n,n}}) = 1 + o_p(1) \tag{143}$$

under P_n^{sym}, as $n \to \infty$. □

The approximation (139) is very good, but having second order terms can complicate inference. In comparison, the linear one-period (Euler) approximation (15) has substantial conceptual value, but it requires a post-limit adjustment, as required in Theorem 2.

10 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 (powers of volatility, leverage effect, ANOVA) are provided.

The theory relies heavily on the interplay between stable convergence and measure change, and on asymptotic expansions for martingales.

A number of unsolved questions remain. The approach provided 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. Finally, while compound Poisson jumps are easy to screen out, there is no theory as yet for how to handle the case of infinitely many jumps.

Department of Statistics, The University of Chicago, Chicago, IL60637–1514, U.S.A.;
mykland@galton.uchicago.edu; http://galton.uchicago.edu/∼mykland.

and

Department of Finance, The University Illinoius at Chicago, Chicago, IL60607, 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.

——— (2004): "Disentangling Diffusion from Jumps," *Journal of Financial Economics*, 74, 487–528.

Aït-Sahalia, Y. and J. Jacod (2007): "Volatility Estimators for Discretely Sampled Lévy Processes," *Annals of Statistics*, 35, 335–392.

Aït-Sahalia, Y. and P. A. Mykland (2003): "The Effects of Random and Discrete Sampling When Estimating Continuous-Time Diffusions," *Econometrica*, 71, 483–549.

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 (2000): "Great realizations," *Risk*, 13, 105–108.

——— (2001): "The Distribution of Exchange Rate Realized Volatility," *Journal of the American Statistical Association*, 96, 42–55.

——— (2003): "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579–625.

Barndorff-Nielsen, O., S. Graversen, J. Jacod, M. Podolskij, and N. Shephard (2006): "A central limit theorem for realised bipower variations of continuous semimartingales," in *From Stochastic Calculus to Mathematical Finance, The Shiryaev Festschrift*, ed. by Y. Kabanov, R. Liptser, and J. Stoyanov, Berlin: Springer Verlag, 33–69.

Barndorff-Nielsen, O. E. and N. Shephard (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, B*, 64, 253–280.

——— (2004): "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.

Chernov, M. and E. Ghysels (2000): "A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation," *Journal of Financial Economics*, 57, 407–458.

Cizek, P., W. Härdle, and V. Spokoiny (2007): "Adaptive pointwise estimation in time-inhomogeneous time series models," Tech. rep., Weierstrass Institute, Berlin.

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.

Duffie, D. (1996): *Dynamic Asset Pricing Theory*, Princeton, N.J.: Princeton University Press.

Engle, R. F. (2000): "The Econometrics of Ultra-High Frequency Data," *Econometrica*, 68, 1–22.

Fan, J. (1993): "Local Linear Regression Smoothers and Their Minimax Efficiencies," *Annals of Statistics*, 21, 196–216.

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

Fan, J. and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications*, London, U.K.: Chapman and Hall.

Fan, J., I. Gijbels, and M. King (1997): "Local Likelihood and Local Partial Likelihood in Hazard Regression," *Annals of Statistics*, 25, 1661–1690.

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

Gallant, A. R., C.-T. Hsu, and G. T. Tauchen (1999): "Using Daily Range Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance," *The Review of Economics and Statistics*, 81, 617–631.

Goncalves, S. and N. Meddahi (2005): "Bootstrapping realized volatility," Tech. rep., Université de Montréal.

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

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

Harrison, M. and D. Kreps (1979): "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory*, 20, 381–408.

Harrison, M. and S. Pliska (1981): "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Processes and Their Applications*, 11, 215–260.

Hayashi, T. and N. Yoshida (2005): "On Covariance Estimation of Non-synchronously Observed Diffusion Processes," *Bernoulli*, 11, 359–379.

Heston, S. (1993): "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options," *Review of Financial Studies*, 6, 327–343.

Hjort, N. L. and M. Jones (1996): "Locally parametric nonparametric density estimation," *The Annals of Statistics*, 24, 1619–1647.

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

——— (2006): "Asymptotic Properties of Realized Power Variations and related Functionals of Semimartingales," Tech. rep., Université de Paris VI.

Jacod, J. and P. Protter (1998): "Asymptotic Error Distributions for the Euler Method for Stochastic Differential Equations," *Annals of Probability*, 26, 267–307.

Jacod, J. and A. N. Shiryaev (2003): *Limit Theorems for Stochastic Processes*, New York: Springer-Verlag, second ed.

Kalnina, I. and O. Linton (2007): "Inference about Realized Volatility using Infill Subsampling," Tech. rep., London School of Economics.

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

Kauermann, G. and J. Opsomer (2003): "Local Likelihood Estimation in Generalized Additive Models," *Scandinavian Journal of Statistics*, 30, 317–337.

Kessler, M. (1997): "On the Estimation of an Ergodic Diffusion from Discrete Observations," *Scandinavian Journal of Statistics*, 24, 211–229.

Kloeden, P. E. and E. Platen (1992): *Numerical Solution of Stochastic Differential Equations*, Berlin, Germany: Springer-Verlag.

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.

Lee, S. Y. and P. A. Mykland (2006): "Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics," Tech. rep., Georgia Institute of Technology and The University of Chicago, to appear in *Review of Financial Studies*.

Lehmann, E. (1983): *Theory of Point Estimation*, New York: John Wiley and Sons, second ed.

Linton, O. (2007): "Notes about Gaussian calculus," Presented in Imperical College Financial Econometrics Conference.

Loader, C. R. (1996): "Local likelihood density estimation," *The Annals of Statistics*, 24, 1602–1618.

Mancini, C. (2001): "Disentangling the Jumps of the Diffusion in a Geometric Jumping Brownian Motion," *Giornale dell'Istituto Italiano degli Attuari*, LXIV, 19–47.

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

Mykland, P. A. (1993): "Asymptotic Expansions for Martingales," *Annals of Probability*, 21, 800–818.

——— (1995a): "Embedding and Asymptotic Expansions for Martingales," *Probability Theory and Related Fields*, 103, 475–492.

——— (1995b): "Martingale Expansions and Second Order Inference," *Annals of Statistics*, 23, 707–731.

——— (2006): "A Gaussian Calculus for Inference from High Frequency Data," Tech. rep., University of Chicago.

Mykland, P. A. and L. Zhang (2001): "Inference for volatility type objects and implications for hedging," Tech. rep., Carnegie-Mellon University.

——— (2006): "ANOVA for Diffusions," *Annals of Statistics*, 34, 1931–1963.

Protter, P. (2004): *Stochastic Integration and Differential Equations: A New Approach*, New York: Springer-Verlag, second ed.

Renault, E. and B. J. Werker (2006): "Causality effects in return volatility measures with random times," Tech. rep., Tilburg University.

Rényi, A. (1963): "On Stable Sequences of Events," *Sankyā Series A*, 25, 293–302.

Rootzén, H. (1980): "Limit Distributions for the Error in Approximations of Stochastic Integrals," *Annals of Probability*, 8, 241–251.

Ross, S. (1996): *Stochastic Processes*, New York, N.Y.: Wiley, second ed.

Ross, S. M. (1976): "The arbitrage theory of capital asset pricing," *Journal of Economic Theory*, 13, 341–360.

Sørensen, M. (2007): "Efficient martingale estimating functions for discretely sampled ergodic diffusions," Tech. rep., The University of Copenhagen.

Tibshirani, R. and T. Hastie (1987): "Local Likelihood Estimation," *Journal of the American Statistican Association*, 82, 559–567.

Woerner, J. H. (2004): "Power and multipower variation: Inference for high frequency data," Tech. rep., University of Göttingen.

Zhang, L. (2001): "From Martingales to ANOVA: Implied and Realized Volatility," Ph.D. thesis, The University of Chicago, Department of Statistics.

——— (2005): "Estimating Covariation: Epps Effect and Microstructure Noise," Tech. rep., University of Illinois at Chicago (working paper).

——— (2006): "Efficient Estimation of Stochastic Volatility Using Noisy Observations: A Multi-Scale Approach," *Bernoulli*, 12, 1019–1043.

Zhang, L., P. A. Mykland, and Y. Aït-Sahalia (2005): "A Tale of Two Time Scales: Determining Integrated Volatility with Noisy High-Frequency Data," *Journal of the American Statistical Association*, 100, 1394–1411.

APPENDIX: PROOFS

A Proofs of Theorems 1 and 2

First some further notation. Define

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

$\breve{\sigma}_t^{(r_1,r_2)}$ and $\breve{f}_t(r_1,r_2,r_3)$ are not symmetric in (r_1,r_2). However, since $d\zeta_t = d(\sigma_t)^2 = \sigma_t d\sigma_t + (\sigma_t d\sigma_t)^* + dt$ terms, we obtain from (18) that $d\check{\zeta}_t = \sigma_t^{-1} d\sigma_t + (\sigma_t^{-1} d\sigma_t)^* + dt$ terms. Hence

$$\langle \check{\zeta}^{(r_1,r_2)}, W^{(r_3)} \rangle_t' = \breve{f}_t^{(r_1,r_2,r_3)} + \breve{f}_t^{(r_2,r_1,r_3)}. \tag{A.2}$$

Also

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

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

Proof of Theorem 1. Note that, from (22) and (137)

$$\begin{aligned} \Delta \breve{W}_{t_{n,j+1}} = \Delta W_{t_{n,j+1}} + \breve{f}_{t_{n,j}} \int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{u} dW_u^* \right) dW_t^* \\ + dB\text{-term} \ + \ \text{higher order terms} \, . \end{aligned} \tag{A.4}$$

In the representation (A.4), we obtain, up to $O_p(\Delta t^{5/2})$, .

$$\begin{aligned} &\mathrm{cum}_3(\Delta \breve{W}_{t_{n,j+1}}^{(r_1)}, \Delta W_{t_{n,j+1}}^{(r_2)}, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}}) \\ &\doteq \mathrm{cum}(\sum_{s_2,s_3} \breve{f}_{t_{n,j}}^{(r_1,s_2,s_3)} \int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{u} dW_u^{*(s_3)} \right) dW_t^{*(s_2)}, \Delta W_{t_{n,j+1}}^{(r_2)}, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}}) \\ &= \sum_{s_2,s_3} \breve{f}_{t_{n,j}}^{(r_1,s_2,s_3)} \mathrm{cum}(\int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{u} dW_u^{*(s_3)} \right) dW_t^{*(s_2)}, \Delta W_{t_{n,j+1}}^{(r_2)}, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}}) \\ &= \sum_{s_2,s_3} \breve{f}_{t_{n,j}}^{(r_1,s_2,s_3)} \mathrm{Cov}(\int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{u} dW_u^{*(s_3)} \right) dt \delta^{s_2,r_2}, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}})[2] \\ &= \sum_{s_3} \breve{f}_{t_{n,j}}^{(r_1,r_2,s_3)} \mathrm{Cov}(\int_{t_{n,j}}^{t_{n,j+1}} \left(\int_{t_{n,j}}^{u} dW_u^{*(s_3)} \right) dt, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}})[2] \\ &= \int_{t_{n,j}}^{t_{n,j+1}} dt \sum_{s_3} \breve{f}_{t_{n,j}}^{(r_1,r_2,s_3)} \mathrm{Cov}(\int_{t_{n,j}}^{u} dW_u^{*(s_3)}, \Delta W_{t_{n,j+1}}^{(r_3)} | \mathcal{F}_{t_{n,j}})[2] \\ &= \int_{t_{n,j}}^{t_{n,j+1}} dt \sum_{s_3} \breve{f}_{t_{n,j}}^{(r_1,r_2,s_3)} (u - t_{n,j}) \delta s_3, r_3[2] \\ &= \frac{1}{2} \Delta t_{n,j+1}^2 \breve{f}_{t_{n,j}}^{(r_1,r_2,r_3)} \ [2], \end{aligned} \tag{A.5}$$

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

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

by symmetry, Hence, from (19),

$$\begin{aligned}\kappa^{r_1,r_2,r_3} &= \frac{1}{2}\Delta t^{1/2}_{n,j+1}\breve{f}^{(r_1,r_2,r_3)}_{t_{n,j}}\ [3] + O_p(\Delta t)\\ &= \frac{1}{2}\Delta t^{1/2}_{n,j+1}k^{(r_1,r_2,r_3)}_{t_{n,j}} + O_p(\Delta t).\end{aligned} \tag{A.7}$$

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

$$\begin{aligned}\mathrm{Cov}(\Delta X^{(r_1)}_{t_{n,j+1}}, \Delta X^{(r_2)}_{t_{n,j+1}}|\mathcal{F}_{t_{n,j}}) &= \Delta t_{n,j+1}\zeta^{(r_1,r_2)}_{t_{n,j}} + E(\int_{t_{n,j}}^{t_{n,j+1}}(\zeta^{(r_1,r_2)}_u - \zeta^{(r_1,r_2)}_{t_{n,j}})du|\mathcal{F}_{t_{n,j}})\\ &= \Delta t_{n,j+1}\zeta^{(r_1,r_2)}_{t_{n,j}} + E(\int_{t_{n,j}}^{t_{n,j+1}} du\int_{t_{n,j}}^{u}\tilde{\zeta}^{(r_1,r_2)}_v dv|\mathcal{F}_{t_{n,j}})\\ &= \Delta t_{n,j+1}\zeta^{(r_1,r_2)}_{t_{n,j}} + \frac{1}{2}\Delta t^2_{n,j+1}\tilde{\zeta}^{(r_1,r_2)}_{t_{n,j}} + O_p(\Delta t^3),\end{aligned} \tag{A.8}$$

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

$$\kappa^{r_1,r_2} = \delta^{r_1,r_2}_{t_{n,j}} + O_p(\Delta t). \tag{A.9}$$

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 $\Delta X_{t_{n,j+1}}$ given $\mathcal{F}_{t_{n,j}}$, on the log scale as

$$\begin{aligned}\log f_{n,j+1}(x) = \log\phi(x;\delta^{r_1,r_2}) &+ \frac{1}{3!}\kappa^{r_1,r_2,r_3}h_{r_1r_2r_3}(x)\\ &+ \frac{1}{2}(\kappa^{r1,r2} - \lambda^{r1,r2})h_{r1r2}(x) + \frac{1}{4!}\kappa^{r_1,r_2,r_3,r_4}h_{r_1r_2r_3r_4}(x)\\ &+ \kappa^{r_1,r_2,r_3}\kappa^{r_4,r_5,r_6}h_{r_1r_2r_3r_4r_5r_6}(x)\frac{[10]}{6!}\\ &+ \frac{1}{72}(\kappa^{r_1,r_2,r_3}h_{r_1r_2r_3}(x))^2 + O_p(\Delta t^{3/2})\end{aligned} \tag{A.10}$$

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 $h_{r1r2r3} = h_{r1}h_{r2}h_{r3} - h_{r1}\delta_{r2,r3}[3]$, with $h_{r1} = \delta_{r1,r2}x^{r2}$. Observe that

$$Z_{r1} = h_{r1}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}) = \delta_{r1,r2}\Delta\breve{W}^{r2}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}. \tag{A.11}$$

Under the approximating measure, therefore, the vector consisting of elements Z_{r1} is conditionally normally distributed with mean zero and covariance matrix $\delta_{r1,r2}$.

It follows that

$$h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}) = Z_{r1}Z_{r2}Z_{r3} - Z_{r1}\delta_{r2,r3}[3] \tag{A.12}$$

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

$$\mathrm{Cov}_n(h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}), h_{r_4r_5r_6}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2})|\mathcal{F}_{t_{n,j}}) = \delta_{r1,r4}\delta_{r2,r5}\delta_{r3,r6}[6] \tag{A.13}$$

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

$$\begin{aligned}
&\mathrm{Var}_n(\frac{1}{3!}\kappa^{r_1,r_2,r_3}h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2})|\mathcal{F}_{t_{n,j}}) \\
&\quad= \frac{1}{36}\kappa^{r_1,r_2,r_3}\kappa^{r_4,r_5,r_6}\mathrm{Cov}_n(h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}), h_{r_4r_5r_6}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2})|\mathcal{F}_{t_{n,j}}) \\
&\quad= \frac{1}{6}\kappa^{r_1,r_2,r_3}\kappa^{r_4,r_5,r_6}\delta_{r1,r4}\delta_{r2,r5}\delta_{r3,r6} \\
&\quad= \Delta t_{n,j+1}\frac{1}{24}k^{r_1,r_2,r_3}_{t_{n,j}}k^{r_4,r_5,r_6}_{t_{n,j}}\delta_{r1,r4}\delta_{r2,r5}\delta_{r3,r6} + O_p(\Delta t^{3/2})
\end{aligned} \tag{A.14}$$

by symmetry of the κ's. Thus

$$\sum_{t_{n,j+1}\le t}\mathrm{Var}_n(\frac{1}{3!}\kappa^{r_1,r_2,r_3}h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2})|\mathcal{F}_{t_{n,j}})\xrightarrow{p}\int_0^t\frac{1}{24}k_t^{r_1,r_2,r_3}k_t^{r_4,r_5,r_6}\delta_{r1,r4}\delta_{r2,r5}\delta_{r3,r6}dt \tag{A.15}$$

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

$$\sum_{t_{n,j+1}\le t}\mathrm{Cov}_n(\ h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}),\ h_{r_4}(\Delta X_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}|\mathcal{F}_{t_{n,j}})\xrightarrow{p}0. \tag{A.16}$$

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

$$\breve{M}_n^{(0)} = \sum_{j=0}^{n}\frac{1}{3!}\kappa^{r_1,r_2,r_3}h_{r_1r_2r_3}(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}) \tag{A.17}$$

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 (23)). We now note that, in the notation of (A.10),

$$\log\frac{dP^*}{dP_n^*} = \sum_{j=0}^{n-1}(f_{n,j+1} - \log\phi)\,(\Delta\breve{W}_{t_{n,j+1}}/(\Delta t_{n,j+1})^{1/2}). \tag{A.18}$$

By the same reasoning as above, the terms other than $\breve{M}_n^{(0)}$ and its discrete time quadratic variation (A.15), 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 20. The proof of Theorem 1 uses the Edgeworth expansion (A.10). 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 (137) 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

$$\log \frac{dP^*}{dP_n^*} = M_n^{(0)} - \frac{1}{2}\Gamma_0 + o_p(1) \tag{A.19}$$

where $M_n^{(0)}$ is as defined in equation (23). 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,

$$\begin{aligned}
E^* g(Z_n) &= E_n^* g(Z_n) \exp\{M_n^{(0)} - \frac{1}{2}\Gamma_0\}(1 + o(1)) \\
&\to Eg(Z) \exp\{M - \frac{1}{2}\Gamma_0\} \\
&= E_\infty^* g(b_1 + c_1 \Gamma_0^{1/2} V_1 + c_2 V_2) \exp\{\Gamma_0^{1/2} V_1 - \frac{1}{2}\Gamma_0\} \\
&= \int_{-\infty}^{\infty} E_\infty^* g(b_1 + c_1 \Gamma_0^{1/2} v + c_2 V_2) \exp\{\Gamma_0^{1/2} v - \frac{1}{2}\Gamma_0\} (2\pi)^{-1/2} \exp\{-\frac{1}{2} v^2\} dv \\
&= \int_{-\infty}^{\infty} E_\infty^* g(b_1 + c_1 \Gamma_0^{1/2} (u + \Gamma_0^{1/2}) + c_2 V_2) (2\pi)^{-1/2} \exp\{-\frac{1}{2} u^2\} du \quad (u = v - \Gamma_0^{1/2}) \\
&= E_\infty^* g(Z + c_1 \Gamma_0)
\end{aligned} \tag{A.20}$$

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

PROOF OF THEOREM 7. The exact same derivations that are used in the proof of Theorem 1 show that

$$\log \frac{dP_n^{**}}{dP_n^*} = \log \frac{dP^*}{dP_n^*} + o_p(1) \tag{A.21}$$

Thus proves the result. ∎

PROOF OF REMARK 19. By Itô's formula, $d(W_t^{*,r_2} - W_{t_{n,j}}^{*,r_2})(W_t^{*,r_3} - W_{t_{n,j}}^{*,r_3}) = (W_t^{*,r_2} - W_{t_{n,j}}^{*,r_2}) dW_t^{*,r_3}[2] + \delta^{r_2,r_3} dt$, so that $\Delta W_{t_{n,j+1}}^{*,(r_2)} \Delta W_{t_{n,j+1}}^{*,(r_3)} - \frac{1}{2}\delta^{r_2,r_3}\Delta t_{n,j+1} = \left(\int_{t_{n,j}}^{u} dW_u^{*(r_2)}\right) dW_t^{*(r_3)}[2]$. The scheme given in Remark 19 therefore amounts to replacing f in (139) by $\frac{1}{6}\sigma k$. This in turn amounts to replacing $\breve{f}$ by $\frac{1}{6}k$. However, to first order, dP^*/dP_n^{**} only depends on $\breve{f}$ through k, and therefore, $dP^*/dP_n^{**} = dP^*/dP_n^{\text{sym}}(1 + o_p(1))$. This shows the result. ∎

B Proof of Theorems 3 and 6

COMMON INITIAL DEVELOPMENT.

Let $Z_n^{(1)}$ and $Z_n^{(2)}$ be given by (44) and (109), respectively. Set

$$\begin{aligned}\Delta Z^{(1)}_{n,t_{n,j+1}} &= \frac{1}{2}\Delta X^T_{t_{n,j+1}}(\zeta^{-1}_{t_{n,j}} - \zeta^{-1}_{\tau_{n,i-1}})\Delta X_{t_{n,j+1}}\Delta t^{-1}_{n,j+1} \\ \Delta Z^{(2)}_{n,t_{n,j+1}} &= \frac{1}{2}\Delta X^T_{t_{n,j+1}}(\zeta^{-1}_{\tau_{n,i-1}})\Delta X_{t_{n,j+1}}\left(\Delta t^{-1}_{n,j+1} - \left(\frac{\Delta\tau_{n,i}}{M_i}\right)^{-1}\right)\end{aligned} \tag{B.22}$$

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

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

$$\begin{aligned}E_{P_n^*}(\Delta Z^{(1)}_{n,t_{n,j+1}}|\mathcal{X}_{n,t_{n,j}}) &= -\frac{1}{2}\mathrm{tr}(A_j) \\ E_{P_n^*}(\Delta Z^{(2)}_{n,t_{n,j+1}}|\mathcal{X}_{n,t_{n,j}}) &= -\frac{1}{2}\mathrm{tr}(C_j) = -\frac{1}{2}(p + \mathrm{tr}(A_j))B_j\end{aligned} \tag{B.23}$$

and

$$\begin{array}{l}\text{conditional covariance of} \\ \Delta Z^{(1)}_{n,t_{n,j+1}} \text{ and } \Delta Z^{(2)}_{n,t_{n,j+1}}\end{array} = \frac{1}{2}\begin{pmatrix}\mathrm{tr}(A_j^2) & \mathrm{tr}(A_jC_j) \\ \mathrm{tr}(A_jC_j) & \mathrm{tr}(C_j^2)\end{pmatrix} \tag{B.24}$$

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

$$M_n^{(1)} = Z^{(1)} + (1/2)\sum_j \mathrm{tr}(A_j) \text{ and } M_n^{(2)} = Z^{(2)} + (1/2)\sum_j \mathrm{tr}(C_j). \tag{B.25}$$

If $\langle\cdot,\cdot\rangle^{\mathcal{G}}$ represents discrete time predictable quadratic variation on the grid $\mathcal{G}$, then equation (B.24) yields

$$\begin{pmatrix}\langle M_n^{(1)}, M_n^{(1)}\rangle^{\mathcal{G}} & \langle M_n^{(1)}, M_n^{(2)}\rangle^{\mathcal{G}} \\ \langle M_n^{(1)}, M_n^{(2)}\rangle^{\mathcal{G}} & \langle M_n^{(2)}, M_n^{(2)}\rangle^{\mathcal{G}}\end{pmatrix} = \frac{1}{2}\sum_j \begin{pmatrix}\mathrm{tr}(A_j^2) & \mathrm{tr}(A_jC_j) \\ \mathrm{tr}(A_jC_j) & \mathrm{tr}(C_j^2)\end{pmatrix}. \tag{B.26}$$

PROOF OF THEOREM 3. First note that, by analogy to the development in Zhang (2001),

Mykland and Zhang (2006), Zhang, Mykland, and Aït-Sahalia (2005), and Zhang (2006),

$$\begin{aligned}\langle M_n^{(1)}, M_n^{(1)}\rangle^{\mathcal{G}} &= \frac{1}{2}\sum_j \mathrm{tr}(\zeta^{-2}_{\tau_{n,i-1}}(\zeta_{t_{n,j}} - \zeta_{\tau_{n,i-1}})^2) \\ &= \frac{1}{2}\sum_j \mathrm{tr}(\zeta^{-2}_{\tau_{n,i-1}}(\langle\zeta,\zeta\rangle_{t_{n,j}} - \langle\zeta,\zeta\rangle_{\tau_{n,i-1}})) + o_p(1) \\ &= \frac{1}{2}\sum_j \mathrm{tr}(\zeta^{-2}_{\tau_{n,i-1}}\langle\zeta,\zeta\rangle'_{\tau_{n,i-1}})(t_{n,j} - \tau_{n,i-1}) + o_p(1) \\ &= \frac{1}{2}\int_0^T \mathrm{tr}(\zeta_t^{-2}\langle\zeta,\zeta\rangle'_t)dK(t) + o_p(1) \\ &= \Gamma_1 + o_p(1), \end{aligned} \tag{B.27}$$

where K is the ADD given by equation (42).

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

$$\sup_j \mathrm{tr}(A_j^2) \to 0 \text{ as } n \to \infty. \tag{B.28}$$

Since also,

$$\text{for } r > 2,\ |\mathrm{tr}(A_j^r)| \le \mathrm{tr}(A_j^2)^{r/2}, \tag{B.29}$$

it follows that

$$\begin{aligned}\log\frac{dQ_n}{dP_n^*} &= Z_n^{(1)} + \frac{1}{2}\sum_i \sum_{t_{n,j}\in(\tau_{n,i-1},\tau_{n,i}]} (\log\det\zeta_{t_{n,j}} - \log\det\zeta_{\tau_{n,i-1}}) \\ &= Z_n^{(1)} + \frac{1}{2}\sum_j \log\det(I + A_j) \\ &= Z_n^{(1)} + \frac{1}{2}\sum_j (\mathrm{tr}(A_j) - \mathrm{tr}(A_j^2)/2 + \mathrm{tr}(A_j^3)/3 + ...) \\ &= M_n^{(1)} - \frac{1}{4}\sum_j \mathrm{tr}(A_j^2) + \frac{1}{6}\sum_j \mathrm{tr}(A_j^3) + ... \\ &= M_n^{(1)} - \frac{1}{2}\langle M_n^{(1)}, M_n^{(1)}\rangle^{\mathcal{G}} + o_p(1) \end{aligned} \tag{B.30}$$

At this point, let $\langle M_n^{(1)}, M_n^{(1)}\rangle$ be the quadratic variation of the continuous martingale that coincides at points $t_{n,j}$ with the discrete time martingale leading up to the end point $M_n^{(1)}$. By a standard quarticity argument (as in the proof of Remark 2 in Mykland and Zhang (2006)), (B.27)-(B.29) and the conditional normality of $\Delta Z^{(1)}_{n,t_{n,j+1}}$ 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.

PROOF OF THEOREM 6. The joint stable convergence of $M_n^{(1)}$ and $M_n^{(1)}$ follows analogously to the proof of Theorem 3 by referring to (B.26) and observing that, by (B.27)-(B.29),

$$\begin{aligned}\sum_j \mathrm{tr}(C_j^2) &= \sum_j \mathrm{tr}(I_p)B_j^2 + \sum_j \mathrm{tr}(A_j)B_j^2 + \sum_j \mathrm{tr}(A_j^2)B_j^2 \\ &= p\sum_j B_j^2 + o_p(1) \\ &= 2\Gamma_2 + o_p(1), \end{aligned} \tag{B.31}$$

and

$$\begin{aligned}\sum_j \mathrm{tr}(A_jC_j) &= \sum_j \mathrm{tr}(A_j)B_j + \sum_j \mathrm{tr}(A_j^2)B_j \\ &= \sum_j \mathrm{tr}(A_j)B_j + o_p(1) \\ &= o_p(1) \end{aligned} \tag{B.32}$$

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

$$\log\frac{dR_n}{dP_n^*} = \log\frac{dR_n}{dQ_n} + \log\frac{dQ_n}{dP_n^*}, \tag{B.33}$$

we obtain similarly that

$$\begin{aligned}\log\frac{dR_n}{dQ_n} &= Z_n^{(2)} + \frac{p}{2}\sum_j \log(1+B_j) \\ &= M_n^{(2)} + \Gamma_3 + o_p(1), \end{aligned} \tag{B.34}$$

whence the Theorem is proved. ∎