Inference for Volatility-Type Objects and Implications for Hedging*

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 studies inference for volatility type objects and its implications for the hedging of options. It considers the nonparametric estimation of volatilities and instantaneous covariations between diffusion type processes. This is then linked to options trading, where we show that our estimates can be used to trade options without reference to the specific model. The new options "delta" becomes an additive modification of the (implied volatility) Black-Scholes delta. The modification, in our example, is both substantial and statistically significant. In the inference problem, explicit expressions are found for asymptotic error distributions, and it is explained why one does not in this case encounter a bias-variance tradeoff, but rather a variance-variance tradeoff. Observation times can be irregular.

Some key words and phrases: volatility estimation, statistical uncertainty, small interval asymptotics, mixing convergence, option hedging

*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.

1 INTRODUCTION

Volatility has become a popular topic in the statistics and the econometrics literature. However, most of these studies remain at the stage of estimating volatility and only a few mention both volatility estimation and option hedging. In contrast to the existing literature, we do not focus on issues like option mispricing with different volatility estimates. Rather, this paper seeks to investigate the instantaneous association between two volatility estimates, *realized volatility and option-implied volatility*, and investigate its impact on interval inference for the delta in options hedging. In the process, we state general theorems about the estimation of instantaneous covariations.

The literature on estimation of realized volatility mainly consists of three schemes: parametric, semi-parametric, and non-parametric. Most investigators have adopted parametric assumptions on the data generating process. ARCH (Engle (1982)), GARCH (Bollerslev (1986)), and various stochastic volatility models (Hull and White (1987); Wiggins (1987); Polson et al. (1994)) are just a few examples among the vast literature. The apparent evolution of volatility modeling reflects the need for reconciling the model and the features of the data. For example, the extension of ARCH to GARCH intends to incorporate the heteroscedasticity in the data (Bollerslev (1986)), stochastic volatility models are developed to account for the volatility smile, skewness and kurtosis, and the generalized MA(1) by Bai et al. (2000a,b) uses a semi-parametric approach to capture the high kurtosis in exchange-rate data that could not be adequately explained by GARCH. In addition to the rich parametric literature in volatility estimation, the attention to non-parametric approaches is also rising in the recent decade. The non-parametric approach generally includes (1) chopping the returns data into blocks based on time, then summing intra-period squared returns (Merton (1980); Poterba and Summers (1986); French et al. (1987); Andersen et al. (2001)); (2) rolling regression approach (Officer (1973); Fama and MacBeth (1973); Foster and Nelson (1996)); (3) summing absolute returns (Alizadeh et al. (2000)). For implied volatility, simultaneous equations estimators, weighted average estimators and others have been reported (see, Latané and Rendleman (1976); Beckers (1981) , for example). As reviewed above, substantial efforts have been put on statistical quantification of volatility itself (either realized or implied, separately), however, little effort is extended to the association between implied and realized volatility, and to its implications on option pricing and hedging.

Option pricing and hedging is initiated from the seminal work of Black and Scholes (1973) and Merton (1973). Since then, scholars have developed models to price and hedge various derivative securities, e.g. options on different underlying assets, futures, swaps, debt, more recently mortgage-backed securities and serial defaults. This body of work is mostly theory-oriented. The role of data or the numerically oriented procedures stay at the level of calibrating the model or of implementing the model in the situation of no closed-form solution. In other words, statistical uncertainty is rarely considered in option research.

In a series of papers, Mykland (1998, 1999, 2000, 2001) started to question the lack of communication between statistical uncertainty and financial engineering (abbreviated with FE) research, and pointed out that "neglecting statistics in option pricing is not just a bad form", in fact, naive implementation of the FE theories could lead to losing money in practice. To account for the statistical uncertainty in option pricing and hedging, Mykland (1998) has proposed a non-continuous model that on the one hand makes the task of statistical quantification possible, on the other hand his model converges to a diffusion whose properties are more familiar in option pricing and hedging literature.

On the empirical side, investigators have considered the impact of volatility estimation on option pricing. For example, deRoon and Veld (1996) looked at the mispricing of Dutch index warrants, using the historical standard deviation and implied volatility of the previous day, respectively, as the input to the option valuation model; Chu and Freund (1996) , and more recently Hardle and Hafner (2000), considered the volatility estimate based on GARCH model, and found that the use of GARCH model for volatility can reduce mispricing of an option, also Karolyi (1993) used a Bayesian approach to model volatility for option valuation. All these studies focused on comparing the mispricing with the Black and Scholes model when different volatility estimates are used.

The current work continues our efforts in stressing statistical quantifications in option pricing and hedging. Our work is different from most of the past studies in the literature, in the sense that we do not compare the mispricing of an option with different volatility input. Instead, we emphasize the association between realized volatility and option-implied volatility, and we make inference on the delta in a hedging situation. In particular, with the help of small-interval asymptotics and martingale decomposition techniques, we investigate the estimation properties and thus set confidence interval for the estimators of volatility, the delta (as in the delta hedge), and leverage and so on. Moreover, we have adopted a non-parametric estimation scheme, which frees us from various model assumptions on the system. For example,we do not need to specify a volatility model in the present study.

The inferential part of our results, which use a rolling sample scheme, permit unequal observation times, and has explicit forms for asymptotic variances. We also focus on the case where the underlying (unobserved) process is continuous. This permits a transparent handling of proofs using stochastic calculus. In particular, we present a natural decomposition for the estimation error of the volatility-type objects. This decomposition appears to fall into the traditional bias-variance trade-off, however, it becomes instead a variance-variance trade-off, *cf.* the discussion after Theorem 1. The inference problem studied is related to that of Foster and Nelson (1996), though our scope and results are different (see also the note after Corollary 1).

The organization is as follows. Section 2 describes the general inferential problem for volatility-style objects, for example, instantaneous covariation between returns and implied volatility. Section 3 discusses the application to options and how this leads to a regression problem. Section 4 presents the limiting distributions of the relevant estimation errors in Theorems 1 - 2. Section 5 focuses on the implication of our estimation results, in particular, the implications for pointwise and joint confidence intervals for the delta in a hedging situation. Finally, proofs are in the Appendix.

2 GENERAL SETUP

2.1 *Ito processes*

We shall be concerned with Ito processes, and their instantaneous variations and covariations.

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

$$X_t = \int_0^t v_u du + \int_0^t \sigma_u dW_u,$$

where W is a standard Brownian Motion. Note that W 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.

We are interested in the volatility and instantaneous covariation of Ito processes. To study this, one would start with the cumulative quadratic variation $< X, X >_t$ or covariation $< X, Y >_t$, as defined by Jacod and Shiryaev (1987) or Karatzas and Shreve (1991).

The volatility of the process X is then $\sigma_t^2 =< X, X >'_t$. The more general object is the instantaneous covariation $< X, Y >'_t$, so we shall mostly state general theorems about the latter. Note that the existence of the volatility follows from the Ito process assumption. Similarly, the absolute continuity of $< X, Y >_t$ follows from the Ito process assumption and from the Kunita-Watanabe Inequality (see, for example, p. 51 of Protter (1995)).

2.2 *The inference problem*

Considering now the general problem of finding $< X, Y >'_t$, note first that if the two processes X and Y were observed continuously, there would be no need for inference. The instantaneous covariation could be calculated exactly.

As it is, however, observations on diffusion process data are almost necessarily discrete. We suppose that there is an interval of observation $[0, T]$, and our processes are observed at a non-random partition $0 \leq t_1^{(n)} \leq t_2^{(n)} \leq \cdots \leq t_k^{(n)} = T$.

To mimic the continuous time $< X, Y >_t$, we let $[X, Y]_t$ represent the quadratic covariation of

X and Y at the discrete-time scale. In other words, if

$$\Delta X_{t_i^{(n)}} \triangleq X_{t_{i+1}^{(n)}} - X_{t_i^{(n)}}, \quad \Delta Y_{t_i^{(n)}} \triangleq Y_{t_{i+1}^{(n)}} - Y_{t_i^{(n)}},$$

then

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

Recall that $< X,Y >_t = \lim_{n\to\infty}[X,Y]_t$, where the convergence is uniform in probability (UCP), see Jacod and Shiryaev (1987) and Protter (1995) for details.

The limit is taken as the number of observation points $k = k_n \to \infty$, with the mesh $\delta^{(n)} = \max_t |\Delta t^{(n)}| \to 0$. Most of the time, we omit, for simplicity, the partition number (n).

To estimate the continuous quantity $< X,Y >'_t$, we use an approximation similar to the above, namely

$$<\widehat{X,Y}>'_t \triangleq \frac{1}{h} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \Delta X_{t_i^{(n)}} \Delta Y_{t_i^{(n)}},$$

in other words, $<\widehat{X,Y}>'_t \approx ([X,Y]_t - [X,Y]_{t-h})/h$. As $n \to \infty$, $h = h_n \to 0$. Further discussion of this procedure is given is Section 4.

The approach of letting the observation points become dense on $[0,T]$ is known as small interval asymptotics. We shall also use this approach to find limit laws for statistical errors, when approximating $< X,Y >'_t$ by $<\widehat{X,Y}>'_t$. This is described in Section 4.1.

This type of asymptotics leads to mixed normal limit laws jointly with the underlying data processes. Thus, we end this section with a definition.

DEFINITION (*Mixing convergence*): We let $\mathcal{X}$ be the (typically multidimensional) data generating process. We say $f^{(n),\mathcal{X}} \xrightarrow{\mathcal{L}} N(0,M)$ (*mixing*) if there exists a standard normal random vector W independent of $\mathcal{X}$, such that $(\mathcal{X}, f^{(n),\mathcal{X}})$ converge jointly in law to $((\mathcal{X}), M^{1/2}W)$, where $f_t^{(n),\mathcal{X}}$ is a function of $(\mathcal{X}_s)_{s\le t}$, $M^{1/2}$ is measurable with respect to process $\mathcal{X}$. $M^{1/2}$ is the square root of the symmetric, semi-positive definite matrix M.

There are two types of mixing, *mixing-past*, where the independence is of $(\mathcal{X}_s)_{s\le t}$ only, and *mixing-global*, where the independence is of $(\mathcal{X}_s)_{s\le T}$.

Interchangeably, we write $f^{(n),\mathcal{X}} \stackrel{\mathcal{L}.mixing}{\longrightarrow} N(0,M)$, or $f^{(n),\mathcal{X}} \stackrel{\mathcal{L}}{\longrightarrow} M^{1/2}W$, where W is standard normal random vector.

3 REGRESSION

3.1 *A generalized one factor model for options*

Following the findings in Mykland and Zhang (2001), and in Zhang (2001), we shall particularly be interested in the relationship between the price $\{V_t\}$ of an option, the price of the underlying stock $\{S_t\}$, and the cumulative implied volatility $\{\Xi_t\}$ of the option. Note that $V_t = C(S_t, r(T-t), \Xi_t)$, where C is the Black-Scholes (1973) - Merton (1973) formula expressed in cumulative terms. A regression relationship that accounts for the extent to which implied volatility can be hedged in the underlying stock is given by

$$\begin{aligned} d\Xi_t &= \rho_t dS_t + dZ_t, \\ dZ_t &= -\zeta_t dt. \end{aligned} \tag{3.1}$$

This is a generalization of the usual one-factor model. Further discussion of its trading aspect can be found in Mykland and Zhang (2001). We here, however, are mainly concerned with the question of inference for ρ_t. The connection to instantaneous covariation is as follows

$$\rho_t = \frac{<\Xi, S>'_t}{<S, S>'_t}. \tag{3.2}$$

Given this generalized one-factor setup in Equation (3.1), we have shown in Mykland and Zhang (2001) and Zhang (2001) that under the no-arbitrage rule, the delta hedge ratio can be written as

$$\Delta = C_S + \rho C_\Xi \tag{3.3}$$

where C is as defined as above and subscript refers to derivatives.

3.2 *Estimation*

When the scheme from Section 2 is used for the situation described in Section 3.1, it becomes what is known as rolling regression. This approach has been used frequently by econometricians since the 60's (see Fama and MacBeth (1973), also see Foster and Nelson (1996) for recent developments) when dealing with time-varying parameters.

The estimator for ρ is

$$\hat{\rho}_t = \frac{\widehat{<\Xi, S>}_t}{\widehat{<S, S>}_t} = \frac{\sum_{t-h \leq t_i^{(n)} < t_{i+1}^{(n)} \leq t} (\Delta \Xi_{t_i})(\Delta S_{t_i})}{\sum_{t-h \leq t_i^{(n)} < t_{i+1}^{(n)} \leq t} (\Delta S_{t_i})^2}. \tag{3.4}$$

As we shall see in the following sections, the estimation error of $\widehat{<\Xi, S>}_t$ (as well as $\widehat{<S, S>}_t$) can be decomposed into two parts, which are of order $O_p(\sqrt{h})$ and $O_p(\sqrt{\frac{\overline{\Delta t}^{(n)}}{h}})$ respectively. By stochastic Taylor expansion, the estimation error of ρ can be expressed as

$$\begin{aligned} \hat{\rho}_t - \rho_t &= \frac{1}{<S, S>'_t}[\widehat{<\Xi, S>}'_t - <\Xi, S>'_t] \\ &\quad - \frac{\rho_t}{<S, S>'_t}[\widehat{<S, S>}'_t - <S, S>'_t] + o_p(\sqrt{h} + \sqrt{\frac{\overline{\Delta t}^{(n)}}{h}}) \end{aligned}$$

Before we proceed to the asymptotic property of the estimation error associated with $\widehat{<\Xi, S>}_t$ and with $\hat{\rho}_t$, we first review under what paradigm and under what assumptions the asymptotics is considered.

4 STATISTICAL PROPERTIES

4.1 *Paradigm for asymptotic operations*

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

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

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

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

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

(5) there is a trade-off between h_n and $\overline{\Delta t^{(n)}}$, see the coming theorems. $\overline{\Delta t^{(n)}}$ is the average observation interval, equal to $\frac{T}{k}$.

The above (1) and (2) suggest that, as n increases, we can observe the underlying data process more and more frequently. This observation refinement is not nested in a sense that the set $\{t_0^{(n_1)}, t_1^{(n_1)}, t_2^{(n_1)}, \cdots, t_{k_{n_1}}^{(n_1)}\}$ is not necessarily contained in the set $\{t_0^{(n_2)}, t_1^{(n_2)}, t_2^{(n_2)}, \cdots, t_{k_{n_2}}^{(n_2)}\}$ for $n_1 < n_2$. It only means that with n increasing, the mesh of our observation intervals decreases, in a way that the number of observations in the estimation window increases, as indicated by (4). The requirement (3) indicates that the bandwidth h_n also shrinks with n. We shall show in the coming section 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. From now on, we use h and h_n interchangeably.

4.2 *Notations and assumptions*

ASSUMPTION A (*Homogenous partition*):

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 $\max_i(\Delta t_i^{(n)}) = \delta(n)$.

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

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

(3) $[H_{(n)}^{(2)}(t) - H_{(n)}^{(2)}(t-h)]/h \longrightarrow H^{(2)'}(t)$ as $h \longrightarrow 0$, where the convergence is uniform in t.

When the partitions are evenly spaced, $H^{(2)}(t) = t$ and $H^{(2)'}(t) = 1$. In the more general case, note that the left hand side of (2) is bounded by $t\delta(n)/\overline{\Delta t^{(n)}}$, while the left hand side of (3) 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 (1). The assumptions in (2) and (3) are, therefore, about a unique limit point, and about interchanging limits and differentiation.

For continuous Ito processes X and Y, write $dX_t = dX_t^{DR} + dX_t^{MG} = \tilde{X}_t dt + dX_t^{MG}$, $dY_t =$

$dY_t^{DR} + dY_t^{MG} = \tilde{Y}_t dt + dY_t^{MG}$, and

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

Assumptions on the processes (B-D) are imposed on the pair (X, Y):

Assumption B (*Smoothness*). *$B(X, Y)$: X, Y and $< X, Y >'$ are Ito processes. Also, the following items are in $C^1[0, T]$ almost surely*

(i) the respective quadratic variations of X, Y and $< X, Y >'$

(ii) the drift part of $< X, Y >_t'$ (D_t^{XY})

(iii) the drift parts of X (X^{DR}) and of Y (Y^{DR})

Note in (i) that the quadratic variation of $< X, Y >'$ is the same as $< R^{XY}, R^{XY} >$. The same should be observed about Assumption D below.

Assumption C (*Integrability*). $C(X, Y)$:

(i) $E \sup_{s \in [0,T]} |< X, X >_s'| < \infty$, and similarly for $< Y, Y >'$.

(ii) $E \sup_{s \in [0,T]} | \tilde{D}_s^{XX} | < \infty$, and similarly for $\tilde{D}^{YY}$.

Assumption D (*Non-vanishing volatility*). $D(X, Y)$:

$\inf_{t \in [0,T]} < R^{XY}, R^{XY} >_t' > 0$ *almost surely*

Assumption E (*Structure of the filtration*):

The data $(\mathcal{X}_t)$ is measurable with respect to a filtration generated by a finite number of Brownian Motions.

4.3 *Asymptotic distribution of the estimation error: main theorem*

Under the paradigm and assumptions listed in Section 4.1 and 4.2, we now consider the asymptotic property of the estimation errors $\widehat{< X, Y >}_t' - < X, Y >_t'$ and $\hat{\rho}_t - \rho_t$. We summarize the results in two theorems, whose proofs are provided in the Appendix. First, however, two quantities that constitute a natural decomposition of the estimation error,

$$B_{1,t}^{XY} = \frac{1}{h}(< X, Y >_t - < X, Y >_{t-h}) - < X, Y >_t'$$

$$B_{2,t}^{XY} = \frac{[2]}{h} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_s - X_{t_i^{(n)}}) dY_s$$

where [2] indicates symmetric representation s.t. $[2] \int X dY = \int X dY + \int Y dX$.

Theorem 1 *Suppose that X, Y, Z, and V are continuous Ito processes. Let B_1 and B_2 be defined as above. Also suppose we decompose $< X, Y >_t'$ into a martingale part (R_t^{XY}) and a drift part (D_t^{XY}). Under Assumptions A, $B(X,Y)$, $C(X,Y)$ and $D(X,Y)$, we have (a)-(b). If the same conditions are imposed on Z and V, (c) -(d) also hold.*

(a) $\widehat{< X, Y >}_t' - < X, Y >_t' = B_{1,t}^{XY} + B_{2,t}^{XY}$, *where* $B_{1,t}^{XY} = O_p(\sqrt{h})$, $B_{2,t}^{XY} = O_p(\sqrt{\frac{\overline{\Delta t}^{(n)}}{h}})$.

(b) In order for $B_{1,t}^{XY}$ and $B_{2,t}^{XY}$ to have the same order, $O(h) = O(\sqrt{\overline{\Delta t}^{(n)}})$. In this case, $B_{1,t}^{XY}$ and $B_{2,t}^{XY}$ are both of order $O_p((\overline{\Delta t}^{(n)})^{\frac{1}{4}})$.

(c) jointly and mixing,

$$h^{-1/2} \begin{bmatrix} B_{1,t}^{XY} \\ B_{1,t}^{ZV} \end{bmatrix} \xrightarrow{\mathcal{L}} N(0, M_1), \qquad (\frac{\overline{\Delta t_i}}{h})^{-1/2} \begin{bmatrix} B_{2,t}^{XY} \\ B_{2,t}^{ZV} \end{bmatrix} \xrightarrow{\mathcal{L}} N(0, M_2),$$

$$\begin{aligned} \textit{where } M_1 &= \frac{1}{3} \begin{bmatrix} < R^{XY}, R^{XY} >_t' & < R^{XY}, R^{ZV} >_t' \\ < R^{XY}, R^{ZV} >_t' & < R^{ZV}, R^{ZV} >_t' \end{bmatrix}, \\ \textit{and } M_2 &= H^{(2)'}(t) \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \end{aligned}$$

where

$$a_{11} = < X, X >_t' < Y, Y >_t' + (< X, Y >_t')^2,$$

$$a_{12} = a_{21} = < X, Z >_t' < Y, V >_t' + < X, V >_t' < Y, Z >_t',$$

$$a_{22} = < Z, Z >_t' < V, V >_t' + (< Z, V >_t')^2.$$

The convergence in law is mixing-past. Subject to Assumption E, it is also mixing-global.

(d) the asymptotic distributions of $B_{1,t}$ and $B_{2,t}$ are conditionally independent, given the data either up to time t or up to T, depending on whether Assumption E is used in (c). Also, $\forall t \neq t'$, $B^{XY}_{i,t}$ and $B^{ZV}_{i,t'}$ are conditionally independent given the data, under Assumption E. □

Under regularity conditions, Theorem 1(a) suggests that we can decompose the estimation error of the instantaneous quadratic covariation ($< X, Y >'_t$) into two parts: B^{XY}_1 and B^{XY}_2. From their mathematical expressions (see the beginning of Section 4.3), one perhaps would guess that we had a bias-variance trade-off regarding the estimation error of $< X, Y >'_t$, with B^{XY}_1 serving as a bias term, and B^{XY}_2 serving as a variance term. This would indeed have been the case in the traditional non-parametric estimation (e.g. density estimation). However, there is a difference between traditional and our nonparametrics: the former mainly deals with a smooth quantity, whereas the latter deals with a non-smooth quantity (namely $< X, Y >'_t$).

It turns out that to first order both B^{XY}_1 and B^{XY}_2 are variance terms. As shown in the proof in the Appendix, we can express B^{XY}_1 as

$$B^{XY}_{1,t} = \underbrace{\frac{1}{h}\int_{t-h}^{t}((t-h)-u)dR^{XY}_u}_{\text{martingale: variance term}} + \underbrace{\frac{1}{h}\int_{t-h}^{t}((t-h)-u)dD^{XY}_u}_{\text{bias term}}$$

where the variance term dominates when $< X, Y >'_t$ is not smooth, and the bias term becomes the only term when $< X, Y >'_t$ is smooth (i.e. $R^{XY}_t = 0$ when $< X, Y >'_t$ is smooth). In Theorem 1, $< X, Y >'_t$ is an Ito process, hence the first-order term of B_1 is dominated by a martingale component. Meanwhile, the first order of B^{XY}_2 is also a martingale term, which does not vanish even if $< X, Y >'_t$ is smooth (see the proof in appendix for details). Therefore, we are faced with a variance-variance trade-off in the estimation error of $< X, Y >'_t$.

Theorem 1(b) says that the order of B_1 is determined by the smoothing bandwidth h alone, whereas the order of B_2 depends on the number of observations used for estimation purpose at each time t (i.e. the number of observations in $(t-h, t]$). It is optimal to select h with the order of square root of the average observation interval, optimal in the sense of minimizing the asymptotic variance in the estimation error in part (a).

The asymptotic distributions in (c) are normal mixtures, after M_1 and M_2 are estimated from the data. A more explicit representation would be

$$h^{-1/2}\begin{bmatrix} B^{XY}_{1,t} \\ B^{ZV}_{1,t} \end{bmatrix} \xrightarrow{\mathcal{L}} M^{1/2}_{1,t}\mathcal{E}_1$$

and

$$(\frac{\overline{\Delta t^{(n)}}}{h})^{-1/2}\begin{bmatrix} B^{XY}_{2,t} \\ B^{ZV}_{2,t} \end{bmatrix} \xrightarrow{\mathcal{L}} M^{1/2}_{2,t}\mathcal{E}_2$$

where $\mathcal{E}_1$ and $\mathcal{E}_2$ are bivariate normal independent of each other. It is worth to point out that $M_{1,t}$ and $M_{2,t}$ depend on the data, whereas the $\mathcal{E}s$ are independent of data.

One here encounters the issue of conditional distribution versus unconditional distribution. *Conditional on data*, $M_{1,t}$ and $M_{2,t}$ are observable in a world of continuous records or approximately observable in a discrete-record world. Thus if h is proportional of $\sqrt{\overline{\Delta t}^{(n)}}$, and $\frac{\sqrt{\overline{\Delta t}^{(n)}}}{h} \to c$ as n increases, we can then, for example, construct an approximate 95% conditional confidence set for $< X, Y >'_t$ by $\widehat{< X, Y >}'_t \pm 1.96 h^{1/2}\sqrt{\hat{M}^{(1,1)}_{1,t} + c^2\hat{M}^{(1,1)}_{2,t}}$, where $M^{(1,1)}_{i,t}$ means the (1,1) element in the matrix of $M_{i,t}$. *Unconditionally*, the confidence set is generally different due to dependence between $\mathcal{E}$ and the data. Our findings on the independence between $\mathcal{E}$ and the data make our unconditional confidence set and conditional confidence set the same.

Theorem 1(d) suggests that the quadratic covariation between $B_{1,t}$ and $B_{2,t}$ is of higher order, so is the covariation between $B_{i,t}$ and $B_{i,t'}$ for $t \neq t'$. In the limit, $B_{1,t}$ and $B_{2,t}$ (also $B_{i,t}$ and $B_{i,t'}$ for $t \neq t'$) become uncorrelated, which is the same as independent given the Gaussian findings in (c).

Remarks:

1. Notice that $< X, Y >'_t$ is a random quantity, NOT a constant. The latter is the frequentist's typical notation of a parameter. In this paper, we borrow the terminology "estimation" and "confidence set", and use them in a broader way. The alternative would be to use "prediction" and "prediction set", but this tends to confuse because of the connotations of forecasting future data.

2. The results in Theorem 1 involve the following order of operation: as a first step, the convergence is joint with the underlying data processes (see the definition in Section 2.2) $\{(\Xi_t, S_t)\}$;

then, conditional on the observable (i.e. the whole data processes), M_1 and M_2 can be estimated, making the limit in Theorem 1(c) a mixture normal. Similarly, we can discuss asymptotic bias, variance, and independence after the joint convergence and then the conditioning operations.

4.4 *Estimation of volatility and of regression coefficients*

Suppose we set both X and Y equal to $\log(S)$, then Theorem 1 tells us the asymptotic distribution of realized volatility $< \log S, \log S >'_t$.

Corollary 1 *Suppose that the stock price S is a continuous Ito process. Let $X_t = logS_t$, $\sigma_t^2 =< X, X >'_t$, $\hat{\sigma}_t^2 = \widehat{< X, X >}'_t$. Under Assumptions A, $B(X,X)$, $C(X,X)$, $D(X,X)$, and the assumption about the order of h in Theorem 1(b),*

$$\hat{\sigma}_t^2 - \sigma_t^2 = \frac{1}{h}\int_{t-h}^{t}(t-h-u)dR_u^{XX} + [2]\frac{1}{h}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_u - X_{t_i^{(n)}})dX_u^{MG} + o_p((\overline{\Delta t}^{(n)})^{\frac{1}{4}})$$

Furthermore, if $\frac{\sqrt{\overline{\Delta t}^{(n)}}}{h} \to c$ (nonrandom), conditional on data, $(\overline{\Delta t}^{(n)})^{-1/4}(\hat{\sigma}_t^2 - \sigma_t^2)$ is asymptotically distributed $N(0, V_{\hat{\sigma}^2-\sigma^2})$ (mixing), where

$$V_{\hat{\sigma}^2-\sigma^2} = \frac{1}{3c} < \sigma^2, \sigma^2 >'_t + 2cH^{(2)'}(t)\sigma_t^4 \tag{4.5}$$

The nature of the mixing depends on Assumption E about the data filtration in Theorem 1(c). □

Note that the connection of the first term in Equation (4.5) to Theorem 1 is that $< R^{XX}, R^{XX} >'_t = \langle < X, X >', < X, X >' \rangle'_t =< \sigma^2, \sigma^2 >'_t$. This corrects the expressions in Theorem 2 in Foster and Nelson (1996), when considering the continuous-time limit in their Equation (9) (p. 149).

Theorem 2 *Suppose S and Ξ are continuous Ito processes. Let $\hat{\rho}_t = \frac{\widehat{<\Xi,S>}'_t}{\widehat{<S,S>}'_t}$. Subject to the assumptions applied in Theorem 1 with X=Ξ, Y=Z=V=S, and $O(h) = O(\sqrt{\overline{\Delta t}^{(n)}})$, we have*

(a) representation:

$$\hat{\rho}_t - \rho_t = \frac{1}{< S, S >'_t}[B_1^{\Xi S} - \rho_t B_1^{SS}] + \frac{1}{< S, S >'_t}[B_2^{\Xi S} - \rho_t B_2^{SS}] + o_p((\overline{\Delta t}^{(n)})^{1/4})$$

(b) asymptotic distribution:

if $\frac{\sqrt{\overline{\Delta t}^{(n)}}}{h} \to c$ *(nonrandom), conditional on data,* $(\overline{\Delta t}^{(n)})^{-1/4}(\hat{\rho}_t - \rho_t)$ *is asymptotically distributed* $N(0, V_{\hat{\rho}-\rho})$*, where*

$$V_{\hat{\rho}-\rho} = \frac{<\rho,\rho>'_t}{3c} + cH^{(2)'}(t)(\frac{<\Xi,\Xi>'_t}{<S,S>'_t} - \rho_t^2) \tag{4.6}$$

The convergence in law is mixing, with the past or globally, depending on whether Assumption E is used. □

According to Theorem 2, $\hat{\rho}_t - \rho_t$ has the order of $(\overline{\Delta t}^{(n)})^{1/4}$, where $\overline{\Delta t}^{(n)}$ is the average length of sampling interval. We can arrange the first-order term of $\hat{\rho}_t - \rho_t$ into two parts, one is related to the B_1's and the other related to the B_2's. In the limit, conditional on the whole data process, the estimation error of ρ_t follows a mixture normal distribution, with mean 0 and variance equal to the estimate of $V_{\hat{\rho}-\rho}$. Equation (4.6) indicates that under-smoothing (i.e. c is greater) or over-smoothing would blow up the asymptotic variance. For example, an under-smoothing would reduce $<\rho,\rho>'_t/(3c)$ while increasing $cH^{(2)'}(t)(\frac{<\Xi,\Xi>'_t}{<S,S>'_t} - \rho_t^2)$. This implies that an optimal rate c can be reached in order to minimize the asymptotic variance of the estimation error of ρ_t. The same thing goes for $\hat{\sigma}_t^2$.

Both for $\hat{\sigma}^2$ and $\hat{\rho}$ an optimal choice of c can be found. For example, for $\hat{\rho}$, it would appear that the optimal rate is given with

$$c^2 = c_t^2 = \frac{1}{3}\frac{<\rho,\rho>'_t}{H^{(2)'}(t)(\frac{<\Xi,\Xi>'_t}{<S,S>'_t} - \rho_t^2)}$$

which can then be estimated from the data. The optimal asymptotic variance is then

$$V_{\hat{\rho}-\rho} = 2\left[\frac{<\rho,\rho>'_t}{3}H^{(2)'}(t)(\frac{<\Xi,\Xi>'_t}{<S,S>'_t} - \rho_t^2)\right]^{1/2}$$

We have not investigated how a data-dependent choice of c would affect our theoretical results, which assume nonrandom c.

Remarks:

1. The *mixture* normal result in Theorem 2 mainly comes from our estimation mechanism, where we have used an increasing number of data records in a finite amount of time to deliver the estimator.

2. The convergence holds at each time t, but not as a process. In other words, $\hat{\rho} - \rho$ does not converge as a process, because as $n \to \infty$, $\hat{\rho}_t - \rho_t$ and $\hat{\rho}_{t'} - \rho_{t'}$ become independent for $t \neq t'$, and in the normal stochastic process paradigm, there is no such process consisting of independent elements at each time t. Such a process would be continuous white noise, and the derivative of (the non-differentiable) Brownian Motion.

3. When estimating $< \sigma^2, \sigma^2 >'$, $< \rho, \rho >'$, or, in the broader case of Theorem 1, $< R^{XY}, R^{ZV} >'$ $= \langle < X, Y >', < Z, V >' \rangle'$, a consistent estimate can be obtained by plugging in the estimated quantities for σ^2, ρ, or $< X, Y >'$. One can no longer, however, use the original grid $0 = t_0 < t_1 < ... < t_k = T$ when computing the "outer" $< \cdot, \cdot >'$, but rather a sub-grid or some other partition that is coarser than the original grid, and which permits consistent estimation at each point of the coarser partition. We have not investigated the precise theoretical requirements in this paper, but this is the procedure which lays behind the error bounds in Figure 1 in next section.

5 IMPLICATIONS

5.1 *Implications for the hedge ratio*

Following Equation (3.3) in Section 3.1, $\Delta = C_S + \rho C_\Xi$, where Δ stands for the delta hedge (i.e. the number of stocks to hold for offsetting the risk in option). This implies that the estimation error of the hedge ratio is given by

$$\hat{\Delta} - \Delta = C_\Xi(\hat{\rho} - \rho). \tag{5.7}$$

Hence, our asymptotic results on ρ can help setting a confidence region for Δ. In addition, tests of hypothesis $H_0 : \rho = 0$ vs. $H_a : \rho \neq 0$ tells us whether or not our hedge ratio Δ is significantly different from the Black-Scholes hedge C_S. Finally, our result provides a way of hedging without knowing the model for S. This is not affected by the fact that we use the Black-Scholes-Merton functional form. It does, however, assume the generalized one-factor model in Equation (3.1).

Figure 1: 90% Confidence Interval for Relative Hedge, S&P 500 on Feb. 17, 1994

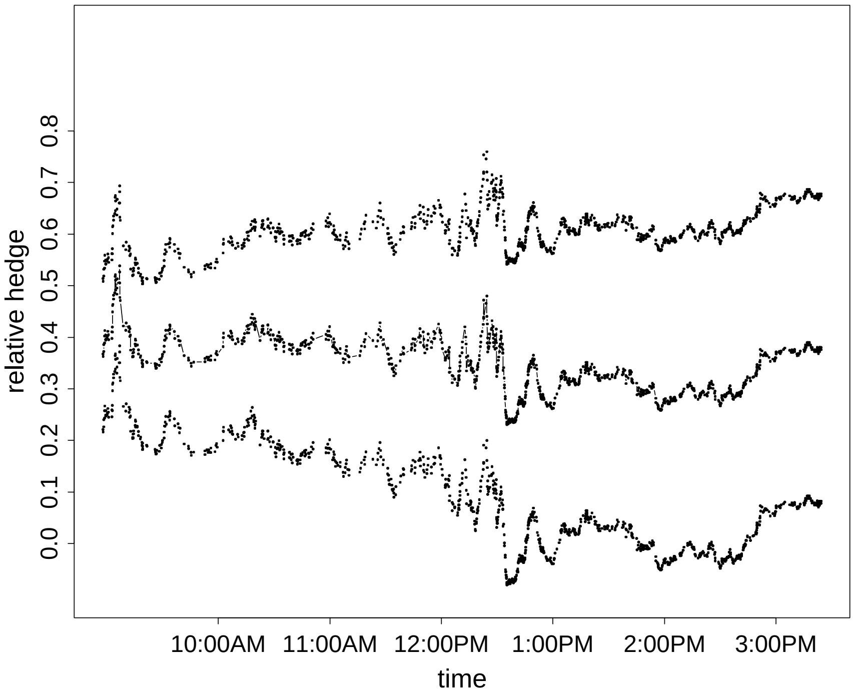

Figure 1 is one example of applying Theorem 2 in option hedging. Using the data from S&P 500 index and option, we can investigate how relative hedge, as well as its 90% confidence interval, evolves across one day. In this application, the relative hedge denotes the ratio of our one-factor delta relative to the Black-Scholes delta ($\frac{\Delta}{C_S}$). As we can see from Figure 1, even the upper bound of the 90% CI of the relative hedge is smaller than 1, indicating that the Black-Scholes delta over-hedged, at least on February 17, 1994. Notice that the confidence interval in Figure 1 is pointwise.

5.2 *Other considerations on confidence sets for ρ*

In the previous section, we considered how to make inference on ρ and then on Δ at each time t. In a real market, making decisions at each possible observation time is too expensive (due to the transaction cost incurred by each hedging action) and too dangerous (due to the uncertainty coming from estimation error, data discreteness, and unexpected news, for example). Therefore, it would be more reasonable to make a hedging decision based on information from several time periods.

Because the delta hedge is closely related to ρ (at least in the generalized one-factor case as we have assumed in this section), we concentrate on ρ at this moment. Instead of focusing on the distribution of ρ at one time t, we now consider simultaneous confidence set for ρ_{s_i} at several times $i = 1, 2, \cdots, m$.

Let $U_n(t) = (\overline{\Delta t^{(n)}})^{-\frac{1}{4}} \frac{1}{\sqrt{V_{\hat{\rho}-\rho}(t)}} (\hat{\rho}_t - \rho_t)$, let $1-\alpha$ be the simultaneous coverage probability, and $1-\gamma$ be the coverage probability at a specific time point, then

$$
\begin{aligned}
1-\alpha &= P\left[\cap_{i=1}^m \{|U_n(s_i)| \le z_{\gamma/2}\}\right] \\
&= \prod_{i=1}^m P\{|U_n(s_i)| \le z_{\gamma/2}\} \qquad (5.8)\\
&\approx (1-\gamma)^m \qquad (5.9)
\end{aligned}
$$

where (5.8) is because $\hat{\rho}_{s_i} - \rho_{s_i}$ and $\hat{\rho}_{s_j} - \rho_{s_j}$ are asymptotically independent for $i \neq j$. Two issues are worth to be pointed out: 1) for fixed α, bigger m leads to smaller γ. This may lead to a true question of bias-variance tradeoff, and this remains to be investigated. If one makes inference on more time periods jointly while maintaining the acceptable total uncertainty, one has to suffer from the wider estimation error at each individual time point; 2) for γ small, (5.9) is close to the multiple comparison result given by Bonferroni Inequality.

Alternatively, we can consider the average coverage, that is,

$$
\begin{aligned}
&\text{fraction of times that CI covers } \rho \\
&= \frac{1}{m}\sum_{i=1}^m I\left(|U_n(s_i)| \le z_{\alpha/2}\right) \to 1-\alpha \quad \text{as } m \to \infty, n \to \infty
\end{aligned}
$$

Both approaches to constructing joint confidence sets can be particularly useful from the viewpoint of risk management.

6 APPENDIX

6.1 *Supporting convergence theorems*

It should be emphasized that the results in this sub-section are straightforward applications of standard limit theory for stochastic processes, as discussed, for example, in the book by Jacod and Shiryaev (1987). Similar results to the ones below exist in many forms in the literature. Because of our application, however, we needed rather specific formulations, and this led us to state and prove the results below.

Theorem A. 1 (Broad Framework Convergence Theorem):

Suppose X and $M^{(n)}$, respectively, are a continuous multidimensional martingale and a sequence of continuous martingales. The martingales are with respect to filtration $\mathcal{F}_{t\leq T}$, where $\mathcal{F}_t = \sigma(X_s, s \leq t)$. Also $M_s^{(n)} = 0, \forall s \leq t - h_n$. Let $\Psi^{(n)}$ be a sequence of time changes, where

$$\Psi^{(n)}(s) = \begin{cases} s & s \leq t - h_n \\ [s-(t-h_n)]h_n + (t-h_n) & t - h_n < s \leq t - h_n + 1 \\ t & t - h_n + 1 < s \leq t + 1 \\ s - 1 & t + 1 < s \leq T + 1 \end{cases}.$$

Let $\tilde{X}_s^{(n)} = X_{\Psi^{(n)}(s)}$, and let

$$\tilde{Y}_s^{(n)} = \begin{cases} 0 & s \leq t - h_n \\ h_n^{-\frac{\alpha}{2}}(M^{(n)}_{[s-(t-h_n)]h_n+(t-h_n)} - M^{(n)}_{t-h_n}) & t - h_n < s \leq t - h_n + 1 \\ h_n^{-\frac{\alpha}{2}}(M_t^{(n)} - M^{(n)}_{t-h_n}) & t - h_n + 1 < s \leq T + 1 \end{cases}$$

Assume

1) $h_n \downarrow 0$ as $n \uparrow \infty$,

2) $h_n^{-\alpha}(<M^{(n)}, M^{(n)}>_{[s-(t-h_n)]h_n+(t-h_n)} - <M^{(n)}, M^{(n)}>_{t-h_n}) \xrightarrow{P} \eta_t^2 f_t(s-t), \forall s \geq t$,

3) $f_t(s)$ is nonrandom and continuously differentiable, with $f_t(0) = 0$, and η_t is random variable measurable with respect to $\mathcal{F}_t$.

Then, $(\tilde{X}_t^{(n)}, \tilde{Y}_t^{(n)})_{0\leq t\leq T+1}$ is C-tight. Moreover, any limit $(\tilde{X}, \tilde{Y})_{0\leq t\leq T+1}$ of a convergent subse-

quence of this sequence satisfies

$$\begin{array}{rcl} \tilde{X}_s & = & X_{\Psi(s)} \\ \tilde{Y}_s & = & \begin{cases} 0 & \text{for } s < t \\ \eta_t \int_t^{s\wedge(t+1)} (f_t'(u-t))^{1/2} d\tilde{W}_u & \text{for } s \geq t \end{cases} \end{array}$$

where $\Psi(s) = \begin{cases} s & s \leq t \\ t & t < s \leq t+1 \\ s-1 & t+1 < s \leq T+1 \end{cases}$ *, and* $\tilde{W}$ *is a Brownian motion on* $[t, t+1]$.

Proof: (for simplicity, write h instead of h_n in the next proof.)

As $n \to \infty$, $\Psi^{(n)}(s) \to \Psi(s)$, where Ψ is another time change. By definition of $\tilde{X}^{(n)}$, we have

$$\tilde{X}_s^{(n)} \longrightarrow X_{\Psi(s)} = \tilde{X}_s, \forall s \leq T+1 \tag{6.1.1}$$

As a matter of fact, $\tilde{X}^{(n)}$ converges to $\tilde{X}$ locally uniformly (a.s.) since for small h,

$$\begin{array}{rl} & \sup_{s\leq T+1} \mid \tilde{X}_s^{(n)} - \tilde{X}_s \mid \\ \leq & \sup_{s\leq t-h} \mid \tilde{X}_s^{(n)} - \tilde{X}_s \mid + \sup_{t-h<s\leq t} \mid \tilde{X}_s^{(n)} - \tilde{X}_s \mid + \sup_{t<s\leq t-h+1} \mid \tilde{X}_s^{(n)} - \tilde{X}_s \mid \\ & + \sup_{t-h+1<s\leq t+1} \mid \tilde{X}_s^{(n)} - \tilde{X}_s \mid + \sup_{t+1<s\leq T+1} \mid \tilde{X}_s^{(n)} - \tilde{X}_s \mid \\ = & \sup_{t-h<s\leq t} \mid X_{[s-(t-h)]h+(t-h)} - X_s \mid + \sup_{t<s\leq t-h+1} \mid X_{[s-(t-h)]h+(t-h)} - X_t \mid \\ \leq & \sup_{|u-v|\leq 2h} |X_u - X_v| + \sup_{|u-v|\leq h} \mid X_u - X_v \mid \\ \longrightarrow & 0 \quad \text{as } X \text{ is continuous and } u, v < T+1 \end{array}$$

so $\tilde{X}^{(n)} \to \tilde{X}$ in D(R).

Similarly, $\sup_{s\leq T+1} \mid < \tilde{X}^{(n)}, \tilde{X}^{(n)} >_s - < X, X >_{\Psi(s)} \mid \to 0$, thus by Jacod and Shiryaev (1987) (abbreviated with J&S hereafter) VI proposition 1.17 (p. 292)

$$< \tilde{X}^{(n)}, \tilde{X}^{(n)} > \to < \tilde{X}, \tilde{X} > \text{ in D(R)} \tag{6.1.2}$$

By definition of $\tilde{Y}^{(n)}$ and assumption 2),

$$< \tilde{Y}^{(n)}, \tilde{Y}^{(n)} >_s \xrightarrow{P} \begin{cases} 0 & s \leq t \\ \eta_t^2 f_t(s-t) & t \leq s < t+1 \\ \eta_t^2 f_t(1) & t+1 \leq s \leq T+1 \end{cases} \quad \text{as } n \to \infty. \tag{6.1.3}$$

$$\text{Jointly, } < \tilde{X}^{(n)}, \tilde{Y}^{(n)} >_s \xrightarrow{P} 0 \tag{6.1.4}$$

(6.1.3) is true for all s, hence true for a subset in $[t, t+1]$. Since $[\tilde{Y}^{(n)}, \tilde{Y}^{(n)}]$ is nondecreasing and has continuous limit, J&S Theorem VI 3.37 (p. 318) yields that the convergence is in law (D(R)). By using continuity and equation (6.1.2), $< \tilde{X}^{(n)}, \tilde{X}^{(n)} >= [\tilde{X}^{(n)}, \tilde{X}^{(n)}]$ is C-tight, and $< \tilde{Y}^{(n)}, \tilde{Y}^{(n)} >= [\tilde{Y}^{(n)}, \tilde{Y}^{(n)}]$ is C-tight. So the sequence $\{([\tilde{X}^{(n)}, \tilde{X}^{(n)}], [\tilde{Y}^{(n)}, \tilde{Y}^{(n)}])\}$ is C-tight by J&S Corollary VI 3.33 (p. 317). Invoking J&S Theorem VI. 4.13 (p. 322), we have the sequence $(\tilde{X}^{(n)}, \tilde{Y}^{(n)})$ is tight.

Now, given any subsequence, we can find further subsequence such that

$(\tilde{X}^{(n)}, \tilde{Y}^{(n)}) \to (\tilde{X}, \tilde{Y})$. (6.1.2)-(6.1.4) and J&S corollary VI. 6.7 (p. 342) lead to

$$\begin{aligned} &((\tilde{X}^{(n)}, \tilde{Y}^{(n)}), [\tilde{X}^{(n)}, \tilde{X}^{(n)}], [\tilde{Y}^{(n)}, \tilde{Y}^{(n)}], [\tilde{X}^{(n)}, \tilde{Y}^{(n)}]) \\ &\qquad \xrightarrow{\mathcal{L}} \quad ((\tilde{X}, \tilde{Y}), [\tilde{X}, \tilde{X}], [\tilde{Y}, \tilde{Y}], [\tilde{X}, \tilde{Y}]) \end{aligned}$$

where $[\tilde{X}, \tilde{X}]_s = [X, X]_{\Psi(s)}, [\tilde{X}, \tilde{Y}] = 0,\ [\tilde{Y}, \tilde{Y}] = \begin{cases} 0 & s \le t \\ \eta_t^2 f_t(s-t) & t \le s < t+1 \\ \eta_t^2 f_t(1) & t+1 \le s \le T+1 \end{cases}$ This implies that $\tilde{X}$ and $\tilde{Y}$ are continuous local martingales. The latter follows from Proposition IX. 1.17 in J&S by using continuity of $M^{(n)}$.

$$\text{If } f' > 0, \text{ let } \tilde{W}_s = \begin{cases} 0 & s \le t \\ \frac{1}{\eta_t} \int_t^s (\frac{d}{du} f(u-t))^{-\frac{1}{2}} d\tilde{Y}_u & t \le s \le t+1 \end{cases} \tag{6.1.5}$$

If f' is not always positive, create $\tilde{W}_s$ as in Vol III of Gikhman and Skorokhod (1979). By definition (6.1.5), $< \tilde{W}, \tilde{W} >= s - t$ for $t \le s \le t+1$. By Levy's Theorem (J&S II Theorem 4.4, p. 102), $\tilde{W}$ is a Brownian Motion on $[t, t+1]$, and it has increments independent of $\tilde{\mathcal{F}}_t$, which is defined as $\sigma(\tilde{X}_u, u \le t)$. Since $\tilde{X}_s = X_s$ for $s \le t$ and $\tilde{X}_s = X_t$ for $t \le s \le t+1$, it follows that $\tilde{W}$ is independent of X over $[0, t+1]$. Hence the joint convergence to $(\tilde{X}, \tilde{Y})$ is uniquely defined, and is independent of subsequence. By inverting equation (6.1.5), we obtain

$$\tilde{Y}_s = \begin{cases} 0 & \text{for } s < t \\ \eta_t \int_t^{s \wedge (t+1)} (f_t'(u-t))^{1/2} d\tilde{W}_u & \text{for } s \ge t \end{cases} \tag{6.1.6}$$

∎

Theorem A. 2 (Convergence Theorem with Independence of the Past):

Following the same setup and assumptions as in Theorem A.1, also assume $T = t$, we have

$$(X_{u,0\le u\le t}, h_n^{-\frac{\alpha}{2}}(M_t^{(n)} - M_{t-h_n}^{(n)})) \xrightarrow{\mathcal{L}} (X_{u,0\le u\le t}, \eta_t\sqrt{f_t(1)}Z),$$

where Z is standard normal independent of the X-process.

Proof:

In formula (6.1.6), f' is nonrandom and the Brownian Motion $\tilde{W}$ has the independent increment property, hence $\tilde{\tilde{Y}}_{t+1} = \int_t^{t+1}(f_t'(u-t))^{1/2}d\tilde{W}_u$ is Gaussian and independent of $\tilde{\mathcal{F}}_t$. Also $< \tilde{\tilde{Y}}, \tilde{\tilde{Y}} >_{t+1} = \int_t^{t+1} f_t'(u-t)du = \int_0^1 f_t'(u)du = f_t(1)$. So $\tilde{\tilde{Y}}_{t+1} \sim N(0, f_t(1))$, independent of the $\tilde{X}$-process. Then $\tilde{Y}_{t+1} \overset{\mathcal{L}}{=} \eta_t(f_t(1))^{1/2}Z$, where Z is standard normal, independent of $\tilde{X}$-process. From definition (6.1.1), $\tilde{X}_s = X_s, \forall 0 \le s \le t$, hence in the end,

$(X_{u,0\le u\le t}, h_n^{-\frac{\alpha}{2}}(M_t^{(n)} - M_{t-h_n}^{(n)})) \xrightarrow{\mathcal{L}} \left(X_{u,0\le u\le t}, \eta_t(f_t(1))^{1/2}Z\right)$, where Z is independent of X-process. ∎

In the case $T > t$, one needs additional regularity conditions, we here give one version. Also, this extra condition may not be needed from the point of view of estimating σ^2 or ρ at point t.

Theorem A. 3 (Convergence Theorem with Independence of both Past and Future):

Following the same setup and assumptions as in Theorem A.1, also assume $\mathcal{F}_t$ is generated by $(W_t^{(1)}, W_t^{(2)}, \ldots, W_t^{(q)})_{0\le t\le T}$, where the W's are independent Brownian Motions. Then we have

$$(X_{u,0\le u\le t}, h_n^{-\frac{\alpha}{2}}(M_t^{(n)} - M_{t-h_n}^{(n)})) \xrightarrow{\mathcal{L}} (X_{u,0\le u\le T}, \eta_t\sqrt{f_t(1)}Z),$$

where Z is standard normal independent of the X-process.

Proof:

Let $\tilde{\mathcal{F}}_t = \sigma(W_{\Psi(t)}^{(i)}, i = 1, 2, \cdots, q; \tilde{W}_t)$ in Theorem 1, and $X_t = (W_t^{(1)}, \cdots, W_t^{(q)})$. Since $[\tilde{W}, W^{(i)}]_t = 0$, $\tilde{W}$ is independent of X. Therefore the results of Theorem 3 hold. ∎

6.2 *Supporting lemmas and corollaries*

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

In analogy with the definition of $H_{(n)}^{(2)}(t)$ in Assumption A, we also define $H_{(n)}^{(j)}(t)$ for $j \geq 1$:

$$H_{(n)}^{(j)}(t) = \frac{\sum_{t_{i+1}^{(n)} \leq t} (\Delta t_i^{(n)})^j}{(\overline{\Delta t^{(n)}})^{j-1}}.$$

By the same argument given just after Assumption A, $[H_{(n)}^{(j)}(t) - H_{(n)}^{(j)}(t - h_n)]/h_n$ is bounded, and hence every sequence (in n) has a convergent subsequence. For clarity of exposition, we shall act as if the sequence itself converges as $n \to \infty$, and call the limit $H^{(j)'}(t)$. Wherever this is used, it is easy to see that the relevant argument (which is always about stochastic order) goes through without the existence of a limit.

Also, for convenience, we disaggregate Assumptions B and C as follows:

Assumption B (*Smoothness*):

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

$B.2(X, Y)$: the drift part of $< X, Y >_t'$ (D_t^{XY}) is in $C^1[0, T]$.

$B.3(X)$: the drift part of X (X^{DR}) is in $C^1[0, T]$.

Assumption C (*Integrability*):

$C.1(X, Y)$: $E \sup_{s \in [0,T]} |< X, Y >_s'| < \infty$.

$C.2(X, Y)$: $E \sup_{s \in [0,T]} | \tilde{D}_s^{XY} | < \infty$.

Assumption $B(X, Y)$ is equivalent to $B.1(X, X)$, $B.1(Y, Y)$, $B.1(R^{XY}), R^{XY})$, $B.2(X, Y)$, $B.3(X)$, and $B.3(Y)$. Similarly, $C(X, Y)$ is equivalent to $C.1(X, X)$, $C.1(Y, Y)$, $C.2(X, X)$ and $C.2(Y.Y)$. Corresponding statements involving covariations of X and Y follow by the Kunita-Watanabe inequalities (Protter (1995), pp. 61-62).

Notice that we shall be using the following notations

$$\Upsilon^X(h) = \sup_{t-h\le u\le s\le t} |X_u - X_s| \quad (6.2.1)$$

$$\Upsilon^{XY}(h) = \sup_{t-h\le u\le s\le t} |< X,Y >'_u - < X,Y >'_s| \quad (6.2.2)$$

Assumption $B.1(X,Y)$ implies $\Upsilon^{XY}(h) \to 0$. Moreover, from condition $C.1(XX)$ and $C.2(XX)$, Burkholder's Inequality yields that $E\Upsilon^X(h) = o(1)$ in h.

Lemma 1 Suppose X, Y, and Z are Ito processes . Subject to assumptions A, $B.1[(X,X),(Z,Z),(X,Z)]$, $B.3[(X)(Z)]$ and $C.1[(X,X),(Z,Z)]$, we have the following for any constant $k > 0$,

(i):

$$\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(< X >_u - < X >_{t_i})(u-t_i)^k Y_u du = O_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h})$$

$$\begin{aligned}&\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_u - X_{t_i})^2(u-t_i)^k Y_u du\\ &= \frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(< X >_u - < X >_{t_i})(u-t_i)^k Y_u du + o_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h})\end{aligned}$$

(ii)

$$\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_u - X_{t_i^{(n)}})(Z_u - Z_{t_i^{(n)}})(u-t_i)^k Y_u du = O_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h})$$

$$\begin{aligned}&\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_u - X_{t_i^{(n)}})(Z_u - Z_{t_i^{(n)}})(u-t_i)^k Y_u du\\ &= \frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(< X,Z >_u - < X,Z >_{t_i^{(n)}})(u-t_i)^k Y_u du\\ &\quad + o_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h})\end{aligned}$$

$\square$

Proof of Lemma 1:

(i) By Itô's Lemma,

$$\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_u-X_{t_i})^2(u-t_i)^kY_udu$$

$$= \frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[<X>_u-<X>_{t_i}+2\int_{t_i}^{u}(X_v-X_{t_i})dX_v](u-t_i)^kY_udu$$

$$= \underbrace{\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[<X>_u-<X>_{t_i}](u-t_i)^kY_udu}_{I}$$

$$+\underbrace{\frac{2}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[\int_{t_i}^{u}(X_v-X_{t_i})dX_v](u-t_i)^kY_udu}_{II}$$

Now we show that both I and II are of oder $O_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h})$. First,

$$\begin{aligned}
|I| &\le \frac{1}{k+2}\sup_{0\le u\le t}<X>'_u\sup_{0\le u\le t}|Y_u|\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}(\Delta t_i)^{k+2} \\
&\underset{\sim}{\text{assumption A}} \frac{H^{(k+2)'}(t)}{k+2}\sup_{0\le u\le t}<X>'_u\sup_{0\le u\le t}|Y_u|\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h} \\
&= O_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h}) \qquad (6.2.3)
\end{aligned}$$

where Equation (6.2.3) follows from assumption $B.1(X,X)$ and the continuity of Y.

For II, we write X as the sum of X^{MG} and X^{DR},

$$II = \underbrace{\frac{2}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[\int_{t_i}^{u}(X_v-X_{t_i})dX_v^{DR}](u-t_i)^kY_udu}_{II_1}$$

$$+\underbrace{\frac{2}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[\int_{t_i}^{u}(X_v^{DR}-X_{t_i}^{DR})dX_v^{MG}](u-t_i)^kY_udu}_{II_2}$$

$$+\underbrace{\frac{2}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[\int_{t_i}^{u}(X_v^{MG}-X_{t_i}^{MG})dX_v^{MG}](u-t_i)^kY_udu}_{II_3}$$

Recall that $dX_v^{DR}=\tilde{X}_vdv$,

$$\begin{aligned}
|II_1| &\le \frac{2}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[\int_{t_i}^{u}|(X_v-X_{t_i})\tilde{X}_v|dv](u-t_i)^k|Y_u|du\\
&\le \sup_{0\le u\le t}|Y_u|\sup_{0\le u\le t}|\tilde{X}_u|\Upsilon^X(h)\frac{2}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(u-t_i)^{k+1}du\\
&\underset{\sim}{\text{assumption A}} \frac{2}{k+2}\sup_{0\le u\le t}|Y_u|\sup_{0\le u\le t}|\tilde{X}_u|\Upsilon^X(h)H^{(k+2)\prime}(t)\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h}\\
&= o_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h}) \qquad (6.2.4)
\end{aligned}$$

where Equation (6.2.4) follows from assumption $B.3(X)$ and the continuity of X and Y. Similar approach leads to $II_2=o_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h})$.

Let

$$L_t=\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}(\Delta t_i)^k\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}\left|\int_{t_i}^{u}(X_v^{MG}-X_{t_i}^{MG})dX_v^{MG}\right|du,$$

We have,

$$\begin{aligned}
E|L_t| &= \frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}(\Delta t_i)^k\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}E\left|\int_{t_i}^{u}(X_v^{MG}-X_{t_i}^{MG})dX_v^{MG}\right|du\\
&\le \frac{c}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}(\Delta t_i)^k\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}E\biggl(\int_{t_i}^{u}(X_v^{MG}-X_{t_i}^{MG})^2d<X^{MG}>_v\biggr)^{1/2}du\\
&\le \frac{c}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}(\Delta t_i)^k\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}\left(E\int_{t_i}^{u}(X_v^{MG}-X_{t_i}^{MG})^2dv\right)^{1/2}\\
&\qquad\cdot\left(E\sup_{u\in(0,t]}<X>_u'\right)^{1/2}du\\
&= \frac{c}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}(\Delta t_i)^k\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}\left(\int_{t_i}^{u}E(<X>_v-<X>_{t_i})dv\right)^{1/2}
\end{aligned}$$

$$\cdot\left(E\sup_{u\in(0,t]} < X >'_u\right)^{1/2} du$$

$$\leq \quad \frac{c^*}{h^2} E \sup_{u\in(0,t]} < X >'_u \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} (\Delta t_i)^{k+2}$$

where the first two inequalities follow from Burkholder's inequality and Hölder's Inequality respectively, and the subsequent equality follows from Fubini's Theorem and the result $E(X_v^{MG} - X_{t_i}^{MG})^2 = E(< X >_v - < X >_{t_i})$. Thus $L_t = O_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h})$ by Markov's inequality, under assumptions A and $C.1(X, X)$.

Let

$$N_t = \frac{1}{h^2} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} Y_{t_i} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \int_{t_i}^{u} (X_v^{MG} - X_{t_i}^{MG}) dX_v^{MG} (u - t_i)^k du$$

Applying integration by parts, we get

$$N_t = \frac{1}{h^2} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} Y_{t_i} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_v^{MG} - X_{t_i}^{MG}) dX_v^{MG} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (u - t_i)^k du$$

$$- \frac{1}{h^2} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} Y_{t_i} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} [\int_{t_i}^{u} (v - t_i)^k dv](X_u^{MG} - X_{t_i}^{MG}) dX_u^{MG}$$

$$= \quad \frac{1}{k+1}\frac{1}{h^2} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} Y_{t_i} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} [(\Delta t_i)^{k+1} - (u - t_i)^{k+1}](X_u^{MG} - X_{t_i}^{MG}) dX_u^{MG}$$

therefore,

$$< N >_t = \frac{1}{(k+1)^2}\frac{1}{h^4} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} Y_{t_i}^2 \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} [(\Delta t_i)^{k+1} - (u - t_i)^{k+1}]^2$$

$$\cdot(X_u^{MG} - X_{t_i}^{MG})^2 d < X >_u$$

$$\leq \quad \frac{1}{(k+1)^2} \sup_{u\in(0,t]} Y_u^2 \sup_{u\in(0,t]} < X, X >'_u \cdot \frac{1}{h^4} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} [(\Delta t_i)^{k+1}$$

$$- (u - t_i)^{k+1}]^2 (X_u^{MG} - X_{t_i}^{MG})^2 du \qquad (6.2.5)$$

Using similar approach as in L_t, we have

$$\begin{aligned}
& E\frac{1}{h^4}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}\left[(\Delta t_i)^{k+1}-(u-t_i)^{k+1}\right]^2(X_u^{MG}-X_{t_i}^{MG})^2du \\
= \ & \frac{1}{h^4}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}\left[(\Delta t_i)^{k+1}-(u-t_i)^{k+1}\right]^2 E(X_u^{MG}-X_{t_i}^{MG})^2du \\
\le \ & E\sup_{u\in(0,t]} < X,X >'_u \frac{a}{h^4}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}(\Delta t_i)^{2k+4} \\
= \ & o(\frac{(\overline{\Delta t}^{(n)})^{2k+2}}{h^2})
\end{aligned}$$

under assumption A and $C.1(X,X)$, where a is some constant. Thus Equation (6.2.5) has order $o_p(\frac{(\overline{\Delta t}^{(n)})^{2k+2}}{h^2})$ by Markov's inequality, under assumptions A, $B.1(X,X)$, $C.1(X,X)$ and continuity of Y. And so $N_t = o_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h})$.

Hence,

$$|II_3| \le 2\Upsilon^Y(h)|L_t| + 2|N_t| = o_p(\frac{(\overline{\Delta t}^{(n)})^{k+1}}{h}) \tag{6.2.6}$$

Therefore (i) follows from Equations (6.2.3), (6.2.4), and (6.2.6).

(ii) Using Itô's Lemma,

$$\begin{aligned}
& \frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_u-X_{t_i^{(n)}})(Z_u-Z_{t_i^{(n)}})(u-t_i)^kY_udu \\
= \ & \frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(< X,Z >_u - < X,Z >_{t_i})(u-t_i)^kY_udu \\
& +\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[\int_{t_i}^{u}(X_v-X_{t_i})dZ_v](u-t_i)^kY_udu \\
& +\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[\int_{t_i}^{u}(Z_v-Z_{t_i})dX_v](u-t_i)^kY_udu
\end{aligned}$$

then the results can be derived by using same argument as in part (i), under assumptions A, $B.1(XX)(ZZ)(XZ)$, $C.1(XX)(ZZ)$, and $B.3(X)(Z)$. ∎

Lemma 2 Suppose $\{X_t\}$, $\{Y_t\}$ and $\{Z_t\}$ are Itô processes. Also suppose $Z_t \in C^1[0,T]$. Let each Itô process be represented as the sum of its martingale part and drift part (i.e. $X_t = X_t^{DR} + X_t^{MG}$, $Y_t = Y_t^{DR} + Y_t^{MG}$). Subject to assumptions A, $B.1[(X,X),(Y,Y)]$, $B.3[(X)(Y)]$ and $C.1(X,X)$, the following holds, for any nonnegative integer m:

(i)

$$\frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}})(Z_u - Z_{t_i^{(n)}})^m dY_u$$
$$= \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}})(Z_u - Z_{t_i^{(n)}})^m dY_u^{MG} + o_p\left(\frac{\overline{(\Delta t^{(n)})}^{m+1/2}}{h^{3/2}}\right)$$

where

$$\frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}})(Z_u - Z_{t_i^{(n)}}) dY_u^{MG} = O_p\left(\frac{\overline{(\Delta t^{(n)})}^{m+1/2}}{h^{3/2}}\right)$$

(ii)

$$\frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta Z_{t_i^{(n)}})^m \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}}) dY_u$$
$$= \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta Z_{t_i^{(n)}})^m \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}}) dY_u^{MG} + o_p\left(\frac{\overline{(\Delta t^{(n)})}^{m+1/2}}{h^{3/2}}\right)$$

where

$$\frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \Delta Z_{t_i^{(n)}} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}}) dY_u^{MG} = O_p\left(\frac{\overline{(\Delta t^{(n)})}^{m+1/2}}{h^{3/2}}\right)$$

$\square$

Proof of Lemma 2:

(i) treat the martingale part and the drift part separately.

$$\frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}})(Z_u - Z_{t_i^{(n)}})^m dY_u$$

$$
\begin{aligned}
&= \frac{1}{h^2} \sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}})(Z_u - Z_{t_i^{(n)}})^m dY_u^{MG} \quad \leftarrow \mathbf{I} \\
&+ \frac{1}{h^2} \sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}})(Z_u - Z_{t_i^{(n)}})^m dY_u^{DR} \quad \leftarrow \mathbf{II}
\end{aligned}
$$

Write $dZ_t = \tilde{Z}_t dt$, first we can obtain $I = O_p(\frac{(\overline{\Delta t}^{(n)})^{m+1/2}}{h^{3/2}})$ because of the following,

$$
\begin{aligned}
<I> &= \frac{1}{h^4} \sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}})^2 (Z_u - Z_{t_i^{(n)}})^{2m} d <Y^{MG}>_u \\
&\le \sup_{u\in[0,t]} |<Y>_u'| \sup_{u\in[0,t]} \{(\tilde{Z}_u)^{2m}\} \\
&\quad \cdot \frac{1}{h^4} \sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u - X_{t_i^{(n)}})^2 (u - t_i^{(n)})^{2m} du \\
&= O_p(\frac{(\overline{\Delta t}^{(n)})^{2m+1}}{h^3})
\end{aligned}
$$

by $Z_u \in C^1[0,t]$, assumption $B.1(Y,Y)$, and by Lemma 1(i) following assumptions A, $B.1(X,X)$, $C.1(X,X)$, and $B.3(X)$.

Next we consider the order of the drift part, II. Recall the notation $dY_u^{DR} = \tilde{Y}_u du$ and $dX_u^{DR} = \tilde{X}_u du$. Applying Minkovski's inequality, we get

$$
\begin{aligned}
|II| &\le |\frac{1}{h^2} \sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u^{DR} - X_{t_i^{(n)}}^{DR})(Z_u - Z_{t_i^{(n)}})^m dY_u^{DR}| \\
&\quad + |\frac{1}{h^2} \sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (X_u^{MG} - X_{t_i^{(n)}}^{MG})(Z_u - Z_{t_i^{(n)}})^m dY_u^{DR}| \\
&\le \frac{1}{m+2} \sup_{u\in[0,t]} |\tilde{Y}_u| \sup_{u\in[0,t]} |\tilde{Z}_u|^m \sup_{u\in[0,t]} |\tilde{X}_u| \frac{1}{h^2} \sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t} (\Delta t_i)^{m+2} \\
&\quad + \sup_{u\in[0,t]} |\tilde{Y}_u| \sup_{u\in[0,t]} |\tilde{Z}_u|^m \frac{1}{h^2} \sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t} (\Delta t_i)^m \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} |X_u^{MG} - X_{t_i^{(n)}}^{MG}| du \qquad (6.2.7)
\end{aligned}
$$

Now let

$$G_t = \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i)^m \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} |X_u^{MG} - X_{t_i^{(n)}}^{MG}| du,$$

by Fubini's Theorem,

$$\begin{aligned}
E|G_t| &= \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \frac{(\Delta t_i)^m}{h^2} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} E|X_u^{MG} - X_{t_i^{(n)}}^{MG}| du \\
&\le \frac{c}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i)^m \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} E(< X^{MG} >_u - < X^{MG} >_{t_i})^{1/2} du \qquad (6.2.8) \\
&\le E\sqrt{\sup_{u \in [0,t]} < X^{MG} >'_u} \frac{c'}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i)^{m+3/2} \\
&\le \sqrt{E \sup_{u \in [0,t]} < X^{MG} >'_u} \frac{c'}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i)^{m+3/2} \qquad (6.2.9) \\
&= O(\frac{(\overline{\Delta t^{(n)}})^{m+1/2}}{h})
\end{aligned}$$

under assumptions A and $C.1(X,X)$. Equation (6.2.8) follows from Burkholder's inequality with some constant c, Equation (6.2.9) follows from Jensen's inequality. Then $G_t = o_p(\frac{(\overline{\Delta t}^{(n)})^{m+1/2}}{h^{3/2}})$ based on Markov's inequality.

Therefore, Equation (6.2.7) is of order $o_p(\frac{(\overline{\Delta t}^{(n)})^{m+1/2}}{h^{3/2}})$ under the continuously differentiability condition of Z, and the assumptions A, $C.1(X,X)$, and $B.3[(X)(Y)]$. Hence the result follows, given A, $B.1[(X,X)(Y,Y)]$, $C.1(X,X)$, and $B.3[(X)(Y)]$.

(ii) Similar to (i). ∎

Lemma 3 Suppose X, Y, and Z are Itô processes. Then under assumptions A and $B.1[(X,X)$, $(X,Z)]$,

$$(i) \quad \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} [< X >_u - < X >_{t_i}](u - t_i)^k Y_u du$$

$$\sim \frac{1}{k+2} \frac{\overline{\Delta t}^{(k+1)}}{h} H^{(k+2)'}(t) < X >'_t Y_t$$

$$(ii) \quad \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} [< X, Z >_u - < X, Z >_{t_i}](u - t_i)^k Y_u du$$

$$\sim \frac{1}{k+2} \frac{\overline{\Delta t}^{(k+1)}}{h} H^{(k+2)'}(t) < X, Z >'_t Y_t$$

□

Proof of Lemma 3:

(i) Let

$$\begin{aligned} H_1 &\triangleq \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \Big[(< X >_u - < X >_{t_i})(u - t_i)^k \\ &\quad - < X >'_u (u - t_i)^{k+1} \Big] Y_u du \\ H_2 &\triangleq \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \left[\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} < X >'_u (u - t_i)^{k+1} Y_u du - < X >'_{t_i} Y_{t_i} \frac{(\Delta t_i)^{k+2}}{k+2} \right] \\ H_3 &\triangleq \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \left(< X >'_{t_i} Y_{t_i} - < X >'_t Y_t \right) \frac{(\Delta t_i)^{k+2}}{k+2} \end{aligned}$$

Now we show that $H_1 = o_p(\frac{\overline{\Delta t}^{(k+1)}}{h}), H_2 = o_p(\frac{\overline{\Delta t}^{(k+1)}}{h}), H_3 = o_p(\frac{\overline{\Delta t}^{(k+1)}}{h})$.

For $\xi \in (t_i, t_{i+1})$

$$\begin{aligned} H_1 &= \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} (< X >'_\xi - < X >'_u)(u - t_i)^{k+1} Y_u du \\ &\le \frac{1}{k+2} \frac{1}{h^2} \Upsilon^{XX}(h) \sup_{0 \le u \le t} |Y_u| \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i)^{k+2} \\ &= o_p(\frac{\overline{\Delta t}^{(k+1)}}{h}) \end{aligned}$$

under assumptions A and $B.1(X, X)$ and the continuity of Y. Recall that

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

Again, Assumption $B.1(X, Y)$ implies $\Upsilon^{XY}(h) \to 0$.

$$H_2 = \frac{1}{h^2} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} \underbrace{(< X >'_u Y_u - < X >'_{t_i} Y_{t_i})}_{V_u - V_{t_i}} (u - t_i)^{k+1} du$$

$$\begin{aligned}
&\leq \frac{1}{k+2}\Upsilon^V(h)\frac{1}{h^2}\sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t}(\Delta t_i)^{k+2} \\
&= o_p(\frac{\overline{\Delta t}^{(k+1)}}{h})
\end{aligned}$$

under Assumption A and $B.1(X,X)$. Notice that $\Upsilon^V(h)=o_p(1)$, because that Y_t is continuous, also $< X >'_t$ is continuous by assumption $B.1(X,X)$, thus $V_t =< X >'_t Y_t$ is continuous.

$$\begin{aligned}
H_3 &= \frac{1}{h^2}\sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t}\underbrace{\left(< X >'_{t_i} Y_{t_i} - < X >'_t Y_t\right)}_{V_{t_i}-V_t}\frac{(\Delta t_i)^{k+2}}{k+2} \\
&\leq \frac{1}{k+2}\Upsilon^V(h)\frac{1}{h^2}\sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t}(\Delta t_i)^{k+2} \\
&\overset{\text{assumption A}}{=} o_p(\frac{\overline{\Delta t}^{(k+1)}}{h})
\end{aligned}$$

by assumption A and $B.1(X,X)$. Therefore,

$$\begin{aligned}
&\frac{1}{h^2}\sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[< X >_u - < X >_{t_i}](u-t_i)^k Y_u du \\
&= \frac{1}{h^2}\sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t}< X >'_t Y_t\frac{(\Delta t_i)^{k+2}}{k+2} + H_1 + H_2 + H_3 \\
&\overset{\text{assumption A}}{\sim} \frac{1}{k+2}\frac{\overline{\Delta t}^{(k+1)}}{h}H^{(k+2)'}(t) < X >'_t Y_t
\end{aligned}$$

(ii) follow from similar argument as part (i), with extra assumption $B.1(X,Z)$. ∎

Corollary 2 Suppose X, Y, Z, V are Itô processes, Let

$$H^{(2)}_{n,<X,Y>,<Z,V>}(t) = \frac{1}{\overline{\Delta t}^{(n)}}\sum_{t_{i+1}^{(n)}\leq t}\Delta < X,Y >_{t_i^{(n)}}\Delta < Z,V >_{t_i^{(n)}}$$

Then under assumptions A and B.1[(X,Y), (Z,V)],

(i) $H^{(2)}_{n,<X,Y>,<Z,V>}(t) - H^{(2)}_{n,<X,Y>,<Z,V>}(t-h)$
$= \frac{1}{\overline{\Delta t}^{(n)}} < X,Y >'_t< Z,V >'_t \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t}(\Delta t_i^{(n)})^2 + o_p(h)$

(ii) $H^{(2)'}_{<X,Y>,<Z,V>}(t)$ exists, and $H^{(2)'}_{<X,Y>,<Z,V>}(t) = H^{(2)'}(t) < X,Y >'_t< Z,V >'_t$

□

Proof of Corollary 2:

(i)

$$
\begin{aligned}
& H^{(2)}_{n,<X,Y>,<Z,V>}(t) - H^{(2)}_{n,<X,Y>,<Z,V>}(t-h) \\
= \; & \frac{1}{\overline{\Delta t}^{(n)}} < X,Y >'_t < Z,V >'_t \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i)^2 \\
& + \frac{1}{\overline{\Delta t}^{(n)}} \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} \Delta < X,Y >_{t_i} \left[\Delta < Z,V >_{t_i} - < Z,V >'_t (\Delta t_i) \right] \\
& + \frac{1}{\overline{\Delta t}^{(n)}} < Z,V >'_t \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i) \left[\Delta < X,Y >_{t_i} - < X,Y >'_t (\Delta t_i) \right] \\
\le \; & \frac{1}{\overline{\Delta t}^{(n)}} < X,Y >'_t < Z,V >'_t \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i)^2 \\
& + \frac{1}{\overline{\Delta t}^{(n)}} \sup_{u \in (0,t]} < X,Y >'_u \Upsilon^{ZV}(h) \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i)^2 \\
& + \frac{1}{\overline{\Delta t}^{(n)}} < Z,V >'_t \Upsilon^{XY}(h) \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i)^2 \\
= \; & \frac{1}{\overline{\Delta t}^{(n)}} < X,Y >'_t < Z,V >'_t \sum_{t-h \le t_i^{(n)} < t_{i+1}^{(n)} \le t} (\Delta t_i)^2 + o_p(h)
\end{aligned}
$$

under assumptions A and $B.1[(X,Y),(Z,V)]$.

(ii) follows from assumption A directly. ∎

6.3 *Proof of theorems and corollary*

Proof of Theorem 1:

(a)

$$
< \widehat{X,Y} >'_t - < X,Y >'_t
$$

$$
\begin{aligned}
&= \frac{1}{h}\left(\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t} \Delta X_{t_i^{(n)}}\cdot \Delta Y_{t_i^{(n)}}\right) - <X,Y>'_t \\
&= \frac{1}{h}(<X,Y>_t - <X,Y>_{t-h} \\
&\quad + [2]\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_s - X_{t_i^{(n)}})dY_s) - <X,Y>'_t \\
&= \underbrace{\frac{1}{h}(<X,Y>_t - <X,Y>_{t-h}) - <X,Y>'_t}_{B_{1,t}^{XY}} \\
&\quad + \underbrace{\frac{[2]}{h}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_s - X_{t_i^{(n)}})dY_s}_{B_{2,t}^{XY}}
\end{aligned}
$$

where the second equality follows from Itô's Lemma. We begin by considering the order of the $B_{2,t}^{XY}$. By Lemma 2 (ii) under assumptions A, $B.1[(XX),(YY)]$, $C.1(XX)$ and $B.3[(X),(Y)]$, $B_{2,t}^{XY} = O_p(\sqrt{\frac{\overline{\Delta t}^{(n)}}{h}})$. We next consider the order of $B_{1,t}^{XY}$ in the following.

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, then,

$$
\begin{aligned}
B_{1,t}^{XY} &= \frac{1}{h}\int_{t-h}^{t} <X,Y>'_u du - <X,Y>'_t \\
&= \frac{1}{h}\int_{t-h}^{t}(<X,Y>'_u - <X,Y>'_t)du \\
&= \frac{1}{h}\int_{t-h}^{t}((t-h)-u)d<X,Y>'_u \qquad \text{(integration by parts)} \\
&= \underbrace{\frac{1}{h}\int_{t-h}^{t}((t-h)-u)dR_u^{XY}}_{B_{1,t}^{XY,MG}} + \underbrace{\frac{1}{h}\int_{t-h}^{t}((t-h)-u)dD_u^{XY}}_{B_{1,t}^{XY,DR}}
\end{aligned}
$$

As shown, we refer to the first term as $B_{1,t}^{XY,MG}$-the martingale part of $B_{1,t}^{XY}$, and the second term as $B_{1,t}^{XY,DR}$-the drift part of $B_{1,t}^{XY}$. Note that, naturally, $B_{1,t}^{XY,DR} = O_p(h)$ under assumption $B.2(X,Y)$.

$$
\begin{aligned}
<B_1^{XY,MG}, B_1^{ZV,MG}>_t &= \frac{1}{h^2}\int_{t-h}^{t}(t-h-u)^2 d<R^{XY},R^{ZV}>_u \\
&= \frac{1}{3}h <R^{XY},R^{ZV}>'_t + o_p(h) \qquad (6.1)
\end{aligned}
$$

Note that $o_p(h)$ is from the following

$$\begin{aligned}&\frac{1}{h^2}\int_{t-h}^{t}(t-h-u)^2(<R^{XY},R^{ZV}>'_t - <R^{XY},R^{ZV}>'_u)du\\ &\qquad\le \frac{h}{3}\Upsilon^{R^{XY},R^{ZV}}(h) = o_p(h)\end{aligned}$$

by assumption $B.1(R^{XY},R^{ZV})$. Hence $B_1^{XY,MG} = O_p(\sqrt{h})$ by $B.1(R^{XY},R^{XY})$. Since $B_{1,t}^{XY,DR} = O_p(h)$, it follows that $B_{1,t}^{XY} = O_p(\sqrt{h})$

(b) Equate $O_p(\sqrt{h}) = O_p(\sqrt{\frac{\overline{\Delta t}^{(n)}}{h}})$, it follows that $O_p(h) = O_p(\sqrt{\overline{\Delta t}^{(n)}})$.

(c) The asymptotic distribution of $B_{1,t}^{XY}$ follows from (6.1) in (a) by Theorems A.2 or A.3 in Appendix 6.1, depending on assumption E. Now we consider the order of $B_{2,t}^{XY}$.

$$\begin{aligned}B_{2,t}^{XY} &= \underbrace{\frac{[2]}{h}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_s - X_{t_i^{(n)}})dY_s^{MG}}_{B_{2,t}^{XY,MG}}\\ &\quad + \underbrace{\frac{[2]}{h}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_s - X_{t_i^{(n)}})dY_s^{DR}}_{B_{2,t}^{XY,DR}}\end{aligned}$$

and then

$$\begin{aligned}&<B_2^{XY,MG}, B_2^{ZV,MG}>_t\\ &= \frac{[2]}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_s - X_{t_i^{(n)}})(Z_s - Z_{t_i^{(n)}})d<Y^{MG},V^{MG}>_s\\ &\quad + \frac{[2]}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_s - X_{t_i^{(n)}})(V_s - V_{t_i^{(n)}})d<Y^{MG},Z^{MG}>_s\\ &\sim \frac{\overline{\Delta t}^{(n)}}{h}[H^{(2)'}_{<X,Z>,<Y,V>}(t) + H^{(2)'}_{<X,V>,<Y,Z>}(t)] + o_p(\frac{\overline{\Delta t}^{(n)}}{h})\end{aligned}$$

by Lemma 1, Lemma 3 and Corollary 2.

In particular, $<B_2^{XY}, B_2^{XY}>_t = \frac{\overline{\Delta t}^{(n)}}{h}[H^{(2)'}_{<X,X>,<Y,Y>}(t) + H^{(2)'}_{<X,Y>,<X,Y>}(t)]$ in the limit. Hence the asymptotic distribution of B_2^{XY} follows from Theorems A.1 - A.3 in Appendix 6.1.

(d) We here will show $< B_1^{XY}, B_2^{XY} >_t = O_p(\frac{\overline{\Delta t}^{(n)}}{\sqrt{h}})$

$$
\begin{aligned}
& < B_1^{ZV,MG}, B_2^{XY,MG} >_t \\
= \;& < \frac{1}{h}\int_{t-h}^{t}((t-h)-s)dR_s^{ZV}, \frac{[2]}{h}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_s^{MG}-X_{t_i^{(n)}}^{MG})dY_s^{MG} > \\
= \;& \frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_s^{MG}-X_{t_i^{(n)}}^{MG})((t-h)-s)d< R^{ZV}, Y^{MG} >_s \\
& + \frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(Y_s^{MG}-Y_{t_i^{(n)}}^{MG})((t-h)-s)d< R^{ZV}, X^{MG} >_s
\end{aligned}
$$

Now suffice to consider one of the above two terms, we will examine the first one. Let $dG_s = [s-(t-h)]d < R^{ZV}, Y^{MG} >_s$, integration by parts yields,

$$
\begin{aligned}
& \frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_s^{MG}-X_{t_i^{(n)}}^{MG})((t-h)-s)d< R^{ZV}, Y^{MG} >_s \\
= \;& -\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(X_s^{MG}-X_{t_i^{(n)}}^{MG})dG_u \\
= \;& -\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}(\Delta X_{t_i}^{MG})(\Delta G_{t_i}) + \frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}} G_s dX_s^{MG} \\
= \;& \underbrace{-\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}(\Delta X_{t_i}^{MG})[\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(u-(t-h))d< R^{ZV}, Y^{MG} >_u]}_{I} \\
& \underbrace{+\frac{1}{h^2}\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}[\int_{t_i}^{s}(u-(t-h))d< R^{ZV}, Y^{MG} >_u]dX_s^{MG}}_{II} \\
= \;& O_p(\frac{\overline{\Delta t^{(n)}}}{\sqrt{h}})
\end{aligned}
$$

because

$$
I \;\le\; \frac{1}{h^2}\sqrt{\sum_{t-h\le t_i^{(n)}<t_{i+1}^{(n)}\le t}(\Delta X_{t_i}^{MG})^2 \cdot \sum_{t_{i+1}^{(n)}\le t}[\int_{t_i^{(n)}}^{t_{i+1}^{(n)}}(u-(t-h))d< R^{ZV}, Y^{MG} >_u]^2}
$$

$$
\begin{aligned}
&\leq \sup_{0\leq u\leq t} < R^{ZV}, Y^{MG} >_u{}' \frac{1}{h^2}\sqrt{[X^{MG}]_t - [X^{MG}]_{t-h}}\sqrt{\sum_{t_{i+1}^{(n)}\leq t} h^2(\Delta t_i)^2} \\
&= O_p(\frac{\sqrt{\Delta t}}{\sqrt{h}})
\end{aligned}
$$

$$
\begin{aligned}
< II > = \frac{1}{h^4} & \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} [\int_{t_i}^{s}(u-(t-h))d < R^{ZV}, Y^{MG} >_u]^2 d < X^{MG} >_s \\
\leq \quad & (\sup_{0\leq u\leq t} < R^{ZV}, Y^{MG} >_u')^2 \sup_{0\leq u\leq t} < X^{MG} >_u' \\
& \cdot\frac{1}{h^4} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} \int_{t_i^{(n)}}^{t_{i+1}^{(n)}} [\int_{t_i}^{s}(u-(t-h))du]^2 ds \\
= \quad & (\sup_{0\leq u\leq t} < R^{ZV}, Y^{MG} >_u')^2 \sup_{0\leq u\leq t} < X^{MG} >_u' \\
& \cdot\frac{1}{h^4} \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} \{\frac{1}{20}(\Delta t_i)^5 + \frac{1}{4}(\Delta t_i)^4[t_i-(t-h)] + \frac{1}{3}(\Delta t_i)^3[t_i-(t-h)]^2\} \\
\leq \quad & (\sup_{0\leq u\leq t} < R^{ZV}, Y^{MG} >_u')^2 \sup_{0\leq u\leq t} < X^{MG} >_u' \\
& \cdot \sum_{t-h\leq t_i^{(n)}<t_{i+1}^{(n)}\leq t} \{\frac{(\Delta t_i)^5}{20h^4} + \frac{(\Delta t_i)^4}{4h^3} + \frac{(\Delta t_i)^3}{3h^2}\} \\
\overset{\text{Assumption A}}{\sim} \quad & (\sup_{0\leq u\leq t} < R^{ZV}, Y^{MG} >_u')^2 \sup_{0\leq u\leq t} < X^{MG} >_u' \\
& \cdot\{\frac{(\overline{\Delta t^{(n)}})^4}{20h^3} H^{(5)'}(t) + \frac{(\overline{\Delta t^{(n)}})^3}{4h^2} H^{(4)'}(t) + \frac{(\overline{\Delta t^{(n)}})^2}{3h^2} H^{(3)'}(t)\} \\
= \quad & O_p(\frac{(\overline{\Delta t^{(n)}})^2}{h})
\end{aligned}
$$

by assumption $B.1[R^{ZV}, Y), (X, X)]$, and the order selection of $h^2 = O(\overline{\Delta t^{(n)}})$.

The independence for $t \neq t'$ follows by the same methods as in Theorem A.1 and A.3. ∎

Proof of Corollary 1:

The result follows directly from Theorem 1. ∎

Proof of Theorem 2:

By Taylor expansion on $\frac{1}{\widehat{<S,S>}}$ and result in Theorem 1 (a),

$$\begin{aligned}
\hat{\rho}_t - \rho_t &= \frac{\widehat{<\Xi,S>}'_t}{\widehat{<S,S>}'_t} - \frac{<\Xi,S>'_t}{<S,S>'_t} \\
&= \frac{1}{<S,S>'_t}[\widehat{<\Xi,S>}'_t - <\Xi,S>'_t] - \frac{\rho_t}{<S,S>'_t}[\widehat{<S,S>}'_t - <S,S>'_t] + o_p(\sqrt{h}) \\
&= \frac{1}{<S,S>'_t}[B_1^{\Xi S} - \rho_t B_1^{SS}] + \frac{1}{<S,S>'_t}[B_2^{\Xi S} - \rho_t B_2^{SS}] + o_p(\sqrt{h}) \qquad (6.2)
\end{aligned}$$

From Theorem 1, we also know that asymptotically,

$$h^{-1/2}\begin{bmatrix} B_{1,t}^{\Xi S} \\ B_{1,t}^{SS} \\ B_{2,t}^{\Xi S} \\ B_{2,t}^{SS} \end{bmatrix} \xrightarrow{L} N(0, M_3)$$

where

$$M_3 = \begin{bmatrix} \frac{1}{3}\begin{bmatrix} <R^{\Xi S}>'_t & <R^{\Xi S}, R^{SS}>'_t \\ <R^{\Xi S}, R^{SS}>'_t & <R^{SS}>'_t \end{bmatrix} & 0 \\ 0 & cH^{(2)'}(t)\begin{bmatrix} <\Xi>'_t<S>'_t + (<\Xi,S>'_t)^2 & 2<\Xi,S>'_t<S>'_t \\ 2<\Xi,S>'_t<S>'_t & 2(<S>'_t)^2 \end{bmatrix} \end{bmatrix}$$

Straightforward calculation following (6.2) and M_3 gives,

$$\begin{aligned}
V_{\hat{\rho}_t-\rho_t} &= \frac{1}{3(<S>'_t)^2}[<R^{\Xi S}>'_t + \rho_t^2 <R^{SS}>'_t - 2\rho_t <R^{\Xi S}, R^{SS}>'_t] \\
&\quad + \frac{H^{(2)'}(t)}{(<S>'_t)^2}\frac{\overline{\Delta t}^{(n)}}{h^2}[<\Xi>'_t<S>'_t + (<\Xi,S>'_t)^2 + 2\rho^2(<S>'_t)^2 \\
&\qquad - 4\rho_t <\Xi,S>'_t<S>'_t] \\
&= \frac{1}{3}<\rho>'_t + (\frac{1}{<S>'_t})^2 H^{(2)'}(t)\frac{\overline{\Delta t}^{(n)}}{h^2}[<\Xi>'_t<S>'_t - (<\Xi,S>'_t)^2] \\
&= \frac{1}{3}<\rho>'_t + cH^{(2)'}(t)[\frac{<\Xi>'_t}{<S>'_t} - \rho_t^2]
\end{aligned}$$

Notice that we use $<X>$ to represent $<X,X>$ for simplicity, where X can be any process. ∎

References

Alizadeh, S., Brandt, M., and Diebold, F. (2000). Range-based estimation of stochastic volatility models. NBER working paper.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of The American Statistical Association*, 96 (453):42–55.

Bai, X., Russell, J. R., and Tiao, G. C. (2000a). Beyond merton's utopia (i): effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data. working paper.

Bai, X., Russell, J. R., and Tiao, G. C. (2000b). Beyond merton's utopia (ii). working paper.

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

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Chu, S.-H. and Freund, S. (1996). Volatility estimation for stock index options: A garch approach. *Quarterly Review of Economics and Finance*, 36(4):431–450.

deRoon, F. and Veld, C. (1996). An empirical investigation of the factors that determine the pricing of dutch index warrants. *European Financial Management*, 2(1):97–112.

Engle, R. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of u.k. inflation. *Econometrica*, 50:987–1008.

Fama, E. F. and MacBeth, J. D. (1973). Risk, return, and equilibium: Empirical tests. *Journal of Political Economy*, 81:607–636.

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

French, K., Schwert, W., and Stambaugh, R. (1987). Expected stock return volatility. *Journal of Financial Economics*, 19:3–29.

Gikhman, I. I. and Skorokhod, A. V. (1979). *The Theory of Stochastic Processes III.* Berlin; New York: Springer-Verlag.

Hardle, W. and Hafner, C. M. (2000). Discrete time option pricing with flexible volatility estimation. *Finance and Stochastics*, 4(2):189–207.

Hull, J. and White, A. (1987). The pricing of options on assets with stochastic volatilities. *Journal of Finance*, 42:281–300.

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

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

Karolyi, G. A. (1993). A bayesian approach to modeling stock return volatility for option valuation. *Journal of Financial and Quantitative Analysis*, 28(4):579–594.

Latané, H. and Rendleman, R. J. (1976). Standard deviations of stock price ratios implied in option prices. *Journal of Finance*, 31:369–381.

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

Merton, R. (1980). On estimating the expected return on the market. *Journal of Financial Economics*, 8:323–361.

Mykland, P. 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. (1998). Options pricing and statistical uncertainty. Technical report no. 430, Department of Statistics, University of Chicago.

Mykland, P. A. (1999). Financial options and statistical prediction intervals. In revision for *Annals of Statistics*.

Mykland, P. A. (2000). Conservative delta hedging. *Annals of Applied Probability*, 10:664–683.

Mykland, P. A. (2001). The interpolation of options. Technical Report no. 478, Department of Statistics, University of Chicago. Accepted by *Finance and Stochastics* subject to revision.

Officer, R. R. (1973). The variability of the market factor of the new york stock exchange. *Journal of Business*, 46:434–453.

Polson, N., Jacquier, E., and Rossi, P. (1994). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics*, 12:371–418.

Poterba, J. and Summers, L. (1986). The persistence of volatility and stock market fluctuations. *The American Economic Review*, 76:1142–1151.

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

Wiggins, J. B. (1987). Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics*, 19:351–372.

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