

# A Gaussian Calculus for Inference from High Frequency Data \*

Per A. Mykland

First version: February 15, 2006.

This version: September 29, 2007

## Abstract

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

*Some key words and phrases:* consistency, cumulants, contiguity, continuity, discrete observation, efficiency, Itô process, likelihood inference, realized volatility, stable convergence

---

\*Per Mykland is Professor, Department of Statistics, University of Chicago, Chicago, IL 60637. E-mail: mykland@pascal.uchicago.edu. I would like to thank Mark Podolskij for suggesting Example 3. I would also like to thank the referees, who helped greatly to improve the paper. The support of the National Science Foundation under grants DMS 06-04758 and SES 06-31605 is gratefully acknowledged.

## 1. Introduction

Recent years have seen an explosion of literature in the area of estimating volatility on the basis of high frequency data. The concepts go back to stochastic calculus, 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 give the mixed normal distribution of the error in approximation. References include Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002), Zhang (2001) and Mykland and Zhang (2006).

Further econometric literature includes, in particular, Gallant, Hsu, and Tauchen (1999), Chernov and Ghysels (2000), Andersen, Bollerslev, Diebold, and Labys (2001, 2003), Dacorogna, Gençay, Müller, Olsen, and Pictet (2001), Oomen (2004) and Goncalves and Meddahi (2005). Problems that are attached to the estimation of covariations between two processes are discussed in Hayashi and Yoshida (2005). 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 that happens in the presence of observation error, but we are not planning to address this question here.

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

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

As applications of the theory, we consider first (Section 5.1) the problem of estimating quarticity. We shall see that one can improve somewhat on the efficiency of the Barndorff-Nielsen and Shephard (2002) estimator, though the latter also holds up pretty well. Then, in Section 5.2 we shall revisit the ANOVA problem from Zhang (2001) and Mykland and Zhang (2006), this time in the setting of several regressors. We shall see that with the theory in hand, one can use classical regression theory to carry out an ANOVA. In fact, the amount of smoothing needed is over a *finite* number of observations. This is a proposition which, I think, would have attracted long odds if not announced as a theorem, and it is evidence of the power of the theorems from Section 4. Finally, we discuss the

problem of setting error bars on the Hayashi-Yoshida estimator of quadratic covariation (Section 5.3).

Some disclaimers. First of all, we are not claiming to have invented the conditional Gaussian assumption; it is used by a big fraction of the theory. Second, we do not claim that it can solve all problems; for example, it does not help if one wishes to estimate the leverage effect. Third, the results in the conditionally Gaussian case may not carry over to the general case, they only provide suggestions for what may occur. We *do*, however, assume that the quadratic variations are themselves random processes, to avoid excessive oversimplification.

Finally, this is not an attempt at a comprehensive study of what one can do with the Gaussian case; there are much too many tools available and open problems for that. In particular, we do not consider the case where observations have error. Our hope is that this study will encourage further use in high frequency data of the ideas and results that are available for Gaussian situations.

## 2. The Model, and Some Immediate Conclusions.

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

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

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

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

$$N = n_{N,1} + \dots + n_{N,p}. \quad (2)$$

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

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

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

We now describe the setting for parametric inference.

**Assumption 2.** (*A conditionally Gaussian system.*) *We let  $P$  be a probability distribution on the form (1) for which  $\mu_t^{(k)} = 0$  for all  $k$ . We assume that we can take the quadratic variations and*

covariations  $\langle X^{(k)}, X^{(l)} \rangle_t$  to be  $\mathcal{F}_0$ -measurable. As is customary, in this case, we call  $(X_t^{(1)}, \dots, X_t^{(p)})$  a Brownian martingale. Note that “nonrandom” is a special case of “ $\mathcal{F}_0$ -measurable”. We let  $P_\omega$  denote the regular conditional probability distribution given  $\mathcal{F}_0$ , and note that the  $(X_t)$  process is Gaussian under  $P_\omega$ , for (almost) every  $\omega$ . We also suppose that the observation times  $t_{k,i}$  are nonrandom, but irregular.

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

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

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

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

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

The log likelihood is then, as usual, given by

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

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

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

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

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

(cf., for example, the derivation in Chapter 4 of Mardia, Kent, and Bibby (1979)). Thus, two immediate conclusions.

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

$$\langle \widehat{X^{(k)}}, \widehat{X^{(k)}} \rangle_T = \sum_i \hat{\kappa}_{k,i;k,i} = \sum_i (X_{t_{k,i}}^{(k)} - X_{t_{k,i-1}}^{(k)})^2. \quad (7)$$

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

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

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

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

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

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

### 3. Warm-Up: The Quantification of Error in the Estimators.

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

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

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

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

Again, there are some immediate consequences of this.

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

$$\begin{aligned} \text{Var}_\omega \left( \langle \widehat{X^{(k)}}, \widehat{X^{(k)}} \rangle_{\mathbb{T}} \right) &= 2 \sum_i \kappa_{k,i;k,i}^2 \\ &= 2 \sum_i \left( E_\omega (X_{t_{k,i}}^{(k)} - X_{t_{k,i-1}}^{(k)})^2 \right)^2 \\ &= \frac{2}{3} \sum_i E_\omega \left( (X_{t_{k,i}}^{(k)} - X_{t_{k,i-1}}^{(k)})^4 \right), \end{aligned} \quad (10)$$

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

$$\frac{2}{3} \sum_i (X_{t_{k,i}}^{(k)} - X_{t_{k,i-1}}^{(k)})^4. \quad (11)$$

This, of course, is the quarticity estimate of Barndorff-Nielsen and Shephard (2002). Again, the estimator follows from a simple conditionally Gaussian likelihood. Its asymptotic validity has been shown under more general Itô process assumptions by Barndorff-Nielsen and Shephard (2002) and Mykland and Zhang (2006) (in the latter, see Remark 2 (p. 16) and the proof on p. 25).

We shall show in Section 5.1, however, that this may not be the best estimator available. This will be one of the applications of our theory in Section 4 below.

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

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

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

where

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

A perhaps slightly more interpretable expression is to write

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

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

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

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

so that even if  $t_{k,i} = t_{l,i}$  (which is to say that there is no asynchronicity, so that the second term in (14) vanishes), there is no directly obtainable estimate of the quantity in (14). We shall now see how this type of issue can be remedied. We return to the problem of the Hayashi-Yoshida estimator in Section 5.3.

## 4. Main Results: Locally Constant Volatility

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

### 4.1. Setup

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

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

and

$$v_N = O(N). \quad (18)$$

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

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

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

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

We shall need the quadratic variation of the above set of time points: the ‘‘Asymptotic Quadratic Variation of Time’’ (‘‘AQVT’’)  $H(t)$  is defined by

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

provided the limit exists. From (3),

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

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

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

The assumption is more restrictive than the one made in Mykland and Zhang (2006), but is still quite broad. An easy extension can make  $G$  random, covering, in particular, the case of Section 5.1 of Barndorff-Nielsen, Hansen, Lunde, and Shephard (2006).

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

$$\nu_{k,l,i} = \text{Cov}_\omega(X_{\theta_{N,i}}^{(k)} - X_{\theta_{N,i-1}}^{(k)}, X_{\theta_{N,i}}^{(l)} - X_{\theta_{N,i-1}}^{(l)}). \quad (21)$$

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

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

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

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

$$\log \frac{dP}{dP^N} = \frac{1}{2} \sum_{\iota} \sum_{\theta_{N,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}]} \left( \log \det(\bar{\nu}_\iota) - \log \det(\nu_j) - \Delta X_{\theta_j}^* (\nu_j^{-1} - \bar{\nu}_\iota^{-1}) \Delta X_{\theta_j} \right). \quad (23)$$

Note that

$$\log \frac{dP}{dP^N} = \log \frac{dP_\omega}{dP_\omega^N}. \quad (24)$$

## 4.2. Main Contiguity Theorem

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

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

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

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

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

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

$$\log \frac{dP}{dP^N} = Z_N + \frac{1}{2}\Gamma_{ZZ} + o_p(1). \quad (27)$$

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

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

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

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

$$\begin{aligned} \Gamma_{\xi\xi} &= \lim_{N \rightarrow \infty} \text{Var}_\omega^N(N^{1/2}(\hat{\xi}_N - \xi_N)) \\ \text{and } \Gamma_{\xi Z} &= \lim_{N \rightarrow \infty} \text{Cov}_\omega^N(N^{1/2}(\hat{\xi}_N - \xi_N), Z_N). \end{aligned} \quad (28)$$

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

$$\lim_{N \rightarrow \infty} E_\omega^N \sum_\iota g(N^{1/2}(\hat{\xi}_{N,\iota} - \xi_{N,\iota})) = 0, \quad (29)$$

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

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

in law, for almost every  $\omega$ .

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

$$\Gamma_{\xi\xi} = \lim_{N \rightarrow \infty} N \sum_{\iota=1}^{v_N} \text{Var}_{N\omega}(\hat{\xi}_{N,\iota}), \quad (31)$$

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

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

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

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

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

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

### 4.3. When the drift $\mu$ is not zero

We first consider consistency.

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

Under Assumptions 3 and 5, Girsanov's Theorem (see, for example, Chapter 5.5 of Karatzas and Shreve (1991)) yields that there is a probability measure  $P$  satisfying Assumption 2 so that for almost all  $\omega$

$$Q_{\omega} \text{ and } P_{\omega} \text{ are mutually absolutely continuous.} \quad (32)$$

This means that consistency holds under  $Q_{\omega}$  if and only if it holds under  $P_{\omega}$ . Thus, for instance, the final sentence in Theorem 3 remains true also under  $Q_{\omega}$ . (One still verifies the conditions under  $P$  and  $P^N$ ).

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

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

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

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

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

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

Specifically, we obtain

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

PROOF OF THEOREM 1. The two parts of the proof are similar since the main fact we use about  $Z_N$  is that

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

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

From Theorem 4 it follows directly that

**Theorem 5.** *Assume the conditions of Theorem 2, but replace Assumption 2 by Assumption 5. Also assume (33). Then the (stable) convergence in law (30) holds also under  $Q_\omega$ , for almost every  $\omega$ .*

The trick to both show and understand the theorem is to note that for every bounded and continuous  $g$ , and for the indicator function  $I_A$  of every  $A \in \mathcal{F}_T$ ,

$$\begin{aligned} E_\omega^Q I_{Ag} \left( N^{1/2}(\hat{\xi}_N - \xi_N) \right) &= E_\omega^P \frac{dQ_\omega}{dP_\omega} I_{Ag} \left( N^{1/2}(\hat{\xi}_N - \xi_N) \right) \\ &\rightarrow E_\omega^P \frac{dQ_\omega}{dP_\omega} I_{Ag} (N(\Gamma_{\xi Z}, \Gamma_{\xi\xi})) \\ &= E_\omega^Q I_{Ag} (N(\Gamma_{\xi Z}, \Gamma_{\xi\xi})), \end{aligned} \quad (35)$$

by uniform integrability, whence Theorem 5 is shown.

Note that the above description is mainly a spelling out of the statement in the second paragraph of the proof of Theorem 2 in Zhang, Mykland, and Ait-Sahalia (2005) (p. 1410). The original idea appears to go back to Rootzén (1980).

**Remark 3.** *In the more general case where the limit in (33) can be nonzero, the stable convergence in Theorem 2 will be to a normal limit with smaller variance but extra bias, subject to suitable regularity conditions. We have not further discussed this case since such estimators are clearly inefficient. An example is given in the following. Note that in this more general case, Theorem 5 continues to hold, but one has to be careful to interpret the result in a very precise manner, cf. also here the following example.*

**Example 3.** *(A Technical Example). We shall see examples of application of the above results in the following Section 5. In this example, we shall see a technical illustration of the theory which will serve to put our theorems in perspective. Assume that we are dealing with a scalar process ( $p = 1$ ), and take  $N = n_1 (= v_N M)$ . We also suppose that all time increments are equidistant:  $t_{N,i+1} - t_{N,i} = \Delta t = T/N$ . In this setting,  $\hat{\xi}_N = \sqrt{N} \sum_i (X_{t_{N,i+1}} - X_{t_{N,i}})^3$  is a consistent estimator of  $\xi = 0$  (yes, it is a silly example). The estimator satisfies the conditions of Theorem 2, but  $\Gamma_{ZW^{(1)}}$  is nonzero. In fact, it is easy to see that  $\sqrt{N}(\hat{\xi}_N - \xi)$  converges stably (under  $P$  and  $P_\omega$ ) to a normal distribution with mean  $3T \int_0^T \sigma_t^3 dW_t$  and variance  $6T^2 \int_0^T \sigma_t^6 dt$ . In the non-stable case (Theorem 2) this translates into a normal distribution with mean zero and variance  $15T^2 \int_0^T \sigma_t^6 dt$ .*

In relation to Theorem 5, suppose that under  $Q$ , the process  $X_t$  has the form  $dX_t = \mu_t dt + \sigma_t dV_t$ , with  $V_t$  as a Brownian motion. As  $W_t$  is a  $P$ -Brownian motion, one obtains the relationship  $dW_t = (\mu_t/\sigma_t)dt + dV_t$ . The stated limit under  $P$  remains true under  $Q$ , but one should note that the bias takes the form

$$3T \int_0^T \sigma_t^3 dW_t = 3T \int_0^T \sigma_t^3 [(\mu_t/\sigma_t)dt + dV_t]. \quad (36)$$

Thus Theorem 5 remains valid at the same time as the limit depends on  $\mu$ !

## 5. Some Applications of the Theory

### 5.1. Estimating Quarticity

We now return to Example 1. The process  $X_t$  is observed at times  $0 = t_{N,0} < t_{N,1} < \dots < t_{N,n_1} = T$ . We take  $N = n_1$ .

Our object is to estimate the variance of  $\widehat{\langle X, X \rangle}_T$ . By the theory in Jacod and Protter (1998), Barndorff-Nielsen and Shephard (2002), and Mykland and Zhang (2006),  $N^{1/2} \left( \widehat{\langle X, X \rangle}_T - \langle X, X \rangle_T \right)$  is asymptotically normal under  $P_\omega$ , with mean zero and variance

$$\xi = 2T \int_0^T \sigma_t^4 dH(t), \quad (37)$$

where  $H(t)$  is the AQVT (equation (19)).

To form our grid points, we take the  $\theta_i$ 's to be identical with the  $t_i$ 's. For simplicity of argument, we assume that  $N$  is a multiple of  $M$ . We now describe the system under  $P^N$ . In each of  $v_N$  intervals  $[\tau_{l-1}, \tau_l]$ , we get  $M$  iid observations of  $\Delta X_{t_j}$ , which are normal with mean zero and variance  $\bar{\nu}_l$ .

Under  $P^N$ , the variance we are looking for is

$$\xi_N = N \text{Var}_\omega^N \left( \widehat{\langle X, X \rangle}_T \right) = 2NM \sum_{l=1}^{v_N} \bar{\nu}_l^2, \quad (38)$$

Up to high order, it is enough to estimate this quantity, since

$$\begin{aligned} \xi_N &= \frac{2N}{M} \sum_{i=1}^{v_N} \left( \int_{\tau_{l-1}}^{\tau_l} \sigma_t^2 dt \right)^2 \\ &= \xi + o_p(N^{-1/2}) \end{aligned} \quad (39)$$

under Assumption 3.

The question under  $P^N$  is how to estimate  $\bar{\nu}_l^2$  from the  $M$  iid observations of  $\Delta X_{t_j}$  ( $t_j \in (\tau_{l-1}, \tau_l]$ ). Both likelihood theory and sufficiency suggests that the estimator must be of the form constant  $\times \hat{\nu}_l^2$ , where

$$\hat{\nu}_l = \frac{1}{M} \sum_{t_j \in (\tau_{l-1}, \tau_l]} \Delta X_{t_j}^2. \quad (40)$$

In fact, following Example 2.2.1 (p. 84-85) in Lehmann (1983), the estimator

$$\frac{M}{M+2} \hat{\nu}_l^2 \quad (41)$$

is a  $P_\omega^N$ -unbiased minimum variance (UMVU) estimator for  $\bar{\nu}_l^2$ . Our proposed estimator for  $N \text{Var}_\omega(\widehat{\langle X, X \rangle}_T)$  is therefore

$$\hat{\xi}_N = N \widehat{\text{Var}}(\widehat{\langle X, X \rangle}_T) = \frac{2NM^2}{M+2} \sum_{i=1}^{v_N} \hat{\nu}_l^2. \quad (42)$$

This estimator is UMVU under  $P_\omega^N$ . Note that for  $M = 1$ , our estimator coincides with the estimator from Barndorff-Nielsen and Shephard (2002), cf. (our) equation (11).

To further assess the quality of our estimator (42), note that

$$\begin{aligned}
\text{Var}_\omega^N(\hat{\xi}_N) &= \left(\frac{2NM^2}{M+2}\right)^2 \sum_{\iota=1}^{v_N} \text{Var}_\omega^N\{\hat{v}_\iota^2\}. \\
&= \left(\frac{2NM^2}{M+2}\right)^2 \frac{1}{M^4} \text{Var}((\chi_M^2)^2) \sum_{\iota=1}^{v_N} \bar{v}_\iota^4 \\
&= \frac{4N^2(M+3)M}{M+2} \sum_{\iota=1}^{v_N} \bar{v}_\iota^4 \\
&= \frac{4N^2(M+3)M}{M+2} \frac{1}{M} \left(\frac{T}{N}\right)^3 \int_0^T \sigma_t^8 dH(t) + o_p(N^{-1}) \\
&= \frac{4(M+3)T^3}{M+2} \frac{1}{N} \int_0^T \sigma_t^8 dH(t) + o_p(N^{-1})
\end{aligned} \tag{43}$$

by using the methodology from Section 5 of Zhang (2006).

To apply Theorem 2, we see by the same methods as above (cf. also the proof of Theorem 1) that  $\text{Cov}_\omega^N\left(Z_N, N^{1/2}(\hat{\xi}_N - \xi_N)\right) = o_p(N^{-1/2})$ , and that the Lindeberg condition is obviously satisfied. Hence we have obtained

**Theorem 6.** *Assume the conditions of Theorem 1. Then, under  $P_\omega$ ,  $N^{1/2}(\hat{\xi}_N - \xi)$  converges in law to a normal distribution with mean zero and variance*

$$\frac{4(M+3)}{M+2} T^3 \int_0^T \sigma_t^8 dH(t). \tag{44}$$

Several comments are in order. First of all, picking  $M > 1$  does improve on the estimator from Barndorff-Nielsen and Shephard (2002) (which corresponds to  $M = 1$ ). The improvement, however, is not huge: the variance is reduced by up to 25 % as  $M \rightarrow \infty$ . In other words, the Barndorff-Nielsen and Shephard (2002) estimator emerges relatively well from the comparison.

There are some open questions here. One is that we here use a fixed starting point for dividing the observations into groups of size  $M$ . One should, of course, start group 1 at points 0 up to  $M - 1$ , and then average across the results. By a sufficiency argument, this will improve the performance of the estimator. The second is that it looks like it would be optimal to let  $M \rightarrow \infty$  with  $N$ . This would typically induce bias, and there is probably a bias-variance tradeoff to be explored here.

Finally, note that setting errors for estimators is not necessarily the same as estimating the size of the error. This issue was the underlying problem in the observed-vs.-estimated-expected-information debate, which went back to Fisher and Neyman, see, for example, Cox (1958) and Efron and Hinkley (1978). The debate created a slew of important results in the 1980s and 1990s, such

as Barndorff-Nielsen (1986, 1991), Jensen (1992, 1995, 1997), McCullagh (1984, 1987), McCullagh and Tibshirani (1990), Pierce and Peters (1994), Reid (1988), Skovgaard (1986, 1991), and many others. (For those who are familiar with the area: thanks for all the fun, and apologies for not citing everyone!) In other words, showing that  $\widehat{\text{Var}}(\langle X, X \rangle_T)$  has low variance is not the same as showing that it is a good statistic with which to set a confidence interval.

## 5.2. ANOVA with Multiple Regression and Finite Smoothing

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

$$dY_t = \sum_{i=1}^p f_s^{(i)} dX_s^{(i)} + dZ_t, \text{ with } \langle X^{(i)}, Z \rangle_t = 0 \text{ for all } t \text{ and } i. \quad (45)$$

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

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

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

$$\bar{\nu}_{ZZ, \tau_l} = E_\omega^N \Delta Z_{t_j}^2 = \bar{\nu}_{YY, \tau_l} - \bar{\nu}_{XY, \tau_l}^* \bar{\nu}_{XX, \tau_l}^{-1} \bar{\nu}_{XY, \tau_l} \quad (46)$$

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

$$\begin{aligned} \hat{\xi}_{\tau_l, N} &= \frac{M}{M-p} \sum_{t_j \in (\tau_{l-1}, \tau_l]} \widehat{\Delta Z}_{t_j}^2 \\ \text{and } \xi_{\tau_l, N} &= E_\omega^N \left( \sum_{t_j \in (\tau_{l-1}, \tau_l]} \Delta Z_{t_j}^2 \right) \end{aligned} \quad (47)$$

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

$$E_\omega^N \left( \hat{\xi}_{\tau_l, N} \mid \Delta X_{t_j}, \text{ all } j = 1, \dots, n_1 \right) = \xi_{\tau_l, N}, \quad (48)$$

and, in fact, more generally,

$$(M-p)\frac{\hat{\xi}_{\iota,N}}{\xi_{\iota,N}}, \iota = 1, \dots, v_N \text{ are iid } \chi_{M-p}^2 \text{ under } P_\omega^N \text{ given all } \Delta X_{t_j}, j = 1, \dots, n_1. \quad (49)$$

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

$$\widehat{\langle Z, Z \rangle}_T = \frac{M}{M-p} \sum_{j=1}^{n_1} \widehat{\Delta Z}_{t_j}^2 = \sum_{\iota=1}^{v_N} \hat{\xi}_{\iota,N}. \quad (50)$$

From (46) and (48), and by Assumption 3,

$$\begin{aligned} E_\omega^N(\widehat{\langle Z, Z \rangle}_T) &= E_\omega^N \left( \sum_{j=1}^{n_1} \Delta Z_{t_j}^2 \right) \\ &= \int_0^T \zeta_t^{(YY)} dt - \sum_{\iota=1}^{v_N} \left( \int_{\tau_{\iota-1}}^{\tau_\iota} \zeta_t^{(XY)} dt \right)^* \left( \int_{\tau_{\iota-1}}^{\tau_\iota} \zeta_t^{(XX)} dt \right)^{-1} \left( \int_{\tau_{\iota-1}}^{\tau_\iota} \zeta_t^{(XY)} dt \right) \\ &= \int_0^T \zeta_t^{(YY)} dt - \frac{T}{v_N} \sum_{\iota=1}^{v_N} \left( \zeta_{\tau_{\iota-1}}^{(XY)} + \frac{v_N}{T} \int_{\tau_{\iota-1}}^{\tau_\iota} (\tau_\iota - t) d\zeta_t^{(XY)} \right)^* \\ &\quad \times \left( (\zeta_{\tau_{\iota-1}}^{(XX)})^{-1} - \frac{v_N}{T} (\zeta_{\tau_{\iota-1}}^{(XY)})^* \int_{\tau_{\iota-1}}^{\tau_\iota} (\tau_\iota - t) d\zeta_t^{(XX)} \zeta_{\tau_{\iota-1}}^{(XY)} \right) \left( \zeta_{\tau_{\iota-1}}^{(XY)} + \frac{v_N}{T} \int_{\tau_{\iota-1}}^{\tau_\iota} (\tau_\iota - t) d\zeta_t^{(XY)} \right) \\ &\quad + o_p(N^{-1/2}) \\ &= \langle Z, Z \rangle_T + o_p(N^{-1/2}). \end{aligned} \quad (51)$$

Similarly, from (48)-(49),

$$\begin{aligned} \text{Var}_\omega^N(\widehat{\langle Z, Z \rangle}_T) &= \frac{2}{M-p} \sum_{\iota=1}^{v_N} \xi_{\iota,N}^2 \\ &= n_1^{-1} 2 \frac{M}{M-p} \int_0^T (\langle Z, Z \rangle'_t)^2 dt + o_p(n_1^{-1}) \end{aligned} \quad (52)$$

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

To match notation with Theorem 2, finally let  $X^{(p+1)} = Y$ . Obtain from (48)-(49) and the

normality of the observations that

$$\begin{aligned}
& \text{Cov}_\omega^N \left( \sum_{t_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } \Delta X_{t_j}^* (\nu_j^{-1} - \bar{\nu}_\ell^{-1}) \Delta X_{t_j}, \hat{\xi}_{N,\ell} \mid \mathcal{X} \right) \\
&= \text{Cov}_\omega^N \left( \sum_{t_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } \Delta Z_{t_j} (\nu_j^{-1} - \bar{\nu}_\ell^{-1})^{(YY)} \Delta Z_{t_j}, \hat{\xi}_{N,\ell} \mid \mathcal{X} \right) \\
&= \frac{M}{M-p} \sum_{t_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } \text{Cov}_\omega^N \left( (\nu_j^{-1} - \bar{\nu}_\ell^{-1})^{(YY)} \Delta Z_{t_j}^2, (1 - h_j) \Delta Z_{t_j}^2 \right) \\
&= 2 \frac{M}{M-p} \bar{\nu}_{ZZ,\ell}^2 \sum_{t_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } (\nu_j^{-1} - \bar{\nu}_\ell^{-1})^{(YY)} (1 - h_j). \tag{53}
\end{aligned}$$

At this point, note that by symmetry,  $E_\omega^N(1 - h_j)$  is independent of  $j$ , and so  $E_\omega^N(1 - h_j) = E_\omega^N \text{tr}(I - H)/M = (M - p)/M$ . Thus, from (48) and (53),

$$\text{Cov}_\omega^N \left( \sum_{t_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } \Delta X_{t_j}^* (\nu_j^{-1} - \bar{\nu}_\ell^{-1}) \Delta X_{t_j}, \hat{\xi}_{N,\ell} \right) = 2 \bar{\nu}_{ZZ,\ell}^2 \sum_{t_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } (\nu_j^{-1} - \bar{\nu}_\ell^{-1})^{(YY)}. \tag{54}$$

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

$$\begin{aligned}
\text{Cov}_\omega^N(\hat{\xi}_N, Z_N) &= 2 \sum_{\ell} \bar{\nu}_{ZZ,\ell}^2 \sum_{t_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } (\nu_j^{-1} - \bar{\nu}_\ell^{-1})^{(YY)} \\
&= o_p(N^{-1/2}). \tag{55}
\end{aligned}$$

Finally, the Lindeberg condition in Theorem 2 is satisfied from (49). We have therefore shown the following corollary to this theorem:

**Theorem 7.** (*Multiple ANOVA in the Gaussian case*). Let  $\langle \widehat{Z}, \widehat{Z} \rangle_T$  be as defined in (50). Under the conditions of Theorem 1,

$$n_1^{1/2} (\langle \widehat{Z}, \widehat{Z} \rangle_T - \langle Z, Z \rangle_T) \tag{56}$$

converges in law under  $P_\omega$  to a normal distribution with mean zero and variance

$$2 \frac{M}{M-p} \int_0^T (\langle Z, Z \rangle_t')^2 dt. \tag{57}$$

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

$$2 \int_0^T (\langle Z, Z \rangle_t')^2 dt \tag{58}$$

(as in the single series case in Jacod and Protter (1998) and Barndorff-Nielsen and Shephard (2002)). Of course, the expression in (57) converges to that of (58) as  $M \rightarrow \infty$ . Comparing the two sets of results, one is lead to conjecture that there is a bias-variance tradeoff where  $M$  should go very slowly to infinity  $n_1$ , but exploring that is beyond the scope of this paper.

### 5.3. Estimating the Variability in the Hayashi-Yoshida estimator

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

$$\begin{aligned} \text{Var}_\omega \left( \langle \widehat{X}^{(k)}, \widehat{X}^{(l)} \rangle_{\mathbb{T}} \right) &= \sum_{i,j:(t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset} (\kappa_{k,i;k,i} \kappa_{l,j;l,j} + (\kappa_{k,i;l,j})^2) \\ &+ \sum_{(i,j,g,h) \in B_n} \kappa_{k,i;l,h} \kappa_{k,g;l,j}, \end{aligned} \quad (59)$$

where  $B_n$  is given by (15). We are seeking to consistently estimate this variance.

There is, obviously, a local likelihood estimate along the lines of Section 5.1. To provide some variety, we here provide a moment based estimate in the style of Barndorff-Nielsen and Shephard (2002). In this section, we only consider consistency of the estimate, though, of course, the same principles as before can be used to find asymptotic normality.

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

$$E_\omega(\hat{\kappa}_{k,i;l,h} \hat{\kappa}_{k,g;l,j}) = \kappa_{k,i;l,h} \kappa_{k,g;l,j} + \kappa_{k,i;l,j} \kappa_{k,g;l,h} + \kappa_{k,i;k,g} \kappa_{l,h;l,j} \quad (60)$$

Thus, if we set

$$\begin{aligned} \hat{V}_1 &= \sum_{i,j:(t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset} \hat{\kappa}_{k,i;k,i} \hat{\kappa}_{l,j;l,j} \\ \hat{V}_2 &= \sum_{(i,j,g,h) \in B_n} \hat{\kappa}_{k,i;l,h} \hat{\kappa}_{k,g;l,j}, \end{aligned} \quad (61)$$

we obtain

$$\begin{aligned} V_1 = E_\omega \hat{V}_1 &= \sum_{i,j:(t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset} (\kappa_{k,i;k,i} \kappa_{l,j;l,j} + 2(\kappa_{k,i;l,j})^2) \\ V_2 = E_\omega \hat{V}_2 &= 2 \sum_{(i,j,g,h) \in B_n} \kappa_{k,i;l,h} \kappa_{k,g;l,j}. \end{aligned} \quad (62)$$

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

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

Note that it is sufficient to obtain a consistent estimate for

$$V_3 = \sum_{i,j:(t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset} (\kappa_{k,i;l,j})^2 \quad (63)$$

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

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

$$\begin{aligned} D_{k,l} &= \{i : (t_{k,i-1}, t_{k,i}) \subseteq (\tau_{l-1}, \tau_l)\} \\ D_{k,l,l} &= \{(i, j) : (t_{k,i-1}, t_{k,i}) \cup (t_{l,j-1}, t_{l,j}) \subseteq (\tau_{l-1}, \tau_l) \text{ and } (t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset\} \\ M_{k,i} &= \# \text{ of intervals } (\theta_{N,h-1}, \theta_{N,h}) \subseteq (t_{k,i-1}, t_{k,i}) \\ M_{k,i;l,j} &= \# \text{ of intervals } (\theta_{N,h-1}, \theta_{N,h}) \subseteq (t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \\ m_{k,l} &= \sum_{i \in D_{k,l}} M_{k,i} \\ m_{k,l,l} &= \sum_{(i,j) \in D_{k,l,l}} M_{k,i;l,j}. \end{aligned} \quad (64)$$

Now note that under the approximate measure  $P^N$

$$\begin{aligned} E_\omega^N \sum_{(i,j) \in (D_{k,l} \times D_{l,l} - D_{k,l,l})} \hat{\kappa}_{k,i;k,i} \hat{\kappa}_{l,j;l,j} &= \bar{v}_{k,l,l}^2 \sum_{(i,j) \in (D_{k,l} \times D_{l,l} - D_{k,l,l})} M(k, i; k, i) M(l, j; l, j) \\ &= \bar{v}_{k,l,l}^2 (m_{k,l} m_{l,l} - m_{k,l,l}). \end{aligned} \quad (65)$$

On the other hand, set

$$\begin{aligned} \bar{D}_{k,l,l} &= \{(i, j) : t_{k,i-1}, t_{l,j-1} \in [\tau_{l-1}, \tau_l) \text{ and } (t_{k,i-1}, t_{k,i}) \cap (t_{l,j-1}, t_{l,j}) \neq \emptyset\} \\ \bar{m}_{k,l,l}^{(2)} &= \sum_{(i,j) \in \bar{D}_{k,l,l}} M_{k,i;l,j}^2. \end{aligned} \quad (66)$$

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

$$V_3 = \sum_l \bar{v}_{k,l,l}^2 \bar{m}_{k,l,l}^{(2)} + o(N^{-1}). \quad (67)$$

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

$$\hat{V}_3 = \sum_l \frac{\bar{m}_{k,l,l}^{(2)}}{m_{k,l} m_{l,l} - m_{k,l,l}} \sum_{(i,j) \in (D_{k,l} \times D_{l,l} - D_{k,l,l})} \hat{\kappa}_{k,i;k,i} \hat{\kappa}_{l,j;l,j}, \quad (68)$$

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

Take the grid  $\mathcal{G}_N = \{\theta_{N,j}\}$  to consist of all points of the form  $t_{k,i}$  or  $t_{l,j}$ . Thus  $w_N \leq n_k + n_l$ . Fix  $M$  as a sufficiently large integer (more about the exact condition below). We let  $\tau_1$  be one of the first  $M$  points of  $\mathcal{G}_N$ , and then let the following  $\tau_l$  be every  $M$ 'th point in  $\mathcal{G}_N$ . (We can without impacting the asymptotics ignore the first and last interval, which may not have  $M$  points).

One can therefore, finally, use the estimate

$$\hat{\text{Var}}_\omega \left( \langle \widehat{X^{(k)}}, \widehat{X^{(l)}} \rangle_T \right) = \hat{V}_1 + \frac{1}{2} \hat{V}_2 - \hat{V}_3. \quad (69)$$

As an example of grid, one can let  $\mathcal{G}_N = \{\theta_{N,j}\}$  consist of all points of the form  $t_{k,i}$  or  $t_{l,j}$ . Thus  $w_N \leq n_k + n_l$ . Fix  $M$  as a sufficiently large integer. We let  $\tau_1$  be one of the first  $M$  points of  $\mathcal{G}_N$ , and then let the following  $\tau_l$  be every  $M$ 'th point in  $\mathcal{G}_N$ . (We can without impacting the asymptotics ignore the first and last interval, which may not have  $M$  points). In line with general principles of sufficiency, one should, of course, average  $\hat{V}_3$  over the  $M$  choices of initial point  $\tau_{N,1}$ . This does not, of course, affect consistency. Otherwise, the same comments as in Section 5.1 apply.

## 6. Proof of Theorem 1

PROOF OF THEOREM 1. Set  $U_j = \bar{\nu}_l^{-1/2} \nu_j^{-1} \bar{\nu}_l^{1/2} - I$ , where  $\bar{\nu}_l^{1/2}$  is the symmetric square root of  $\bar{\nu}_l$ . By standard Gaussian arguments,

$$\begin{aligned} E_\omega^N(\Delta X_{\theta_j}^* (\nu_j^{-1} - \bar{\nu}_l^{-1}) \Delta X_{\theta_j}) &= \text{tr}(U_j) \\ \text{and } \text{Var}_\omega^N(\Delta X_{\theta_j}^* (\nu_j^{-1} - \bar{\nu}_l^{-1}) \Delta X_{\theta_j}) &= 2\text{tr}(U_j^2). \end{aligned} \quad (70)$$

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

$$\text{tr}(U_j^r) = \text{tr}((-\bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})^r) + r \text{tr}((-\bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})^{r+1}) + \dots \quad (71)$$

For  $r \geq 2$ ,

$$|\text{tr}((-\bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})^r)| \leq |\text{tr}((-\bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})^2)|^{r/2} \quad (72)$$

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

$$\text{tr}((-\bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})^2) = \sum_{r,s} ((-\bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})_{rs})^2,$$

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

$$\sum_j \text{tr}((-\bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})^2) = 2\Gamma_{ZZ} + o_p(1). \quad (73)$$

Meanwhile, from (72), we obtain that for  $r > 2$ ,

$$\left| \sum_j \operatorname{tr}((- \bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})^r) \right| \leq \sup_j |\operatorname{tr}((- \bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})^2)|^{(r-2)/2} \sum_j \operatorname{tr}((- \bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})^2).$$

Thus (71) yields that

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

where  $f(x) = 2x^{1/2} + 3x + \dots = o(1)$  as  $x \rightarrow 0$ . Since  $\sup_j |\operatorname{tr}((- \bar{\nu}_l^{-1/2} V_j \bar{\nu}_l^{-1/2})^2)| = o_p(1)$ , we obtain from (73) that

$$\sum_j \operatorname{tr}(U_j^2) = 2\Gamma_{ZZ} + o_p(1). \quad (74)$$

Also, since

$$\sum_{\theta_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } \operatorname{tr}(\bar{\nu}_l^{-1} V_j) = 0 \quad (75)$$

by construction, we obtain

$$\begin{aligned} \sum_{\iota} \sum_{\theta_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } \operatorname{tr}(U_j) &= \sum_j \operatorname{tr}(U_j^2) + o_p(1) \\ &= 2\Gamma_{ZZ} + o_p(1) \end{aligned} \quad (76)$$

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

$$\begin{aligned} \frac{1}{2} \sum_{\iota} \sum_{\theta_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } (\log \det(\bar{\nu}_l) - \log \det(\nu_j)) &= \frac{1}{2} \sum_j \log \det(I + U_j) \\ &= \frac{1}{2} \sum_j (\operatorname{tr}(U_j) - \frac{1}{2} \operatorname{tr}(U_j^2)) + o_p(1) \\ &= \frac{1}{2} \Gamma_{ZZ} + o_p(1). \end{aligned} \quad (77)$$

This shows the Theorem. ■

It remains to prove

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

$$\begin{aligned} \sum_{\iota} \sum_{\theta_{N,j} \in (\tau_{N,\ell-1}, \tau_{N,\ell}] } f_{\tau_{N,\ell-1}}(\nu_{k,l,j} - \bar{\nu}_{k,l,\ell})(\nu_{m,n,j} - \bar{\nu}_{m,n,\ell}) \\ = \frac{1}{6} \frac{M-1}{M^2} T^2 v_N^{-2} \int_0^T f_t H'(t)^2 d\langle \zeta^{(k,l)}, \zeta^{(m,n)} \rangle_t + o_p(N^{-2}). \end{aligned} \quad (78)$$

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

First, suppose that

$$\sum_{j=1}^{w_N} \left( \theta_{N,j} - \theta_{N,j-1} - \frac{T}{w_N} \right)^2 = o(N^{-1}). \quad (79)$$

Write

$$\begin{aligned} \nu_{k,l,j} - \bar{\nu}_{k,l,\ell} &= \int_{\theta_{N,j-1}}^{\theta_{N,j}} \zeta_u^{(k,l)} du - \frac{1}{M} \int_{\tau_{N,\ell-1}}^{\tau_{N,\ell}} \zeta_u^{(k,l)} du \\ &= \int_{\theta_{N,j-1}}^{\theta_{N,j}} (\zeta_u^{(k,l)} - \zeta_{\theta_{N,j-1}}^{(k,l)}) du \\ &\quad + (\theta_{N,j} - \theta_{N,j-1}) (\zeta_{\theta_{N,j-1}}^{(k,l)} - \zeta_{\tau_{N,\ell-1}}^{(k,l)}) \\ &\quad - \frac{1}{M} \int_{\tau_{N,\ell-1}}^{\tau_{N,\ell}} (\zeta_u^{(k,l)} - \zeta_{\tau_{N,\ell-1}}^{(k,l)}) du \\ &\quad + \left( (\theta_{N,j} - \theta_{N,j-1}) - \frac{\tau_{N,\ell} - \tau_{N,\ell-1}}{M} \right) \zeta_{\tau_{N,\ell-1}}^{(k,l)} \\ &= \int_{\theta_{N,j-1}}^{\theta_{N,j}} (\theta_{N,j} - u) d\zeta_u^{(k,l)} \\ &\quad + (\theta_{N,j} - \theta_{N,j-1}) (\zeta_{\theta_{N,j-1}}^{(k,l)} - \zeta_{\tau_{N,\ell-1}}^{(k,l)}) \\ &\quad - \frac{1}{M} \int_{\tau_{N,\ell-1}}^{\tau_{N,\ell}} (\tau_{N,\ell} - u) d\zeta_u^{(k,l)} \\ &\quad + \left( (\theta_{N,j} - \theta_{N,j-1}) - \frac{\tau_{N,\ell} - \tau_{N,\ell-1}}{M} \right) \zeta_{\tau_{N,\ell-1}}^{(k,l)} \\ &= \int_{\tau_{N,\ell-1}}^{\tau_{N,\ell}} g_{N,j}(u) d\zeta_u^{(k,l)} \\ &\quad + \left( (\theta_{N,j} - \theta_{N,j-1}) - \frac{\tau_{N,\ell} - \tau_{N,\ell-1}}{M} \right) \zeta_{\tau_{N,\ell-1}}^{(k,l)} \end{aligned} \quad (80)$$

where

$$g_{N,j}(u) = I_{\{\theta_{N,j-1} < u \leq \theta_{N,j}\}} (\theta_{N,j} - u) + I_{\{\tau_{N,\ell-1} < u \leq \theta_{N,j-1}\}} (\theta_{N,j} - \theta_{N,j-1}) - \frac{1}{M} I_{\{\tau_{N,\ell-1} < u \leq \tau_{N,\ell}\}} (\tau_{N,\ell} - u).$$

The last term in (80) is ignorable in view of assumption (79). Thus,

$$\begin{aligned}
& \sum_{\iota} \sum_{\theta_{N,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}] } f_{\tau_{N,\iota-1}}(\nu_{k,l,j} - \bar{\nu}_{k,l,\iota})(\nu_{m,n,j} - \bar{\nu}_{m,n,\iota}) \\
&= \sum_{\iota} f_{\tau_{N,\iota-1}} \int_{\tau_{N,\iota-1}}^{\tau_{N,\iota}} \sum_{\theta_{N,j} \in (\tau_{N,\iota-1}, \tau_{N,\iota}] } g_{N,j}(u)^2 d\langle \zeta^{(k,l)}, \zeta^{(m,n)} \rangle_u + o_p(N^{-2}) \\
&= c_M M^{-2} T^2 v_N^{-2} \sum_{\iota} f_{\tau_{N,\iota-1}} \langle \zeta^{(k,l)}, \zeta^{(m,n)} \rangle'_{\tau_{N,\iota-1}} (\tau_{N,\iota} - \tau_{N,\iota-1}) + o_p(N^{-2}) \\
&= c_M M^{-2} T^2 v_N^{-2} \int_0^T f_t d\langle \zeta^{(k,l)}, \zeta^{(m,n)} \rangle_t + o_p(N^{-2}) \tag{81}
\end{aligned}$$

where

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

This shows the result under assumption (79). (Note that under this assumption,  $H'(t) = 1$  identically.)

For the more general case, set

$$G(t) = \int_0^t \frac{1}{H'(s)} ds. \tag{82}$$

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

$$\int_0^{\tilde{T}} \tilde{f}_u d\langle \tilde{\zeta}^{(k,l)}, \tilde{\zeta}^{(m,n)} \rangle_t = \int_0^T f_t H'(t)^2 d\langle \zeta^{(k,l)}, \zeta^{(m,n)} \rangle_t, \tag{83}$$

the result of the lemma follows. ■

## 7. Conclusion

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

A big caveat is that the analysis does not help in cases where there is leverage effect, or more generally where one can not do the analysis conditionally on the volatility processes. The current method, however, suggests estimators that can be analyzed under more general conditions.

**REFERENCES**

- Aldous, D. J., Eagleson, G. K., 1978. On mixing and stability of limit theorems. *Annals of Probability* 6, 325–331.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2000. Great realizations. *Risk* 13, 105–108.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2001. The distribution of exchange rate realized volatility. *Journal of the American Statistical Association* 96, 42–55.
- Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P., 2003. Modeling and forecasting realized volatility. *Econometrica* 71, 579–625.
- Barndorff-Nielsen, O. E., 1986. Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* 73, 307–322.
- Barndorff-Nielsen, O. E., 1991. Modified signed log likelihood ratio. *Biometrika* 78, 557–563.
- Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2006. Designing realized kernels to measure ex-post variation of equity prices in the presence of noise. Tech. rep.
- Barndorff-Nielsen, O. E., Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, B* 64, 253–280.
- Billingsley, P., 1995. *Probability and Measure*, 3rd Edition. Wiley, New York.
- Chernov, M., Ghysels, E., 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.
- Cox, D. R., 1958. Some problems connected with statistical inference. *Ann. Math. Statist.* 29, 357–372.
- Dacorogna, M. M., Gençay, R., Müller, U., Olsen, R. B., Pictet, O. V., 2001. *An Introduction to High-Frequency Finance*. Academic Press, San Diego.
- Efron, B., Hinkley, D. V., 1978. Assessing the accuracy of the maximum likelihood estimator: Observed versus expected fisher information. *Biometrika* 65, 457–482.
- Foster, D., Nelson, D., 1996. Continuous record asymptotics for rolling sample variance estimators. *Econometrica* 64, 139–174.
- Gallant, A. R., Hsu, C.-T., Tauchen, G. T., 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., Meddahi, N., 2005. Bootstrapping realized volatility. Tech. rep., Université de Montréal.
- Hájek, J., Sidak, Z., 1967. Theory of Rank Tests. Academic Press, New York.
- Hall, P., Heyde, C. C., 1980. Martingale Limit Theory and Its Application. Academic Press, Boston.
- Hayashi, T., Yoshida, N., 2005. On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli* 11, 359–379.
- Jacod, J., Protter, P., 1998. Asymptotic error distributions for the euler method for stochastic differential equations. *Annals of Probability* 26, 267–307.
- Jacod, J., Shiryaev, A. N., 2003. Limit Theorems for Stochastic Processes, 2nd Edition. Springer-Verlag, New York.
- Jensen, J., 1995. Saddlepoint Approximations. Oxford University Press, Oxford, U.K.
- Jensen, J. L., 1992. The modified signed likelihood statistic and saddlepoint approximations. *Biometrika* 79, 693–703.
- Jensen, J. L., 1997. A simple derivation of  $r^*$  for curved exponential families. *Scandinavian Journal of Statistics* 24, 33–46.
- Karatzas, I., Shreve, S. E., 1991. Brownian Motion and Stochastic Calculus. Springer-Verlag, New York.
- Lehmann, E., 1983. Theory of Point Estimation, 2nd Edition. John Wiley and Sons, New York.
- Mardia, K. V., Kent, J., Bibby, J., 1979. Multivariate Analysis. Academic Press, London.
- McCullagh, P., 1984. Local sufficiency. *Biometrika* 71, 233–244.
- McCullagh, P., 1987. Tensor Methods in Statistics. Chapman and Hall, London, U.K.
- McCullagh, P., Tibshirani, R., 1990. A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society ser. B* 52, 325–344.
- Mykland, P. A., Zhang, L., 2001. Inference for volatility type objects and implications for hedging. Tech. rep., Carnegie-Mellon University.
- Mykland, P. A., Zhang, L., 2006. ANOVA for diffusions. *Annals of Statistics* 34, 1931–1963.
- Oomen, R., 2004. Properties of realized variance for a pure jump process: Calendar time sampling versus business time sampling. Tech. rep., The University of Warwick, Warwick Business School.
- Pierce, D., Peters, D., 1994. Higher-order asymptotics and the likelihood principle: One-parameter models. *Biometrika* 81, 1–10.

- Protter, P., 2004. *Stochastic Integration and Differential Equations: A New Approach*, 2nd Edition. Springer-Verlag, New York.
- Reid, N., 1988. Saddlepoint methods and statistical inference. *Statistical Science* 3, 213–227.
- 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.
- Skovgaard, I., 1986. A note on the differentiation of cumulants of log likelihood derivatives. *International Statistical Review* 54, 29–32.
- Skovgaard, I., 1991. *Analytic Statistical Models*. Institute of Mathematical Statistics, Hayward.
- Weisberg, S., 1985. *Applied Linear Regression*, 2nd Edition. Wiley, New York.
- Zhang, L., 2001. *From martingales to ANOVA: Implied and realized volatility*. Ph.D. thesis, The University of Chicago, Department of Statistics.
- Zhang, L., 2006. Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli* 12, 1019–1043.
- Zhang, L., Mykland, P. A., Aït-Sahalia, Y., 2005. A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association* 100, 1394–1411.