ANOVA for Diffusions*

Per A. Mykland
Department of Statistics
The University of Chicago
Chicago, IL 60637

Lan Zhang
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213

Abstract

The paper examines the relationship among Ito processes from the angle of quadratic variation. The proposed methodology, "ANOVA for diffusions", allows drawing inference for a time interval, rather than for single time points. One of its applications is in fitness of modeling a stochastic process, it also helps quantify and characterize the trading (hedging) error in the case of financial applications.

The reason why the ANOVA permits conclusions over a time interval is that the asymptotic errors of the residual quadratic variation converge as a process (in time). A main conceptual finding was the clear cut effect of the two sources behind the asymptotics. The variation component (mixed Gaussian) comes only from the discretization error (in time discrete sampling). On the other hand, the bias depends only on the choice of estimator of the quadratic variation. This two-sources principle carries over to other criteria of goodness of fit, for example, the coefficient of determination.

Some key words and phrases: ANOVA, statistical uncertainty, Goodness of Fit, discrete sampling, nonparametric estimation, small interval asymptotics, stable convergence, option hedging, volatility model.

*This research was supported in part by National Science Foundation grants DMS 96-26266 and DMS 99-71738. The authors would like to thank The Bendheim Center for Finance at Princeton University, where part of the research was carried out. This manuscript was prepared using computer facilities supported by The Department of Statistics Fund at Carnegie Mellon University, and by The University of Chicago Block Fund.

1. Introduction. We consider in the following the regression relationship between two Ito processes Ξ_t and S_t,

$$d\Xi_t = \rho_t dS_t + dZ_t, \quad 0 \leq t \leq T, \tag{1.1}$$

where Z_t is a residual process. We suppose that the processes S_t and Ξ_t are observed at discrete sampling points $0 = t_0 < ... < t_k = T$. With the advent of high frequency financial data, this type of regression has been a growing topic of interest in the literature, cf. Section 2.2.

Our purpose in the following is to assess nonparametrically what is the smallest possible residual sum of squares in this regression. Specifically, denote by $< Z, Z >_T$ the sum of squares of the process Z under the idealized condition of continuous observation. We wish to estimate

$$\min_{\rho} < Z, Z >_T, \tag{1.2}$$

where the minimum is over all adapted regression processes ρ.

Our problem is "orthogonal" to that of estimation of the regression coefficient ρ_t. Depending on the goal of inference, statistical estimates $\hat{\rho}_t$ of the regression coefficient can be obtained using methods that are either local in space or in time, as discussed in Section 2.2. In addition, is is also common in financial contexts to use calibration ("implied quantities", see Section 5).

The importance of the question we ask in this paper is this. Once you know how to estimate (1.2), you also know how to assess the goodness of fit of any given estimation method for ρ_t. You also know more about the appropriateness of a one regressor model of the form (1.1). We return to the goodness of fit questions in Section 4.

The motivating application for the system (1.1) is that of statistical risk management in financial markets. We suppose that S_t and Ξ_t are the discounted values of two securities. At each time t, a financial institution is short one unit of the security represented by Ξ, and at the same time seeks to offset as much risk as possible by holding ρ_t units of security S. Denote the gain/loss, up to time t, from following this "risk-neutral" procedure by Z_t. This is then given by (1.1).

The reason why the problem (1.2) connects to an ANOVA is that if the processes Ξ and S were continuously observed, it would be possible to determine ρ_t uniquely (see (2.4)) to both minimize $< Z, Z >$, and also make the processes S and Z orthogonal at all times t, *i.e.*, $< S, Z >_t = 0$ in the sense of quadratic covariation. This gives rise to an ANOVA decomposition of the form

$$\underbrace{< \Xi, \Xi >_t}_{\text{total SS}} = \underbrace{\int_0^t \rho_u^2 d < S, S >_u}_{\text{SS explained}} + \underbrace{< Z, Z >_t}_{\text{RSS}}, \tag{1.3}$$

The main theorems in the current paper are concerned with the asymptotic behavior of the estimated RSS, as more observations are available within a fixed time window. There will be some choice in how to select the estimator $<\widehat{Z, Z}>_t$. We consider, therefore, a class of such estimators $<\widehat{Z, Z}>_t$. No matter which of our estimators is used, we get the decomposition

$$<\widehat{Z, Z}>_t - < Z, Z >_t \;\approx\; \text{bias}_t \;+\; ([Z, Z]_t - < Z, Z >_t)\,, \tag{1.4}$$

to first order asymptotically, where $[Z, Z]$ is the sum of squares of the (unseen) process Z at the sampling points, cf. the definition (2.1) below.

A primary conceptual finding in (1.4) is the clear cut effect of the two sources behind the asymptotics. The form of the bias depends only on the choice of estimator of the quadratic variation. On the other hand, the variation component is common for all the estimators under study, it comes only from the discretization error (in time discrete sampling) under the assumption of ρ known.

The organization is as follows, in Section 2, we establish the framework for ANOVA, and we examine a class of estimators of the residual quadratic variation. Our main results, in Section 3, provides the distributional properties of the estimation errors for RSS. See Theorem 1-Theorem 2. In Section 4, we discuss the statistical application of the main theorems. Parametric and nonparametric estimation are compared in the context of residual analysis. The goodness of fit of a model is addressed. In Section 5, we present a financial example where ANOVA can be implemented. Broad issues, including the analysis of variation versus analysis of variance, the moderate level of aggregation versus long run, the *actual* probability distribution versus the *risk neutral* probability distribution in the derivative valuation setting, are discussed in Section 4.3-4.4 and 5. After concluding in Section 6, we give proofs in Sections 7-8.

2. ANOVA for diffusions: framework. ANOVA is one of the main tools in statistics to assess the relationship among the variables of interest. It operates by breaking down the variation within a data collection. As is the case in the ordinary linear regression, estimation and ANOVA answer different questions: the former tells the estimated value of parameters, the latter tells how well the model fits. The same comparison holds for a diffusion case where the response and explanatory variables are continuous-time processes. ANOVA for diffusions does not focus on the inference *at independent time points*, rather, it can be aggregated *over a certain time span* to determine the goodness of fit relating processes.

2.1. *Ito processes, and quadratic variation.* We shall be concerned with quadratic variation ("q.v.") and co-variation both in continuous and discrete time, the latter reflecting the actual times of observation. We suppose that there is an interval $[0,T]$, and that the processes $\{X,Y\}$ are observed at a non-random partition $0 = t_0^{(n)} \le t_1^{(n)} \le t_2^{(n)} \le \cdots \le t_k^{(n)} = T$.

We let $[X,Y]_t$ denote the quadratic covariation of X and Y at the discrete-time scale, with the expression

$$[X,Y]_t = \sum_{t_{i+1}^{(n)} \le t} (\Delta X_{t_i^{(n)}})(\Delta Y_{t_i^{(n)}}), \tag{2.1}$$

where $\Delta X_{t_i^{(n)}} = X_{t_{i+1}^{(n)}} - X_{t_i^{(n)}}$ and $\Delta Y_{t_i^{(n)}} = Y_{t_{i+1}^{(n)}} - Y_{t_i^{(n)}}$. If X and Y are continuous semimartingales, such as the Ito processes defined below, the quadratic variation $< X,Y >_t$ at continuous time-scale is given as the limit of $[X,Y]_t$ as the number of observation points $k = k_n \to \infty$, with the mesh $\delta^{(n)} = \max_t |\Delta t^{(n)}| \to 0$. The convergence of $[X,Y]$ to $< X,Y >$ is uniform in probability (UCP). See Jacod and Shiryaev (1987), Theorem I.4.47 (p. 52), and Protter (1995), Theorem II.23 (p. 61), for details. Most of the time, we omit, for simplicity, the partition number (n).

The diffusion processes we shall work with are also known as Ito processes, as follows. Note that we suppose that there is an underlying filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, P)_{0\le t\le T}$.

DEFINITION 1. By saying that X is an *Ito process*, we mean that X can be represented as a smooth process plus a local martingale,

$$X_t = X_0 + \int_0^t \tilde{X}_u du + \int_0^t \sigma_u^X dW_u^X, \tag{2.2}$$

where W is a standard $(\mathcal{F}_t)$-Brownian Motion, X_0 is $\mathcal{F}_0$ measurable, and the coefficients $\tilde{X}_t$ and σ_t^X are $\mathcal{F}_t$-adapted, with $\int_0^t |\tilde{X}_u| du < +\infty$ and $\int_0^t (\sigma_u^X)^2 du < +\infty$. We also write

$$X_t = X_0 + X_t^{DR} + X_t^{MG} \tag{2.3}$$

as shorthand for the Doob-Meyer decomposition in (2.2). □

Note that W^X is typically different for different Ito processes. If W^X is the Brownian Motion appearing in the above equation, then the relationship between W^X and W^Y can be arbitrary.

Following similar notations, the (cumulative) quadratic variation and quadratic covariation can be expressed as

$$\begin{aligned} < X, X >_t &= \int_0^t (\sigma_u^X)^2 du, \\ < X, Y >_t &= \int_0^t \sigma_u^X \sigma_u^Y d < W^X, W^Y >_u . \end{aligned}$$

Both quadratic variation and covariation are absolutely continuous, where the latter follows from the Ito process assumption and from the Kunita-Watanabe Inequality (see, for example, p. 51 of Protter (1995)). For more details about the definitions, see Jacod and Shiryaev (1987) or Karatzas and Shreve (1991).

We shall often have occasion to suppose that $< X, Y >_t'$ is itself an Ito process. For ease of notation, we then write its Doob-Meyer decomposition as

$$d < X, Y >_t' = dD_t^{XY} + dR_t^{XY} = \tilde{D}_t^{XY} dt + dR_t^{XY}.$$

Note that the quadratic variation of $< X, Y >'$ is the same as $< R^{XY}, R^{XY} >$.

2.2. *Estimation schemes.*. Following model (1.1), with Ξ and S being Ito processes, we are interested in estimating the quadratic variation $< Z, Z >_t$ of the residual. Clearly, one also has to deal with the estimation of the instantaneous quantity ρ, as it enters into the estimation of $< Z, Z >$. Specifically, by model (1.1),

$$\rho_t = \frac{d < \Xi, S >_t}{d < S, S >_t}. \tag{2.4}$$

In a non-continuous world, where Ξ and S can only be observed over grid times, the most straightforward estimator of ρ is,

$$\hat{\rho}_t = \frac{\widehat{< \Xi, S >}_t'}{\widehat{< S, S >}_t'} = \frac{\sum_{t-h_n \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \left(\Delta \Xi_{t_i^{(n)}}\right) \left(\Delta S_{t_i^{(n)}}\right)}{\sum_{t-h_n \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \left(\Delta S_{t_i^{(n)}}\right) \left(\Delta S_{t_i^{(n)}}\right)}. \tag{2.5}$$

For simplicity, this is the one we shall use in the following.

Asymptotics for estimators on the form $\widehat{<X,Y>}'_t$ and hence for $\hat{\rho}_t$, is given by Foster and Nelson (1996) and Mykland and Zhang (2002). Let $\overline{\Delta t}^{(n)}$ be the average observation interval, assumed to converge to zero. If $<X,Y>'_t$ is an Ito process with nonvanishing volatility, then it is optimal to take $h_n = O((\overline{\Delta t^{(n)}})^{1/2})$, and $(\overline{\Delta t}^{(n)})^{1/4}(\widehat{<X,Y>}'_t - <X,Y>'_t)$ converges (for each fixed t) to a (conditional on the data) normal distribution with mean zero and random variance. The asymptotic distributions are (conditionally) independent for different times t. If $<X,Y>'_t$ is smooth, on the other hand, the rate becomes $(\overline{\Delta t}^{(n)})^{1/3}$ rather than $(\overline{\Delta t}^{(n)})^{1/4}$, and the asymptotic distribution contains both bias and variance. The same applies to the estimator $\hat{\rho}_t$.

The scheme given in (2.5) is only one of many for estimating $<X,Y>'_t$. In particular, Genon-Catalot et al. (1992) use wavelets for this purpose, and determine rates of convergence and limit distributions under the assumption that $<X,Y>'_t$ is deterministic and has smoothness properties.

Another important literature in this area seeks to estimate $<X,Y>'_t$ as a function of underlying state variables, see in particular Florens-Zmirou (1993),Hoffmann (1999) and Jacod (2000). The typical setup is that $U = (X, Y, ...)$ is a Markov process, so that $<X,Y>'_t = f(U_t)$ for some function f, and the problem is to estimate f. If all coefficients in the Markov diffusion are smooth of order s, and subject to regularity conditions, the function f can be estimated with a rate of convergence of $(\overline{\Delta t}^{(n)})^{s/(1+2s)}$.

The convergence obtained for the estimator of f under Markov assumptions is considerably faster than what can be obtained for (2.5). It does, however, rely on stronger assumptions than what we shall be working with in the following. Since we shall only be interested in ρ_t as a (random) function of time, our development does not require a Markov specification, and in particular does not require full knowledge of what potential state variables might be.

Back to the estimation of the quadratic variation $<Z,Z>$ of residuals. Given the discrete data of (Ξ, S), there exist different schemes to estimate the residual variation.

One scheme is to start with model (1.1). One first estimates $\Delta Z_{t_i^{(n)}}$ through relation $\Delta\hat{Z}_{t_i^{(n)}} = \Delta\Xi_{t_i^{(n)}} - \hat{\rho}_{t_i^{(n)}}(\Delta S_{t_i^{(n)}})$, where all increments are from time $t_i^{(n)}$ to $t_{i+1}^{(n)}$, and then obtains the quadratic variation (q.v. hereafter) of $\hat{Z}$. This gives an estimator of $< Z, Z >$ as

$$[\hat{Z},\hat{Z}]_t = \sum_{t_{i+1}^{(n)} \le t} (\Delta \hat{Z}_{t_i^{(n)}})^2 = \sum_{t_{i+1}^{(n)} \le t} [\Delta \Xi_{t_i^{(n)} \wedge t} - \hat{\rho}_{t_i^{(n)} \wedge t} (\Delta S_{t_i^{(n)} \wedge t})]^2 \tag{2.6}$$

where the notation of square brackets (discrete time-scale q.v.) is invoked, since $\Delta Z_{t_i^{(n)}}$ is the increment over discrete times.

Alternatively, one can directly analyze the ANOVA version of the model, (1.3), where $d < Z, Z >_t = d < \Xi, \Xi >_t - \rho_t^2 d < S, S >_t$. This yields a second estimator of $< Z, Z >_t$,

$$\widehat{< Z, Z >}_t^{(1)} = \sum_{t_{i+1}^{(n)} \le t} [(\Delta \Xi_{t_i^{(n)} \wedge t})^2 - \hat{\rho}^2_{t_i^{(n)} \wedge t} (\Delta S_{t_i^{(n)} \wedge t})^2]. \tag{2.7}$$

In general, any convex combination of these two,

$$\widehat{< Z, Z >}_t^{(\alpha)} = (1-\alpha)[\hat{Z},\hat{Z}]_t + \alpha \widehat{< Z, Z >}_t^{(1)} \tag{2.8}$$

would seem like a reasonable scheme for estimating $< Z, Z >_t$, and this is the class of estimators that we shall consider. Particular properties will be seen to attach to $\widehat{< Z, Z >}_t^{(1/2)}$, which we shall also denote by $\widetilde{< Z, Z >}_t$,

$$\widetilde{< Z, Z >}_t = [\Xi, \hat{Z}]_t. \tag{2.9}$$

Note that (2.9) also has a direct motivation from the continuous model. Since $< S, Z >_t = 0$, (1.1) yields that $< \Xi, Z >_t = < Z, Z >_t$.

Our primary goal is to investigate the statistical properties of the estimator $\widehat{< Z, Z >}_t^{(\alpha)}$ of the residual q.v. $< Z, Z >$, and in particular those of ($[\hat{Z},\hat{Z}]$, and $\widetilde{< Z, Z >}_t$). Asymptotic properties are naturally studied with the help of small interval asymptotics.

2.3. *Paradigm for asymptotic operations.* The asymptotic property of the estimation error is considered under the following paradigms and assumptions.

For a sequence of partitions of $[0, T]$, $0 = t_0^{(n)} \le t_1^{(n)} \le \cdots \le t_k^{(n)} = T, n = 1, 2, 3, \cdots$, we assume that as $n \to \infty$,

(i) the number of observations $k = k_n \to \infty$

(ii) the mesh $\delta^{(n)} \to 0$. The mesh is the maximum distance between the $t_i^{(n)}$'s,

(iii) the bandwidth $h_n \to 0$,

(iv) the number of observations between $t - h_n$ and t goes to infinity,

(v) there is a trade-off between h_n and $\overline{\Delta t^{(n)}}$.

The above (i) and (ii) suggest that, as n increases, we can observe the underlying data process more frequently. This observation refinement is not nested in that the set $\{t_0^{(n_1)}, t_1^{(n_1)}, \cdots, t_{k_{n_1}}^{(n_1)}\}$ is not necessarily contained in the set $\{t_0^{(n_2)}, t_1^{(n_2)}, \cdots, t_{k_{n_2}}^{(n_2)}\}$ for $n_1 < n_2$. The requirement (iii)-(iv) indicate that estimation window (or bandwidth h_n) shrinks with n while the number of observations within the window increases. In (v), $\overline{\Delta t^{(n)}}$ is the average observation interval, equal to $\frac{T}{k}$. As mentioned above, Mykland and Zhang (2002) showed that as n increases, how fast h_n and $\overline{\Delta t^{(n)}}$ decay respectively has a trade-off in terms of the asymptotic variance of the estimation error in ρ. It is optimal to take $h_n = O(\sqrt{\overline{\Delta t^{(n)}}})$. From now on, we use h and h_n interchangeably.

Specifically, we suppose the following.

ASSUMPTION A (*Quadratic variation of time:*) *For each $n \in N$, we have a sequence of non-random partitions $\{t_i^{(n)}\}$, $\Delta t_i^{(n)} = t_{i+1}^{(n)} - t_i^{(n)}$. Let $\max_i(\Delta t_i^{(n)}) = \delta(n)$. Suppose that*

(i) $\delta(n) \to 0$ as $n \to \infty$, and $\delta(n)/\overline{\Delta t^{(n)}} = O(1)$.

(ii) $H_{(n)}(t) = \frac{\sum_{t_{i+1}^{(n)} \le t} (\Delta t_i^{(n)})^2}{\overline{\Delta t^{(n)}}} \to H(t)$ as $n \to \infty$, where $H(t)$ is continuously differentiable.

(iii) $[H_{(n)}(t) - H_{(n)}(t - h_n)]/h_n \to H'(t)$ as $h_n \to 0$, where the convergence is uniform in t.

When the partitions are evenly spaced, $H(t) = t$ and $H'(t) = 1$. In the more general case, the left hand side of (ii) is bounded by $t\delta(n)/\overline{\Delta t^{(n)}}$, while the left hand side of (iii) is bounded by $\delta(n)^2/(\overline{\Delta t^{(n)}}h) + \delta(n)/\overline{\Delta t^{(n)}}$. In all our results, h is bigger than $\overline{\Delta t^{(n)}}$, and hence both the left hand sides are bounded because of (i). The assumptions in (ii) and (iii) are, therefore, about a unique limit point, and about interchanging limits and differentiation.

2.4. *Assumptions on the process structure.* The following assumptions are imposed on the relevant Ito processes.

ASSUMPTION B(X) (*Smoothness*): *X is an Ito process. Also, $< X, X >'_t$ and $\tilde{X}_t$ are continuous, almost surely.*

REMARK 1. The addition of Assumption B to an Ito process X, and similar smoothness assumptions in results below, is mainly a matter of convenience of proof. The assumption can in many instances be dropped, at the cost of more involved technical arguments. □

ASSUMPTION C(X) (*Non-vanishing volatility*): $\inf_{t \in [0,T]} < X, X >'_t > 0$ *almost surely*

ASSUMPTION D (*Description of the filtration*): *There is a continuous multidimensional P-local martingale $\mathcal{X} = (\mathcal{X}^{(1)}, \cdots, \mathcal{X}^{(p)})$, any p, so that $\mathcal{F}_t$ is the smallest sigma-field containing $\sigma(\mathcal{X}_s, s \leq t)$ and $\mathcal{N}$, where $\mathcal{N}$ contains all the null sets in $\sigma(\mathcal{X}_s, s \leq T)$.*

REMARK 2. The final statement in Assumption D assures that the "usual conditions" (Jacod and Shiryaev (1987), p. 2, Karatzas and Shreve (1991), p. 10) are satisfied. The main implication, however, is on our mode of convergence, as follows. □

2.5. *The limit for the discretization error.* As mentioned in the introduction, the error $\widehat{< Z, Z >}_t - < Z, Z >_t$ can be decomposed in bias and pure discretization error $[Z, Z]_t - < Z, Z >_t$. We here discuss the limit result for the latter, following Jacod and Protter (1998).

PROPOSITION 1. (Discretization Theorem). Let Z be an Ito process for which $\int_0^T (< Z, Z >')_t^2 dt < \infty$ a.s. $\int_0^T \tilde{Z}_t^2 dt < \infty$ a.s. Subject to assumptions A and D,

$$\frac{1}{\sqrt{\Delta t^{(n)}}} ([Z, Z]_t - < Z, Z >_t) \xrightarrow{\mathcal{L}.stable} \int_0^t \sqrt{2H'(u)} < Z, Z >'_u dW_u,$$

where W is a standard Brownian Motion, independent of the underlying process $\mathcal{X}^{(i)}$. □

The symbol $\xrightarrow{\mathcal{L}.stable}$ denotes *stable* convergence of the process, as defined in Rényi (1963) and Aldous and Eagleson (1978); see also Rootzén (1980) and Section 2 of Jacod and Protter (1998).

In the case of an equidistant grid, the result coincides with applicable part of Theorems 5.1 and 5.5 in Jacod and Protter (1998), and the proof is essentially the same (see Section 7). On an abstract form, results of this type appear to go back to Rootzén (1980).

Note that the conditions on $< Z, Z >'$ and $\tilde{Z}$ are the same as in the equidistant case, due to the Lipschitz continuity of H.

Some further discussion of this proposition, and some further results in this direction that we shall use, are contained in Section 7.

3. ANOVA for diffusions: main distributional results.

3.1. *Distribution of* $\frac{[\hat{Z},\hat{Z}]_t - <Z,Z>_t}{\sqrt{\overline{\Delta t}^{(n)}}}$. Recall that the square bracket $[Z, Z]$ and the angled bracket $< Z, Z >$ represent the quadratic variation of Z at discrete and continuous time-scale, respectively.

THEOREM 1. . Assume condition A, with $\frac{\sqrt{\overline{\Delta t}^{(n)}}}{h} \to c$ as $n \to \infty$, and $0 < c < \infty$. Suppose S, Ξ, ρ, $< S, S >'$, $< \Xi, S >'$, $< R^{SS}, R^{SS} >'$, $< R^{\Xi S}, R^{\Xi S} >'$, and $< R^{\Xi\Xi}, R^{\Xi\Xi} >'$ are Ito processes, each satisfying condition B. Also assume condition $C(S)$. Let estimator $\Delta \hat{Z}_{t_i^{(n)}} = \Delta \Xi_{t_i^{(n)}} - \hat{\rho}_{t_i^{(n)}} \cdot \Delta S_{t_i^{(n)}}$, with S, Ξ and ρ satisfying model (1.1). Then,

$$\begin{aligned} &(\overline{\Delta t}^{(n)})^{-\frac{1}{2}}([\hat{Z}, \hat{Z}]_t - < Z, Z >_t) \\ &= \int_0^t V_{\hat{\rho}-\rho} d < S, S >_u + (\overline{\Delta t}^{(n)})^{-\frac{1}{2}}([Z, Z]_t - < Z, Z >_t) + o_p(1) \end{aligned} \tag{3.10}$$

uniformly in t, where

$$\int_0^t V_{\hat{\rho}-\rho} d < S, S >_u = \frac{1}{3c} \int_0^t < \rho, \rho >'_u d < S, S >_u + c \int_0^t H'(u) d < Z, Z >_u \tag{3.11}$$

□

REMARK 3. . The consequence of Theorem 1 (same for Theorem 2 below) is that,

$$D_t + \frac{1}{\sqrt{\overline{\Delta t}^{(n)}}} ([Z, Z]_t - < Z, Z >_t) + o_p(1)$$

converges in law (stably) to

$$D_t + \int_0^t \sqrt{2H'(u)} < Z, Z >'_u dW_u,$$

$o_p(1)$ term goes away by Lemma VI 3.31 (p. 316) in Jacod and Shiryaev (1987). □

Notice that Theorem 1, together with Proposition 1, says that the estimator $[\hat{Z}, \hat{Z}]_t$ converges to $< Z, Z >_t$ at the order of square root of the average sampling interval. In the limit, the error term consists of a non-negative bias $\int_0^t V_{\hat{\rho}-\rho} d < S, S >_u$, due to the estimation uncertainty $< \hat{Z}, \hat{Z} > - < Z, Z >$, and a mixture Gaussian, due to the discretization $[Z, Z]_t - < Z, Z >_t$. The non-negativeness of the asymptotic bias is because the q.v.'s ($< \rho, \rho >$, $< S, S >$, $< Z, Z >$) are non-decreasing processes. Furthermore, (3.11) displays a bias-bias tradeoff, thus an optimal c for smoothing can be reached to minimize the asymptotic bias, though we have not investigated the effect of having a random c. The discretization term is independent of the smoothing factor.

3.2. *Distribution of* $< \widetilde{Z, Z} >_t - < Z, Z >_t$.

THEOREM 2. . Suppose that S, Ξ, and ρ are Ito Processes satisfying model (1.1). Assume condition A and $C(S)$, with $\frac{\sqrt{\overline{\Delta t}^{(n)}}}{h} \to c$ as $n \to \infty$, and $0 < c < \infty$. Also assume each of the processes Ξ, S, ρ, $< \Xi, S >'$, $< S, S >'$, $< R^{SS}, R^{SS} >'$, $< R^{\Xi S}, R^{SS} >'$, and $< R^{\Xi S}, R^{\Xi S} >'$) satisfies condition B, and processes $< \Xi, \rho >'$ and $< S, \rho >'$ are continuous. Then, uniformly in t

$$\frac{< \widetilde{Z, Z} >_t - < Z, Z >_t}{\sqrt{\overline{\Delta t}^{(n)}}} = \frac{1}{2c} \int_0^t < \Xi, S >'_u \, d\rho_u + (\overline{\Delta t}^{(n)})^{-\frac{1}{2}} ([Z, Z]_t - < Z, Z >_t) + o_p(1). \tag{3.12}$$

□

Remark 3 applies similarly.

Unlike $[\hat{Z}, \hat{Z}]$, the asymptotic (conditional) bias associated with $< \widetilde{Z, Z} >_t$ does not necessarily have a positive or negative sign. Moreover, we are no longer faced with a bias-bias tradeoff due to the position of c in (3.12). In this case, the role of smoothing in asymptotic bias shall be discussed in Section 3.3.

3.3. *Distribution of* $< \widehat{Z, Z} >_t^{(\alpha)} - < Z, Z >_t$. From (2.8),

$$< \widehat{Z, Z} >_t^{(\alpha)} = (1 - 2\alpha)[\hat{Z}, \hat{Z}]_t + 2\alpha < \widetilde{Z, Z} >_t,$$

it follows from the assumptions of Theorems 1-2 that, if one sets

$$\text{bias}_t^{(\alpha)} = \frac{\alpha}{c} \int_0^t < \Xi, S >'_u \, d\rho_u + (1 - 2\alpha) \int_0^t V_{\hat{\rho}-\rho} d < S, S >_u, \tag{3.13}$$

Then

$$\frac{\widehat{<Z,Z>}_t^{(\alpha)} - <Z,Z>_t}{\sqrt{\overline{\Delta t}^{(n)}}} = \text{bias}_t^{(\alpha)} + (\overline{\Delta t}^{(n)})^{-\frac{1}{2}}([Z,Z]_t - <Z,Z>_t) + o_p(1)$$

$$\overset{\mathcal{L}.stable}{\longrightarrow} \text{bias}_t^{(\alpha)} + \int_0^t \sqrt{2H'(u)} <Z,Z>'_u \, dW_u, \quad (3.14)$$

As a summary for any linear combination of the estimators in Theorem 1- 2, $\alpha \in [0,1]$, the convergence in (3.14) is in law as a process, and the limiting Brownian Motion W is independent of the entire data process. See the definition of stable convergence in Section 2.5 above.

The "variance" term $(\overline{\Delta t}^{(n)})^{-\frac{1}{2}}([Z,Z]_t - <Z,Z>_t)$ is the same for any estimator in the linear-combination class, they are all asymptotically perfectly correlated. The common asymptotic, conditional variance is independent of the smoothing bandwidth. It remains unclear whether the common asymptotic variance could, perhaps, be a lower bound under the nonparametric setting (see Bickel et al. (1993) for a comprehensive discussion). This needs further investigation.

For the bias, on the other hand, the estimation procedure plays an important role, as the bias term varies with α. Also the smoothing effect enters the bias terms. From Theorem 1-2, excessive over-smoothing (smaller c) or under-smoothing (bigger c) can explode the bias of $\widehat{<Z,Z>}^{(\alpha)}$, for $\alpha \neq \frac{1}{2}$, thus (conditional) bias may be minimized optimally. When $\alpha = \frac{1}{2}$, it is not obvious how to deal with bias-bias tradeoff. One might theoretically be able to reduce the bias for $\widetilde{<Z,Z>}$ (i.e. $\widehat{<Z,Z>}^{(1/2)}$) by choosing the smallest possible bandwidth h. This thought should, however, be taken with caution. It is not obvious whether the magnitude of the higher order terms in the earlier results would remain negligible if the estimation window h were to decrease faster than the order $\sqrt{\Delta t}$. Also in the case of financial practice, one may want to choose a c such that both the asymptotic bias and hedging error are relatively small.

Table 1 in next page shows that assuming constant ρ, $\widetilde{<Z,Z>}_t$ will be the best choice among the three. When ρ is random, none of the estimation schemes in section 2.2 is obviously superior to the others.

4. Goodness of fit.

The purpose of ANOVA is to assess the goodness of fit of a regression model on the form (1.1).

Table 1: The effect of constant ρ on the bias components

Estimator	Asymptotic Bias
$[\hat{Z}, \hat{Z}]_t$	$c \int_0^t H'(u) d < Z, Z >_u$
$<\widehat{Z, Z}>_t^{(1/2)}$	$-c \int_0^t H'(u) d < Z, Z >_u$
$<\widetilde{Z, Z}>_t$	0

We here illustrate the use of Theorems 1 and 2 by considering two different questions of this type. In the first section, we discuss how to assess the fit of a parametric estimator for ρ. Afterward, we focus on the question of how good is the one regressor model itself, independently of estimation techniques. This is already measured by the quantity $< Z, Z >_T$, but can be further studied by considering confidence bands for $< Z, Z >_t$ as a process, and by an analogue to the coefficient of determination. Finally, we discuss the question of the relationship between this ANOVA and the analysis of variance that is used in the standard regression setting.

4.1. *The assessment of parametric models.* In the following we suppose that a parametric model is fit to the data, and ρ is estimated as a function of the parameter. Since we have a nonparametric estimate for $< Z, Z >$, we can compare to this a parametric estimate for the residual sum of squares to see how good the parametric model is in capturing the true regression of Ξ on S.

Specifically, we suppose that data from the multidimensional process X_t is observed at the grid points. X_t has among its components at least S_t and Ξ_t, but there are possibly also other processes that are components in X_t. The parametric model is of the form $P_{\theta,\psi}$, $\theta \in \Theta$, $\psi \in \Psi$, where the modeling is such that diffusion coefficients are functions of θ, while drift coefficients can be functions of both θ and ψ. It is thus reasonable to suppose that as $\bar{\Delta}t \to 0$, $\hat{\theta}$ converges in probability to a nonrandom parameter value θ_0, and that

$$(\bar{\Delta}t)^{-1/2}(\hat{\theta} - \theta_0) \to \eta N(0, 1)$$

in law, stably, where η is a function of the data and the $N(0, 1)$ term is independent of the data. θ_0 is the true value of the parameter if the model does contain the true probability, but is otherwise also taken to be a defined parameter.

Under $P_{\theta,\psi}$, the regression coefficient ρ_t is of the form $\beta_t(\theta)$. Most commonly, $\beta_t(\theta) = b(X_t;\theta)$ for a nonrandom functional b.

We now ask whether the true regression coefficient can be correctly estimated with the model at hand. In other words, we wish to test the null hypothesis H_0 that $\beta_t(\theta_0) = \rho_t$.

For the ANOVA analysis, define the theoretical residual by

$$dV_t = d\Xi_t - \beta_t(\theta_0)dS_t, \quad V_0 = 0,$$

and the observed one by

$$\Delta\hat{V}_{t_i} = \Delta\Xi_{t_i} - \beta_{t_i}(\hat{\theta})\Delta S_{t_i}, \quad \hat{V}_0 = 0.$$

Under the null hypothesis, $< V, V > = < Z, Z >$ and so a natural test statistic is of the form

$$U = (\bar{\Delta}t)^{-1/2}([\hat{V},\hat{V}]_T - \widehat{< Z, Z >}_T)$$

We now derive the null distribution for U, using the results above.

As an intermediate step, define the discretized theoretical residual

$$\Delta V^d_{t_i} = \Delta\Xi_{t_i} - \beta_{t_i}(\theta_0)\Delta S_{t_i}, \quad V^d_0 = 0$$

Subject to obvious regularity conditions,

$$\begin{aligned}
[\hat{V},\hat{V}]_T - [V^d, V^d]_T &= -2\sum_{t_{i+1}\le t}(\beta_{t_i}(\hat{\theta}) - \beta_{t_i}(\theta_0))\Delta V^d_{t_i}\Delta S_{t_i} \\
&\quad + \sum_{t_{i+1}\le t}(\beta_{t_i}(\hat{\theta}) - \beta_{t_i}(\theta_0))^2\Delta S^2_{t_i} \\
&= -2(\hat{\theta} - \theta_0)\sum_{t_{i+1}\le t}\frac{\partial\beta_{t_i}}{\partial\theta}(\theta_0)\Delta V^d_{t_i}\Delta S_{t_i} + O_p(\bar{\Delta}t) \\
&= -2(\hat{\theta} - \theta_0)\int_0^T \frac{\partial\beta_t}{\partial\theta}(\theta_0)d < V, S >_t + O_p(\bar{\Delta}t).
\end{aligned}$$

Also, under the conditions in Proposition 2 in Sect 7 and Lemma 3 in Section 8, $[V^d, V^d]_T = [V,V]_T + o_p(\bar{\Delta}t^{1/2})$ as $\bar{\Delta}t \to 0$ (since $< V^d, V^d >_t = < V, V >_t + o_p(\bar{\Delta}t^{1/2})$, and $< V^d, V^d >'_t \approx < V^d, V >'_t \approx < V, V >'_t$).

Hence, under the conditions of Theorem 1 or 2,

$$\begin{aligned} U &= -2(\bar{\Delta}t)^{-1/2}(\hat{\theta}-\theta_0)\int_0^T \frac{\partial \beta_t}{\partial \theta}(\theta_0) d < V, S >_t \\ &\quad +(\bar{\Delta}t)^{-1/2}\{([V,V]_T - [Z,Z]_T)\} \\ &\quad -\text{bias}_T + o_p(1) \end{aligned}$$

where bias_T has the same meaning as in Section 2. If the null hypothesis is satisfied, therefore,

$$U \to N(0,1) \times 2\eta \int_0^T \frac{\partial \beta_t}{\partial \theta}(\theta_0) d < V, S >_t -\text{bias}_T$$

in law, stably. The variance and bias can be estimated from the data. This, then, provides the null distribution for U.

Another approach is to use U to measure how close the parametric residual $< V, V >$ is to the lower bound $< Z, Z >$. To first order,

$$\begin{aligned} (\bar{\Delta}t)^{1/2} U &\to < V, V >_T - < Z, Z >_T \\ &= \int_0^T (\beta_t(\theta_0) - \rho_t)^2 d < S, S >_t \end{aligned}$$

The behavior of $U - (\bar{\Delta}t)^{-1/2}(< V, V >_T - < Z, Z >_T)$ depends on the joint limiting distribution of $([V,V]_T - < V, V >_T) - ([Z,Z]_T - < Z, Z >_T)$ and $(\bar{\Delta}t)^{-1/2}(\hat{\theta} - \theta_0)$. The former can be provided by Theorem A.2 (or Section 5 of Jacod and Protter 1998), but further assumptions are needed to obtain the joint distribution. A study of this is beyond the scope of this paper.

4.2. *Confidence bands.* In addition to pointwise confidence intervals for $\widehat{< Z, Z >}_t^{(\alpha)}$, we can construct joint confidence band for the estimated quadratic variation $\widehat{< Z, Z >}^{(\alpha)}$ of residuals, because $\widehat{< Z, Z >}^{(\alpha)}$ converges as a process by earlier Theorems.

One proceeds as follows. As a process on $[0, T]$,

$$(\overline{\Delta t^{(n)}})^{-\frac{1}{2}} \left(\widehat{< Z, Z >}_t^{(\alpha)} - < Z, Z >_t \right) \xrightarrow{\mathcal{L}} \text{bias}_t^{(\alpha)} + L_t,$$

Under all estimation schemes in the linear combination class, we have, by Theorem 1-2 and subsequent results on $\widehat{< Z, Z >}_t^{(\alpha)}$,

$$L_t = \int_0^t \sqrt{2H'(u)} < Z, Z >'_u dW_u,$$

where W is a standard Brownian Motion independent of the complete data filtration. Now condition on $\mathcal{F}_T$: by the stable convergence, L_t is then a Gaussian process, with $< L, L >_t$ nonrandom. Now use the change-of-time construction by Dambis (1965) and Dubins and Schwartz (1965) to obtain $L_t = W^*_{<L,L>_t}$, where W^* is a new Brownian Motion conditional on $\mathcal{F}_T$. It then follows that

$$\begin{aligned} \max_{0\le t\le T} L_t &= \max_{0\le t\le 2\int_0^T H'(u)(<Z,Z>'_u)^2 du} W^*_t \\ \min_{0\le t\le T} L_t &= \min_{0\le t\le 2\int_0^T H'(u)(<Z,Z>'_u)^2 du} W^*_t \end{aligned}$$

Now write $L_n(t) = (\overline{\Delta t^{(n)}})^{-\frac{1}{2}} \left(\widehat{< Z, Z >_t}^{(\alpha)} - < Z, Z >_t \right) - \text{bias}_t^{(\alpha)}$, we have

$$\begin{aligned} P\left(|L_n(t)| \le c, \text{ for all } t \in [0,T]\right) &\to P\left(|L(t)| \le c, \text{ for all } t \in [0,T]\right) \\ &= P\left(\min_{0\le t\le \tau} W^*_t \ge -c, \max_{0\le t\le \tau} W^*_t \le c \right) \end{aligned}$$

Choose $c = c_\tau$ such that

$$P\left(\min_{0\le t\le \tau} W^*_t \ge -c_\tau, \max_{0\le t\le \tau} W^*_t \le c_\tau | \tau \right) = 1 - \alpha,$$

with $\tau = 2\int_0^T H'(u)(< Z >'_u)^2 du$. To find c_τ, one can refer to Karatzas and Shreve (1991) Section 2.8 for the distributions of the running minimum and maximum of a Brownian Motion. This completes our construction of a global confidence band.

4.3. *The Coefficient of Determination.* In analogy with standard linear regression, one can define the R^2 by

$$R_t^2 = 1 - \frac{< Z, Z >_t}{< \Xi, \Xi >_t}.$$

This is the quantity one would have observed if the whole path of the processes Ξ and S had been available. If observations are on a grid, it is natural to use

$$\hat{R}_t^2 = 1 - \frac{\widehat{< Z, Z >_t}}{[\Xi, \Xi]_t}.$$

Under the assumptions of Section 2, the distribution of $\hat{R}_t^2$ can be found by

$$\begin{aligned}
&(\overline{\Delta t}^{(n)})^{-1/2}(\hat{R}_t^2 - R_t^2) \\
&= -(\overline{\Delta t}^{(n)})^{-1/2}\frac{1}{<\Xi,\Xi>_t}\Big((<\widehat{Z,Z}>_t - <Z,Z>_t) - (1-R_t^2)([\Xi,\Xi]_t - <\Xi,\Xi>_t)\Big) \\
&\quad + o_p(1) \\
&= -(\overline{\Delta t}^{(n)})^{-1/2}\frac{1}{<\Xi,\Xi>_t}\big(([Z,Z]_t - <Z,Z>_t) - (1-R_t^2)([\Xi,\Xi]_t - <\Xi,\Xi>_t)\big) \\
&\quad - \mathrm{bias}_t^{(\alpha)} + o_p(1),
\end{aligned}$$

where $\mathrm{bias}_t^{(\alpha)}$ is that corresponding to estimator $<\widehat{Z,Z}>_t = <\widehat{Z,Z}>_t^{(\alpha)}$ (we use $<\widehat{Z,Z}>_t$ and $<\widehat{Z,Z}>_t^{(\alpha)}$ interchangeably, unless otherwise stated).

A straightforward generalization of Proposition 1 yields that $(\overline{\Delta t}^{(n)})^{-1/2}([Z,Z]_t - <Z,Z>_t, [\Xi,\Xi]_t - <\Xi,\Xi>_t)_{0\le t\le T}$ converges (stably) with the integrand from Definition **??** given by

$$g_t = 2H'(t)\begin{pmatrix} (<Z,Z>_t')^2 & (<Z,\Xi>_t')^2 \\ (<Z,\Xi>_t')^2 & (<\Xi,\Xi>_t')^2 \end{pmatrix},$$

and equation (1.1) yields that $<Z,\Xi>_t' = <Z,Z>_t'$. It follows that

$$\begin{aligned}
&(\overline{\Delta t}^{(n)})^{-1/2}(\hat{R}_t^2 - R_t^2) \\
&\xrightarrow{\mathcal{L}.stable} -\frac{R_t^2}{<\Xi,\Xi>_t}\int_0^t \sqrt{2H'(u)}<Z,Z>_u'\, dW_u \\
&\qquad - \frac{1-R_t^2}{<\Xi,\Xi>_t}\int_0^t \sqrt{2H'(u)[(<\Xi,\Xi>_u')^2 - (<Z,Z>_u')^2]}dW_u^* \\
&\qquad - \mathrm{bias}_t^{(\alpha)},
\end{aligned}$$

where W and W^* are independent Brownian motions. For fixed t, the limit is conditionally normal, with mean $-\mathrm{bias}_t^{(\alpha)}$ and variance

$$\frac{1}{<\Xi,\Xi>_t^2}\left(R_t^4\int_0^t 2H'(u)(<Z,Z>_u')^2du + (1-R_t^2)^2\int_0^t 2H'(u)[(<\Xi,\Xi>_u')^2 - (<Z,Z>_u')^2]du\right),$$

which can be readily estimated.

4.4. *Variance versus variation: Which ANOVA?.* The formulation we have given in (1.3) is in terms of quadratic variation. This raises the question how our analysis relates to the traditional meaning of ANOVA, namely a decomposition of variance.

There are several answers to this, some concerning the broad setting provided by model (1.1), and they are discussed presently in Sections 4.3 and 4.4. More specific structure is provided by financial applications, and a discussion is provided in Section 4.5.

In model (1.1), the variation in Z can come from both drift and martingale. As in (2.3),

$$Z_t = Z_0 + Z_t^{DR} + Z_t^{MG}. \tag{4.1}$$

The analysis we have presented concerns most directly the variation in Z_t^{MG}, in that $\mathrm{var}(Z_t^{MG}) = E(< Z, Z >_t)$, where it should be noted that $< Z, Z >_t = < Z^{MG}, Z^{MG} >_t$. Hence, if the Z^{DR} term is identically zero, the analysis of variation is also an exact analysis of variance, in terms of expectations.

On a deeper level, when considering *conditionality*, the quadratic variation of Z may be a more relevant target in an analysis than the variance of Z, as the latter is a conditional quantity and hence less representative of data. Following the Dambis (1965)/Dubins and Schwartz (1965) representation, $Z_t^{MG} = V_{<Z,Z>_t}$, where V is a standard Brownian motion (on a different time scale). $< Z, Z >_t = < Z^{MG}, Z^{MG} >_t$, therefore, contains information about the actual amount of variation that has occurred in the process Z_t^{MG}. Using the quadratic variation is, in this sense, analogous to using observed information in a likelihood setting (see, for example, Efron and Hinkley (1978)). The analogy is valid also on the technical level: if one forms the dual likelihood (Mykland (1995)) from score function Z_t^{MG}, the observed information is, indeed, $< Z, Z >_t$.

4.5. *A tale of tradeoffs: ANOVA when the drift is nonzero.* So what if the drift Z^{DR} in (4.1) *is* nonzero? Is $< Z, Z >_t$ any longer relevant to the variation in Z_t itself?

Our contention is that $< Z, Z >_T$ is the correct measure of variation for small and intermediate sized intervals $[0, T]$, but that bias may take over when T is large.

There is, however, a tradeoff between bias and, on the other hand, conditionally correct inference. As pointed out in Section 4.3, the estimators $\widehat{< Z, Z >}_T$ are conditionally more appropriate than a straight estimate of $\mathrm{var}(Z_T - Z_0)$, and one may wish to accept more bias (as T grows bigger) in return for the greater conditionality offered by our "ANOVAriation". In fact, in the sense of approximate likelihoods as in Mykland (2001), $< Z, Z >_T$ remains the observed information.

In terms of the size of T, the martingale term in (4.1) will dominate for small T, whereas the drift term will dominate for large T. Specifically, no matter whether $T \to 0$ or $T \to +\infty$, $Z_T^{MG} = O_p(T^{1/2})$ and $Z_T^{DR} = O_p(T)$. The application under our consideration generally concerns non-large level of time aggregation, hence the current analysis of variation, rather than analysis of variance, is both the feasible analysis and the one that captures the dominant effect.

Of course, one would ideally like to take account of both terms, but there is the added fact that it is so much easier to estimate $< Z, Z >_t$, and hence the spread of Z_t^{MG}, than the spread in Z_t^{DR}. As we have seen in this work, and as is clear from other works (such as Jacod and Shiryaev (1987), Theorem I.4.47 (p. 52), and Protter (1995), Theorem II.23 (p. 61)), $< Z^{MG}, Z^{MG} >_t$ can be consistently estimated on a fixed interval, with the paradigm from Section 2.3. On the other hand, even with parametric model assumptions and even with observations continuous in time, the drift can only be accurately estimated when T is large. This is a consequence of Girsanov's Theorem (see, e.g., Chapter 3.5 of Karatzas and Shreve (1991)). The long-run (large T) behavior merits separate discussion and will not be addressed in present work.

4.6. *Financial applications: An instance where variance and variation relate exactly.* The case from Section 4.3, where Z itself is a martingale, or where one is interested in Z^{MG} only, is quite common in finance. We show two examples of this; a conceptual one in this section, and one that involves data in the next section.

In either example, one is interested in testing whether the residual Z is zero, or in quantifying the distribution of the residual under the so called *Risk Neutral* or *Equivalent Martingale Measure* P^*.

If P is the true, actual, probability distribution under which data is collected, P^*, is by contrast, the following. It must be equivalent to P, in the sense of mutual absolute continuity, and satisfy that the discounted value of all traded securities must be P^*-martingales. The value of financial assets, consequently, are expectations under P^*. For further details, refer to Harrison and Krebs (1979),Harrison and Pliska (1981),Delbaen and Schachermayer (1995a,b),Duffie (1996), Hull (1999). Note that, in the simplest case, where we take the short term risk free interest rate r to be constant over time, the discounted value at time t of a security X_t is $\exp(-rt)X_t$, and the initial price at time 0 of this security is $E^*[exp(-rt)X_t]$, where E^* is expectation taken with respect to P^*.

If the residual Z relates to the value of a security, one is often interested in its behavior under P^* rather than under P. Specifically we shall see that one is interested in Z^{MG*}, where this is the martingale part in the Doob-Meyer decomposition (4.1), *when taken under* P^*;

$$Z_t = Z_0 + Z_t^{DR*} + Z^{MG*} \text{ w.r.t. } P^*. \tag{4.2}$$

The quadratic variation $< Z, Z >=< Z^{MG}, Z^{MG} >=< Z^{MG*}, Z^{MG*} >$ is the same under P and P^*, but under the latter distribution, it refers to the behavior of Z^{MG*} rather than Z^{MG}.

The simplest case is the following.

EXAMPLE 1. *Suppose that Ξ and S are both discounted securities prices, and that one seeks to offset risk in Ξ by holding ρ units of S. The residual is then, itself, the discounted value of the unhedged part of Ξ. Under P^*, therefore, Z is a martingale, $Z_t = Z_0 + Z_t^{MG*}$.*

A deeper example is encountered below in Section 4,where we analyze implied volatilities.

In both these cases, to put a value of the risk involved in Z^{MG*}, one is interested in bounds on the quadratic variation $< Z, Z >$, under P^*. This will help, for example, in pricing spread options on Z.

How do our results for probability P relate to P^*? They simply carry over, unchanged, to this probability distribution. Theorems 1 and 2 remain valid by absolute continuity of P^* under P. In the case of limiting results, such as those in Propositions 1-2 (in Section 7) and the development for goodness of fit in Section 3, we also invoke the mode of stable convergence in Definition **??**, together with the fact that dP^*/dP is measurable with respect to the underlying σ-field $\mathcal{F}_T$.

Finally, if one wishes to test a null hypothesis H_0 that Z^{MG*} is constant, then H_0 is equivalent to asking whether $< Z, Z >_T$ is zero (whether under P or P^*). This can again be answered with our distributional results above. In the case of Example 1, the H_0 of fully offsetting the risk in Ξ also tests if Z itself is constant.

5. Introducing ANOVA to volatility research.

Following market practice, as well as the results of Mykland and Zhang (2001, 2002), we instead investigate the question from Example 1 from the point of view of *volatility*. The goal is to investigate the association between the process of implied volatility and that of realized volatility, which originate from different data sources.

We are interested in the variation of a non-dividend paying stock S_t. To be specific, suppose the stock price follows

$$dS_t = \mu_t S_t dt + \sigma_t S_t dW_t \tag{5.1}$$

where the drift term μ_t and the diffusion term σ_t can be stochastic and time-varying, and W is a standard Brownian Motion. It will be assumed that all quantities in (5.1) are adapted to the underlying filtration $(\mathcal{F}_t)$. The actual distribution governing (5.1) remains denoted by P.

The *realized* volatility of the underlying asset is $\sigma_t^2 =< \log S, \log S >_t'$, the derivative of the quadratic variation of $\log S$. An estimate of σ_t^2 based on the past data can then be obtained using, for example, $< \log \widehat{S, \log} S >_t'$, following the discussion in (2.1) and (2.5).

The *implied volatility* provides another way to study the variation in the underlying security S. The computation of implied volatility involves inverting a pricing formula for an derivative security as detailed below (see (5.3)).

Specifically, consider a European option with payoff $f(S_T)$ at the maturity time T. Under the Geometric Brownian Motion paradigm (GBM) for S, where μ and σ in (5.1) are assumed constant, the price of this European option at time $t, t \leq T$, can be written as $C(S_t, \sigma^2(T-t), r(T-t))$, where

$$C(S, \Xi, R) = \exp(-R) E^* f(S \exp(R - \Xi/2 + \sqrt{\Xi} Z)), \tag{5.2}$$

Z is standard normal under P^* (the Equivalent Martingale Measure discussed in Section 4.5). The option maturity T and payoff form f are given by the option contract.

Notice that equation (5.2) is the Black and Scholes (1973) and Merton (1973) formula (abbreviated with BS). The most common instance would be the call option, where $f(s) = (s-K)^+$ with a pre-determined strike price K.

Suppose that at time t, the actual market price of this option is given by V_t. The *cumulative* *implied volatility at time t* is defined to be the unique solution Ξ_t of

$$V_t = C(S_t, \Xi_t, r(T-t)). \tag{5.3}$$

An estimate of Ξ_t can be obtained given the paired data $(S, V)_t$. In above we have assumed S, V and Ξ are *Ito processes*, whose definition and regularity conditions are provided in Section 2.1.

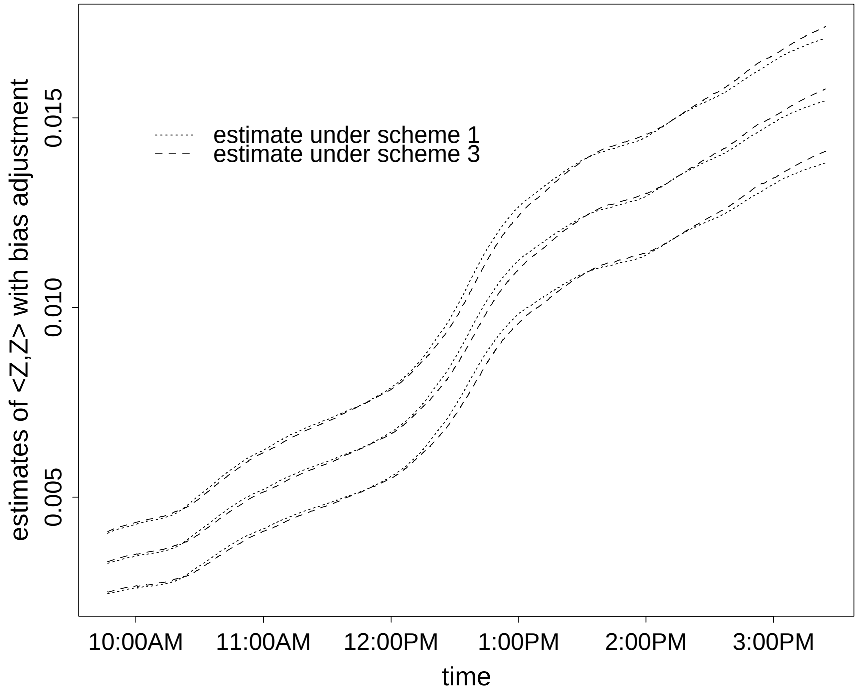
Note that our notion of implied volatility is **always** on the cumulative scale (from t to expiration T). This is in contrast to much of the literature, which considers implied instantaneous volatility (*e.g.*, Beckers (1981), Engle and Mustafa (1992), Bick (1995), Pena et al. (1999), and Rubinstein (1994)).

It should also be noted that implied volatility has a deeper meaning than what is generally supposed. Originally, implied (instantaneous) volatility is based on the GBM assumption for stock prices, which runs into real-case conflicts. However, substantial empirical work show that implied volatilities contain the information about future variability of S in a way that the past realized data cannot capture (Lamoureux and Lastrapes (1993), Jorion (1995), Christensen and Prabhala (1998), Gwilym and Buckle (1997), Blair and Taylor (2001)). In the conceptual front, implied volatility comes up in connection with bounds-based pricing (Mykland (2003, 2000)) and has a theoretical connection to actual volatility as seen in Zhang (2001) and in Mykland and Zhang (2001, 2002).

There are two purposes to carrying out an ANOVA of Ξ in a regression on S, as in (1.1). On the one hand, one can ask whether the model (5.1) is of one factor type as in Example 1, in the sense that the option V_t can be fully hedged in the stock S_t. This is a generalization of the usual one factor setup, a common assumption in the literature (Duffie (1996), Derman and Kani (1994)). As discussed in Mykland and Zhang (2001), this translates into asking whether the residual has the form H_0: $dZ_t = -\zeta_t dt$. Note that the null hypothesis is on the form Z^{MG*} = constant, or $< Z, Z >_T = 0$, as described in Section 4.5.

On the other hand, if H_0 is rejected, one wishes to know the magnitude of the hedging error. If discounted securities are given as $S^*_t = \exp(-rt)S_t$ and $V^*_t = \exp(-rt)V_t$, note that S^*_t can replace S_t in (1.1) without changing the analysis on $< Z, Z >$ (see discussion in Section 4.5). Also,

Figure 1: 90% Confidence Interval for $< Z, Z >$ of S&P 500 on Feb. 17, 1994, under P^*



following Mykland and Zhang (2001),

$$dV_t^* = (C_S + C_\Xi \rho) dS_t^* + C_\Xi dZ_t^{MG*}$$

under P^*. Under this distribution, therefore, the magnitude of Z^{MG*}, characterized by that of $< Z, Z >$, determines the amount of hedging error.

As a direct application, Theorem 1- 2 tell us how to estimate and to set confidence interval under P^* for the quadratic variation $< Z, Z >$ of the residual, and then determine the adequacy of a volatility model with one factor. As an example, we apply ANOVA for diffusions to the intra-day (tick-by-tick) data in SP&P 500 options and index. Figure 1 shows the pattern of intra-day, bias-corrected estimates $< Z, Z >^{(\alpha)}$, with $\alpha = 0$ (scheme 1) and $\frac{1}{2}$ (scheme 3) of Section 2.2. As shown in the figure, the bias-adjusted estimates of $< Z, Z >^{(\alpha)}$ behave similarly under different estimation schemes, in particular, they all significantly deviate from zero. This provides a strong evidence against one-factor volatility structure (or, rejecting H_0: $dZ_t = -\zeta_t dt$), including the state-dependent volatility model (see Duffie (1996), Derman and Kani (1994), for example).

The confidence interval in Figure 1 is pointwise. The discussion on joint confidence band for

$<\widehat{Z,Z}>$ is given in Section 4.1.

6. conclusion. This paper provides a methodology to analyze the association between diffusion-type processes. Its contribution is both conceptual and applied.

Under the general framework of nonparametric, one-factor regression, we obtain the distribution results for estimation errors in residual variation $< Z, Z >$ and then in a measure of *goodness of fit* R^2. The limiting distributions identify two sources of uncertainty, one from the discreteness nature of the data process, the other from the estimation procedure. Interestingly, among the class of estimators $<\widehat{Z,Z}>^{(\alpha)}$ under consideration, $\alpha \in [0,1]$, discrete-time sampling only impacts the "variance" component, on the other hand, different estimation schemes lead to different biases in the asymptotics.

ANOVA for diffusions permits inference over a time horizon. This is because the error terms in the quadratic variation $<\widehat{Z,Z}>^{(\alpha)}$ of residuals, and hence error terms in the estimated *Coefficient of Determination* R^2, converge as a process (in time) , whereas the errors in the estimated regression parameters $\hat{\rho}_t$ are asymptotically independent from one time point to the next. This feature of time aggregation makes ANOVA a natural procedure to determine the adequacy of an adopted model. Also, the ANOVA is better posed in that the rate of the convergence is the square of the rate for $\hat{\rho}_t - \rho_t$.

The "ANOVA for diffusions" approach is appealing also from the position of applications in finance. As in Example 1, it can test whether a financial derivative can be fully hedged in another asset. In the event of non-perfect hedging, Theorem 1-2 tell us how to quantify the amount of hedging error as well as its distribution. As in the example of Section 4.5, the current ANOVA approach helps in testing a volatility model regarding the number of factors involved, without assuming a specific functional form for the process $\{S_t\}$ and $\{\sigma_t^2\}$. Along the same line, the nonparametric feature of the current ANOVA permits a trading decision based on "data". The implementation of ANOVA is a dynamic one.

Another feature embedded with the financial application is that: the "ANOVA for diffusion" turns out to be a key messenger between the actual distribution P and the *Risk Neutral* distribution P^*. As detailed in Sections 4 and 5, the distributional results of residual variation are based on data (a realization of P), meanwhile it provides inference for hedging and valuing a financial derivative under P^*.

7. Convergence in law – Proofs and further results.

7.1. *Cumulative processes.*

In the following, we deal with processes that are exemplified by $[Z, Z] - < Z, Z >$. We mostly follow Jacod and Protter (1998).

Proof of Proposition 1. The applicable parts of the proof of the cited Theorems 5.1 and 5.5 of Jacod and Protter (1998) carry over directly under Assumption A. When modifying the proofs, as appropriate, t_* replaces $[tn]/n$, δ_n, replaces n^{-1}, and so on. For example, the right hand side of their equation (5.10) (p. 290) becomes $K\delta_n^2$. The main change due to the non-equidistant case occurs in part (iii) of Jacod and Protter's Lemma 5.3 (p. 291-292), where in the definition of α_n, $\frac{t}{2}B_r^{ik}$ should be replaced by $(H(t_r + t) - H(t_r))B_r^{ik}$. Assumption A is clearly sufficient. ∎

Note that the result extends in an obvious fashion to the case of multidimensional $Z = (Z^{(1)}, ..., Z^{(p)})$. Also, instead of studying $[Z, Z] - < Z, Z >$, one case, like Jacod and Protter (1998), state the result for $\int_0^t (Z_u^{(i)} - Z_*^{(i)}) dZ_u^{(j)}$.

In the sequel, we shall also be using a triangular array form of Proposition 1, cf. the end of the proofs of both Theorems 1 and 2.

Proposition 2. (Triangular array version of the discretization theorem). Let Z be a vector Ito process for which $\int_0^T || < Z, Z >_t' ||^2 dt < \infty$ and $\int_0^T \left|\left|\tilde{Z}_t\right|\right|^2 dt < \infty$ a.s. Also suppose that $Z^{(n)}$, $i = 1, 2, ...$ are Ito processes satisfying the same requirement, uniformly. Suppose that the (vector) Brownian motion W is the same in the Ito process representations of Z and of all the $Z^{(n)}$, i.e.,

$$dZ_u^{(n),MG} = \sigma_u^{(n)} dW \quad \text{and} \quad dZ_u^{MG} = \sigma_u dW. \tag{7.4}$$

Suppose that

$$\int_0^T \left|\left|\sigma_u^{(n)} - \sigma_u\right|\right|^4 du = o_p(1). \tag{7.5}$$

Then, subject to Assumption A, the processes $\frac{1}{\sqrt{\Delta t^{(n)}}} \int_0^t (Z_u^{(i,n)} - Z_*^{(i,n)}) dZ_u^{(j,n)}$ converge jointly with the processes $\frac{1}{\sqrt{\Delta t^{(n)}}} \int_0^t (Z_u^{(i)} - Z_*^{(i)}) dZ_u^{(j)}$ to the same limit. □

If one requires stable convergence, one just imposes Assumption (D), cf. Theorem 11.2 (pp.

338) and Theorem 15.2(c) (p. 496) of Jacod (1979).

PROOF OF PROPOSITION 2. This is mainly a takeoff on the development on p. 292 and the beginning of p. 293 in Jacod and Protter (1998), and the further development in (their) Theorem 5.5 is straightforward. Again we recollect that H (from Assumption A) is Lipschitz continuous.

Note that to match the end of the proof of Theorem 5.1, we really need $o_p(\delta_n^4)$. This, of course, follows by appropriate use of subsequences of subsequences. ∎

8. Proofs of main results. In order to motivate the analysis of $\hat{\rho}_t - \rho_t$, we here give a decomposition for

8.1. *Notation.*

In the following proofs, we sometimes write $< X, X >_t$ as $< X >_t$, and $< X, X >_t'$ as $< X >_t'$ for simplicity.

Also J& S is the shorthand for the reference Jacod and Shiryaev (1987).

In the cases where adapted processes X and Y are càdlàg, and Ito process except at grid points, we take $< X, Y >_t$ to mean the quadratic variation that comes from the continuous part only. See the definition of $\tilde{C}^1[0, T]$, as defined below.

Also, for convenience, we adopt the following shorthand for smoothness assumptions for Ito processes:

ASSUMPTION B (*Smoothness*):

$B.0(X)$: X is in $C^1[0, T]$.

$B.1(X, Y)$: $< X, Y >_t$ is in $C^1[0, T]$.

$B.2(X)$: the drift part of X (X^{DR}) is in $C^1[0, T]$.

Assumption $B(X)$ is equivalent to $B.1(X, X)$ and $B.2(X)$.

If $\{X^{(n))}\}, \{Y^{(n)}\}$, etc, are sequences of processes rather than fixed ones, we replace $C^1[0, T]$ in the smoothness conditions by $\tilde{C}^1[0,T]$, to be defined presently. Hence, for example, $B.1(X^{n)}, Y^{(n)})$ means that $< X^{n)}, Y^{(n)} >$ is in $\tilde{C}^1$.

A sequence $\{X^{(n)}\}$ of processes is said to be $\tilde{C}^1 = \tilde{C}^1[0,T]$ if each $X^{(n)}$ is continuously differentiable in each grid interval $[t_i^{(n)}, t_{i+1}^{(n)})$, where $\tilde{X}^{(n)}(t_i)$ is the right hand derivative at t_i. For $t_{k_n}^{(n)} = T$, we take the interval to be $[t_{k_n-1}^{(n)}, T]$. We also require that $\sup_{n,t} |\tilde{X}_t^{(n)}| < \infty$. All of the above is, of course, a.s.

We shall be using the following notations

$$\Upsilon^X(h) = \sup_{t-h \le u \le s \le t} |X_u - X_s| \tag{8.1}$$

$$\Upsilon^{XY}(h) = \sup_{t-h \le u \le s \le t} |< X, Y >_u' - < X, Y >_s'| \tag{8.2}$$

Assumption $B.1(X,Y)$ implies $\Upsilon^{XY}(h) \to 0$. Moreover, $\Upsilon^X(h) = o_p(1)$ in h, when X is an Ito process. More precise orders are given in Lemma 2.

8.2. *Proof of results: lemmas and corollaries.*

Lemma 1. Let $M_{i,n}(t)$, $0 \le t \le T$, $i = 1, \cdots, k_n$, $k_n = O((\overline{\Delta t}^{(n)})^{-1})$, be a collection of continuous local martingales. Suppose that

$$\sup_{1 \le i \le k_n} < M_{i,n}, M_{i,n} >_T = O_p((\overline{\Delta t}^{(n)})^{\beta}),$$

Then, for any $\epsilon > 0$,

$$\sup_{1 \le i \le k_n} \sup_{0 \le t \le T} |M_{i,n}(t)| = O_p((\overline{\Delta t}^{(n)})^{\frac{\beta}{2} - \epsilon})$$

□

Proof of Lemma 1:

Let $\alpha > 4$, to be determined later.

We shall use

$$\sup_{1 \le i \le k_n} \sup_t |M_{i,n}(t)|^\alpha \le \sup_t \sum_{i=1}^{k_n} |M_{i,n}(t)|^\alpha \tag{8.1}$$

and the right hand side (r.h.s. from now on) of Equation (8.1) is L-dominanted (see, Jacod and Shiryaev (1987), p. 35) by

$$A_n(t) = c_\alpha \sum_{i=1}^{k_n} < M_{i,n}, M_{i,n} >_t^{\alpha/2}, \tag{8.2}$$

following Burkholder's inequality (see section 3 of Chapter VII of Dellacherie and Meyer (1982), Barlow et al. (1986), and Protter (1995)), where c_α is the constant from this inequality.

From Equation (8.1) and Lenglart's inequality (Jacod and Shiryaev (1987), p. 35),

$$P\left(\sup_{1\leq i\leq k_n} \sup_t |M_{i,n}(t)|^\alpha \geq K_1\right) \leq P\left(\sup_t \sum_{i=1}^{k_n} |M_{i,n}(t)|^\alpha \geq K_1\right) \leq \frac{K_2}{K_1} + P\left(A_n(T) \geq K_2\right) \tag{8.3}$$

The cited result requires the r.h.s. of Equation (8.1) and $(A_n(T))$ to be integrable, but since both processes are continuous, this requirement can be removed by localization.

To bound the r.h.s. of Equation (8.3), note that

$$\begin{aligned} A_n(T) &\leq c_\alpha k_n \sup_{1\leq i\leq k_n} < M_{i,n}, M_{i,n} >_T^{\alpha/2} \\ &= O_p((\overline{\Delta t}^{(n)})^{\frac{\alpha\beta}{2}-1}) \end{aligned}$$

by the assumption of the lemma. Now set $K_i = K_i'(\overline{\Delta t}^{(n)})^{\frac{\alpha\beta}{2}-1}$, for $i = 1, 2$. Equation (8.3) then yields that

$$\begin{aligned} &P\left(\sup_{1\leq i\leq k_n} \sup_t |M_{i,n}(t)| \geq (K_1')^{\frac{1}{\alpha}} (\overline{\Delta t}^{(n)})^{\frac{\beta}{2}-\frac{1}{\alpha}}\right) \\ \leq\ & \frac{K_2'}{K_1'} + P\left(O_p(1) \geq K_2'\right) \end{aligned}$$

By choosing K_2' large and K_1' even larger, it follows that

$$\sup_{1\leq i\leq k_n} \sup_t |M_{i,n}(t)| = O_p((\overline{\Delta t}^{(n)})^{\frac{\beta}{2}-\frac{1}{\alpha}}).$$

Since α can be arbitrarily large, the result follows. ∎

Lemma 2. Let X and Y be Ito processes, and let Assumption A hold. If X satisfies Assumption $B.1(X,X)$ and $B.2(X)$, then for any $\epsilon > 0$,

$$\Upsilon^X(\eta) = O_p((\eta + \delta^{(n)})^{\frac{1}{2}-\epsilon}).$$

Similarly, if X and Y satisfy Assumption $B.1(< X,Y >', < X,Y >')$ and $B.2(< X,Y >')$, then for any $\epsilon > 0$,

$$\Upsilon^{XY}(\eta) = O_p\left((\eta + \delta^{(n)})^{\frac{1}{2}-\epsilon}\right),$$

where Υ^X and Υ^{XY} are the same as defined at the beginning of the appendix, and $\delta^{(n)}$ is the mesh of the partition. □

Proof of Lemma 2:

$$\begin{aligned}\Upsilon^X(\eta) &= \sup_{0\le u,s\le T;|u-s|\le\eta} |X_u - X_s| \\ &\le \eta \sup_{0\le s\le T} |\tilde{X}_s| + \Upsilon^{X^{MG}}(\eta)\end{aligned}$$

For X^{MG}, set

$$M_{i,n}(s) = \begin{cases} X^{MG}_{t_i+\eta+\delta^{(n)}} - X^{MG}_{t_i} & s \ge t_i + \eta + \delta^{(n)} \\ X^{MG}_s - X^{MG}_{t_i} & t_i \le s \le t_i + \eta + \delta^{(n)} \\ 0 & s \le t_i \end{cases}$$

Let t_i be the closest grid point below $min(u,s)$.

Then

$$\begin{aligned}|X^{MG}_u - X^{MG}_s| &\le |X^{MG}_u - X^{MG}_{t_i}| + |X^{MG}_s - X^{MG}_{t_i}| \\ &= |M_{i,n}(u)| + |M_{i,n}(s)|\end{aligned}$$

since $\max(u,s) \le \min(u,s) + \eta \le t_i + \delta^{(n)} + \eta$. Hence, $\Upsilon^{X^{MG}}(\eta) \le 2\sup_i \sup_s |M_{i,n}(s)|$.

Now

$$\sup_i < M_{i,n} >_T \le \sup_s < X >'_s (\eta + \delta^{(n)}) = O_p(\eta + \delta^{(n)})$$

by $B.1(X,X)$.

Hence $\sup_i \sup_s |M_{i,n}(s)| = O_p((\eta + \delta^{(n)})^{\frac{1}{2}-\epsilon})$ for any $\epsilon > 0$, by Lemma 1. Further by $B.2(X)$, the conclusion follows. Similarly, the result follows for $\Upsilon^{XY}(\eta)$. ■

Lemma 3. Suppose X, Y and Z are Ito processes. Subject to assumptions A, $B.0(Y)$, $B[(X),(Z)]$, and $B.1(X,Z)$, we have the following for any $\epsilon > 0$, for either $h \to 0$ or $h = t$, with constant $k > 0$

(i)

$$\sup_t \frac{1}{h^2} \left| \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \left(\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i})(Z_u - Z_{t_i})(u - t_i)^k dY_u - \frac{1}{k+2} < X, Z >'_{t_i} \tilde{Y}_{t_i} (\Delta t_i)^{k+2} \right) \right|$$

$$= o_p\left(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h}\right)$$

(ii) In particular,

$$\sup_t \left| \frac{1}{\overline{\Delta t}^{(n)}} \sum_{t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i})(Z_u - Z_{t_i}) dY_u - \frac{1}{2} \int_0^t < X, Z >'_u \tilde{Y}_u dH_u \right| \xrightarrow{P} 0$$

and for $k > 0$,

$$\sup_t \left| \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i})(Z_u - Z_{t_i})(u - t_i)^k dY_u \right| = o_p\left(\left(\overline{\Delta t}^{(n)}\right)^{k+1} h \right)$$

where $\tilde{Y}_u = dY_u/du$. □

PROOF OF LEMMA 3:

Without loss of generality, it is enough to show the result for $X = Z$. This is because one can prove the results for X, Z and $X + Z$, and then proceed via the polarization identity. The conditions imposed also means that the assumptions of Lemma 3 are also satisfied for $X + Z$.

We here prove (i). The first part of (ii) follows since $< X >'$ and Y are continuous and H_u is Lipschitz continuous (under Assumption A). The second part of (ii) follows by simple order considerations.

(a) We first show that

$$\sup_t \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \left| \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (< X >_u - < X >_{t_i})(u - t_i)^k dY_u - \frac{1}{k+2} < X >'_{t_i} \tilde{Y}_{t_i} (\Delta t_i)^{k+2} \right|$$

$$= o_p\left(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h}\right)$$

Notice that the l.h.s. of the above equation is equal to

$$\sup_t \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \left| (<X>_u - <X>_{t_i})(u-t_i)^k \tilde{Y}_u - <X>'_{t_i} \tilde{Y}_{t_i}(u-t_i)^{k+1} \right| du \tag{8.4}$$

Let

$$f_n(u) = (u-t_i)^k [(<X>_u - <X>_{t_i})\tilde{Y}_u - <X>'_{t_i} \tilde{Y}_{t_i}(u-t_i)],$$

by adding and subtracting $(u-t_i)^k(<X>_u - <X>_{t_i})\tilde{Y}_{t_i}$ in $f_n(u)$, we get

$$\begin{aligned}
& |f_n(u)| \\
\le\ & (u-t_i)^k \left[\Upsilon^{\tilde{Y}}(\delta^{(n)}) | <X>_u - <X>_{t_i} | + |\tilde{Y}_{t_i}[(<X>_u - <X>_{t_i}) - <X>'_{t_i} (u-t_i)]| \right] \\
\le\ & \left(\delta^{(n)}\right)^{k+1} \left[\Upsilon^{\tilde{Y}}(\delta^{(n)}) \sup_u <X>'_u + \sup_i |\tilde{Y}_{t_i}| \Upsilon^{<X>'}(\delta^{(n)}) \right] \\
=\ & o_p\left((\overline{\Delta t}^{(n)})^{k+1} \right)
\end{aligned}$$

uniformly in u, by assumption $B.0(Y)$ and $B.1(X,X)$. Hence (8.4) is $o_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h})$.

(b) Let

$$II_t = \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \left[(X_u - X_{t_i})^2 - (<X>_u - <X>_{t_i}) \right] (u-t_i)^k dY_u$$

We now show

$$\sup_t |II_t| = o_p\left(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h}\right) \tag{8.5}$$

By Itô's Lemma,

$$\begin{aligned}
|II_t| &= \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \left[2 \int_{t_i}^{u} (X_v - X_{t_i}) dX_v \right] (u-t_i)^k dY_u \\
&= \underbrace{\frac{2}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \int_{t_i}^{u} (X_v - X_{t_i}) dX_v^{DR} (u-t_i)^k dY_u}_{II_{t,1}} \\
&\quad \underbrace{+ \frac{2}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} [\int_{t_i}^{u} (X_v - X_{t_i}) dX_v^{MG}] (u-t_i)^k dY_u}_{II_{t,2}}
\end{aligned}$$

where $X = X^{MG} + X^{DR}$ was used above.

Recall that $dX_v^{DR} = \tilde{X}_v dv$, then

$$|II_{t,1}| \leq \sup_{0\leq u\leq t} |\tilde{X}_u| \sup_{0\leq u\leq t} |\tilde{Y}_u| \Upsilon^X(\delta^{(n)}) \frac{2}{h^2} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (u-t_i)^{k+1} du = o_p(\frac{(\delta^{(n)})^{k+1}}{h})$$

by $B.0(Y)$, $B.2(X)$ and the continuity of X.

For $II_{t,2}$, we first let $f_k(u)$ denote any sum of terms of the form $c(\Delta t_i)^\alpha (u-t_i)^\beta$, where $\alpha+\beta = k$ and α, $\beta \geq 0$. By integration by parts,

$$\begin{aligned} & \frac{1}{h^2} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} [\int_{t_i}^{u} (X_v - X_{t_i}) dX_v^{MG}](u-t_i)^k du \\ = & \frac{1}{k+1} \frac{1}{h^2} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i}) \underbrace{[(\Delta t_i)^{k+1} - (u-t_i)^{k+1}]}_{f_{k+1}(u)} dX_u^{MG} \qquad (8.6) \end{aligned}$$

(8.6) has q.v. bounded by

$$\begin{aligned} & \frac{1}{(k+1)^2} \frac{1}{h^4} \sup_i \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i})^2 f_{k+1}^2(u) d < X >_u \\ \leq & \frac{1}{(k+1)^2} \frac{1}{h^4} \sup_i < X >'_{t_i} \sup_t \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (\Upsilon^X(\delta^{(n)})^2 f_{2k+2}(u) du \\ = & O_p(\frac{(\overline{\Delta t}^{(n)})^{2k+3-\epsilon}}{h^3}) \qquad (8.7) \end{aligned}$$

by Lemma 2 under Assumption A, $B.1(X,X)$ and $B.2(X)$. Following Lemma 1 and $B.0(Y)$.

$$\sup_t |II_{t,2}| = O_p(\frac{(\overline{\Delta t}^{(n)})^{k+3/2-\epsilon}}{h^{3/2}}).$$

∎

DEFINITION 2. Suppose X and Y are continuous Ito processes. Let

$$B_{1,i,t}^{XY} = \begin{cases} \frac{1}{h} \int_{t_i-h}^{t\wedge t_i} ((t_i - h) - u) d < X, Y >'_u & t \geq t_i - h \\ 0 & \text{otherwise} \end{cases} \qquad (8.8)$$

and

$$B_{2,i,t}^{XY} = \begin{cases} \frac{[2]}{h} \sum_{t_i-h\le t_j\le t_{j+1}\le t_i\wedge t} \int_{t_j}^{t_{j+1}} (X_s - X_{t_j})dY_s & t \ge t_i - h \\ 0 & \text{otherwise} \end{cases} \tag{8.9}$$

where [2] indicates symmetric representation s.t. $[2]\int XdY = \int XdY + \int YdX$. □

Note that by integration by parts via Ito's Lemma,

$$B_{1,i,t_i}^{XY} = \frac{1}{h}(<X,Y>_{t_i} - <X,Y>_{t_i-h}) - <X,Y>'_{t_i}$$

and hence $\widehat{<X,Y>}'_{t_i} - <X,Y>'_{t_i} = B_{1,i,t_i}^{XY} + B_{2,i,t_i}^{XY}$.

LEMMA 4. Under Assumptions A, $B(X,Y,<X,Y>')$, and the order selection of $h^2 = O(\overline{\Delta t}^{(n)})$, for any $\epsilon > 0$,

$$\sup_{0\le t_i\le T} \left|B_{1,i,t_i}^{XY}\right| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{4}-\epsilon}) \quad \text{and} \quad \sup_{0\le t_i\le T} \left|B_{2,i,t_i}^{XY}\right| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{4}-\epsilon})$$

In particular,

$$\sup_{0\le t_i\le T} |\widehat{<X,Y>}'_{t_i} - <X,Y>'_{t_i}| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{4}-\epsilon})$$

In addition, under condition $B(<R^{XY}, R^{ZV}>')$.

$$\sup_{t_i} \left| <B_{1,i}^{XY}, B_{1,i}^{ZV}>_{t_i} - \frac{h}{3} <R^{XY}, R^{ZV}>'_{t_i} \right| = O_p(h^{3/2-\epsilon}), \tag{8.10}$$

also,

$$\sup_{t} \left| <B_{2,i}^{XY}, B_{2,i}^{ZV}>_t - \frac{1}{h^2} \sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t} (<X,Z>'_{t_i}<Y,V>'_{t_i} + <X,V>'_{t_i}<Y,Z>'_{t_i})(\Delta t_i)^2 \right|$$
$$= o_p((\overline{\Delta t}^{(n)})^{\frac{1}{2}}) \tag{8.11}$$

and

$$\sup_{t_i} \left| <B_{1,i}^{ZV}, B_{2,i}^{XY}>_{t_i} \right| = O_p(\frac{\overline{\Delta t}^{(n)}}{\sqrt{h}}) \tag{8.12}$$

□

PROOF OF LEMMA 4:

(i) Suppose we decompose $< X, Y >'_t$ into a martingale part (R_t^{XY}) and a drift part (D_t^{XY}) which is differentiable with respect to t,

$$B_{1,i,t}^{XY} = \underbrace{\frac{1}{h}\int_{t_i-h}^{t}((t-h)-u)dR_u^{XY}}_{B_{1,i,t}^{XY,MG}} + \underbrace{\frac{1}{h}\int_{t_i-h}^{t}((t-h)-u)dD_u^{XY}}_{B_{1,i,t}^{XY,DR}} \tag{8.13}$$

Note that, naturally, under assumption $B.2(< X, Y >')$

$$\sup_i |B_{1,i,t_i}^{XY,DR}| = O_p(h). \tag{8.14}$$

Under Assumptions A and $B.1(R^{XY}, R^{XY})$, hence using $\sup_{0\le u\le T} < R^{XY} >'_u = O_p(1)$,

$$\begin{aligned}\sup_i < B_{1,i}^{XY}, B_{1,i}^{XY} >_T &\le \sup_{0\le u\le T} < R^{XY} >'_u \sup_i \frac{1}{h^2}\int_{t_i-h}^{t_i}(t_i-h-u)^2 du \\ &= O_p(h) = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{2}})\end{aligned}$$

So $\sup_i \sup_t |B_{1,i,t}^{XY,MG}| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{4}-\epsilon})$ by Lemma 1, and hence $\sup_i \left|B_{1,t_i}^{XY}\right| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{4}-\epsilon})$.

Next we examine $\sup_i |B_{2,t_i}^{XY}|$,

$$\begin{aligned}\left|\frac{1}{h}\sum_{t_i-h\le t_j\le t_{j+1}\le t_i}\int_{t_j}^{t_{j+1}}(X_s - X_{t_j})dY_s\right| &\le \underbrace{\left|\frac{1}{h}\sum_{t_i-h\le t_j\le t_{j+1}\le t_i}\int_{t_j}^{t_{j+1}}(X_s - X_{t_j})dY_s^{MG}\right|}_{F_1} \\ &+ \underbrace{\left|\frac{1}{h}\sum_{t_i-h\le t_j\le t_{j+1}\le t_i}\int_{t_j}^{t_{j+1}}(X_s - X_{t_j})dY_s^{DR}\right|}_{F_2}\end{aligned}$$

Consider F_1 first. Let t_* be the largest $t_j \le t$. Set

$$M_{i,n}^{XY}(t) = \begin{cases} M_{i,n}^{XY}(t_i) & t \ge t_i \\ \frac{1}{h}\sum_{t_i-h\le t_j\le t_{j+1}\le t}\int_{t_j}^{t_{i+1}}(X_s - X_{t_j})dY_s^{MG} + \frac{1}{h}\int_{t_*}^{t}(X_s - X_{t_*})dY_s^{MG} & t_i - h \le t \le t_i \\ 0 & t \le t_i - h \end{cases} \tag{8.15}$$

Then

$$
\begin{aligned}
< M_{i,n}, M_{i,n} >_T &\leq \frac{1}{h^2} \sum_{t_i - h \leq t_j \leq t_{j+1} \leq t_i} \int_{t_j}^{t_j+1} (X_s - X_{t_j})^2 ds \sup_{0 \leq s \leq T} < Y, Y >'_s \\
&\leq \frac{1}{h} \left(\Upsilon^X (\delta^{(n)}) \right)^2 \sup_{0 \leq s \leq T} < Y, Y >'_s \\
&= \frac{1}{h} O_p((\overline{\Delta t}^{(n)})^{1-\epsilon}) = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{2}-\epsilon})
\end{aligned}
$$

by Lemma 2, under Assumption A, $B.1[(X,X),(Y,Y)]$ and $B.2(X)$. So $\sup_i |M_{i,n}| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{4}-\epsilon})$ by Lemma 1. It then follows that

$$
\sup_i |B_{2,t_i}^{XY,MG}| = \sup_i |\frac{[2]}{h} \sum_{t_i - h \leq t_j \leq t_{j+1} \leq t_i} \int_{t_j}^{t_j+1} (X_s - X_{t_j}) dY_s^{MG}| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{4}-\epsilon}) \tag{8.16}
$$

For F_2,

$$
\sup_i |F_2| \leq \Upsilon^X (\delta^{(n)}) \sup_s |\tilde{Y}_s| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{2}-\tilde{\epsilon}})
$$

by Lemma 2, under Assumption $B.1(X,X)$ and $B.2[(X)(Y)]$. Therefore,

$$
\sup_i |B_{2,t_i}^{XY,DR}| = \sup_i |\frac{[2]}{h} \sum_{t_i - h \leq t_j \leq t_{j+1} \leq t_i} \int_{t_j}^{t_j+1} (X_s - X_{t_j}) dY_s^{DR}| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{2}-\tilde{\epsilon}}) \tag{8.17}
$$

Hence the result of the lemma follows.

(ii)

Following the definition of $B_{1,i,t}^{XY}$ in (8.8),

$$
\begin{aligned}
& \sup_{t_i} |< B_1^{XY,MG}, B_1^{ZV,MG} >_{t_i} - \frac{h}{3} < R^{XY}, R^{ZV} >'_{t_i} | \\
= & \sup_{t_i} |\frac{1}{h^2} \int_{t_i - h}^{t_i} (t_i - h - u)^2 (< R^{XY}, R^{ZV} >'_u - < R^{XY}, R^{ZV} >'_{t_i}) du \\
\leq & \frac{h}{3} \Upsilon^{< R^{XY}, R^{ZV} >'}(h)
\end{aligned}
$$

Hence (8.10) follows by Lemma 2, under Assumption $B.1(< R^{XY}, R^{ZV} >', < R^{XY}, R^{ZV} >')$ and $B.2(< R^{XY}, R^{ZV} >')$.

Equation (8.11) follows directly from Lemma 3.

Following the proof of Theorem 1d in Mykland and Zhang (2002),

$$< B_{1,i}^{ZV}, B_{2,i}^{XY} >_t = \underbrace{\frac{1}{h^2} \sum_{t_i-h \le t_j \le t_{j+1} \le t} \int_{t_j}^{t_{j+1}} (X_s - X_{t_j})(t_i - h - s) d < R^{ZV}, Y >_s}_{Q_t^{X,Y}}$$
$$+ \underbrace{\frac{1}{h^2} \sum_{t_i-h \le t_j \le t_{j+1} \le t} \int_{t_j}^{t_{j+1}} (Y_s - Y_{t_j})(t_i - h - s) d < R^{ZV}, X >_s}_{Q_t^{Y,X}}$$

for any $t_i - h \le t \le t_i$, otherwise stopped at t_i. Enough to show that $sup_i |Q_{t_i}^{X,Y}| = o_p(\sqrt{\Delta t^{(n)}})$.

Let

$$Q*_t^{X,Y} = \frac{1}{h^2} \sum_{t_i-h \le t_j \le t_{j+1} \le t} \int_{t_j}^{t_{j+1}} (X_s - X_{t_j})(t_i - h - s)(< R^{ZV}, Y >'_s - < R^{ZV}, Y >'_{t_j}) ds.$$

$$\begin{aligned} Q_t^{X,Y} &= \frac{1}{h^2} \sum_{t_i-h \le t_j \le t_{j+1} \le t} < R^{ZV}, Y >'_{t_j} \int_{t_j}^{t_{j+1}} (X_s - X_{t_j})(t_i - h - s) ds + Q*_t^{X,Y} \\ &= \frac{1}{h^2} \sum_{t_i-h \le t_j \le t_{j+1} \le t} < R^{ZV}, Y >'_{t_j} \int_{t_j}^{t_{j+1}} \left[\int_s^{t_{j+1}} (t_i - h - u) du \right] dX_s + Q*_t^{X,Y} \\ &= \frac{1}{h^2} \sum_{t_i-h \le t_j \le t_{j+1} \le t} < R^{ZV}, Y >'_{t_j} \int_{t_j}^{t_{j+1}} \left[\int_s^{t_{j+1}} (t_i - h - u) du \right] dX_s^{MG} \\ &+ \frac{1}{h^2} \sum_{t_i-h \le t_j \le t_{j+1} \le t} < R^{ZV}, Y >'_{t_j} \int_{t_j}^{t_{j+1}} \left[\int_s^{t_{j+1}} (t_i - h - u) du \right] dX_s^{DR} + Q*_t^{X,Y} \end{aligned} \quad (8.18)$$

where the last equality follows from integration by parts. Note that the first term on the r.h.s. of (8.18) has q.v.

$$\frac{1}{h^4} \sum_{t_i-h \le t_j \le t_{j+1} \le t} (< R^{ZV}, Y >'_{t_j})^2 \int_{t_j}^{t_{j+1}} \left[\int_s^{t_{j+1}} (t_i - h - u) du \right]^2 d < X >_s$$
$$\le \sup_s < X >'_s (\sup_j < R^{ZV}, Y >'_{t_j})^2 \frac{(\delta^{(n)})^2}{h}$$

Also, the second term on the r.h.s. of (8.18) is bounded by

$$\frac{\delta^{(n)}}{2} \sup_t |\tilde{X}_s| \sup_j |< R^{ZV}, Y >'_{t_j}|$$

and the third term $Q*_t^{X,Y}$ is bounded by $\Upsilon^X(\delta^{(n)}) \Upsilon^{<R^{ZV},Y>'}(\delta^{(n)})$. Thus by Lemma 2, $Q_{t_i}^{X,Y}$ is of order $O_p(\frac{\overline{\Delta t^{(n)}}}{\sqrt{h}})$ uniformly in t, under Assumption A, and $B.1[(X,X),(R^{ZV},Y)]$, $B.2(X)$ and and those in Lemma 2. ∎

Corollary 1. (Linearization of $\hat{\rho}$). Suppose Ξ, S, and ρ are Ito processes, where ρ and $\hat{\rho}$ are as defined in Section 2. Define

$$\begin{aligned} L_{t_i} &= \frac{1}{<S,S>'_{t_i}}[(\widehat{<\Xi,S>}'_{t_i} - <\Xi,S>'_{t_i}) - \rho_{t_i}(\widehat{<S,S>}'_{t_i} - <S,S>'_{t_i})] \\ &= \frac{1}{<S,S>'_{t_i}}\sum_{j=1}^{2}(B^{\Xi S}_{j,t_i} - \rho_{t_i}B^{SS}_{j,t_i}). \end{aligned}$$

Then, under conditions A, $B(\Xi, S, <\Xi,S>', <S,S>')$, and $C(S)$, for any $\epsilon > 0$, we have

$$\sup_{t_i\in[0,T]} |\hat{\rho}_{t_i} - \rho_{t_i} - L_{t_i}| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{2}-\epsilon}) \tag{8.19}$$

In particular, from Lemma 4,

$$\sup_{t_i\in[0,T]} |\hat{\rho}_{t_i} - \rho_{t_i}| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{4}-\epsilon})$$

□

Proof of Corollary 1:

Recall that

$$\hat{\rho}_{t_i} - \rho_{t_i} = \frac{\widehat{<\Xi,S>}'_{t_i}}{\widehat{<S,S>}'_{t_i}} - \frac{<\Xi,S>'_{t_i}}{<S,S>'_{t_i}} \tag{8.20}$$

By Taylor expansion of $f(\widehat{<S,S>}'_{t_i})$ at $<S,S>'_{t_i}$, where $f(\widehat{<S,S>}'_{t_i}) = \frac{1}{\widehat{<S,S>}'_{t_i}}$

$$\begin{aligned} &\hat{\rho}_{t_i} - \rho_{t_i} \\ &= \frac{\widehat{<\Xi,S>}'_{t_i} - <\Xi,S>'_{t_i}}{<S,S>'_{t_i}} - \frac{\widehat{<S,S>}'_{t_i} - <S,S>'_{t_i}}{(<S,S>'_{t_i})^2}\widehat{<\Xi,S>}'_{t_i} \\ &\quad + \frac{(\widehat{<S,S>}'_{t_i} - <S,S>'_{t_i})^2}{\xi^3_{t_i}}\widehat{<\Xi,S>}'_{t_i} \\ &= L_{t_i} - \underbrace{\frac{1}{(<S,S>'_{t_i})^2}(\widehat{<S,S>}'_{t_i} - <S,S>'_{t_i})(\widehat{<\Xi,S>}'_{t_i} - <\Xi,S>'_{t_i})}_{R_1} \\ &\quad + \underbrace{\frac{1}{\xi^3_{t_i}}(\widehat{<S,S>}'_{t_i} - <S,S>'_{t_i})^2 <\Xi,S>'_{t_i}}_{R_2} \\ &\quad + \underbrace{\frac{1}{\xi^3_{t_i}}(\widehat{<S,S>}'_{t_i} - <S,S>'_{t_i})^2(\widehat{<\Xi,S>}'_{t_i} - <\Xi,S>'_{t_i})}_{R_3} \end{aligned}$$

where ξ_{t_i} is between $<\widehat{S,S}>'_{t_i}$ and $<S,S>'_{t_i}$.

Now $\forall c'>0$, on set $A=\{\inf_{u\in[0,T]}<S,S>'_u\geq c'\}$,

$$\begin{aligned}
& I_A\cdot\sup_{t_i}|L_{t_i}| \\
\leq\ & I_A\cdot\sup_{t_i}\left|\frac{<\widehat{\Xi,S}>'_{t_i}-<\Xi,S>'_{t_i}}{<S,S>'_{t_i}}\right|+I_A\cdot\sup_{t_i}\left|\rho_{t_i}\frac{<\widehat{S,S}>'_{t_i}-<S,S>'_{t_i}}{<S,S>'_{t_i}}\right| \\
=\ & O_p((\overline{\Delta t}^{(n)})^{1/4-\epsilon}) \qquad (8.21)
\end{aligned}$$

Equation (8.21) follows from Lemma 4, under Assumption A, $B.1[(SS),(\Xi\Xi)]$, and $B.2[(S),(\Xi)]$.

Similarly $\sup_{t_i}|R_1|=O_p((\overline{\Delta t}^{(n)})^{1/2-\epsilon})$ on set A. Next under the assumptions of A, $B.1[(SS),(\Xi\Xi)]$, and $B.2(S)$ (note that $B.1(\Xi S)$ follows from $B.1(S,S)$ and the continuity of ρ, and the relationship $d\Xi=\rho dS+dZ$.

$$\begin{aligned}
I_A\cdot\sup_{t_i}|R_2| &\leq I_A\cdot\frac{1}{\inf_{t_i}|\xi_{t_i}|^3}\sup_{t_i}<\Xi,S>'_{t_i}\sup_{t_i}(<\widehat{S,S}>'_{t_i}-<S,S>'_{t_i})^2 \\
&= O_p((\overline{\Delta t}^{(n)})^{1/2-\epsilon}) \qquad (8.22)
\end{aligned}$$

Equation (8.22) follows from Lemma 4 and the next result: $\sup_i\frac{1}{|\xi_{t_i}|}=O_P(1)$, which is proved in the following.

Since ξ_{t_i} is between $<\widehat{S,S}>'_{t_i}$ and $<S,S>'_{t_i}$, on set A we have

$$\begin{aligned}
\inf_{t_i}|\xi_{t_i}| &\geq \inf\left[<S,S>'_{t_i}-|<\widehat{S,S}>'_{t_i}-<S,S>'_{t_i}|\right] \\
&\geq \inf_{t_i}<S,S>'_{t_i}-\sup_{t_i}|<\widehat{S,S}>'_{t_i}-<S,S>'_{t_i}| \\
&\geq c'-\sup_{t_i}|<\widehat{S,S}>'_{t_i}-<S,S>'_{t_i}| \qquad (8.23)
\end{aligned}$$

hence,

$$\begin{aligned}
& P(\sup_{t_i}\frac{1}{|\xi_{t_i}|}\geq K)=P(\inf_{t_i}|\xi_{t_i}|\leq\frac{1}{K}) \\
\leq\ & P(\{\inf_{t_i}|\xi_{t_i}|\leq\frac{1}{K}\}\cap A)+P(\bar{A}) \\
\overset{(8.23)}{\leq}\ & P(\{c'-\sup_{t_i}|<\widehat{S,S}>'_{t_i}-<S,S>'_{t_i}|\leq\frac{1}{K}\}\cap A)+P(\bar{A}) \\
\leq\ & P(\{\sup_{t_i}|<\widehat{S,S}>'_{t_i}-<S,S>'_{t_i}|\geq c'-\frac{1}{K}\})+P(\bar{A})
\end{aligned}$$

$\forall\epsilon > 0$, since by assumption $C(S)$,

$$P\{\inf_{t\in[0,T]} < S,S >_t' < c'\} \to P\{\inf_{t\in[0,T]} < S,S >_t' = 0\} = 0 \text{ as } c' \to 0,$$

we can choose $c' s.t. P(\bar{A}) = P\{\inf_{t\in[0,T]} < S,S >_t' < c'\} \le \frac{\epsilon}{2}$.

From Lemma 4, $\sup_{t_i} |\widehat{< S,S >}_{t_i}' - < S,S >_{t_i}' | = o_p(1)$, thus we can choose K,where $\frac{1}{K} < c'$, $s.t. P(\sup_{t_i} |\widehat{< S,S >}_{t_i}' - < S,S >_{t_i}' | \ge c' - \frac{1}{K}) \le \frac{\epsilon}{2}$, therefore, $P(\sup_{t_i} \frac{1}{|\xi_{t_i}|} \ge K) \le \epsilon$, that is, $\sup_{t_i} \frac{1}{|\xi_{t_i}|} = O_p(1)$.

Similarly, $\sup_{t_i} |R_3| = O_p((\overline{\Delta t}^{(n)})^{3/4-3/r})$ on set A.

$$I_A \cdot \sup_{t_i} |\hat{\rho}_{t_i} - \rho_{t_i} - L_{t_i}| = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{2}-\epsilon})$$

$\forall\epsilon > 0, \forall\delta > 0$, we can choose c' s.t. $P(\bar{A}) \le \delta$, thus

$$\begin{aligned} P(\frac{1}{(\overline{\Delta t}^{(n)})^{\frac{1}{2}-\epsilon}} \sup_{t_i} |\hat{\rho}_{t_i} - \rho_{t_i} - L_{t_i}| > \epsilon) &\le P(\frac{1}{(\overline{\Delta t}^{(n)})^{\frac{1}{2}-\epsilon}} I_A \sup_{t_i} |\hat{\rho}_{t_i} - \rho_{t_i} - L_{t_i}| > \epsilon) + P(\bar{A}) \\ &\to 0 + P(\bar{A}) \text{ as } n \to \infty \\ &\le \delta \end{aligned}$$

since δ is arbitrary, $\frac{1}{(\overline{\Delta t}^{(n)})^{\frac{1}{2}-\epsilon}} \sup_{t_i} |\hat{\rho}_{t_i} - \rho_{t_i} - L_{t_i}| = O_p(1)$ ∎

8.3. *Proof of the first theorem.*

Proof of Theorem 1:

Note that the assumptions of the theorem imply those of Corollary 1.

(i) By the definition of $\hat{Z}$,

$$\begin{aligned} &\sum_{t_{i+1}^{(n)} \le t} \Delta < \hat{Z},\hat{Z} >_{t_i^{(n)}\wedge t} - \sum_{t_{i+1}^{(n)} \le t} \Delta < Z,Z >_{t_i^{(n)}\wedge t} \\ &= \sum_{t_{i+1}^{(n)} \le t} [\Delta < \Xi,\Xi >_{t_i^{(n)}\wedge t} - 2\int_{t_i^{(n)}\wedge t}^{t_{i+1}^{(n)}\wedge t} \hat{\rho}_u d < \Xi,S >_u + \int_{t_i^{(n)}\wedge t}^{t_{i+1}^{(n)}\wedge t} \hat{\rho}_u^2 d < S,S >_u] \\ &\quad - \sum_{t_{i+1}^{(n)} \le t} [\Delta < \Xi,\Xi >_{t_i^{(n)}\wedge t} - 2\int_{t_i^{(n)}\wedge t}^{t_{i+1}^{(n)}\wedge t} \rho_u d < \Xi,S >_u + \int_{t_i^{(n)}\wedge t}^{t_{i+1}^{(n)}\wedge t} \rho_u^2 d < S,S >_u] \\ &= \int_0^t (\hat{\rho}_u - \rho_u)^2 d < S,S >_u \qquad (8.24) \end{aligned}$$

That is, $\frac{<\hat{Z},\hat{Z}>_t - <Z,Z>_t}{\sqrt{\overline{\Delta t}^{(n)}}} = \frac{1}{\sqrt{\overline{\Delta t}^{(n)}}} \int_0^t (\hat{\rho}_u - \rho_u)^2 d <S,S>_u$.

(ii) Now let $\tilde{L}_{i,n}(t) = \tilde{L}_{1,i,n}(t) + \tilde{L}_{2,i,n}(t)$, where for $j = 1, 2$,

$$\tilde{L}_{j,i,n}(t) = \begin{cases} \frac{1}{<S,S>'_{t_i-h}}[B^{\Xi S}_{j,t_i} - \rho_{t_i-h} B^{SS}_{j,t_i}], & t \geq t_i \\ \frac{1}{<S,S>'_{t_i-h}}[B^{\Xi S}_{j,t} - \rho_{t_i-h} B^{SS}_{j,t}], & t_i - h \leq t \leq t_i \\ 0 & t \leq t_i - h \end{cases}$$

We show here that

$$\int_0^t (\hat{\rho}_u - \rho_u)^2 d <S,S>_u = \sum_{t_{i+1}^{(n)} \leq t} \tilde{L}^2_{i,n}(t_i) \tilde{\sigma}^2_{t_i-h} \Delta t_i + O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-\epsilon}) \tag{8.25}$$

uniformly in t, for any $\epsilon > 0$.

First observe that

$$\begin{aligned} & |\int_0^t (\hat{\rho}_u - \rho_u)^2 <S,S>'_u du - \sum_i (\hat{\rho}_{t_i} - \rho_{t_i})^2 <S,S>'_{t_i} \Delta t_i| \\ = & |\sum_i \int_{t_i}^{t_{i+1}} [(\hat{\rho}_u - \rho_u)^2 - (\hat{\rho}_{t_i} - \rho_{t_i})^2] <S,S>'_u du \\ & + \sum_i (\hat{\rho}_{t_i} - \rho_{t_i})^2 \int_{t_i}^{t_{i+1}} (<S,S>'_u - <S,S>'_{t_i}) du| \\ \leq & \sum_i \int_{t_i}^{t_{i+1}} |(\hat{\rho}_u - \rho_u + \hat{\rho}_{t_i} - \rho_{t_i})(\rho_{t_i} - \rho_u)| <S,S>'_u du + \sup_i |\hat{\rho}_{t_i} - \rho_{t_i}|^2 \Upsilon^{SS}(\delta^{(n)}) t \\ \leq & 2 \sup_i \sup_u |\hat{\rho}_u - \rho_u| \Upsilon^{\rho}(\delta^{(n)}) <S,S>_t + \sup_i |\hat{\rho}_{t_i} - \rho_{t_i}|^2 \Upsilon^{SS}(\delta^{(n)}) t \\ = & O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-\epsilon} + (\overline{\Delta t}^{(n)})^{1-\epsilon}) \end{aligned} \tag{8.26}$$

by Corollary 1 and Lemma 2 with Assumption A, $B.1[(\rho,\rho),(R^{SS},R^{SS})]$, and $B.2[(<S,S>'),(\rho)]$.

Then, use the notation that $<S,S>'_t = \sigma^2_t S^2_t = \tilde{\sigma}^2_t$. We now show

$$\sup_t |\sum_{t_{i+1}^{(n)} \leq t} L^2_{t_i} \tilde{\sigma}^2_{t_i} \Delta t_i - \sum_{t_{i+1}^{(n)} \leq t} \tilde{L}^2_{i,n}(t_i) \tilde{\sigma}^2_{t_i-h} \Delta t_i| = O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-\epsilon}) \tag{8.27}$$

where L_{t_i} is the same as defined in Corollary 1.

We know

$$L_{t_i}\tilde{\sigma}_{t_i} - \tilde{L}_{i,n}(t_i)\tilde{\sigma}_{t_i-h} = \sum_{j=1}^{2}(\rho_{t_i-h} - \rho_{t_i})B^{SS}_{j,t_i}\frac{1}{\tilde{\sigma}_{t_i-h}} + \sum_{j=1}^{2}(B^{\Xi S}_{j,t_i} - \rho_{t_i}B^{SS}_{j,t_i})[\frac{1}{\tilde{\sigma}_{t_i}} - \frac{1}{\tilde{\sigma}_{t_i-h}}]$$

thus,

$$\begin{aligned}\sup_i |L_{t_i}\tilde{\sigma}_{t_i} - \tilde{L}_{i,n}(t_i)\tilde{\sigma}_{t_i-h}| &\leq \sup_{0\leq t\leq T}\frac{1}{\tilde{\sigma}_t}\Upsilon^{\rho}(h)\sum_{j=1}^{2}\sup_i |B^{SS}_{j,t_i}| \\ &+ \sup_{0\leq t\leq T}\frac{1}{\tilde{\sigma}_t^2}\Upsilon^{\tilde{\sigma}}(h)\sum_{j=1}^{2}\left[\sup_i |B^{\Xi S}_{j,t_i}| + \sup_{0\leq t\leq T}|\rho_t|\sup_i |B^{SS}_{j,t_i}|\right] \\ &= O_p((\overline{\Delta t}^{(n)})^{\frac{1}{2}-2\epsilon}) \qquad (8.28)\end{aligned}$$

by Lemma 2 and Lemma 4, under Assumption A, $B(S, \Xi, \rho, < S, S >', < \Xi, S >')$, and $C(S)$. Notice that $B(\tilde{\sigma})$ is satisfied given $B(S, < S, S >')$, due to $< \tilde{\sigma}^2, \tilde{\sigma}^2 >'_t = 4\tilde{\sigma}_t^2 < \tilde{\sigma}, \tilde{\sigma} >'_t$ and the definition of $\tilde{\sigma}^2$.

Now let

$$\begin{aligned}a_i &= (L_{t_i}\tilde{\sigma}_{t_i} - \tilde{L}_{i,n}(t_i)\tilde{\sigma}_{t_i-h})^2 \\ b_i &= L_{t_i}\tilde{\sigma}_{t_i}(L_{t_i}\tilde{\sigma}_{t_i} - \tilde{L}_{i,n}(t_i)\tilde{\sigma}_{t_i-h}) \\ c_i &= (L_{t_i}\tilde{\sigma}_{t_i})^2 - (\tilde{L}_{i,n}(t_i)\tilde{\sigma}_{t_i-h})^2\end{aligned}$$

Then, Equation (8.28) and Corollary 1 yield that $\sup_i |a_i| = O_p((\overline{\Delta t}^{(n)})^{1-4\epsilon})$ and $\sup_i |b_i| = O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-3\epsilon})$.

And so $\sup_i |c_i| = \sup_i |a_i - 2b_i| = O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-3\epsilon})$.

Based on above,

$$|\sum_{t_{i+1}^{(n)}\leq t} L_{t_i}^2\tilde{\sigma}_{t_i}^2\Delta t_i - \sum_{t_{i+1}^{(n)}\leq t}\tilde{L}_{i,n}^2(t_i)\tilde{\sigma}_{t_i-h}^2\Delta t_i| = |\sum_{t_{i+1}^{(n)}\leq t} c_i\Delta t_i| \leq \sup_i |c_i|T = O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-3\epsilon})$$

uniformly in $t \in [0, T]$. Hence Equation (8.27) follows.

$$
\begin{aligned}
\int_0^t (\hat{\rho}_u - \rho_u)^2 d < S,S >_u &\overset{(8.26)}{=} \sum_{t_{i+1}^{(n)} \le t} (\hat{\rho}_{t_i} - \rho_{t_i})^2 < S,S >'_{t_i} \Delta t_i + O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-\epsilon}) \\
&\overset{Cor.1}{=} \sum_{t_{i+1}^{(n)} \le t} L_{t_i}^2 \tilde{\sigma}_{t_i}^2 \Delta t_i + O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-\epsilon}) \\
&\overset{(8.27)}{=} \sum_{t_{i+1}^{(n)} \le t} \tilde{L}_{i,n}^2(t_i) \tilde{\sigma}_{t_i-h}^2 \Delta t_i + O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-3\epsilon})
\end{aligned}
$$

Hence (8.25) follows.

(iii) We here show that

$$
\sum_{t_{i+1}^{(n)} \le t} \tilde{L}_{i,n}^2(t_i) \tilde{\sigma}_{t_i-h}^2 \Delta t_i = \sum_{t_{i+1}^{(n)} \le t} < \tilde{L}_{i,n} >_{t_i} \tilde{\sigma}_{t_i-h}^2 \Delta t_i + O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-\epsilon}) \tag{8.29}
$$

Set

$$
Y^{(i)}(t) = \tilde{L}_{i,n}^2(t) - < \tilde{L}_{i,n} >_t \tag{8.30}
$$

and

$$
Z_{n,t} = \sum_{t_{i+1}^{(n)} \le t} Y^{(i)}(t_i) \tilde{\sigma}_{t_i-h}^2 \Delta t_i + Y^{(i*)}(t) \tilde{\sigma}_{t_*-h}^2 \Delta t_* \tag{8.31}
$$

where t_* is the largest grid point less than t.

Suffice to show that for any $\epsilon > 0$,

$$
\sup_{0 \le t \le T} |Z_{n,t}| = O_p((\overline{\Delta t}^{(n)})^{\frac{3}{4}-\epsilon}) \tag{8.32}
$$

By Lenglart's Inequality (Jacod and Shiryaev (1987) Lemma I.3.30, p.35), it is enough to show that

$$
< Z_n, Z_n >_T = O_p((\overline{\Delta t}^{(n)})^{\frac{3}{2}-\epsilon}) \tag{8.33}
$$

for any $\epsilon > 0$, which is shown in the following.

Since $< \tilde{L}_{i,n}, \tilde{L}_{j,n} >_t = 0$ if $(t_i - h, t_i]$ and $(t_j - h, t_j]$ are disjoint for any $t_i \le T$ and $t_j \le T$, the quadratic variation of the Z_n is considered in the overlapping time interval.

$$< Z_n, Z_n >_T \leq \sup_{0\leq t\leq T} \tilde{\sigma}_t^4 |\sum_i \sum_j (\Delta t_i)(\Delta t_j) \int_{(t_i-h,t_i]\cap(t_j-h,t_j]} d < Y^{(i)}, Y^{(j)} >_u |$$

$$\leq \sup_{0\leq t\leq T} \tilde{\sigma}_t^4 \sum_i \sum_j (\Delta t_i)(\Delta t_j) (\int_{(t_i-h,t_i]\cap(t_j-h,t_j]} d < Y^{(i)} >_u)^{\frac{1}{2}} \cdot (\int_{(t_i-h,t_i]\cap(t_j-h,t_j]} d < Y^{(j)} >_u)^{\frac{1}{2}}$$

$$\leq 2 \sup_{0\leq t\leq T} \tilde{\sigma}_t^4 \sum_{i\leq j} (\Delta t_i)(\Delta t_j) I_{\{i<j:t_i>t_j-h\}} (\int_{(t_j-h,t_i]} d < Y^{(i)} >_u)^{\frac{1}{2}} (\int_{(t_j-h,t_i]} d < Y^{(j)} >_u)^{\frac{1}{2}} \quad (8.34)$$

The above second line follows from Kunita-Watanabe Inequality (p. 61 of Protter, 1995).

By Ito's formula, for $t_i - h < t < t_i$,

$$< Y^{(i)} >'_t = 4\tilde{L}_{i,n}^2(t) < \tilde{L}_{i,n} >'_t \quad (8.35)$$

By Corollary 1 and (8.28),

$$< Y^{(i)} >'_t \leq U_1 < \tilde{L}_{i,n} >'_t \quad (8.36)$$

where U_1 is independent of i, and $U_1 = O_p((\overline{\Delta t}^{(n)})^{\frac{1}{2}-\epsilon})$.

By the Kunita-Watanabe Inequality and Cauchy-Schwarz Inequality,

$$< \tilde{L}_{i,n} >'_t \leq 4 \sup_{0\leq u\leq T} \frac{1}{\tilde{\sigma}_u^4} \sum_{j=1}^{2} \left(< B_{j,i}^{\Xi S} >'_t + \rho_{t_i-h}^2 < B_{j,i}^{SS} >'_t \right) \quad (8.37)$$

where $B_{j,i,t}^{XY}$, $j = 1, 2$, is defined before Lemma 4.

Obviously, on $t_i - h < t < t_i$,

$$< B_{1,i}^{XY} >'_t = \frac{1}{h^2}(t - (t_i - h))^2 < R^{XY} >'_t \quad (8.38)$$

Also, let $M_{i,n}^{XY}(t)$ be the martigale defined in equation (8.15). Then, again by the Kunita-Watanabe Inequality and Cauchy-Schwarz In equality,

$$< B_{2,i}^{XY} >'_t \leq 2 \left(< M_{i,n}^{XY} >'_t + < M_{i,n}^{YX} >'_t \right) \quad (8.39)$$

Note that,

$$< M_{i,n}^{XY} >'_t = \frac{1}{h^2}(X_t - X_{t*})^2 < Y >'_t \quad (8.40)$$

for $t_i - h < t < t_i$, where $t*$ is the largest grid point smaller than t. In view of equations (8.36) and (8.37), it is enough to show that (8.34) is $O_p((\overline{\Delta t}^{(n)})^{3/2-\epsilon})$ in the two cases where

$< Y^{(i)} >_t'$ is replaced by (a) $U_1 < B_{1,i}^{XY} >_t'$, for $(X,Y) = (\Xi, S)$ and (S,S), and (b) $U_1 < M_{i,n}^{XY} >_t'$, for $(X,Y) = (\Xi,S), (S,\Xi)$ and (S,S).

Consider first the case of $< B_{1,i}^{XY} >_t'$. Set

$$N_n = \sup_t \#\{j : t \leq t_j \leq t+h\}$$

and $\delta_-^{(n)} = \min(t_{i+1}^{(n)} - t_i^{(n)})$, note that $N_n = O(\frac{h}{\delta_-^{(n)}}) = O(\frac{h}{\overline{\Delta t}^{(n)}})$ under Assumption A. Since $\sup_{0\leq u\leq T} < R^{XY} >_u' < \infty$, up to a multiplicative $O_p((\overline{\Delta t}^{(n)})^{1/2-\epsilon})$ term, Equation (8.34) becomes

$$\begin{aligned} & \frac{(\overline{\Delta t^{(n)}})^2}{h^2} \sum_{i\leq j: t_j - t_i < h} \left[h^3 - (t_j - t_i)^3\right]^{\frac{1}{2}} [t_i - (t_j - h)]^{3/2} \\ = \; & \frac{(\Delta t)^2}{h^2} h^3 \sum_{i\leq j: t_j - t_i < h} \left[1 - (\frac{t_j - t_i}{h})^3\right]^{1/2} (1 - \frac{t_j - t_i}{h})^{3/2} \\ \leq \; & (\overline{\Delta t^{(n)}})^2 h \#\{(i,j) : t_i \leq t_j \leq t_i + h\} \\ \leq \; & (\overline{\Delta t^{(n)}})^2 h k_n N_n \\ = \; & O(\overline{\Delta t^{(n)}}) \end{aligned} \tag{8.41}$$

we thus establish the fact that the whole term of Equation (8.34) is of order $O_p((\overline{\Delta t}^{(n)})^{3/2-\epsilon})$, under assumption $B.1(R^{XY}, R^{XY})$, which is satisfied for $(X,Y) = (\Xi, S)$ and (S,S) under the assumption of the theorem.

As for $< M_{i,n}^{XY} >_t'$, note that

$$\int_{(t_j - h, t_i]} d < M_{i,n}^{XY} >_u = \int_{(t_j - h, t_i]} d < M_{j,n}^{XY} >_u$$

and that this expression is dominated by

$$\begin{aligned} \frac{1}{h^2} \sup_{0\leq u\leq T} < Y >_u' N \sup_k \int_{t_k}^{t_{k+1}} (X_t - X_{t_k})^2 dt \; & \leq \; \frac{1}{h^2} N \sup_{0\leq u\leq T} < Y >_u' \left(\Upsilon^X(\delta^{(n)})\right)^2 \delta^{(n)} \\ & = \; O_p(N \frac{1}{h^2} (\overline{\Delta t^{(n)}})^{2-\epsilon}) = O_p((\overline{\Delta t^{(n)}})^{\frac{1}{2}-\epsilon}) \end{aligned}$$

by Lemma 2 under Assumption A, $B.1[(X,X),(Y,Y)]$ and $B.2(X)$. It follows that (8.34) is, up to an $O_p(1)$ factor,

$$\begin{aligned} (\Delta t)^2 O_p((\overline{\Delta t^{(n)}})^{\frac{1}{2}-\epsilon}) \sum_{i\leq j: t_j - t_i < h} 1 \; & \leq \; (\Delta t)^2 O_p((\overline{\Delta t^{(n)}})^{\frac{1}{2}-\epsilon}) N \frac{T}{\Delta t} \\ & = \; O_p((\overline{\Delta t^{(n)}})^{1-\epsilon}) \end{aligned}$$

which is what we needed to show. Translated to S and Ξ, the conditions imposed are $B.1[(S,S),(\Xi,\Xi)]$ and $B.2[(S),(\Xi)]$.

(iv) We next show that as $\frac{\sqrt{\Delta t^{(n)}}}{h} \to c$, by $B(< R^{SS}, R^{SS} >', < R^{\Xi S}, R^{SS} >', < R^{\Xi S}, R^{\Xi S} >')$ and $C(S)$.

$$\frac{1}{\sqrt{\Delta t}^{(n)}} \sum_{t_{i+1}^{(n)} \leq t} < \tilde{L}_{i,n} >_{t_i} \tilde{\sigma}^2_{t_i-h}(\Delta t_i) \to \int_0^t V_{\hat{\rho}-\rho}(u) d < S, S >_u \tag{8.42}$$

in probability, uniformly in $t \in [0,T]$, where $V_{\hat{\rho}-\rho}(t) = \frac{<\rho,\rho>'_t}{3c} + cH^{(2)\prime}(t)(\frac{<\Xi,\Xi>'_t}{<S,S>'_t} - \rho_t^2)$. □

Since $\sup_i |< \tilde{L}_{1,i,n}, \tilde{L}_{2,i,n} >_{t_i}| = O_p((\overline{\Delta t^{(n)}})^{\frac{3}{4}})$ from Lemma 4, it follows that we can prove separately that

$$\frac{1}{\sqrt{\Delta t}^{(n)}} \sum_{t_{i+1}^{(n)} \leq t} < \tilde{L}_{1,i,n} >_{t_i} \tilde{\sigma}^2_{t_i-h}(\Delta t_i) \xrightarrow{P} \frac{1}{3c} \int_0^t < \rho, \rho >'_u d < S, S >_u \tag{8.43}$$

and

$$\frac{1}{\sqrt{\Delta t}^{(n)}} \sum_{t_{i+1}^{(n)} \leq t} < \tilde{L}_{2,i,n} >_{t_i} \tilde{\sigma}^2_{t_i-h}(\Delta t_i) \xrightarrow{P} c \int_0^t (\frac{< \Xi, \Xi >'_u}{< S, S >'_u} - \rho_u^2) H^{(2)\prime}(u) d < S, S >_u \tag{8.44}$$

uniformly in t.

It is enough to prove the statement for each t, since the convergence of increasing functions to an increasing function is automatically uniform.

Equation (8.43) follows directly from the approximation (8.10) in Lemma 4. It remains to show (8.44). This is what we do for the rest of (iv).

Let A be an Ito process, which we shall variously take to be $\frac{1}{<S,S>'}$, $-\frac{2\rho}{<S,S>'}$, and $\frac{\rho^2}{<S,S>'}$.

Consider a subproblem of (8.44), that of the convergence of

$$\frac{1}{\sqrt{\Delta t}^{(n)}} \sum_{t_{i+1}^{(n)} \leq t} < B_{2,i}^{XY}, B_{2,i}^{ZV} >_{t_i} A_{t_i-h} \Delta t_i \tag{8.45}$$

By equation (8.11) in Lemma 4, this is (uniformly in t) equal to

$$\frac{1}{\sqrt{\Delta t}^{(n)}} \frac{1}{h^2} \sum_{t_{i+1}^{(n)} \leq t} \sum_{t_i-h \leq t_j \leq t_{j+1} \leq t_i} f(t_j)(\Delta t_j)^2 A_{t_i-h} \Delta t_i + o_p(1)$$

where $f(t) =< X, Z >'_t< Y, V >'_t + < X, V >'_t< Y, Z >'_t$. By interchanging the two summations, (8.45) then becomes

$$\begin{aligned}
& \frac{1}{\sqrt{\overline{\Delta t}^{(n)}}} \frac{1}{h} \sum_{t_{j+2} \le t} f(t_j)(\Delta t_j)^2 \frac{1}{h} \sum_{t_{j+1} \le t_i \le t_j + h} A_{t_i - h} \Delta t_i + o_p(1) \\
= & \frac{1}{\sqrt{\overline{\Delta t}^{(n)}}} \frac{1}{h} \sum_{t_{j+2} \le t} f(t_j) A_{t_j} (\Delta t_j)^2 + o_p(1)
\end{aligned}$$

Since the difference between the last two terms is bounded by

$$\begin{aligned}
& \frac{1}{\sqrt{\overline{\Delta t}^{(n)}}} \frac{1}{h} \sum_{t_{j+2} \le t} |f(t_j)| (\Delta t_j)^2 \Upsilon^A(h) \\
\le & \sup_t |f(t)| \Upsilon^A(h) \frac{\sqrt{\overline{\Delta t}^{(n)}}}{h} H_n(t) = o_p(1)
\end{aligned}$$

by Lemma 2. Hence, (8.45) converges to $c \int_0^t f(u) A_u dH(u) = c \int_0^t f(u) A_u H'(u) du$ by Assumption A and since A and f are bounded and continous. Note that H is absolutely continuous since Lipschitz.

The result (8.44) now follows by aggregating this convergence for the cases of $< B^{\Xi S}_{2,i}, B^{\Xi S}_{2,i} >$ $(A = \frac{1}{<S,S>'})$, $< B^{\Xi S}_{2,i}, B^{SS}_{2,i} >$ $(A = -\frac{2\rho}{<S,S>'})$, and $< B^{SS}_{2,i}, B^{SS}_{2,i} >$ $(A = \frac{\rho^2}{<S,S>'})$.

(v) To finish the theorem, we want to show that

$$\sup_t |[\hat{Z}, \hat{Z}]_t - < \hat{Z}, \hat{Z} >_t - ([Z, Z]_t - < Z, Z >_t)| = o_p((\overline{\Delta t}^{(n)})^{-1/2}). \tag{8.46}$$

Note that $< \hat{Z}, \hat{Z} >'_t - < Z, Z >'_t = (\hat{\rho}_t - \rho_t)^2 < S, S >'$ and similarly for the drift of $\hat{Z}$ and Z. The result then follows from Proposition 2 and Corollary 1.

∎

8.4. *Additional lemmas for Theorem 2, and proof of the theorem.*

LEMMA 5. Let X, Y, and A be Ito processes. Let $h = O(\overline{\Delta t}^{(n)\,1/2})$.

Define

$$V_t^{XY} = \sum_{t_{i+1}^{(n)} \leq t} (\Delta X_{t_i})(\Delta Y_{t_i}) + (X_t - X_{t_*})(Y_t - Y_{t_*})$$
$$U(t) = \frac{1}{h}\int_0^t \sum_i A_u I_{(t_i,t_{i+1}\in[u-h,u])} \Delta V_{t_i} du - \frac{1}{h} \sum_{t_{i+1}^{(n)} \leq t} A_{t_i} \Delta V_{t_i}(h - \Delta t_i) \tag{8.47}$$

Then subject to conditions A, $B.1[(X,X),(A,A),(Y,Y),(A,X),(A,Y)]$ and $B.2[(X),(A),(Y)]$,

$$\sup_t |U(t)| = o_p(\overline{\Delta t}^{(n)1/2}).$$

□

Proof of Lemma 5:

By polarization, one can without loss of generality take $Y = X$.

Note first that

$$\sup_i |\Delta V_{t_i}| = \sup_i (\Delta X_{t_i})^2 = O_p(\overline{\Delta t}^{(n)1-\epsilon}) \tag{8.48}$$

by Lemma 2, under assumptions A, $B.1[(X,X)]$, and $B.2[(X)]$.

In the following, we provide the proof in two steps. We use $x \wedge y$ to denote $\min(x,y)$. Note that $U(t)$ is the sum of the two components considered.

(i) Set

$$\tilde{U}(t) = \frac{1}{h}\int_0^t \sum_i (A_u - A_{t_{i+1}}) I_{(t_i,t_{i+1}\in[u-h,u])} \Delta V_{t_i} du \tag{8.49}$$

Then by exchanging the summation and the integral, and by integration by parts,

$$\tilde{U}(t) = \sum_{t_{i+1}^{(n)} \leq t} \Delta V_{t_i} \frac{1}{h} \int_{t_{i+1}}^{(t_i+h)\wedge t} ((t_i+h)\wedge t - u)\, dA_u.$$

Under Assumptions A and $B.2(A)$, it is straightforward to see that

$$\sup_i \left|\frac{1}{h}\int_{t_{i+1}}^{(t_i+h)\wedge t} ((t_i+h)\wedge t - u) dA_u^{DR} - \frac{1}{2}\tilde{A}_{t_{i+1}} h\right| = O_p(h^2).$$

In view of (8.48), the part of (8.49) that is attributable to A_u^{DR} is therefore $o_p(h)$ uniformly.

Meanwhile, set

$$\begin{aligned} g(i,j,t) &= \frac{1}{h^3} < \int_{t_{i+1}}^{(t_i+h)\wedge t} ((t_i+h)\wedge t-u)dA_u, \int_{t_{j+1}}^{(t_j+h)\wedge t} ((t_j+h)\wedge t-u)dA_u > \\ &= O_p(1) \text{ uniformly in } i,j,t \end{aligned}$$

provided that $B.1(A,A)$ is satisfied.

Hence, by (8.48)

$$< \tilde{U} >_t \leq \sup_i |\Delta V_{t_i}| \sup_j |\Delta V_{t_j}| \sup_{i,j,t} |g(i,j,t)| \sum_{t_{i+1}^{(n)} \leq t} h = o_p(\overline{\Delta t}^{(n)}).$$

By Lenglart's Inequality, it follows that $\sup_t |\tilde{U}(t)| = o_p(h)$, given that $h^2 = O(\overline{\Delta t}^{(n)})$.

(ii) Now set

$$\tilde{\tilde{U}}(t) = \frac{1}{h}\int_0^t \sum_i I_{(t_i,t_{i+1}\in[u-h,u])}\Delta A_{t_i}\Delta V_{t_i} du,$$

which is equal to $\frac{1}{h}\sum_{t_{i+1}^{(n)}\leq t} \Delta A_{t_i}\Delta V_{t_i}((t_i+h)\wedge t - t_{i+1})$. Under our assumptions, this equals in turn $\frac{1}{h}\sum_{t_{i+1}^{(n)}\leq t} \Delta A_{t_i}\Delta V_{t_i}(h-\Delta t_i) + O_p(h^{2-\epsilon})$ following (8.48) and $B.1[(A,A)]$, $B.2[(A)]$. Next we will only argue that

$$\sum_{t_{i+1}^{(n)}\leq t} \Delta A_{t_i}\Delta V_{t_i} = o_p({\overline{\Delta t}^{(n)}}^{1/2}), \tag{8.50}$$

and the result $\frac{1}{h}\sum_{t_{i+1}^{(n)}\leq t} \Delta A_{t_i}\Delta V_{t_i}(\Delta t_i) = o_p({\overline{\Delta t}^{(n)}}^{1/2})$ follows similarly.

First consider the part related to A^{DR},

$$|\sum_{t_{i+1}^{(n)}\leq t} \Delta A_{t_i}^{DR}\Delta V_{t_i}| \leq \sup_u |\tilde{A}_u| \sup_i |\Delta V_{t_i}| t = o_p({\overline{\Delta t}^{(n)}}^{1/2}),$$

by Assumption $B.1[(X,X)]$, $B.2[(A),(X)]$.

Similarly, one can remove the drift part from the X process, under Assumption $B.1[(A,A),(X,X)]$, $B.2[(A),(X)]$.

One is then left to argue that

$$\sum_{t_{i+1}^{(n)}\leq t} \Delta A_{t_i}^{MG}(\Delta X_{t_i}^{MG}) = o_p({\overline{\Delta t}^{(n)}}^{1/2}), \tag{8.51}$$

and this follows as in the proof of Theorem 5.1 of Jacod and Protter (1998), with modifications noted in the proof of (our) Proposition 1. This is since we have assumed $B.1[(A,A),(X,X)]$.

(iii) Combining parts (i) and (ii), (8.50) follows ∎

LEMMA 6. Let Ξ, S, ρ, and Z be the Ito processes defined in earlier sections. Subject to regularity conditions in Lemma 5 with $(X,Y) = (\Xi, S)$, (S,S), or (ρ, S), and with $A = \rho$, ρ^2 or Z,

$$\int_0^t (\hat{\rho}_u - \rho_u) d < \Xi, S >_u = [\Xi, Z]_t - [Z,Z]_t - \frac{h}{3}\int_0^t \rho_u d\left\langle \rho, < S,S >' \right\rangle_u + o_p(h)$$

uniformly in t. □

PROOF OF LEMMA 6:

Note that by Taylor expansion of $\frac{1}{<S,S>'}$ at $\frac{1}{\widehat{<S,S>'}}$, we get

$$\begin{aligned} \frac{\widehat{<\Xi,S>}'_u - \rho_u \widehat{<S,S>}'_u}{<S,S>'_u} &= \underbrace{\frac{\widehat{<\Xi,S>}'_u - \rho_u \widehat{<S,S>}'_u}{\widehat{<S,S>}'_u}}_{\hat{\rho}_u - \rho_u} \\ &\quad + \frac{\widehat{<\Xi,S>}'_u - \rho_u \widehat{<S,S>}'_u}{\left(\widehat{<S,S>}'_u\right)^2}\left(\widehat{<S,S>}'_u - <S,S>'_u\right) \\ &\quad + o_p\left(\sqrt{\Delta t^{(n)}}\right) \end{aligned} \tag{8.52}$$

where the higher order $o_p(\sqrt{\Delta t^{(n)}})$ is uniform in u, by Lemma 4 and Corollary 1.

Thus

$$\begin{aligned} &\int_0^t (\hat{\rho}_u - \rho_u) d < \Xi, S >_u \\ = &\int_0^t \frac{\widehat{<\Xi,S>}'_u - \rho_u \widehat{<S,S>}'_u}{<S,S>'_u} d < \Xi, S >_u \\ &- \int_0^t \frac{\hat{\rho}_u - \rho_u}{\widehat{<S,S>}'_u}\left(\widehat{<S,S>}'_u - <S,S>'_u\right) d < \Xi, S >_u + o_p\left(\sqrt{\Delta t^{(n)}}\right) \end{aligned} \tag{8.53}$$

uniformly in t.

Following Lemma 5, the first term on the r.h.s. of (8.53) becomes

$$\begin{aligned}
&\frac{1}{h}\sum_{t_{i+1}^{(n)}\le t}\rho_{t_i}\Delta S_{t_i}[\Delta\Xi_{t_i}-\rho_{t_i}\Delta S_{t_i}](h-\Delta t_i)+o_p(\sqrt{\overline{\Delta t}^{(n)}})\text{ uniformly in } t,\\
=\;& [\Xi,Z]_t-[Z,Z]_t+o_p(\sqrt{\overline{\Delta t}^{(n)}})\text{ uniformly in } t,
\end{aligned}$$

where the last line is due to

$$\begin{aligned}
\frac{1}{h}\sum_{t_{i+1}^{(n)}\le t}\rho_{t_i}\Delta S_{t_i}\left(\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(\rho_u-\rho_{t_i})dS_u\right)(h-\Delta t_i) &= o_p(\sqrt{\overline{\Delta t}^{(n)}}),\\
\frac{1}{h}\sum_{t_{i+1}^{(n)}\le t}\Delta Z_{t_i}\left(\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(\rho_u-\rho_{t_i})dS_u\right)(h-\Delta t_i) &= o_p(\sqrt{\overline{\Delta t}^{(n)}})
\end{aligned}$$

by similar arguement in handling (8.50), under assumption A, $B.1[(\rho,\rho),(S,S),(Z,Z),(Z,S),(S,\rho)]$, and $B.2[(S),(Z),(\rho)]$.

For the second term on the r.h.s. of (8.53), one can invoke Lemma 4, Corollary 1, and similar arguments in Theorem 1, it becomes

$$\begin{aligned}
&-\sum_{t_{i+1}^{(n)}\le t}\frac{1}{<S,S>'_{t_i-h}}\tilde{L}_{i,n}^{MG}(t_i)(B_{1,t_i}^{SS,MG}+B_{2,t_i}^{SS,MG})\Delta<\Xi,S>_{t_i}+o_p(\sqrt{\overline{\Delta t}^{(n)}})\\
=\;&-\sum_{t_{i+1}^{(n)}\le t}\left(\frac{1}{<S,S>'_{t_i-h}}\right)^2(<B_1^{\Xi S},B_1^{SS}>_{t_i}-\rho_{t_i-h}<B_1^{SS},B_1^{SS}>_{t_i})\Delta<\Xi,S>_{t_i}\\
&-\sum_{t_{i+1}^{(n)}\le t}\left(\frac{1}{<S,S>'_{t_i-h}}\right)^2(<B_2^{\Xi S},B_2^{SS}>_{t_i}-\rho_{t_i-h}<B_2^{SS},B_2^{SS}>_{t_i})\Delta<\Xi,S>_{t_i}+o_p(\sqrt{\overline{\Delta t}^{(n)}})\\
=\;&-\sum_{t_{i+1}^{(n)}\le t}\frac{h}{3}\rho_{t_i}\left\langle\rho,<S,S>'\right\rangle'_{t_i}\Delta t_i+o_p(\sqrt{\overline{\Delta t}^{(n)}})
\end{aligned}$$

where $\tilde{L}_{i,n}^{MG}$ is defined as $\tilde{L}_{i,n}$ with B_1 and B_2 being replaced by B_1^{MG} and B_2^{MG} correspondingly (c.f. Theorem 1 for the definition of $\tilde{L}_{i,n}$, and Lemma 4 for those of B_js, $j=1,2$). Note that in the above last line the conditions used include $B(S,<R^{SS},R^{SS}>',<R^{\Xi S},R^{SS}>')$, the B_2 terms are canceled out because of (8.11) in Lemma 4, and the B_1 terms yield the result by Lemma 2 and the next two equations:

$$\begin{aligned}
&\sum_{t_{i+1}^{(n)}\leq t} A_{t_i-h} < B_1^{XY}, B_1^{SS} >_{t_i} \Delta < \Xi, S >_{t_i} \\
&= \sum_{t_{i+1}^{(n)}\leq t} A_{t_i-h} \frac{h}{3} \left\langle < X, Y >', < S, S >' \right\rangle'_{t_i} \Delta < \Xi, S >_{t_i} + o_p(\sqrt{\Delta t^{(n)}})
\end{aligned}$$

for $(X, Y) = (\Xi, S)$ or (S, S), and with $A_t = (\frac{1}{<S,S>'_t})^2$ or $A_t = (\frac{\rho_t}{<S,S>'_t})^2$, and

$$\left\langle < \Xi, S >', < S, S >' \right\rangle'_{t_i} = \rho_{t_i} \left\langle < S, S >', < S, S >' \right\rangle'_{t_i} + < S, S >'_{t_i} \left\langle \rho, < S, S >' \right\rangle'_{t_i}.$$

Hence the result follows. ∎

PROOF OF THEOREM 2:

Note that by definition,

$$\begin{aligned}
&\Delta \widetilde{< Z, Z >}_{t_i^{(n)}} \\
= &\frac{1}{2}[(\Delta\Xi_{t_i^{(n)}})^2 - 2\hat{\rho}_{t_i^{(n)}}(\Delta\Xi_{t_i^{(n)}})(\Delta S_{t_i^{(n)}}) + \hat{\rho}^2_{t_i^{(n)}}(\Delta S_{t_i^{(n)}})^2 + (\Delta\Xi_{t_i^{(n)}})^2 - \hat{\rho}^2_{t_i^{(n)}}(\Delta S_{t_i^{(n)}})^2] \\
= &\Delta[\Xi, \hat{Z}]_{t_i}
\end{aligned}$$

Because $d < \Xi, Z >_t = d < Z, Z >_t$ by assumption, one gets

$$\begin{aligned}
&\frac{1}{\sqrt{\Delta t^{(n)}}} \left(\widetilde{< Z, Z >}_t - < Z, Z >_t \right) \\
&= \frac{1}{\sqrt{\Delta t^{(n)}}} \sum_{t_{i+1}^{(n)}\leq t} \left(\Delta[\Xi, \hat{Z}]_{t_i} - \Delta < \Xi, Z >_{t_i} \right) \\
&= \underbrace{\frac{1}{\sqrt{\Delta t^{(n)}}}([\Xi, \hat{Z}]_t - < \Xi, \hat{Z} >_t)}_{C_1} + \underbrace{\frac{1}{\sqrt{\Delta t^{(n)}}}(< \Xi, \hat{Z} >_t - < \Xi, Z >_t)}_{C_2}
\end{aligned}$$

First notice that

$$< \Xi, \hat{Z} >'_t = < \Xi, Z >'_t + (\rho_t - \hat{\rho}_t) < \Xi, S >'_t \tag{8.54}$$

Then as $\sqrt{\Delta t^{(n)}}/h \to c$, (8.54) and Lemma 6 show that

$$\begin{aligned}
C_2 &= \frac{1}{\sqrt{\Delta t^{(n)}}} \int_0^t (\rho_u - \hat{\rho}_u) d < \Xi, S >_u \\
&= \frac{[Z, Z]_t - [\Xi, Z]_t}{\sqrt{\Delta t^{(n)}}} + \frac{1}{3c} \int_0^t \rho_u d \left\langle \rho, < S, S >' \right\rangle_u + o_p(1) \text{ uniformly in } t.
\end{aligned}$$

That is,

$$\begin{aligned}
&\frac{1}{\sqrt{\Delta t^{(n)}}}\left(<\widetilde{Z,Z}>_t - <Z,Z>_t\right) \\
&= \frac{[\Xi,\hat{Z}]_t - <\Xi,\hat{Z}>_t + [Z,Z]_t - [\Xi,Z]_t}{\sqrt{\Delta t^{(n)}}} + \frac{1}{3c}\int_0^t \rho_u d\left\langle \rho, <S,S>'\right\rangle_u + o_p(1) \\
&= \frac{[Z,Z]_t - <Z,Z>_t}{\sqrt{\Delta t^{(n)}}} + \frac{1}{3c}\int_0^t \rho_u d\left\langle \rho, <S,S>'\right\rangle_u \\
&\quad + \frac{[\Xi,\hat{Z}-Z]_t - <\Xi,\hat{Z}-Z>_t}{\sqrt{\Delta t^{(n)}}} + o_p(1) \qquad (8.55)
\end{aligned}$$

again since $<\Xi,Z>_t = <Z,Z>_t$.

The second component on the right hand side of (8.55) , $([\Xi,\hat{Z}-Z]_t - <\Xi,\hat{Z}-Z>_t)/\sqrt{\Delta t^{(n)}}$ goes to zero in probability, by Proposition 2, since $<\Xi,\Xi>'_u$, $<\hat{Z}-Z,\hat{Z}-Z>'_u$, and $<\Xi,\hat{Z}-Z>'_u$ satisfy the conditions of this proposition. This is in view of Corollary 1. The argument is similar to that at the end of the proof of Theorem 1.

∎

9.References

Aldous, D. and Eagleson, G. (1978). On mixing and stability of limit theorems. *Annals of Probability*, 6:325–331.

Barlow, M. T., Jacka, S. D., and Yor, M. (1986). Inequalities for a pair of processes stopped at a random time. *Proceedings for London Mathematics Society*, 52:142–172.

Beckers, S. (1981). Standard deviations implied in option prices as predictors of future stock price variability. *Journal of Banking and Finance*, 5:363–381.

Bick, A. (1995). Quadratic-variation-based dynamic strategies. *Management Science*, 41 (4):722–732.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y., and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. Baltimore, MD: The Johns Hopkins University Press.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–659.

Blair, B.J., S. P. and Taylor, S. (2001). Forecasting s & p 100 volatility: the incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics*, 105 (1):5–26.

Christensen, B. and Prabhala, N. (1998). The relation between implied and realized volatility. *Journal of Financial Economics*, 50 (2):125–150.

Dambis, K. (1965). On decomposition of continuous sub-martingales. *Theory of Probability and its Applications*, 10:401–410.

Delbaen, F. and Schachermayer, W. (1995a). The existence of absolutely continuous local martingale measures. *Annals of Applied Probability*, 4:926–945.

Delbaen, F. and Schachermayer, W. (1995b). The no-arbitrage property under a change of numeraire. *Stochastics and Stochastics Reports*, 53 (3-4):213–226.

Dellacherie, D. and Meyer, P. A. (1982). *Probabilities and Potential B.* Amsterdam; New York: North Holland.

Derman, E. and Kani, I. (1994). Riding on a smile. *Risk*, 7 (2):32–39.

Dubins, L. E. and Schwartz, G. (1965). On continuous martingales. *Proceedings of the National Academy of Sciences of the United States of America*, 53:913–916.

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

Efron, B. and Hinkley, D. (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected fisher information. *Biometrika*, 65 (3):457–487.

Engle, R. F. and Mustafa, C. (1992). Implied arch models for option prices. *Journal of Econometrics*, 52:289–311.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability*, 30:790–804.

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

Genon-Catalot, V., Laredo, C., and Picard, D. (1992). Nonparametric-estimation of the diffusion-coefficient by wavelets methods. *Scandinavian Journal of Statistics*, 19 (4):317–335.

Gwilym, O. and Buckle, M. (1997). Forward/forward volatilities and the term structure of implied volatility. *Applied Economics Letters*, 4 (5):325–328.

Harrison, J. and Krebs, D. (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory*, 20:381–408.

Harrison, J. and Pliska, S. (1981). Martingales and stochastic integralsl in the theory of continuous trading. *Stochastic Processes and their Applications*, 11:261–271.

Hoffmann, M. (1999). l_p estimation of the diffusion coefficient. *Bernoulli*, 5:447–481.

Hull, J. (1999). *Options, Futures, & Other Derivatives.* Prentice Hall, 4th edition.

Jacod, J. (1979). *Calcul Stochastique et problèmes de martingales.* Berlin; New York: Springer-Verlag. Lecture notes in mathematics; 714.

Jacod, J. (2000). Non-parametric kernel estimation of the coefficient of a diffusion. *Scand. J. of Statist.*, 27 (1):83–96.

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

Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes.* Berlin; New York: Springer-Verlag.

Jorion, P. (1995). Predicting volatility in the foreign-exchange market. *Journal of Finance*, 50 (2):507–528.

Karatzas, I. and Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus.* Berlin; New York: Springer-Verlag, 2nd edition.

Lamoureux, C. and Lastrapes, W. (1993). Forecasting stock return variance: Toward an understanding of stochastic implied volatilities. *The Review of Financial Studeis*, 6:293–326.

Merton, R. (1973). Theory of rational option pricing. *Bell Journal of Economics and measurement Science*, 4:141–183.

Mykland, P. A. (1995). Dual likelihood. *Ann. of Statist.*, 23:396–421.

Mykland, P. A. (2000). Conservative delta hedging. *Ann. of Appl. Probab.*, 10:664–683.

Mykland, P. A. (2001). Likelihood computations without bartlett identities. *Bernoulli*, 7:473–485.

Mykland, P. A. (2003). The interpolation of options. *Finance and Stochastics*, 7:417–432.

Mykland, P. A. and Zhang, L. (2001). A no-arbitrage relationship between implied and realized volatility. Technical report no. 505, Department of Statistics, University of Chicago.

Mykland, P. A. and Zhang, L. (2002). Inference for volatility-type objects and implications for hedging. Technical report no. 506, Department of Statistics, University of Chicago (submitted).

Pena, I., Rubio, G., and Serna, G. (1999). Why do we smile? on the determinants of the implied volatility function. *Journal of Banking and Finance*, 23 (8):1151–1179.

Protter, P. E. (1995). *Stochastic Integration and Differential Equations*. Berlin; New York: Springer-Verlag, 3rd edition.

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

Rootzén, H. (1980). Limit distribution for the error in approximation of stochastic integrals. *Ann. Probab.*, 2:241–251.

Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance*, 49 (3):771–818.

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