The Estimation of Leverage effect with High Frequency Data

D. Christina Wang
The University of Chicago

Per A. Mykland
The University of Chicago

First version: July 22, 2009
This version: December 27, 2012

Abstract

The leverage effect has become an extensively studied phenomenon which describes the (usually) negative relation between stock returns and their volatility. Although this characteristic of stock returns is well acknowledged, most studies of the phenomenon are based on cross-sectional calibration with parametric models. On the statistical side, most previous works are conducted over daily or longer return horizons, and few of them have carefully studied its estimation, especially with high frequency data. However, estimation of the leverage effect is important because sensible inference is possible only when the leverage effect is estimated reliably. In this paper, we provide nonparametric estimation for a class of stochastic measures of leverage effect. In order to construct estimators with good statistical properties, we introduce a new stochastic leverage effect parameter. The estimators and their statistical properties are provided in cases both with and without microstructure noise, under the stochastic volatility model. In asymptotics, the consistency and limiting distribution of the estimators are derived and corroborated by simulation results. For consistency, a previously unknown bias correction factor is added to the estimators. Applications of the estimators are also explored. This estimator provides the opportunity to study high frequency regression, which leads to the prediction of volatility using not only previous volatility but also the leverage effect. The estimator also reveals a theoretical connection between skewness and the leverage effect, which further leads to the prediction of skewness. Furthermore, adopting the ideas similar to these, it is easy to extend these methods to other important aspects of stock returns, such as volatility of volatility.

Keywords: consistency, discrete observation, efficiency, Itô process, leverage effect, realized volatility, stable convergence, skewness, microstructure noise.

1 Introduction

The leverage effect has become an extensively studied empirical phenomenon in the form of the (usually negative) correlation between (current) returns and (current and future) volatility [Engle and Ng (1993), Zakoian (1994), and Wu and Xiao (2002), etc.]. It is one of the many stylized facts of the security return distribution, along with the well known fat tails, skewness, excess kurtosis, and heteroscedasticity. The discovery of leverage effect closely relates to the study of stochastic volatility. Although for very low frequency data, such as monthly or yearly asset returns, the assumption of homogeneity seems not to be entirely unreasonable [Mandelbrot (1963), Fama (1965) and Officer (1973)], the increasing frequency of observed data in studies suggests heterogeneity in volatility, in other words, time-varying volatility as pointed out by [Engle (1982, 2000), Bollerslev (1986), Andersen and Bollerslev (1998), Andersen, Bollerslev, Diebold, and Ebens (2001)]. This finding has had profound implications in both the theory and practice of financial economics and econometrics. It has inspired new model building, such as the emergence of ARCH models and the later Stochastic Volatility models. Modeling volatility as a separate process allows the study of its relation with the associated return process, which leads to the discovery of asymmetric volatility. Time varying volatility is also of substantial importance in modeling for options pricing, as in Hull and White (1987), Stein and Stein (1991), Heston (1993), and Ball and Roma (1994).

Black (1976) and Christie (1982) were among the first to document the volatility asymmetry, and gave an explanation based on the "Leverage Effect" hypothesis: A drop in the value of the stock (negative return) increases the financial leverage (debt-to-equity ratio), which makes the stock riskier and increases its volatility. Since then, "Leverage Effect" has been taken to be synonymous with asymmetric volatility. Financial leverage itself, however, seems not enough to explain either the large magnitude of the effect of declines in current price on future volatility [Figlewski and Wang (2001)], or the phenomenon that the asymmetry of market index returns is generally larger than that for individual stocks [Kim and Kon (1994), Tauchen and Zhang (1996), and Andersen, Bollerslev, Diebold, and Ebens (2001)]. In another point of view, the asymmetric nature of volatility to return shocks simply reflects time-varying risk premium [Pindyck (1984), French, Schwert, and Stambaugh (1987), and Campbell and Hentschel (1991)]. This explanation is often referred to as the "volatility feedback effect" : If volatility is priced, an anticipated increase in volatility raises the required return on equity, leading to a immediate stock price decline.

Many later works either compare the two effects or seek to argue that they can both be at work [Nelson (1991), Engle and Ng (1993), Glosten, Jagannathan, and Runkle (1993), Bekaert and Wu (1997) and Wu (2001), Hasanhodzic and Lo (2011)]. While there is little agreement concerning the fundamental causes behind the leverage effect, that is not the focus of this paper.

As most early studies are conducted over daily or longer time horizons, it is worthwhile to examine this phenomenon with high frequency data, which provides the opportunity to explore more closely the relation between stock price and its own volatility. Some recent work has demonstrated that the volatility asymmetry still appears over fairly small time intervals. But some new aspects are added as both very good and bad news increase volatility, with the latter having a more severe effect [Barndorff-Nielsen, Kinnebrock, and Shehpard (2008b), Chen and Ghysels (2011)]. Also, the leverage effect decays exponentially as the time lag between return and volatility increases. In the literature, the peak effect is obtained at instantaneous correlation between return and volatility [Bollerslev, Litvinova, and Tauchen (2006), Bouchaud, Matacz, and Potters (2001)][1]. This corresponds to our definition of the leverage effect as being instantaneous. For further discussion of this effect, see the end of Section 2.2.

Although many papers deal with the source or new properties of the leverage effect, few have tried to rigorously estimate it, which is critical for supporting any conclusive claims. There are actually many pitfalls if simple correlation estimators are applied to the estimation of the leverage effect. Those estimators lose consistency with high frequency data [Aït-Sahalia, Fan, and Li (2011)]. It is the purpose of this article to construct nonparametric estimators of leverage effects in a stochastic volatility model. This paper defines a new general form of the leverage effect as the covariance (a covariation, to be precise) between the stock return and a function of its volatility. A first study of the classical equi-distant sampling case without microstructure noise serves as the foundation for study (later in the paper) of more complicated cases and provides some insights into the estimation problem. Even with equi-distant observations without microstructure noise, this paper discovers a bias correction factor which is critical to obtaining consistency. The bias correction factor may also have a substantial impact on the estimated value since it functions as a magnifier, especially when estimates are close to zero. This factor is previously unknown. It is not hard to conjecture that when the situation becomes more complicated as market microstructure noise is presented, the bias

[1]The relative change between time lag and lagged leverage effect should be maintained, even if the consistency may be an concern, since the consistency can be achieved by a bias correction multiplier from our study shown later.

correction factor may play an even more important role in estimation. Indeed, as estimators of the leverage effect for this case are derived, the bias correction factor is found to be bigger than that in the case with uncontaminated continuous price paths. As is emphasized by [Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a) and Renault and Werker (2009)] in their studies, it is more natural to work with irregularly spaced data in practice. Based on the results in equi-distant cases, the extension of estimators to irregularly spaced data can be constructed in a similar way, with some adaptations.

Statistical properties such as consistency and asymptotic distribution are carefully studied in different settings. These asymptotic properties have applications to hypothesis testing (e.g. for model checking) and constructing confidence intervals. To examine the theoretical results in data, simulations, which further corroborate all the theoretical findings, are provided .

There are many ways to apply the estimators of the leverage effect depending on the practical purpose. One way to explore the potential application of the estimators is embedded in the definition of the stochastic parameter of the leverage effect. According to the definition, one specific choice of the function imposed on volatility gives rise to a unique relation between the leverage effect and skewness, which will help to estimate skewness consistently. This may introduce further applications in hedging strategy or new product design (see Neuberger (2011)). Another carefully chosen function of volatility can simplify the estimation of high frequency regression coefficients. This leads to an interesting discovery in volatility prediction. The empirical study with Microsoft stock data (2008-2011) shows strong predictive power of a term containing the leverage effect on the next period volatility. The power is comparable to that of current period volatility which is believed to be the most significant term in volatility prediction.

The main results of this paper will be given in Sections 2 and 5. The data generating mechanism and model setting can be found in Section 2.1. The (stochastic) parameter of the leverage effect is defined in Section 2.2. Based on this, for the case without microstructure noise, the estimator and limit theorems can be found in the following Section 2.3 and Section 2.4. Simulation results are provided in Section 3. Results that corroborate the theorems can be found in Section 3.1. Section 4 studies the case where market microstructure noise is present in the data. The estimator and limit theorems for this case are provided in Section 4.1 and Section 4.2. Simulation results are also provided in Section 5. The extension to irregularly spaced data can be found in Section 6. Section 7 shows the relation between leverage effect and skewness. Section 8 implements an application of the leverage effect in high frequency regression. The details of empirical studies are in Section 9. Section 10 provides the conclusion. Proofs are in the Appendix.

2 Main Results.

2.1 Data Generating Mechanism

In general, we shall work with a broad class of continuous semimartingales, namely *Itô processes*. In econometrics and financial mathematics studies, this is the most popular model for log price processes due to no-arbitrage considerations [Delbaen and Schachermayer (1994, 1995, 1998)].

Definition 1. *A process X_t is called an Itô process provided it satisfies*

$$dX_t = \mu_t dt + \sigma_t dW_t, X_0 = x_0, \tag{1}$$

where μ_t and σ_t are adapted càdlàg locally bounded random processes, and W_t is a Wiener process. The underlying filtration will be called $(\mathcal{F}_t)$. The probability measure will be called P.

The integrated variance process is given as

$$\langle X, X\rangle_t = \int_0^t \sigma_u^2\, du. \tag{2}$$

The process (2) is also known as the quadratic variation of X. We shall sometimes also use the term "integrated volatility."

We further assume that σ_t is also an Itô process (see the next section for discussion of this)

$$d\sigma_t = a_t\, dt + f_t\, dW_t + g_t\, dB_t, \tag{3}$$

where B_t is another Wiener process independent of W_t, and a_t, f_t, and g_t are all assumed to be Itô processes.

Clearly, in this Stochastic Volatility (SV) model, X_t corresponds to the log return process and σ_t is its own volatility process. Both processes have a common driving Wiener process W_t, which accommodates the leverage effect.

To summarize the technical requirements, specify exact conditions as follows:

Assumption 1. *The system satisfies (1) and (3), where X and σ are continuous processes (the continuous modification). We assume that all the coefficients f_t, g_t, a_t, μ_t are locally bounded in absolute* *value. We also assume that σ_t is locally bounded away from zero.*[2]

2.2 The parameter: A definition of leverage effect

As we have seen in the introduction, the literature offers various perspectives on how to specify a parameter for this effect, cf. also the discussion below in this section. This paper concentrates on the estimation of the following stochastic parameter:

DEFINITION 2. *The stochastic parameter of the contemporaneous leverage effect is defined as the quadratic co-variation between X_t and $F(\sigma_t^2)$*

$$\langle X, F(\sigma^2)\rangle_T = \int_0^T 2F'(\sigma_t^2)\sigma_t^2 f_t \, dt, \tag{4}$$

where we suppose that

ASSUMPTION 2. $x \mapsto F(x)$ *is twice continuously differentiable, monotone on* $(0, \infty)$.

The incorporation of the function F allows more flexibility and a wider range of applications when different forms of volatility are of interest, such as log volatility processes which tend to be more stationary over time as implied by many empirical studies. The actual choice of F will depend on the practical purpose and empirical evidence. The inclusion of this function can also help reveal some interesting connection between leverage effects and other statistics. Further interpretation of this specification will be seen in Section 7 and Section 8.

We stop here for a moment to reflect on the definitions. First of all, we work with continuous processes. The interface with jump processes remains to be explored. The latter permits additional concepts of asymmetry, in particular the semivariance of Barndorff-Nielsen, Kinnebrock, and Shehpard (2008b). The connection between semivariance and the leverage effect (and skewness, cf. Section 7) in this paper is an important question which we leave for future investigation. This is necessarily a complex matter, as it involves a different model of the price process (continuous paths vs. jumps).

Once one works with continuous paths, the assumption that the leverage effect is instantaneous is both natural in a semimartingale model for σ_t, and is empirically supported by the finding in Bollerslev, Litvinova, and Tauchen (2006), where it was shown that that the connection between

[2] To get from local boundedness to results that cover the whole time imterval, use arguments as in Chapter 2.4.5 (p. 160-161) of Mykland and Zhang (2012). $|\sigma_t|$ is locally bounded from above by continuity. The assumptions guarantee that the equivalent martingale measure for X exists locally. This is used in the proofs, cf. the beginning of Section A.1.

return and volatility is most significant when the time lag is 0. This does not contradict the fact that the effect can appear at a greater time lag, as documented by Chen and Ghysels (2011).

As far as the Itô process (continuous semimartingale) assumption is concerned, this assumption appears frequently both on the options pricing side [Hull and White (1987), Stein and Stein (1991), Heston (1993), Ball and Roma (1994)], and on the econmetric side [Barndorff-Nielsen and Shephard (2002), Barndorff-Nielsen, Graversen, Jacod, Podolskij, and Shephard (2006), Jacod (2008), Barndorff-Nielsen and Veraart (2009), Aït-Shalia and Jacod (2009), Mykland and Zhang (2011b)]. A parallel development can be carried out under assumptions of fractional Brownian motion. [Comte and Renault (1998), Gloter and Hoffmann (2004), Brockwell and Marquardt (2005), Nualart (2006), Comte, Coutin, and Renault (2010)].

The above is, of course, a set of theoretical considerations. We finally appeal to the results in Section 9 to show that our current definition of volatility asymmetry does find something empirically relevant: we substantially improve the prediction of next-period volatility using the current-period leverage effect.

2.3 Estimation in the Absence of Microstructure Noise

As the first step, we shall work with the equally spaced case for the process X_t; specifically it is observed every $\Delta t_{n,i+1} = \Delta t = \frac{T}{n}$ units of time, at times $0 = t_{n,0} < t_{n,1} < t_{n,2} < \ldots < t_{n,n} = T$. Furthermore, we divide observed values into K_n blocks, with block size $M_n = [c\sqrt{n}]$ (except possibly for the first and last block, which does not matter for the asymptotics), for some constant c. The boundary points are on the grid $\mathcal{H} = \{0 < \tau_{n,1} < \tau_{n,2} < \cdots < \tau_{n,K_n-1} \leq T\}$, where $K_n = [\frac{n}{M_n}]$.

Define[3]

$$\begin{aligned} \langle \widehat{X, F(\sigma^2)} \rangle_T &= 2 \sum_{i=0}^{K_n-2} (X_{\tau_{n,i+1}} - X_{\tau_{n,i}})(F(\hat{\sigma}^2_{\tau_{n,i+1}}) - F(\hat{\sigma}^2_{\tau_{n,i}})) \\ \text{and } \hat{\sigma}^2_{\tau_{n,i}} &= \frac{1}{M_n \times \Delta t} \sum_{t_{n,j} \in (\tau_{n,i}, \tau_{n,i+1}]} (X_{t_{n,j+1}} - X_{t_{n,j}})^2. \end{aligned} \tag{5}$$

The factor 2 in the first equation above might look unnatural to be included. However, it is crucial for the consistency of the estimator. See Remark 3 for a discussion of this previously unknown factor.

[3]One can also consider a kernel estimator of the spot volatility in (5), by applying the methods in Kristensen (2010), with some adaptation. A detailed study is beyond the scope of this paper.

THEOREM 1. *Under Assumptions 1-2, as $n \to \infty$ and T fixed,*

$$n^{1/4}\left(\widehat{\langle X, F(\sigma^2)\rangle}_T - \langle X, F(\sigma^2)\rangle_T\right) \xrightarrow{\mathcal{L}} Z\left(\frac{16}{c}\int_0^T (F'(\sigma_t^2))^2\sigma_t^6\,dt + cT\int_0^T (F'(\sigma_t^2))^2\sigma_t^4(\frac{44}{3}f_t^2 + \frac{32}{3}g_t^2)\,dt\right)^{1/2}, \tag{6}$$

stably in law [4], *where Z is a standard normal random variable and independent of $\mathcal{F}_T$.*

Another natural estimator analogous to $\widehat{\langle X, F(\sigma^2)\rangle}_T$ is

$$\begin{aligned}
\widetilde{\langle X, F(\sigma^2)\rangle}_T &= 2\sum_{i=0}^{K_n-2}(X_{\tau_{i+1,n}} - X_{\tau_{n,i}})(F(\widetilde{\sigma}^2_{\tau_{n,i+1}}) - F(\widetilde{\sigma}^2_{\tau_{n,i}})),\\
\widetilde{\sigma}^2_{\tau_{n,i}} &= \frac{1}{M_n \times \Delta t}\sum_{t_{n,j}\in(\tau_{n,i},\tau_{n,i+1}]}(\Delta X_{t_{n,j+1}} - \overline{\Delta X_{\tau_{n,i+1}}})^2,\\
\text{and } \overline{\Delta X_{\tau_{i+1,n}}} &= \frac{1}{M_n}\sum_{t_{n,j}\in(\tau_{n,i},\tau_{n,i+1}]}\Delta X_{t_{n,j+1}} = \frac{1}{M_n}(X_{\tau_{n,i+1}} - X_{\tau_{n,i}}).
\end{aligned} \tag{7}$$

Noticing the relation between the two estimators when $F(x) = x$: $\widetilde{\langle X, \sigma^2\rangle}_T = \frac{M_n}{M_n-1}\widehat{\langle X, \sigma^2\rangle} - \sum_i \frac{2}{M_n(M_n-1)\Delta t}(\Delta X_{\tau_{n,i+1}})^3$, the following theorem can be easily derived:

THEOREM 2. *Under Assumptions 1-2 as in Theorem 1, as $n \to \infty$ and T fixed,*

$$n^{1/4}\left(\widetilde{\langle X, F(\sigma^2)\rangle}_T - \langle X, F(\sigma^2)\rangle_T\right) \xrightarrow{\mathcal{L}} Z\left(\frac{16}{c}\int_0^T (F'(\sigma_t^2))^2\sigma_t^6\,dt + cT\int_0^T (F'(\sigma_t^2))^2\sigma_t^4(\frac{44}{3}f_t^2 + \frac{32}{3}g_t^2)\,dt\right)^{1/2}, \tag{8}$$

stably in law[5], *where Z is a standard normal random variable and independent of $\mathcal{F}_T$.*

Remark 1. From the limit theorems, it is not hard to see that by properly choosing c, one can minimize the limit variance. The optimal value is

$$c^2 = \frac{16\int_0^T (F'(\sigma_t^2))^2\sigma_t^6 dt}{T\int_0^T (F'(\sigma_t^2))^2\sigma_t^4(\frac{44}{3}f_t^2 + \frac{32}{3}g_t^2)dt}. \tag{9}$$

See Section 2.4 below for an estimator of c[6].

Remark 2. The two estimators have the same asymptotic properties. Even though the centered version gives slightly more symmetric results, it does not behave very differently from the non-centered version.

[4]Suppose that all relevant processes (X_t, σ_t, etc.) are adapted to the filtration ($\mathcal{F}_t$). Let Z_n be a sequence of $\mathcal{F}_t$-measurable random variables. We say that Z_n converges stably in law to Z as $n \to \infty$ if Z is measurable with respect to an extension of $\mathcal{F}_T$ so that for all $A \in \mathcal{F}_T$ and for all bounded continuous g, $EI_A g(Z_n) \to EI_A g(Z)$ as $n \to \infty$. The same definition applies to triangular arrays.

[5]See Footnote 4.

[6]Here and in the continuation of Remark 1 below we assume that the denominator in (9) is nonzero.

In practice, the non-centered version can be applied with less programming effort. Therefore, our later simulation mainly adopts the non-centered version of estimator.

Remark 3. The origin of the factor 2 in the estimator can be found in the proof of Theorem 1. For intuition, however, we give here a verbal explanation of the source of this adjustment constant. Let us consider the case where $F(x) = x$. Since $\hat{\sigma}_t^2$ is a consistent estimator of σ_t^2, then the first multiplication $(X_{\tau_{i+1}} - X_{\tau_i})(\hat{\sigma}^2_{\tau_{i+1}} - \sigma^2_{\tau_i})$ already gives a consistent (though infeasible) estimator of the leverage effect in the interval $(\tau_i, \tau_{i+1}]$. Then one may expect the remainder term $(X_{\tau_{i+1}} - X_{\tau_i})(\hat{\sigma}^2_{\tau_i} - \sigma^2_{\tau_i})$ to have mean zero. However, since $\hat{\sigma}^2_{\tau_i}$ employs data in time interval $(\tau_i, \tau_{i+1}]$, as does $(X_{\tau_{i+1}} - X_{\tau_i})$, the product does not converge to zero but to one half of the leverage effect. To see why it is one half, note that each increment $\frac{\Delta X^2_{t_j}}{\Delta t}$ term is roughly an (inconsistent) estimator of $\sigma^2_{t_j}$. Thus the cross product gives an average of leverage effects over $(\tau_i, t_1], (\tau_i, t_2], \cdots, (\tau_i, \tau_{i+1}]$. If $\langle X, \sigma^2 \rangle'_t$ is considered to be constant over $(\tau_i, \tau_{i+1}]$, that average of leverage effects will give a value of about half of the leverage effect over the entire interval. Hence we have reduced the estimation of leverage effect by half. An adjustment factor of 2 therefore needs to be added to achieve consistency.

2.4 Estimation of asymptotic variance

Let

$$\begin{aligned} G_n^1 &= 2n^{\frac{1}{2}} \sum_{i=1}^{K_n-1} (X_{\tau_{n,i+1}} - X_{\tau_{n,i}})^2 (F(\widehat{\sigma}^2_{\tau_{n,i+1}}) - F(\widehat{\sigma}^2_{\tau_{n,i}}))^2, \\ \text{and } G_n^2 &= 2n^{\frac{1}{2}} M_n \Delta t \sum_{i=1}^{K_n-1} \widehat{\sigma}^2_{\tau_{n,i}} (F(\widehat{\sigma}^2_{\tau_{n,i+1}}) - F(\widehat{\sigma}^2_{\tau_{n,i}}))^2. \end{aligned} \tag{10}$$

By the same methods as in the proof of Theorem 1, we have the following convergences in probability:

$$G_n^1 \xrightarrow{p} \frac{8}{c} \int_0^T (F'(\sigma_t^2))^2 \sigma_t^6 dt + cT \int_0^T (F'(\sigma_t^2))^2 \sigma_t^4 (\frac{28}{3} f_t^2 + \frac{16}{3} g_t^2) dt, \tag{11}$$

$$G_n^2 \xrightarrow{p} \frac{8}{c} \int_0^T (F'(\sigma_t^2))^2 \sigma_t^6 dt + cT \int_0^T (F'(\sigma_t^2))^2 \sigma_t^4 \frac{16}{3} (f_t^2 + g_t^2) dt, \tag{12}$$

$$\text{and } G_n^1 + G_n^2 \xrightarrow{p} \frac{16}{c} \int_0^T (F'(\sigma_t^2))^2 \sigma_t^6 dt + cT \int_0^T (F'(\sigma_t^2))^2 \sigma_t^4 (\frac{44}{3} f_t^2 + \frac{32}{3} g_t^2) dt \tag{13}$$

The equation (13) gives the estimation of the asymptotic variance. With this estimation, a feasible version of the central limit distribution can be derived.

THEOREM 3. *Under Assumptions 1-2, as $n \to \infty$ and T fixed,*

$$\frac{n^{1/4}\left(\widehat{\langle X, F(\sigma^2)\rangle}_T - \langle X, F(\sigma^2)\rangle_T\right)}{\sqrt{G_n^1 + G_n^2}} \xrightarrow{\mathcal{L}} Z_1,$$
$$\text{and } \frac{n^{1/4}\left(\widetilde{\langle X, F(\sigma^2)\rangle}_T - \langle X, F(\sigma^2)\rangle_T\right)}{\sqrt{G_n^1 + G_n^2}} \xrightarrow{\mathcal{L}} Z_1 \tag{14}$$

stably in law[7]*, where Z_1 is a standard normal random variable and independent of $\mathcal{F}_T$.*

Notice that the limiting distribution Z_1 is the same in both limits. In other words, the difference between the two statistics converges to zero in probability. With this feasible CLT one can conduct hypothesis testing and construct confidence interval about the leverage effect parameter.

Remark 1 (continued). The result (13) opens paths to estimating the tuning parameter c in (9). We here outline two approaches.

Method 1: The conceptually simplest possibility is to pick $c = \arg\min\{G_n^1 + G_n^2\}$ over a suitable grid of c's. If the grid is nested and becomes dense as $n \to \infty$, this automatically provides a consistent estimator of c.

Method 2: Since Method 1 is computationally heavy, we here also propose an alternative two step method. Fix an initial value c_0, and compute $(G_n^1 + G_n^2)(c_0)$. On the other hand, we can reduce estimation of $\gamma^2 = \int_0^T (F'(\sigma_t^2))^2 \sigma_t^6\, dt$ to the local estimation of volatility by the methods in Section 4.1 in Mykland and Zhang (2009). Call this latter estimate $\hat{\gamma}^2$. We thus obtain that $c_0^{-1}(G_n^1 + G_n^2)(c_0) - 16\hat{\gamma}^2/c_0^2$ consistently estimates $T \int_0^T (F'(\sigma_t^2))^2 \sigma_t^4 (\frac{44}{3} f_t^2 + \frac{32}{3} g_t^2) dt$. A consistent estimate of c^2 is thus given from (9) as

$$\hat{c}^2 = 16\hat{\gamma}^2 \left(c_0^{-1}(G_n^1 + G_n^2)(c_0) - 16\hat{\gamma}^2/c_0^2\right)^{-1}$$

The two methods can be used together, with the second providing a starting point for seaching for a minimum in Method 1.

3 Simulation Results

All simulation results are based on 10000 sample paths while varying the sample size n, function F and optimal choice of c (path dependent). In the simulation, the properties of the estimator are studied

[7] See Footnote 4.

with the Heston model [Heston (1993)]. In order to corroborate the theoretical limit distribution, the distribution of the statistic in Theorem 1

$$\frac{\frac{1}{\sqrt{M_n \Delta t}}\left(\widehat{\langle X, F(\sigma^2)\rangle}_T - \langle X, F(\sigma^2)\rangle_T\right)}{\left(\frac{16}{cT}\int_0^T F'(\sigma_t^2)^2\sigma_t^6 dt + c\int_0^T (F'(\sigma_t^2))^2\sigma_t^4(\frac{44}{3}f_t^2 + \frac{32}{3}g_t^2)dt\right)^{1/2}},$$

and the statistics in Theorem 3,

$$\frac{n^{1/4}\left(\widehat{\langle X, F(\sigma^2)\rangle}_T - \langle X, F(\sigma^2)\rangle_T\right)}{\sqrt{G_n^1 + G_n^2}} \text{ and } \frac{n^{1/4}\left(\widetilde{\langle X, F(\sigma^2)\rangle}_T - \langle X, F(\sigma^2)\rangle_T\right)}{\sqrt{G_n^1 + G_n^2}},$$

are studied. So if the asymptotics correctly predict small sample behavior, the distributions should be close to the standard normal distribution.

The Heston model used in the simulation is of the form:

$$\begin{aligned} dX_t &= \sqrt{\sigma_t}\, dW_t, \\ d\sigma_t &= \kappa(\theta - \sigma_t)\, dt + \gamma\sqrt{\sigma_t}(\rho\, dW_t + \sqrt{1-\rho^2}\, dB_t), \text{ where } W_t \perp\!\!\!\perp B_t. \end{aligned} \tag{15}$$

3.1 Normality demonstration

In the simulation, the true log price is simulated from the Heston model with broadly realistic parameter values: $\kappa = 5, \theta = 0.04, \gamma = 0.5, \rho = -\sqrt{0.5}$ over 1 trading day. Two different sampling frequencies are studied to examine the small sample behavior. The first is when the data are observed at one minute frequency, which corresponds to sample size 390. The second is when the data are observed at every second, which corresponds to sample size 23400. The results are given in Table 1.

4 Estimation with Microstructure Noise

It is well known that markets are not so ideal that log price processes can be simply represented by pure semimartingales. This has long been thought about as "microstructure", see, for example, Roll (1984), O'Hara (1995), Harris (1990) and Hasbrouck (1996). In the context of high frequency data, such microstructure was originally observed through the so-called signature plot (introduced by Andersen, Bollerslev, Diebold, and Labys (2000), see also the discussion in Mykland and Zhang

	MSE	mean	median	Q_1	Q_3
n=390, T=1/250 infeasible	1.097112	-0.02108	-0.005707	-0.676	0.6496
n=390, T=1/250 feasible	1.051725	0.006831	-0.004835	-0.754	0.7399
n=23400, T=1/250 infeasible	1.009285	-0.006430	0.003727	-0.6982	0.6796
n=23400, T=1/250 feasible	1.002125	0.002964	0.004267	-0.6964	0.6917
n $\to \infty$, fixed T (asymptotic value)	1	0	0	-0.674	0.674

Table 1: The summary statistics do exhibit the target normality. This corroborates the theorems and shows that the asymptotics can predict small sample behavior. For sample size 390, both the mean and median are very close to 0. The MSE is close to 1 and the quartiles are close to the theoretical values from $N(0,1)$. As sample size increases, the MSE decreases further closer to 1.

(2005). This led researchers to investigate a model where the efficient price is latent, and one actually observes

$$Y_t = X_t + \epsilon_t. \tag{16}$$

Several approaches[8] seek to deal with microstructure noise while estimating integral volatility, and they shed light on how to proceed in the estimation of leverage effects in the similar situation. Among these approaches, this paper has focused on pre-averaging. The pre-averaging method [Jacod, Li, Mykland, Podolskij, and Vetter (2009), Podolskij and Vetter (2009a), Mykland and Zhang (2011a)] provides a plausible way to solve the problem with microstructure. Therefore all of the following discussion will be in the framework of pre-averaging and the blocking method will be adjusted as follows:

The contaminated log return process Y_t is observed every $\Delta t_{n,i} = \frac{T}{n}$ units of time, at times $0 = t_{n,0} < t_{n,1} < t_{n,2} < \ldots < t_{n,n} = T$.

Assumption 3.

$$Y_t = X_t + \epsilon_t, \textit{ where } \epsilon_t\textit{'s are i.i.d. } N(0, a^2) \textit{ and } \epsilon_t \perp\!\!\!\perp \textit{ the } W \textit{ and } B \textit{ processes, for all } t. \tag{17}$$

We also assume that ϵ_t's have finite fourth moment, and are independent of both return and volatility processes.

[8]such as Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2006), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a), Reiss (2010), and Xiu (2010), as well as the preaveraging papers cited in the text.

We can also relax these assumptions, cf. the development in Mykland and Zhang (2011a).

Here two nested levels of blocks will be required. The first level of blocks defines the range of pre-averaging and the second one implements a blocking idea similar to that in the case without noise in section 2.3.

Blocks are defined on a much less dense grid of $\tau_{n,i}$, also spanning $[0, T]$, so that

$$\text{block } \# \ i = \{t_{n,j} : \tau_{n,i} \leq t_{n,j} < \tau_{n,i+1}\} \tag{18}$$

(the last block, however, includes T). We define the block size by

$$M_{n,i} = \#\{j : \tau_{n,i} \leq t_{n,j} < \tau_{n,i+1}\}. \tag{19}$$

In principle, the block size $M_{n,i}$ can vary across the trading period $[0, T]$, but for this development we take $M_{n,i} = M_n$: it depends on the sample size n, but not on the block index i.

We then use as an estimated value of the efficient price in the time period $[\tau_{n,i}, \tau_{n,i+1})$:

$$\hat{X}_{\tau_{n,i}} = \frac{1}{M_n} \sum_{t_{n,j} \in [\tau_{n,i}, \tau_{n,i+1})} Y_{t_{n,j}}$$

Treating the estimated efficient price as a new data frame, we proceed as in Section 2.3 but with X_t replaced by $\hat{X}_t$, n by $n' = n/M_n$ (up to rounding), and $t_{n,i}$ by $\tau_{n,i}$. Furthermore, we divide $\hat{X}_t$ values into K_n blocks, with block size $L = L_n = [c\sqrt{n'}]$ (except possibly for the first and last block, which does not matter for the asymptotics), for some constant c. The boundary points are on the grid $\mathcal{G} = \{0 < \lambda_{n,1} < \lambda_{n,2} < \cdots < \lambda_{n,K_n-1} \leq T\} \subset \mathcal{H}$.

4.1 The case with microstructure noise

In the case with microstructure noise, the data blocking mechanism will be similar to that just stated, but less complicated where $M = M_n = [c_1\sqrt{n}]$, $\tau_{n,i} = iM_n\frac{T}{n}$, and $L = L_n = [\frac{cn^{1/4}}{\sqrt{c_1}}]$. The interval between successive observations is now $\Delta t = \Delta t_n = t_{n,j+1} - t_{n,j} = T/n$.

Define

$$\begin{aligned}
\widehat{\langle X, F(\sigma^2)\rangle}_T &= 3 \sum_{i=0}^{K_n-2} (\hat{X}_{\lambda_{n,i+1}} - \hat{X}_{\lambda_{n,i}})(F(\hat{\sigma}^2_{\lambda_{n,i+1}}) - F(\hat{\sigma}^2_{\lambda_{n,i}})), \\
\hat{X}_{\tau_{n,i}} &= \frac{1}{M} \sum_{t_{n,j} \in [\tau_{n,i}, \tau_{n,i+1})} Y_{t_{n,j}}, \\
\text{and } \hat{\sigma}^2_{\lambda_{n,i}} &= \frac{1}{L \times M \times \Delta t} \sum_{\tau_{n,j+1} \in (\lambda_{n,i}, \lambda_{n,i+1}]} (\hat{X}_{\tau_{n,j+1}} - \hat{X}_{\tau_{n,j}})^2.
\end{aligned} \tag{20}$$

Note that the factor 2 in the previous proposed estimator in equation (5) is now changed to 3 instead. This change is due to the pre-averaging method we adopted first in order to asymptotically eliminate the impact of noise on the estimation. The change is consistent with the adjustment to estimate realized volatility by pre-averaging, cf. Jacod, Li, Mykland, Podolskij, and Vetter (2009).

THEOREM 4. *Under Assumptions 1-3, as $n \to \infty$ and T fixed,*

$$\begin{aligned} n^{1/8}\left(\widehat{\langle X, F(\sigma^2)\rangle}_T - \langle X, F(\sigma^2)\rangle_T\right) \xrightarrow{\mathcal{L}} Z\Bigg(c\sqrt{c_1}T\int_0^T (F'(\sigma_t^2))^2\sigma_t^4(\frac{44}{3}f_t^2 + \frac{32}{3}g_t^2)dt \\ + \frac{16\sqrt{c_1}}{c}\int_0^T (F'(\sigma_t^2))^2\sigma_t^6 dt + \frac{96a^2}{cc_1^{3/2}T}\int_0^T (F'(\sigma_t^2))^2\sigma_t^4 dt + \frac{216a^4}{cc_1^{7/2}T^2}\int_0^T (F'(\sigma_t^2))^2\sigma_t^2 dt\Bigg)^{1/2}, \end{aligned} \tag{21}$$

stably in law[9], *where Z is a standard normal random variable and independent of $\mathcal{F}_T$.*

The optimal c and c_1 that minimize the asymptotic variance are derived as follows:

$$c = \sqrt{\frac{-C^2 + 12AD + C\sqrt{C^2 + 12AD}}{9BD}}, \tag{22}$$

$$\text{and } c_1 = \sqrt{\frac{C + \sqrt{C^2 + 12AD}}{2A}} \tag{23}$$

where $A = 16\int_0^T (F'(\sigma_t^2))^2\sigma_t^6 dt$, $B = T\int_0^T (F'(\sigma_t^2))^2\sigma_t^4(\frac{44}{3}f_t^2 + \frac{32}{3}g_t^2)dt$, $C = \frac{96a^2}{T}\int_0^T (F'(\sigma_t^2))^2\sigma_t^4 dt$, and $D = \frac{216a^4}{T^2}\int_0^T (F'(\sigma_t^2))^2\sigma_t^2 dt$.

In practice, c and c_1 can be estimated by minimizing $G_n^1 + G_n^2$ defined in the next section, over a suitable grid of c's and c_1's. If the grid is nested and becomes dense as $n \to \infty$, this automatically provides a consistent estimator of c and c_1.

4.2 Estimation of asymptotic variance

Let

$$\begin{aligned} G_n^1 &= \frac{9}{2}n^{\frac{1}{4}}\sum_{i=1}^{K_n-1}(\hat{X}_{\lambda_{n,i+1}} - \hat{X}_{\lambda_{n,i}})^2(F(\hat{\sigma}^2_{\lambda_{n,i+1}}) - F(\hat{\sigma}^2_{\lambda_{n,i}}))^2, \\ \text{and } G_n^2 &= \frac{9}{2}n^{\frac{1}{4}}L_n M_n \Delta t\sum_{i=1}^{K_n-1}\widehat{\sigma}^2_{\lambda_{n,i+1}}(F(\widehat{\sigma}^2_{\lambda_{n,i+1}}) - F(\widehat{\sigma}^2_{\lambda_{n,i+1}}))^2. \end{aligned} \tag{24}$$

By the same methods as in the proof of Theorem 1, we have the following convergences in probability

$$G_n^1 \xrightarrow{p} \frac{8}{c^2T}\int_0^T (F'(\sigma_t^2))^2\sigma_t^6 dt + \int_0^T (F'(\sigma_t^2))^2\sigma_t^4(\frac{28}{3}f_t^2 + \frac{16}{3}g_t^2)dt$$

[9] See Footnote 4.

$$+\frac{48a^2}{c^2c_1^2T^2}\int_0^T(F'(\sigma_t^2))^2\sigma_t^4dt+\frac{108a^4}{c^2c_1^4T^3}\int_0^T(F'(\sigma_t^2))^2\sigma_t^2dt, \tag{25}$$

$$G_n^2 \xrightarrow{p} \frac{8}{c^2T}\int_0^T(F'(\sigma_t^2))^2\sigma_t^6dt+\int_0^T(F'(\sigma_t^2))^2\sigma_t^4\frac{16}{3}(f_t^2+g_t^2)dt$$
$$+\frac{48a^2}{c^2c_1^2T^2}\int_0^T(F'(\sigma_t^2))^2\sigma_t^4dt+\frac{108a^4}{c^2c_1^4T^3}\int_0^T(F'(\sigma_t^2))^2\sigma_t^2dt, \tag{26}$$

$$\text{and } G_n^1+G_n^2 \xrightarrow{p} \frac{16}{c^2T}\int_0^T(F'(\sigma_t^2))^2\sigma_t^6dt+\int_0^T(F'(\sigma_t^2))^2\sigma_t^4(\frac{44}{3}f_t^2+\frac{32}{3}g_t^2)dt$$
$$+\frac{96a^2}{c^2c_1^2T^2}\int_0^T(F'(\sigma_t^2))^2\sigma_t^4dt+\frac{216a^4}{c^2c_1^4T^3}\int_0^T(F'(\sigma_t^2))^2\sigma_t^2dt. \tag{27}$$

The equation (27) gives the estimation of the asymptotic variance. With this estimation, a feasible version of the central limit distribution can be derived.

THEOREM 5. *Under Assumptions 1-3, as $n\to\infty$ and T fixed,*

$$\frac{n^{1/4}\left(\widehat{\langle X,F(\sigma^2)\rangle}_T-\langle X,F(\sigma^2)\rangle_T\right)}{\sqrt{G_n^1+G_n^2}} \xrightarrow{\mathcal{L}} Z_1 \tag{28}$$

stably in law[10], *where Z_1 is a standard normal random variable and independent of $\mathcal{F}_T$.*

5 Simulation for the case with microstructure noise

Similarly to the case without microstructure noise, the small sample behavior of the asymptotic normality can be demonstrated by simulating the statistics

$$\frac{n^{1/8}\left(\widehat{\langle X,F(\sigma^2)\rangle}_T-\langle X,F(\sigma^2)\rangle_T\right)}{\sqrt{\frac{16\sqrt{c_1}}{c}\int_0^T(F'(\sigma_t^2))^2\sigma_t^6dt+c\sqrt{c_1}T\int_0^T(F'(\sigma_t^2))^2\sigma_t^4(\frac{44}{3}f_t^2+\frac{32}{3}g_t^2)dt+\frac{96a^2}{cc_1^{3/2}T}\int_0^T(F'(\sigma_t^2))^2\sigma_t^4dt+\frac{216a^4}{cc_1^{7/2}T^2}\int_0^T(F'(\sigma_t^2))^2\sigma_t^2dt}}$$

and

$$\frac{n^{1/4}\left(\widehat{\langle X,F(\sigma^2)\rangle}_T-\langle X,F(\sigma^2)\rangle_T\right)}{\sqrt{G_n^1+G_n^2}}.$$

The Heston model is once again applied in the simulation. The parameterization is the same with $\kappa=5,\theta=0.04,\gamma=0.5,\rho=-\sqrt{0.5}$. The true log price process is latent. It is contaminated by market microstructure as in equation (17). The standard deviation of noise is set to be $a=0.0005$. This is also a realistic value in practice. Since the first step of pre-averaging consumes part of the data and reduces the sample size for the second step of estimation, the choices of n are bigger than those in

[10]See Footnote 4.

the case without noise. The frequency is chosen as one second, which produces 23400 observations in each trading day. The results corroborate the theorem and are demonstrated in Table 2.

	MSE	mean	median	Q_1	Q_3
n=5 days, T=1/50 infeasible	1.315581	-0.05502	-0.01236	-0.71910	0.66330
n=5 days, T=1/50 feasible	1.142911	0.02566	-0.02703	-0.79680	0.80940
n=20 days, T=2/25 infeasible	1.193074	-0.03032	-0.003359	-0.6793	0.6578
n=20 days, T=2/25 feasible	1.125859	0.02167	-0.05247	-0.77740	0.76390
n $\to \infty$, fixed T (asymptotic value)	1	0	0	-0.674	0.674

Table 2: Because of the much slower convergence rate, the simulation results are not as good as in the case without microstructure noise. Even so, with reasonably large sample size, the mean and median are still close to 0. The MSE is not very far from 1, and the quartiles are reasonably close to the theoretical values from the standard normal distribution.

Even though only the simulations with $F(x) = x$ are presented, the results with other functions satisfying the condition in definition (2) such as $F(x) = \log x$ have been investigated. The results look very similar and the tables are omitted for the reasons of space.

6 Irregularly Spaced Data

So far our analysis in cases both with and without microstructure noise has been based on measuring prices at regularly spaced intervals. In some ways it is more natural to work with prices measured in tick time and so it would be desirable to extend the above theory to cover irregularly spaced data. This is emphasized by Zhang, Mykland, and Aït-Sahalia (2005), Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a) and Renault and Werker (2009) in their studies. We here use the framework from Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008a).

Assumption 4. *The observation times* $(t_{n,i})$ *satisfy the condition that:*

$$t_{n,i} = G(i\tfrac{T}{n}) = \int_0^{i\frac{T}{n}} G'(s)ds, \qquad i = 0, 1, \cdots, n,$$

where $G : [0, T] \to [0, T]$ *is a strictly increasing, twice differentiable function with* $G(0) = 0$, $G(T) = T$. $G'(s)$ *is locally bounded away from* 0, *and* G'' *is bounded.*

With the change of time under the assumption(T), the stochastic volatility model can be written as:

$$\begin{aligned} dZ_t = dX \circ G(t) &= \mu_{G(t)} G'(t)dt + \sigma_{G(t)}\sqrt{G'(t)}dW_t^e, \\ \text{and } ds_t = d\sigma \circ G(t) &= a_{G(t)}G'(t)dt + f_{G(t)}\sqrt{G'(t)}dW_t^e + g_{G(t)}\sqrt{G'(t)}dB_t^e, \end{aligned} \tag{29}$$

where W^e and B^e are independent Wiener processes.

Proposition 1. *The leverage effect in Definition 2 satisfies* $\langle X, F(\sigma^2)\rangle_T = \langle Z, F(s^2)\rangle_T$.

Proof:

$$\begin{aligned} \langle Z, F(s^2)\rangle_T &= \int_0^T F'(s^2)d\langle s^2, X\rangle_t \\ &= \int_0^T F'(\sigma^2_{G(t)})2\sigma^2_{G(t)} f_{G(t)} G'(t)dt \\ &= \int_0^T F'(\sigma^2_{G(t)})2\sigma^2_{G(t)} f_{G(t)} dG(t) \\ &= \int_0^T F'(\sigma^2_v)2\sigma^2_v f_v dv \\ &= \langle X, F(\sigma^2)\rangle_T. \end{aligned}$$

With all index notation kept the same as in section 2.3, the estimator of $\langle Z, F(\frac{s^2}{G'})\rangle_T$ can be constructed similar to equation (5), with one adaptation:

$$\begin{aligned} \widehat{\langle Z, F(s^2)}\rangle_T &= 2\sum_{i=0}^{K_n-2}(X\circ G((i+1)M_n\Delta t) - X\circ G(iM_n\Delta t))(F(\hat{s}^2_{(i+1)M_n\Delta t}) - F(\hat{s}^2_{iM_n\Delta t})) \\ \text{and } \hat{s}^2_{iM_n\Delta t} &= \frac{1}{\Delta\tau_{n,i+1}} \sum_{\substack{t_{n,j+1}\in \\ (iM_n\Delta t,\ (i+1)M_n\Delta t]}} (X\circ G(t_{n,j+1}) - X\circ G(t_{n,j}))^2. \end{aligned} \tag{30}$$

The CLT, estimation of asymptotic variance, and feasible CLT follow for the estimator $\widehat{\langle Z, F(s^2}\rangle_T$ as in the equi-distant case without microstructure noise. The results for the case with microstructure noise can be derived analogously. Considering the contaminated process $Y\circ G(t) = X\circ G(t) + \epsilon\circ G(t) = Z_t + \epsilon\circ G(t)$, with all index notation kept the same as in section 4, the estimator of the leverage effect can be constructed as follows:

$$\widehat{\langle Z, F(s^2)\rangle}_T = 3\sum_{i=0}^{K_n-2}(\hat{Z}_{\lambda_{n,i+1}} - \hat{Z}_{\lambda_{n,i}})(F(\hat{s}^2_{\lambda_{n,i+1}}) - F(\hat{s}_{\lambda_{n,i}})),$$

$$\hat{Z}_{\tau_{n,j}} = \frac{1}{M}\sum_{\substack{t_{n,p+1}\in\\(jM\Delta t,(j+1)M\Delta t]}}(Y\circ G(t_{n,p+1}) - Y\circ G(t_{n,p})),$$

$$\text{and } \hat{s}^2_{\lambda_{n,i}} = \frac{1}{\Delta\lambda_{n,i+1}}\sum_{\substack{\tau_{n,j+1}\in\\(iLM\Delta t,\ (i+1)LM\Delta t]}}(\hat{Z}_{\tau_{n,j+1}} - \hat{Z}_{\tau_{n,j}})^2. \quad (31)$$

We emphasize that this estimator is feasible (observable), since both contaminated price Y_t and observation times $t_{n,i}$ and $\lambda_{n,i}$ are directly observable from the market.

The CLT, estimator of asymptotic variance and feasible CLT can be derived in a similar manner as when observations are regularly spaced.

7 Leverage Effect and Skewness

From [Mykland and Zhang (2009), Section 2.5], leverage effect($F(x) = x$) and skewness have a close relationship. For equi-distant data, the skewness of returns in high frequency data satisfies (as $n \to \infty$)

$$\frac{n}{T}\lim\sum_{t_{n,i+1}\le T}\Delta X^3_{t_{n,i+1}} \xrightarrow{\mathcal{L}} \frac{3}{2}\langle\sigma^2, X\rangle_T + 3\int_0^T \sigma_t^3(dW_t + \sigma_t^{-1}\mu_t dt) + \left(6\int_0^T \sigma_t^6 dt\right)^{1/2} Z,$$

where Z is a standard normal random variable. This is a biased and inconsistent estimator, but it is interesting to find that leverage effect appears on the right hand side. When the mean is removed in blocks of size M, this empirical skewness converges to the leverage effect plus a mixed normal error:

$$\frac{n}{T}\lim\sum_{t_{n,i+1}\le T}(\Delta X_{t_{n,i+1}} - \text{local mean of } X)^3 \xrightarrow{\mathcal{L}} \frac{3}{2}\langle\sigma^2, X\rangle_T + \left(\frac{M-1}{M}(6 + \frac{18}{M} - \frac{15}{M^2})\int_0^T \sigma_t^6 dt\right)^{1/2} Z$$

M is chosen differently in this paper and in [Mykland and Zhang (2009)]. It is a constant instead (*i.e.*, M does not grow with n). This tells us that in the case where skewness is hard to estimate directly, the consistent estimation proposed by this paper provides an alternative way to estimate skewness.

To further emphasize that we are indeed estimating a form of skewness by the leverage effect, we now consider the *predictable instantaneous skewness:*

$$\text{p-skew} := \frac{n}{T}\sum_{t_{n,i+1}\le T} E\left(\Delta X^3_{t_{n,i+1}}|\mathcal{F}_{t_i}\right).$$

We obtain

PROPOSITION 2. *Subject to regularity conditions, as $n \to \infty$,*

$$p\text{-}skew \xrightarrow{p} \frac{3}{2}\langle \sigma^2, X\rangle_T.$$

It should be noted that since $E\left(\Delta X^3_{t_{n,i+1}}|\mathcal{F}_{t_i}\right)$ is an unobservable quantity, this proposition also does not yield a method of estimation. It does, however, clarify the relationship between skewness and leverage effect.

The existence of a connection between skewness and the leverage effect has previously been noted in Meddahi and Renault (2004), see the discussion following Proposition 3.4 (p. 370).

8 Leverage Effects and Regression Coefficients

The estimation of leverage effects also has an application to estimating the regression coefficient of the volatility on its own log return. On one hand, the existence of leverage effect implies the relation of volatility and the log return as stated below:

$$d\sigma_t^2 = 2f_t dX_t + 2\sigma_t g_t dB_t + (2\sigma_t a_t - 2f_t\mu_t + f_t^2 + g_t^2)dt, \tag{32}$$

$$\text{and } \frac{d\langle X, \sigma_t^2\rangle_t}{d\langle X, X\rangle_t} = 2f_t \tag{33}$$

On the other hand, the leverage effect specified as $\langle X, \log\sigma\rangle$ takes the following form:

$$\frac{2d\langle X, \log\sigma\rangle}{dt} = 2f_t \tag{34}$$

Equations (33) and (34) suggest two ways of applying the estimation of leverage effects ($F(x) = x$ and $F(x) = \frac{1}{2}\log(x)$) to the estimation of the regression coefficient of the volatility process on its own log return process. The second method only involves lower orders of the volatility process, and is thus comparatively robust. We will use this method in the next section.

9 Empirical Study

In the empirical study, we employ Microsoft stock trades data from the New York Stock Exchange (NYSE TAQ). The years under study are 2008 through 2011. Even though the stock is traded between 9:30 am and 4:00pm, the window 9:45am-3:45pm is chosen in the empirical analysis. The reason for choosing this window is that a vast body of empirical studies documents increased return volatility and trading volume at the open and close of the stock market [Chan, Chockalingam, and Lai (2000) and Wood, McInish, and Ord (1985)]. A 15 minute cushion at the open and close may strike a good balance between avoiding abnormal trading activities in the market and preserving enough data points to perform the estimation procedures in a consistent way. On average, there are currently several hundred thousand trades of Microsoft during each trading day. There are frequently multiple trades in each second.

In section 8, we explored intra-day high frequency regression. It is clear that two forms of the leverage effect reveal the relation between volatility and return in the regression model. One way to extrapolate this intra-day behavior to between-day volatility prediction is to include the previous day's return but scaled by a time-varying leverage effect. Technically, all regressors are now in the drift term.

Since we are not trying to discover the best model calibration for volatility prediction, but rather to investigate the predictive power of return scaled by leverage effect, the prediction model is simply a linear regression (or AR(2)). Though this may not be a very sophisticated model, the results can still improve understanding the role of leverage effects in volatility prediction :

$$\int_{t_i}^{t_{i+1}} \sigma_t^2 dt = \alpha_0 + \alpha_1 \int_{t_{i-1}}^{t_i} \sigma_t dt + \alpha_2 \int_{t_{i-2}}^{t_{i-1}} \sigma_t dt + \alpha_3 \Delta X_{t_i -}^2 + \alpha_4 \int_{t_{i-1}}^{t_i} 2 f_t dt \times \Delta X_{t_i} + \epsilon_i \ .$$

- The integrated volatility $\int_{t_i}^{t_{i+1}} \sigma_t^2 dt$ can be estimated by various methods. In this empirical study, the pre-averaging method [Jacod, Li, Mykland, Podolskij, and Vetter (2009)] is adopted.
- $\Delta X_{t_{i-}}$ denotes the overnight log return.
- $\int_{t_{i-1}}^{t_i} 2 f_t dt$ can be estimated by the proposed leverage effect estimator in this paper when setting $F(x) = \frac{1}{2}\log(x)$.
- The inclusion of lagged volatilities and overnight returns is due to the empirical finding of volatility clustering.

In this study, since we do not consider the case with jumps involved, we first remove the days with jump activities applying the jump test from Lee and Mykland (2012).[11] Alternatively, one can also apply the jump tests as in Aït-Shalia and Jacod (2009), Barndorff-Nielsen and Shephard (2004), Barndorff-Nielsen and Shephard (2006), Mancini (2001), Podolskij and Ziggel (2010), and other work by the same authors.

[11] The total numbers of days removed are 21 for 2008, 27 for 2009, 60 for 2010, and 43 for 2011.

We also follow the convention by pre-averaging the data over every 5 minutes in the first step. We explore the volatility prediction over a 2-day period, a longer period than one day, because of the comparatively slower convergence rate of the estimators as discussed in Section 5. In order to check the robustness, we also repeat the regression replacing return scaled the leverage effect by return itself. The prediction results are shown in the Table 3. And the time series plot of the estimated leverage effect is give in Figure 1.

	2008			2009		
	P(> $\|t\|$)	SS explained	vif	P(> $\|t\|$)	SS explained	vif
α_0	0.003909			0.035226		
RV_{t-1}	$1.75 \cdot 10^{-7}$	$7.05 \cdot 10^{-5}*$	2.050335	0.000752	$1.7633 \cdot 10^{-5} * **$	5.603527
RV_{t-2}	0.369974	$1.05 \cdot 10^{-5}*$	1.461129	0.658205	$1.9960 \cdot 10^{-6}$	1.349047
R^2_{t-}	0.0493	$9.198 \cdot 10^{-6}$	1.142936	0.570849	7.09×10^{-7}	1.062036
RLE_{t-1}	0.000454	$2.0697 \cdot 10^{-5}*$	1.551306	0.027969	$4.827 \cdot 10^{-6}*$	4.855324
or R_{t-1}	0.821	$9.0 \cdot 10^{-8}*$	1.003209	0.280060	$1.1866 \cdot 10^{-6}$	1.198528
	2010			2011		
	P(> $\|t\|$)	SS explained	vif	P(> $\|t\|$)	SS explained	vif
α_0	0.000706			0.000747		
RV_{t-1}	0.833737	$1.18 \cdot 10^{-7}$	1.453209	$5.61 \cdot 10^{-5}$	$2.2341 \cdot 10^{-6} * **$	1.622136
R^2_{t-}	0.180635	$2.066 \cdot 10^{-7}$	1.008225	0.215383	$8.6345 \times 10^{-7}*$	1.178455
R^2_{t-}	$4.32 \cdot 10^{-10}$	$1.4117 \cdot 10^{-5} * **$	1.011785	0.270385	$1.7367 \cdot 10^{-7}$	1.034753
RLE_{t-1}	0.5254	$1.172 \cdot 10^{-7}$	1.455641	0.000486	$2.1891 \cdot 10^{-6} * **$	1.463627
or R_{t-1}	0.870061	$7.8 \cdot 10^{-9}$	1.068818	0.102010	$5.0273 \cdot 10^{-7}$	1.049329

Table 3: 2-day ahead volatility prediction results with Microsoft 2007-2010 data. RV_t denotes the estimated integrated volatility at day t; R_{t-} denotes overnight return for day t; RLE denotes the log return scaled by leverage effect at day t (estimated leverage effect $\times$ log return). R_{t-1} denotes the previous period return itself without scaling. "SS explained" denotes the sum of squares *gained* by adding each co-variate in the order presented in the table.

"SS explained" is the main indicator to show whether return scaled by leverage effect has a big contribution to the prediction of leverage effect. The first column of p-valued based on t-test is given just as reference information. Of course, since the co-variates are not independent from the response variable, this statistics cannot really tell whether the corresponding covariate should be included in the model. The third column gives the collinearity diagnostic by variance inflation factors (vif) (see, for example, Weisberg (2004)).

Since almost all vif values are close to 1, one can consider the covariates not to be collinear with each other. The sum of squares explained by the return scaled by the leverage effect (RLE) are

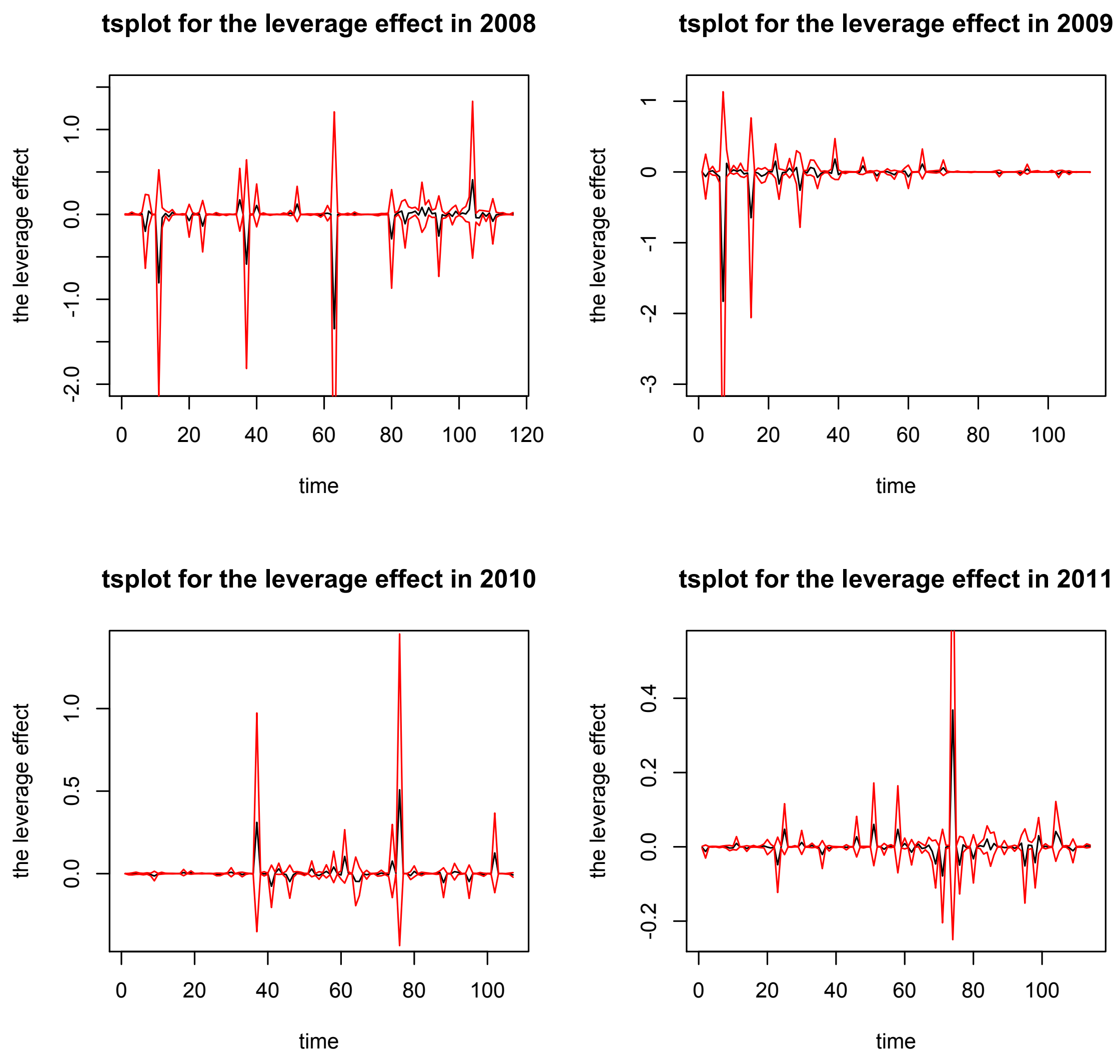

Figure 1: TS-plot of the estimated leverage effects: The black curves present the time series plots of the estimated leverage effects. The red curves are the 95% confidence intervals of the estimated values. The values on the vertical axes are different from one year to another. That is due to the different magnitudes of the estimated leverage effects. Apparently year 2008 and 2009 has the biggest negative leverage effects. This observation coincides with the empirical situation during the financial crisis.

substantial even when this term is included in the model last. Most of these sums of squares are comparable with the sum of squares explained by the previous day's volatility, which is believed to be the most significant factor for volatility prediction [Engle (1982), Bollerslev (1986)]. In all cases, RLE has stronger predictive power than the two-period lagged integrated volatility does. This strongly suggests the inclusion of return scaled by leverage effect into any model trying to predict next-period's volatility. The predictive power of a time-varying leverage effect estimator is consistent with earlier work by Engle and Ng (1993) and Chen and Ghysels (2011), but here appears in a new form. In addition, the previous period return does not contribute to the sum of squares as much as the one scaled by the leverage effect. In some case, the previous period return is not significant while the return scaled by the leverage effect explains significant amount of sum of squares.

10 Conclusion

This paper provides nonparametric estimators of the leverage effect, and analyzes them both theoretically and in simulation. The definition of the stochastic parameter of the leverage effect involves a twice differentiable monotone function. Even though the reliance of the estimation on higher moments of volatility is of concern in practice, the carefully chosen function F can help to reduce the order of moments required and provide robust results. Other benefits of this function can be easily seen in the discussion of the connection between the leverage effect and skewness. While the sum of intra-day cubic returns is not a consistent estimator of skewness, the p-skewness (Section 7) can instead be estimated consistently by the leverage effect estimators proposed in this paper. The related properties of p-skewness can also be studied by the results provided here. Clever choices of the function F can also reduce the work of calculation, such as the estimation of high-frequency regression coefficients. Instead of estimating both leverage effects and also realized volatility, a different form of the leverage effect can serve as the estimated coefficient (Section 8). If the properties of the estimated coefficient are of interest, it is more attractive to apply the method in equation (34), whose statistical properties have already been studied in this paper, than to apply the first ratio statistic in equation (33) whose statistical properties require further efforts to investigate.

The bias correction factors in the estimators contribute to an important finding in this paper and are previously unknown. They not only provide the consistency of the estimation, but also imply that simple covariance estimators tend to underestimate the leverage effect, especially when the values of the leverage effect are close to zero. The amplifying factors play a vital role of bias correction in the estimation.

The empirical studies demonstrate the importance of the leverage effect in volatility prediction. Even though the simple regression (or AR(2)) model is adopted in the study, the explanatory power of RLE is surprisingly high. It is almost of the same magnitude as the predictive power of the current period volatility which is widely considered to be the main source of variation in volatility prediction. This suggests that time-varying leverage effects should be included additionally to explain the variation and clustering in volatility prediction models.

Even though we have provided a way to deal with irregularly spaced data, it is important to study the estimation of leverage effects and the asymptotic properties of estimators when time is endogenous (as in Li, Mykland, Renault, Zhang, and Zheng (2009)). Different methods of dealing with microstructure noise should also be studied and compared with the ones in this paper. Our findings create the important necessary foundation for further analysis both theoretically and empirically, as well as an investigation of how to carry out risk management in the presence of leverage effects.

Finally, as discussed in Section 2.2, many open questions remain in terms of model specification (continuous vs. jumps, semimartingale vs. long range dependence), with reference to the papers cited in that section. In particular, the connection to semivariance remains to be explored.

REFERENCES

Aït-Sahalia, Y., J. Fan, and Y. Li (2011): "The Leverage Effect Puzzle: Disentangling Sources of Bias at High Frequency," Working Paper.

Aït-Shalia, Y. and J. Jacod (2009): "Testing for jumps in a discretely observed process," *Annals of Statistics*, 37, 184–222.

Andersen, T. G. and T. Bollerslev (1998): "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review*, 39, 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001): "The Distribution of Realized Stock Return Volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2000): "Great realizations," *Risk*, 13, 105–108.

Ball, C. A. and A. Roma (1994): "Stochastic Volatility Option Pricing," *The Journal of Financial and Quantitative Analysis*, 29, pp. 589–607.

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

Barndorff-Nielsen, O. and A. Veraart (2009): "Stochastic volatility of volatility in continuous time," CREATES Research Papers.

Barndorff-Nielsen, O. E., P. R. Hansen, A. Lunde, and N. Shephard (2008a): "Designing realized kernels to measure ex-post variation of equity prices in the presence of noise," *Econometrica*, 76, 1481–1536.

Barndorff-Nielsen, O. E., S. Kinnebrock, and N. Shehpard (2008b): "Measuring downside risk - realised semivariance," Dept. of Economics Working Paper 382, University of Oxford.

Barndorff-Nielsen, O. E. and Shephard (2002): "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal Of The Royal Statistical Society Series B*, 64, 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2004): "Power and bipower variation with stochastic volatility and jumps (with discussion)," *Journal of Financial Econometrics*, 2, 1–48.

——— (2006): "Econometrics of testing for jumps in financial economics using bipower variation," *Journal of Financial Econometrics*, 4, 1–30.

Bekaert, G. and G. Wu (1997): "Asymmetric Volatility and Risk in Equity Markets," NBER Working Papers 6022, National Bureau of Economic Research, Inc.

Black, B. (1976): "Studies of Stock Price Volatility Changes," *Proceedings of the 176 Meetings of the American Statistical Association, Business and Economic Statistics*, 177–181.

Bollerslev, T. (1986): "Generalized Autorgeressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

BOLLERSLEV, T., J. LITVINOVA, AND G. TAUCHEN (2006): "Leverage and Volatility Feedback Effects in High-Frequency Data," *Journal of Financial Econometrics*, 4, 353–384.

BOUCHAUD, J.-P., A. MATACZ, AND M. POTTERS (2001): "The leverage effect in financial markets: retarded volatility and market panic," Science & Finance (CFM) working paper archive 0101120, Science & Finance, Capital Fund Management.

BROCKWELL, P. J. AND T. MARQUARDT (2005): "Lévy-driven and fractionally integrated ARMA processes with continuous time parameter," *Statist. Sinica*, 477–494.

CAMPBELL, J. Y. AND L. HENTSCHEL (1991): "No News is Good News: An Asymmetric Model of Changing Volatility in Stock Returns," NBER Working Papers 3742, National Bureau of Economic Research, Inc.

CHAN, K., M. CHOCKALINGAM, AND K. W. L. LAI (2000): "Overnight information and intraday trading behavior: evidence from NYSE cross-listed stocks and their local market information," *Journal of Multinational Financial Management*, 10, 495–509.

CHEN, X. AND E. GHYSELS (2011): "News - good or bad - and the impact on volatility predictions over multiple horizons," *Review of Financial Studies*, 24, 46–81.

CHRISTIE, A. A. (1982): "The stochastic behavior of common stock variances : Value, leverage and interest rate effects," *Journal of Financial Economics*, 10, 407–432.

COMTE, F., L. COUTIN, AND E. RENAULT (2010): "Affine fractional stochastic volatility models," *Recherche*, 13, 1–35.

COMTE, F. AND E. RENAULT (1998): "Long memory in continuous-time stochastic volatility models," *Mathematical Finance*, 8, 291–323.

DELBAEN, F. AND W. SCHACHERMAYER (1994): "A General Version of the Fundamental Theorem of Asset Pricing," *Mathematische Annalen*, 300, 463–520.

——— (1995): "The existence of absolutely continuous local martingale measures," *Annals of Applied Probability*, 5, 926–945.

——— (1998): "The Fundamental Theorem of Asset Pricing For Unbounded Stochastic Processes," *Mathematische Annalen*, 312, 215–250.

ENGLE, R. F. (1982): "Autogregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation," *Econometrica*, 50, 987–1008.

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

ENGLE, R. F. AND V. K. NG (1993): "Measuring and testing the impact of news on volatility," *Journal of Finance*, 48, 1749–1778.

FAMA, E. F. (1965): "The Behavior of Stock-Market Prices," *The Journal of Business*, 38, 34–105.

FIGLEWSKI, S. AND X. WANG (2001): "Is the 'Leverage Effect' a Leverage Effect?" *Review of Financial Studies*, 24, 46–81.

FRENCH, K. R., G. W. SCHWERT, AND R. F. STAMBAUGH (1987): "Expected stock returns and volatility," *Journal of Financial Economics*, 19, 3–29.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993): "On the relation between the expected value and the volatility of the nominal excess return on stocks," *Journal of Finance*, 48, 1779–1801.

Gloter, A. and M. Hoffmann (2004): "Stochastic volatility and fractional Brownian motion," *Stochastic Processes and their Applications*, 113, 143–172.

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

Harris, L. (1990): "Statistical Properties of the Roll Serial Covariance Bid/Ask Spread Estimator," *Journal of Finance*, 45, 579–590.

Hasanhodzic, J. and A. W. Lo (2011): "Black's Leverage Effect Is Not Due To Leverage," Working Paper.

Hasbrouck, J. (1996): "Modeling Market Microstructure Time Series," in *Handbook of Statistics, volume 14*, ed. by C. R. Rao and G. S. Maddala, Amsterdam: North-Holland, 647–692.

Heath, D. (1977): "Interpolation of martingales," *Annals of Probability*, 5, 804–806.

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

Hull, J. and A. White (1987): "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance*, 42, 281–300.

Jacad, J. (2009): "On continuous conditional Gaussian martingales and stable convergence in law," *Seminaire de Probabilites*, 232–246.

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

Jacod, J., Y. Li, P. A. Mykland, M. Podolskij, and M. Vetter (2009): "Microstructure Noise in the Continuous Case: The Pre-Averaging Approach," *Stochastic Processes and Their Applications*, 119, 2249–2276.

Jacod, J. and P. Protter (2011): *Discretization of processes*, Springer Heidelberg Dordrecht London New York.

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

Kim, D. and S. J. Kon (1994): "Alternative Models for the Conditional Heteroscedasticity of Stock Returns," *Journal of Business*, 67, 563–98.

Kristensen, D. (2010): "Nonparametric Filtering of the Realized Spot Volatility: A Kernel-based Approach," *Econometric Theory*, 26.

Lee, S. and P. A. Mykland (2012): "Jumps in Equilibrium Prices and Market Microstructure Noise," *Journal of Econometrics*, 168, 396–406.

Li, Y., P. Mykland, E. Renault, L. Zhang, and X. Zheng (2009): "Realized Volatility when Endogeniety of Time Matters," Working Paper.

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

Mandelbrot, B. (1963): "The Variation of Certain Speculative Prices," *Journal of Business*, 36, 394.

Meddahi, N. and E. Renault (2004): "Temporal aggregation and volatility models," *Journal of Econometrics*, 119, 355–379.

Mykland, P. A. (1994): "Bartlett type identities for martingales," *Annals of Statistics*, 22, 21–38.

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

Mykland, P. A. and L. Zhang (2009): "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1455.

——— (2011a): "Between Data Cleaning and Inference: Pre-Averaging and other Robust Estimators of the Efficient Price," Working Paper, University of Illinois at Chicago and University of Chicago.

——— (2011b): "The Double Gaussian Approximation for High Frequency Data," *Scandinavian Journal of Statistics*, 38, 215–236.

——— (2012): "The Econometrics of High Frequency Data," in *Statistical Methods for Stochastic Differential Equations*, ed. by M. Kessler, A. Lindner, and M. Sørensen, Chapman and Hall/CRC Press.

Nelson, D. B. (1991): "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica*, 59, 347–70.

Neuberger, A. (2011): "Realized Skewness," Working Paper.

Nualart, D. (2006): *Fractional Brownian motion: stochastic calculus and applications*, Eur. Math. Soc., Zürich.

Officer, R. R. (1973): "The Variability of the Market Factor of the New York Stock Exchange," *The Journal of Business*, 46, 434–453.

O'Hara, M. (1995): *Market Microstructure Theorys*, Cambridge, MA: Blackwell Publishers.

Pindyck, R. S. (1984): "Risk, Inflation, and the Stock Market," *American Economic Review*, 74, 335–51.

Podolskij, M. and M. Vetter (2009a): "Bipower-type estimation in a noisy diffusion setting," *Stochastic Processes and Their Applications*, 119, 2803–2831.

——— (2009b): "Understanding limit theorems for semimartingales: a short survey," *Statistica Neerlandica*, 64, 329351.

Podolskij, M. and D. Ziggel (2010): "New tests for jumps in semimartingale models," *Statistical Inference for Stochastic Processes*, 13, 15–41.

Reiss, M. (2010): "Asymptotic equivalence and sufficiency for volatility estimation under microstructure noise," Xiv:1001.3006.

Renault, E. and B. J. Werker (2009): "Causality effects in return volatility measures with random times," *Journal of Econometrics* (forthcoming).

Roll, R. (1984): "A Simple Model of the Implicit Bid-Ask Spread in an Efficient Market," *Journal of Finance*, 39, 1127–1139.

Stein, E. M. and J. C. Stein (1991): "Stock Price Distributions with Stochastic Volatility: An Analytic Approach," *Review of Financial Studies*, 4, 727–752.

Tauchen, G. and L. M. Zhang (1996): "Volume, volatility, and leverage: A dynamic analysis," *Journal of Econometrics*, 74, 177–208.

Weisberg, S. (2004): *Applied Linear Regression*, Wiley/Interscience, third ed.

Wood, R. A., T. H. McInish, and J. K. Ord (1985): "An Investigation of Transactions Data for NYSE Stocks," *Journal of Finance*, 40, 723–39.

Wu, G. (2001): "The Determinants of Asymmetric Volatility," *Review of Financial Studies*, 14, 837–59.

Wu, G. and Z. Xiao (2002): "A generalized partially linear model of asymmetric volatility," *Journal of Empirical Finance*, 9, 287–319.

Xiu, D. (2010): "Quasi-Maximum Likelihood Estimation of Volatility WIth High Frequency Data," *Journal of Econometrics*, 159, 235–250.

Zakoian, J.-M. (1994): "Threshold heteroskedastic models," *Journal of Economic Dynamics and Control*, 18, 931–955.

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

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

APPENDIX: PROOFS

A Proof of Theorem 1

A.1 Preliminaries

In the following, by $\mathcal{F}_j$ we mean $\mathcal{F}_{t_{n,j}}$, and p is a positive integer. Without loss of generality, we will set $\mu_t = 0$, cf. Mykland and Zhang (2009), Section 2.2, as well as our current Assumption 1 and associated Footnote 1. Recall that

$$\langle \widehat{X, F(\sigma^2)}\rangle_T = 2\sum_i (X_{\tau_{n,i+1}} - X_{\tau_{n,i}})(F(\hat\sigma^2_{\tau_{n,i+1}}) - F(\hat\sigma^2_{\tau_{n,i}})) \tag{A.1}$$

$$\text{and } \hat\sigma^2_{\tau_{n,i+1}} = \frac{1}{M\times\Delta t}\sum_{t_{n,j+1}\in(\tau_{n,i},\tau_{n,i+1}]} \Delta X^2_{t_{n,j+1}}. \tag{A.2}$$

LEMMA 1. *, under assumption 1,*

$$\begin{aligned}E((X_{t_{n,k}} - X_{t_{n,j}})^2\sigma^p_{t_{n,m}}|\mathcal{F}_j) &= \sigma^{p+2}_{t_{n,j}}(t_{n,k}-t_{n,j}) + p^2 f^2_{t_{n,j}}\sigma^p_{t_{n,j}}(t_{n,k}-t_{n,j})^2 + p\sigma^p_{t_{n,j}}\langle X,f\rangle_{t_{n,j}}(t_{n,k}-t_{n,j})^2\\ &+ \frac{1}{2}p(p-1)(f^2_{t_{n,j}} + g^2_{t_{n,j}})\sigma^p_{t_{n,j}}(t_{n,k}-t_{n,j})(t_{n,m}-t_{n,j})\\ &+ (p+\frac{1}{2})(f^2_{t_{n,j}} + g^2_{t_{n,j}})\sigma^p_{t_{n,j}}(t_{n,k}-t_{n,j})^2 + a_{t_{n,j}}\sigma^{p+1}_{t_{n,j}}(t_{n,k}-t_{n,j})^2\\ &+ pa_{t_{n,j}}\sigma^{p+1}_{t_{n,j}}(t_{n,k}-t_{n,j})(t_{n,m}-t_{n,j}) + O_p(\Delta t^{5/2}),\end{aligned} \tag{A.3}$$

$$E((X_{t_{n,k}} - X_{t_{n,j}})\sigma^p_{t_{n,m}}|\mathcal{F}_j) = p\sigma^p_{t_{n,j}} f_{t_{n,j}}(t_{n,k}-t_{n,j}) + O_p(\Delta t^{3/2}), \text{ and} \tag{A.4}$$

$$\begin{aligned}E(2(X_{t_{n,j}} - X_{t_{n,i}})\sigma^2_{t_{n,j+1}}(X_{t_{n,k}} - X_{t_{n,j}})\sigma^2_{t_{n,k+1}}|\mathcal{F}_i) &= 16\sigma^4_{t_{n,i}} f^2_{t_{n,i}}(t_{n,k}-t_{n,j+1})(t_{n,j}-t_{n,i})\\ &+ 4\sigma^4_{t_{n,i}}\langle X,f\rangle_{t_{n,i}}(t_{n,k}-t_{n,j+1})(t_{n,j}-t_{n,i}) + O_p(\Delta t^{5/2}),\end{aligned} \tag{A.5}$$

where $t_{n,i} < t_{n,j} < t_{n,k} \le t_{n,m}$, $t_{n,m} - t_{n,j} = O_p(\Delta t^{\frac{1}{2}})$ *and* $\Delta t = \frac{T}{n}$.

Proof of Lemma 1:

1. For (A.3), note first that

$$\begin{aligned}\sigma^p_{t_{n,m}} - \sigma^p_{t_{n,j}} &= p\int_{t_{n,j}}^{t_{n,m}}\sigma_t^{p-1}d\sigma_t + \frac{p(p-1)}{2}\int_{t_{n,j}}^{t_{n,m}}\sigma_t^{p-2}(f_t^2+g_t^2)dt, \text{ and}\\ (X_{t_{n,k}} - X_{t_{n,j}})^2 &= \int_{t_{n,j}}^{t_{n,k}} 2(X_t - X_{t_{n,j}})dX_t + \int_{t_{n,j}}^{t_{n,k}}\sigma_t^2 dt\\ &= \int_{t_{n,j}}^{t_{n,k}} 2(X_t - X_{t_{n,j}})dX_t + \int_{t_{n,j}}^{t_{n,k}}(t_{n,k}-t)d(\sigma_t^2 - \sigma^2_{t_{n,j}}) + \sigma^2_{t_{n,j}}(t_{n,k}-t_{n,j})\\ &= \int_{t_{n,j}}^{t_{n,k}} 2(X_t - X_{t_{n,j}})dX_t + \int_{t_{n,j}}^{t_{n,k}}(t_{n,k}-t)2\sigma_t d\sigma_t + \int_{t_{n,j}}^{t_{n,k}}(t_{n,k}-t)(f_t^2+g_t^2)dt\\ &\quad+\sigma^2_{t_{n,j}}(t_{n,k}-t_{n,j}).\end{aligned}$$

Hence,

$$
\begin{aligned}
& E((X_{t_{n,k}} - X_{t_{n,j}})^2 \sigma^p_{t_{n,m}} | \mathcal{F}_j) \\
= \; & E([\sigma^p_{t_{n,j}} + \frac{p(p-1)}{2} \int_{t_{n,j}}^{t_{n,m}} \sigma_t^{p-2}(f_t^2 + g_t^2)dt][\int_{t_{n,j}}^{t_{n,k}} 2(X_t - X_{t_{n,j}})dX_t + \sigma^2_{t_{n,j}}(t_{n,k} - t_{n,j}) \\
& + \int_{t_{n,j}}^{t_{n,k}} (t_{n,k} - t) 2\sigma_t a_t dt + \int_{t_{n,j}}^{t_{n,k}} (t_{n,k} - t)(f_t^2 + g_t^2)dt] + p \int_{t_{n,j}}^{t_{n,k}} 2(X_t - X_{t_{n,j}})\sigma_t^{p-1} d\langle X, \sigma\rangle_t \\
& + p \int_{t_{n,j}}^{t_{n,k}} 2(t_{n,k} - t)\sigma_t^p d\langle \sigma, \sigma\rangle_t + \sigma^2_{t_{n,j}}(t_{n,k} - t_{n,j}) p \int_{t_{n,j}}^{t_{n,m}} \sigma_t^{p-1} a_t dt | \mathcal{F}_j) + O_p(\Delta t^{5/2}) \\
= \; & E(\sigma^{p+2}_{t_{n,j}}(t_{n,k} - t_{n,j}) + \sigma^{p+1}_{t_{n,j}} a_{t_{n,j}}(t_{n,k} - t_{n,j})^2 + \frac{1}{2}(f^2_{t_{n,j}} + g^2_{t_{n,j}})\sigma^p_{t_{n,j}}(t_{n,k} - t_{n,j})^2 \\
& + \frac{p(p-1)}{2}(f^2_{t_{n,j}} + g^2_{t_{n,j}})\sigma^p_{t_{n,j}}(t_{n,k} - t_{n,j})(t_{n,m} - t_{n,j}) + p f_{t_{n,j}} \int_{t_{n,j}}^{t_{n,k}} 2(X_t - X_{t_{n,j}})\sigma_t^p dt \\
& + p\sigma^p_{t_{n,j}} \int_{t_{n,j}}^{t_{n,m}} 2(X_t - X_{t_{n,j}})(f_t - f_{t_{n,j}})dt + p \int_{t_{n,j}}^{t_{n,k}} 2(t_{n,k} - t)\sigma_t^p (f_t^2 + g_t^2)dt \\
& + p\sigma^{p+1}_{t_{n,j}} a_{t_{n,j}}(t_{n,k} - t_{n,j})(t_{n,m} - t_{n,j}) | \mathcal{F}_j) + O_p(\Delta t^{5/2}) \\
= \; & E(\sigma^{p+2}_{t_{n,j}}(t_{n,k} - t_{n,j}) + \sigma^{p+1}_{t_{n,j}} a_{t_{n,j}}(t_{n,k} - t_{n,j})^2 + (\frac{1}{2} + p)(f^2_{t_{n,j}} + g^2_{t_{n,j}})\sigma^p_{t_{n,j}}(t_{n,k} - t_{n,j})^2 \\
& + \frac{p(p-1)}{2}(f^2_{t_{n,j}} + g^2_{t_{n,j}})\sigma^p_{t_{n,j}}(t_{n,k} - t_{n,j})(t_{n,m} - t_{n,j}) + p^2 f_{t_{n,j}} \int_{t_{n,j}}^{t_{n,k}} (t_{n,k} - t) 2\sigma_t^p f_t dt \\
& + p\sigma^p_{t_{n,j}} \langle X, f\rangle_{t_{n,j}}(t_{n,k} - t_{n,j})^2 + p\sigma^{p+1}_{t_{n,j}} a_{t_{n,j}}(t_{n,k} - t_{n,j})(t_{n,m} - t_{n,j}) | \mathcal{F}_j) + O_p(\Delta t^{5/2}) \\
= \; & \sigma^{p+2}_{t_{n,j}}(t_{n,k} - t_{n,j}) + a_{t_{n,j}}\sigma^{p+1}_{t_{n,j}}(t_{n,k} - t_{n,j})^2 + (p + \frac{1}{2})(f^2_{t_{n,j}} + g^2_{t_{n,j}})\sigma^p_{t_{n,j}}(t_{n,k} - t_{n,j})^2 \\
& + \frac{p(p-1)}{2}(f^2_{t_{n,j}} + g^2_{t_{n,j}})\sigma^p_{t_{n,j}}(t_{n,k} - t_{n,j})(t_{n,m} - t_{n,j}) + p^2 f^2_{t_{n,j}} \sigma^p_{t_{n,j}}(t_{n,k} - t_{n,j})^2 \\
& + p\sigma^p_{t_{n,j}} \langle X, f\rangle_{t_{n,j}}(t_{n,k} - t_{n,j})^2 + p a_{t_{n,j}}\sigma^{p+1}_{t_{n,j}}(t_{n,k} - t_{n,j})(t_{n,m} - t_{n,j}) + O_p(\Delta t^{5/2}).
\end{aligned}
$$

2. For (A.4):

$$
\begin{aligned}
E((X_{t_{n,k}} - X_{t_{n,j}})\sigma^p_{t_{n,m}} | \mathcal{F}_j) &= E(p \int_{t_{n,j}}^{t_{n,k}} \sigma_t^{p-1} d\langle X, \sigma\rangle_t | \mathcal{F}_j) + O_p(\Delta t^{3/2}) \\
&= E(p \int_{t_{n,j}}^{t_{n,k}} \sigma_t^p f_t dt | \mathcal{F}_j) + O_p(\Delta t^{3/2}) \\
&= p\sigma^p_{t_{n,j}} f_{t_{n,j}}(t_{n,k} - t_{n,j}) + O_p(\Delta t^{3/2}).
\end{aligned}
$$

3. (A.5) is the direct consequence of (A.3) and (A.4). This completes the proof.

Later in the derivation of limit theorem, it is easy to see that the proof and calculations strongly depend on Lemma 1, which will be applied recursively.

A.2 Main martingale representation and argument to prove theorem.

The proof of Theorem 1 is now described in this subsection, with supporting details in later subsections. – Construct an approximate martingale (MG) on the grid of the $\tau_{n,i}$'s as follows:

$$
\begin{aligned}
&\frac{1}{\sqrt{M\Delta t}}\langle X, \widehat{F(\sigma^2)}\rangle\rangle_t \\
&= \frac{2}{\sqrt{M\Delta t}}\{\sum_{i=0,\tau_{n,i+1}\le t}^{K_n-2}(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(F(\hat\sigma^2_{\tau_{n,i+1}})-F(\hat\sigma^2_{\tau_{n,i}}))\} \\
&= \frac{2}{\sqrt{M\Delta t}}\{\sum_{i=0,\tau_{n,i+1}\le t}^{K_n-2}(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(F(\hat\sigma^2_{\tau_{n,i+1}})-F(\sigma^2_{\tau_{n,i+1}})+F(\sigma^2_{\tau_{n,i+1}})-F(\sigma^2_{\tau_{n,i}})+F(\sigma^2_{\tau_{n,i}})-F(\hat\sigma^2_{\tau_{n,i}}))\} \\
&= \frac{2}{\sqrt{M\Delta t}}\{\sum_{i=0,\tau_{n,i+1}\le t}^{K_n-2}(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(F(\sigma^2_{\tau_{n,i+1}})-F(\sigma^2_{\tau_{n,i}}))+\sum_{i=0,\tau_{n,i+1}\le t}^{K_n-2}(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(F(\hat\sigma^2_{\tau_{n,i+1}})-F(\sigma^2_{\tau_{n,i+1}})) \\
&-\sum_{i=0,\tau_{n,i+1}\le t}^{K_n-2}(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(F(\hat\sigma^2_{\tau_{n,i}})-F(\sigma^2_{\tau_{n,i}})\} \\
&= \frac{2}{\sqrt{M\Delta t}}\{\sum_{i=0,\tau_{n,i+1}\le t}^{K_n-2}(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(F(\sigma^2_{\tau_{n,i+1}})-F(\sigma^2_{\tau_{n,i}}))+\sum_{i=1,\tau_{n,i+1}\le t}^{K_n-1}(X_{\tau_{n,i}}-X_{\tau_{n,i-1}})(F(\hat\sigma^2_{\tau_{n,i}})-F(\sigma^2_{\tau_{n,i}})) \\
&-\sum_{i=0,\tau_{n,i+1}\le t}^{K_n-2}(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(F(\hat\sigma^2_{\tau_{n,i}})-F(\sigma^2_{\tau_{n,i}}))\} \\
&= \frac{2}{\sqrt{M\Delta t}}\{\underbrace{\sum_{i=0,\tau_{n,i+1}\le t}^{K_n-2}(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(\sigma^2_{\tau_{n,i+1}}-\sigma^2_{\tau_{n,i}})F'(\sigma^2_{\tau_{n,i}})}_{(1)}+\underbrace{\sum_{i=1,\tau_{n,i+1}\le t}^{K_n-1}(X_{\tau_{n,i}}-X_{\tau_{n,i-1}})(\hat\sigma^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})F'(\sigma^2_{\tau_{n,i}})}_{(2)} \\
&\underbrace{-\sum_{i=0,\tau_{n,i+1}\le t}^{K_n-2}(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(\hat\sigma^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})F'(\sigma^2_{\tau_{n,i}})\}}_{(3)}+o_p(1).
\end{aligned}
$$

It follows that

$$
\frac{1}{\sqrt{M\Delta t}}\left(\langle X, \widehat{F(\sigma^2)}\rangle\rangle_t - \langle X, F(\sigma^2)\rangle\rangle_t\right) = \sum_{i=0,\tau_{n,i+1}\le t}\Delta V^n_{\tau_{n,i+1}} + o_p(1),
$$

where, except for a few terms at the edge,

$$
\begin{aligned}
\Delta V^n_{\tau_{n,i+1}} = \frac{2}{\sqrt{M\Delta t}}\{&\underbrace{(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(\sigma^2_{\tau_{n,i+1}}-\sigma^2_{\tau_{n,i}})F'(\sigma^2_{\tau_{n,i}})}_{(1)}+\underbrace{(X_{\tau_{n,i}}-X_{\tau_{n,i-1}})(\hat\sigma^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})F'(\sigma^2_{\tau_{n,i}})}_{(2)} \\
&\underbrace{-(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(\hat\sigma^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})F'(\sigma^2_{\tau_{n,i}})}_{(3)}-\underbrace{\int_{\tau_i}^{\tau_{i+1}}F'(\sigma^2_t)\sigma^2_t f_t\,dt}_{(4)}.
\end{aligned}
$$

And so the martingale increment becomes

$$
\Delta M^n_{\tau_{i+1}} = \Delta V^n_{\tau_{n,i+1}} - E(\Delta V^n_{\tau_{n,i+1}}|\mathcal{F}_i). \tag{A.6}
$$

The martingale up to time t is

$$M_t^n = \sum_{\tau_{n,i+1} \le t} \{\Delta V_{\tau_{n,i+1}}^n - E(\Delta V_{\tau_{n,i+1}}^n | \mathcal{F}_i)\}. \tag{A.7}$$

Although M^n above is observed in discrete time, we can interpolate the martingale into a continuous martingale up to any time t, see, in particular, Heath (1977), as well as Mykland (1995) and the references therein. This interpolation is closely related to Skorokhod embedding in Brownian motion, see, for example, Appendix I of Hall and Heyde (1980). Then we only need to prove the CLT for the interpolated continuous martingale. Therefore we can apply Theorem 2.28 (p. 152) in Mykland and Zhang (2012) to prove the CLT. The conditions of cited theorem will follow from the development in the rest of Appendix A, in particular (A.16) and (A.19), and the approximation in Lemma 2. Alternatively, one can develop a functional argument for the CLT as in Jacad (2009), Jacod and Shiryaev (2003), Jacod and Protter (2011) and Podolskij and Vetter (2009b).

A.3 The aggregate conditional variance

In order to calculate the quadratic variation, we will calculate the aggregate conditional variance of M_t^n

$$\begin{aligned} \sum_{\tau_{i+1} \le t} \mathrm{Var}(\Delta M_{\tau_{i+1}}^n | \mathcal{F}_i) &= \sum_{\tau_{i+1} \le t} E((\Delta M_{\tau_{i+1}}^n)^2 | \mathcal{F}_i) \\ &= \sum_{\tau_{i+1} \le t} \{E((\Delta V_{\tau_{n,i+1}}^n)^2 | \mathcal{F}_i) - (E(\Delta V_{\tau_{n,i+1}}^n | \mathcal{F}_i))^2\}. \end{aligned} \tag{A.8}$$

We will first prove that $\sum_{\tau_{i+1} \le t} (E(\Delta V_{\tau_{n,i+1}}^n | \mathcal{F}_i))^2$ is negligible, of order $O_p(M^2 \Delta t^2)$. Except a few terms at the edge, the following could be established.

$$\begin{aligned} & E(\Delta V_{\tau_{n,i+1}}^n | \mathcal{F}_i) \\ &= \frac{2}{\sqrt{M \Delta t}} E((X_{\tau_{n,i+1}} - X_{\tau_{n,i}})(\sigma_{\tau_{n,i+1}}^2 - \sigma_{\tau_{n,i}}^2) F'(\sigma_{\tau_i}^2) + (X_{\tau_{n,i}} - X_{\tau_{n,i-1}})(\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2) F'(\sigma_{\tau_i}^2) \\ & \quad - (X_{\tau_{n,i+1}} - X_{\tau_{n,i}})(\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2) F'(\sigma_{\tau_i}^2) | \mathcal{F}_i) \\ &= \frac{2F'(\sigma_{\tau_{n,i}}^2)}{\sqrt{M \Delta t}} E\left(\int_{\tau_i}^{\tau_{i+1}} \sigma_t^2 f_t \, dt - (X_{\tau_{n,i+1}} - X_{\tau_{n,i}})(\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2) + (X_{\tau_{n,i}} - X_{\tau_{n,i-1}})(\hat{\sigma}_{\tau_{n,i}}^2 - \sigma_{\tau_{n,i}}^2) | \mathcal{F}_i \right) \\ & \qquad + O_p(M_n \Delta t) \\ &= \frac{2F'(\sigma_{\tau_{n,i}}^2)}{\sqrt{M \Delta t}} E\left(\int_{\tau_i}^{\tau_{i+1}} (\sigma_t^2 f_t - \sigma_{\tau_{n,i}}^2 f_{\tau_{n,i}}) \, dt - \sum_{j=1}^{M} \int_{\tau_{n,i}}^{t_{n,j}} 2(\sigma_t^2 f_t - \sigma_{\tau_{n,i}}^2 f_{\tau_{n,i}}) \, dt | \mathcal{F}_i \right) \\ & \quad + \frac{2F'(\sigma_{\tau_{n,i}}^2)}{\sqrt{M \Delta t}} (X_{\tau_{n,i}} - X_{\tau_{n,i-1}}) E\left(\int_{\tau_{n,i}}^{\tau_{n,i+1}} \sigma_t^2 - \sigma_{\tau_{n,i}}^2 \, dt | \mathcal{F}_i \right) + O_p(M_n \Delta t). \end{aligned} \tag{A.9}$$

We derive that

$$
\begin{aligned}
&E\left(\sum_{\tau_{i+1}\le t}(E(\Delta V^n_{\tau_{n,i+1}}|\mathcal{F}_i))^2\right)\\
&\le \frac{12F'(\sigma^2_{\tau_i})^2}{M\Delta t}\sum_{\tau_{i+1}\le t}E\{(E\int_{\tau_i}^{\tau_{i+1}}(\sigma_t^2 f_t-\sigma^2_{\tau_{n,i}}f_{\tau_{n,i}})\,dt|\mathcal{F}_i))^2+(\sum_{j=1}^{M}E(\int_{\tau_{n,i}}^{t_{n,j}}2(\sigma_t^2 f_t-\sigma^2_{\tau_{n,i}}f_{\tau_{n,i}})\,dt|\mathcal{F}_i))^2\\
&\quad+(X_{\tau_{n,i}}-X_{\tau_{n,i-1}})^2\left(E(\int_{\tau_{n,i}}^{\tau_{n,i+1}}\sigma_t^2-\sigma^2_{\tau_{n,i}}\,dt|\mathcal{F}_i)\right)^2\}+O_p((M_n\Delta t)^2)\\
&\le K(M\Delta t)^2. \qquad\qquad (A.10)
\end{aligned}
$$

With this result, we can calculate the conditional variance by only considering conditional second moments instead.

A.4 Aggregate conditional second moment

$$
\begin{aligned}
&\sum_{\tau_{i+1}\le t}E((\Delta V^n_{\tau_{i+1}})^2|\mathcal{F}_i)\\
=&\frac{4}{M\Delta t}\sum_{\tau_{i+1}\le t}E([(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(\sigma^2_{\tau_{n,i+1}}-\sigma^2_{\tau_{n,i}})F'(\sigma^2_{\tau_i})+(X_{\tau_{n,i}}-X_{\tau_{n,i-1}})(\hat\sigma^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})F'(\sigma^2_{\tau_i})\\
&\quad-(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(\hat\sigma^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})F'(\sigma^2_{\tau_i})-\int_{\tau_i}^{\tau_{i+1}}F'(\sigma_t^2)\sigma_t^2 f_t\,dt]^2|\mathcal{F}_i)\\
=&\frac{4F'(\sigma^2_{\tau_i})^2}{M\Delta t}\sum_{\tau_{i+1}\le t}E((X_{\tau_{n,i+1}}-X_{\tau_{n,i}})^2(\sigma^2_{\tau_{n,i+1}}-\sigma^2_{\tau_{n,i}})^2+(X_{\tau_{n,i}}-X_{\tau_{n,i-1}})^2(\hat\sigma^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})^2\\
&\quad+(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})^2(\hat\sigma^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})^2-2(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})^2(\sigma^2_{\tau_{n,i+1}}-\sigma^2_{\tau_{n,i}})(\hat\sigma^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})\\
&\quad+(\int_{\tau_i}^{\tau_{i+1}}\sigma_t^2 f_t\,dt)^2-2(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(\sigma^2_{\tau_{n,i+1}}-\sigma^2_{\tau_{n,i}})\int_{\tau_i}^{\tau_{i+1}}\sigma_t^2 f_t\,dt|\mathcal{F}_i)+O_p((M_n\Delta t)^2).
\end{aligned}
\qquad (A.11)
$$

Applying Lemma (1), we can prove the following results

$$
\begin{aligned}
&\sum_{\tau_{i+1}\le t}E((X_{\tau_{n,i}}-X_{\tau_{n,i-1}})^2(\hat\sigma^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})^2|\mathcal{F}_i)\\
=&\frac{2}{M_n^2\Delta t}\sum_{\tau_{i+1}\le t}\sigma^6_{\tau_{n,i}}M_n\Delta t+\frac{4}{3}\sum_{\tau_{i+1}\le t}\sigma^4_{\tau_{n,i}}(f^2_{\tau_{n,i}}+g^2_{\tau_{n,i}})M_n\Delta t+O_p((M_n\Delta t)^2),
\end{aligned}
\qquad (A.12)
$$

$$
\begin{aligned}
&\sum_{\tau_{i+1}\le t}E((\int_{\tau_i}^{\tau_{i+1}}\sigma_t^2 f_t\,dt)^2-2(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})(\sigma^2_{\tau_{n,i+1}}-\sigma^2_{\tau_{n,i}})\int_{\tau_i}^{\tau_{i+1}}\sigma_t^2 f_t\,dt|\mathcal{F}_i)\\
=&-\sum_{\tau_{i+1}\le t}\sigma^4_{\tau_{n,i}}f^2_{\tau_{n,i}}M_n\Delta t+O_p((M_n\Delta t)^2),
\end{aligned}
\qquad (A.13)
$$

$$\sum_{\tau_{i+1}\leq t} E((X_{\tau_{n,i+1}}-X_{\tau_{n,i}})^2(\sigma^2_{\tau_{n,i+1}}-\sigma^2_{\tau_{n,i}})^2-2(X_{\tau_{n,i+1}}-X_{\tau_{n,i}})^2(\sigma^2_{\tau_{n,i+1}}-\sigma^2_{\tau_{n,i}})(\hat{\sigma}^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})\ |\mathcal{F}_i)$$
$$=O_p((M_n\Delta t)^2), \tag{A.14}$$

and

$$\begin{aligned}&\sum_{\tau_{i+1}\leq t} E((X_{\tau_{n,i+1}}-X_{\tau_{n,i}})^2(\hat{\sigma}^2_{\tau_{n,i}}-\sigma^2_{\tau_{n,i}})^2|\mathcal{F}_i)\\ =&\sum_{\tau_{i+1}\leq t}\frac{2}{M_n^2\Delta t}\sigma^6_{\tau_{n,i}}M_n\Delta t+\frac{4}{3}\sigma^4_{\tau_{n,i}}(f^2_{\tau_{n,i}}+g^2_{\tau_{n,i}})M_n\Delta t+2\sigma^4_{\tau_{n,i}}f^2_{\tau_{n,i}}M_n\Delta t+O_p((M_n\Delta t)^2),\end{aligned} \tag{A.15}$$

The above 3 equations lead to the following asymptotic conditional variance:

$$\sum_{\tau_{i+1}\leq t}\mathrm{Var}(\Delta M^n_{\tau_{i+1}}|\mathcal{F}_i)\xrightarrow{p}\ \ \tfrac{16}{c^2t}\int_0^t(F'(\sigma_s^2))^2\sigma_s^6\,ds+\int_0^t(F'(\sigma_s^2))^2\sigma_s^4(\tfrac{44}{3}f_s^2+\tfrac{32}{3}g_s^2)\,ds. \tag{A.16}$$

To finalize the proof of the theorem, we need the quadratic variation of the *interpolated* martingale. The following lemma shows that the quadratic variation is the same as the aggregate conditional variance as in equation (A.16)

Lemma 2.
$$[M^n,M^n]_t-\sum_{\tau_{n,i}\leq t}\mathrm{Var}(\Delta M^n_{\tau_{n,i}}|\mathcal{F}_i)=o_p(1) \tag{A.17}$$

Proof of Lemma 2:

$$\begin{aligned}
&E\left([M^n,M^n]_t-\sum_{\tau_{n,i}\leq t}\mathrm{Var}(\Delta M^n_{\tau_{n,i}}|\mathcal{F}_i)\right)^2\\
=\;&E\left(\sum_{\tau_{n,i}\leq t}(\Delta M^n_{\tau_{n,i}})^2-\sum_{\tau_{n,i}\leq t}\mathrm{Var}(\Delta M^n_{\tau_{n,i}}|\mathcal{F}_i)\right)^2\\
=\;&\sum_{\tau_{n,i}\leq t}E\left((\Delta M^n_{\tau_{n,i}})^2-E\left[(\Delta M^n_{\tau_{n,i}})^2|\mathcal{F}_i\right]\right)^2\ \text{(By MG Property, cross product term has expectation 0)}\\
\leq\;&\sum_{\tau_{n,i}\leq t}E(\Delta M^n_{\tau_{n,i}})^4+E\left(E\left[(\Delta M^n_{\tau_{n,i}})^2|\mathcal{F}_i\right]\right)^2\\
\leq\;&\sum_{\tau_{n,i}\leq t}E(\Delta M^n_{\tau_{n,i}})^4\ \text{(By Jensen's inequality for conditional expectation)}\\
\leq\;&E\left(\sup(\Delta M^n_{\tau_{n,i}})^2\times[M^n,M^n]_t\right)\\
\to\;&0,
\end{aligned} \tag{A.18}$$

since $\sigma_{\tau_i},\ f_{\tau_i},\ g_{\tau_i},\ [M^n,M^n]_t$ are assumed to be bounded, $M_n\Delta t=O_p(n^{-1/2})$.

A.5 Elimination of the bias term

The bias in the limit of $\langle X, \widehat{F(\sigma^2)}\rangle_T$ also depends on the limit of $[M^n, W^{(i)}]_t$, where either $W^(i)_t = W_t$ or $W^{(i)}$ is orthogonal to W_t, for any $t \in (0, T]$. For the second case, it is obvious that $[M^n, W^{(i)}]_t = 0$. We only need to study the first case $W_t^{(i)} = W_t$. In this case, since the covariance between $O_p(M_n\Delta t)$ terms and W_t will be of even higher order (at least $O_p((M_n\Delta t)^{3/2})$), those are negligible in the limit. Thus we only need to consider the following aggregate conditional expectation:

$$
\begin{aligned}
&\frac{1}{\sqrt{M_n\Delta t}}\sum_{\tau_{i+1}\le t}\mathrm{Cov}(\Delta M^n_{\tau_{n,i+1}}, \Delta W_{\tau_{n,i+1}}|\mathcal{F}_i)\\
=&\frac{1}{\sqrt{M_n\Delta t}}F'(\sigma^2_{\tau_{n,i}})\sum_{\tau_{i+1}\le t}E(\sum_i \Delta W_{\tau_{n,i+1}}(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})(\sigma^2_{\tau_{n,i+1}} - \sigma^2_{\tau_{n,i}})\\
&+\Delta W_{\tau_{n,i+1}}(X_{\tau_{n,i}} - X_{\tau_{n,i-1}})(\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}}) - \Delta W_{\tau_{n,i+1}}(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})(\hat{\sigma}^2_{\tau_{n,i}} - \sigma^2_{\tau_{n,i}})\\
&+O_p((M_n\Delta t)^{3/2})\\
=&O_p((M_n\Delta t)^{3/2}) \text{ (By Itô's formula and lemma 1).} \qquad \text{(A.19)}
\end{aligned}
$$

Thus, $[M^n, W]_t = O_p(M_n\Delta t)^{3/2})$ for any $t \in (0, T]$.

All the proofs above can be easily extended to the case for any $t \in (0, T]$. Thus, as outlined at the end of Section A.2, with (A.3), (A.19), (A.16), and (A.18), the proof of Theorem 1 is completed by applying the Central Limit Theorem for semimartingales (refer to the version in Mykland and Zhang (2012), Thm 2.28).

B Proof of Theorem 2

With Taylor expansion, the first order difference of the two estimators is:$\sum_{\tau_{i+1}\le t}\frac{2}{M_n(M_n-1)\Delta t}F'(\sigma_{\tau_{n,i}})(\Delta X_{\tau_{n,i+1}})^3$. In order for any term to make a difference in the limit, it must be of order lower than $O_p(M_n\Delta t)$. As in the proof of Proposition 2, we can show that the difference term is of order higher than $O_p(M_n\Delta t)$:

$$\sum_{\tau_{i+1}\le t}\frac{2}{M_n(M_n-1)\Delta t}F'(\sigma_{\tau_{n,i}})E((\Delta X_{\tau_{n,i+1}})^3|\mathcal{F}_i) = O_p(n^{-\frac{1}{2}}).$$

Theorem 2 is easily proved with this result and a slight variation in the proof of Theorem 1.

C Proof of Theorem 3

As we already seen in the proof of theorem 1 that
$\sum_i \frac{1}{M_n\Delta t}E((X_{\tau_{n,i+1}} - X_{\tau_{n,i}})^2(\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})^2|\mathcal{F}_i)\xrightarrow{p}\frac{4}{c^2T}\int_0^T(F'(\sigma_t^2))^2\sigma_t^6dt+\int_0^T(F'(\sigma_t^2))^2\sigma_t^4(\frac{14}{3}f_t^2+\frac{8}{3}g_t^2)dt$.
In order to prove the consistency of equation (11), one can again apply the martingale convergence argument. According to Lemma 2, it suffices to check wether the conditional variance of the martingale converges to 0. By applying Lemma 1, one can obtain:

$$
\begin{aligned}
&\sum_i E((\frac{1}{M_n\Delta t}(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})^2(\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})^2 - \frac{1}{M_n\Delta t}E((X_{\tau_{n,i+1}} - X_{\tau_{n,i}})^2(\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})^2|\mathcal{F}_i))^2|\mathcal{F}_i)\\
=&\sum_i E(\frac{1}{(M_n\Delta t)^2}(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})^4(\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})^4|\mathcal{F}_i) - E(\frac{1}{M_n\Delta t}(X_{\tau_{n,i+1}} - X_{\tau_{n,i}})^2(\hat{\sigma}^2_{\tau_{n,i+1}} - \hat{\sigma}^2_{\tau_{n,i}})^2|\mathcal{F}_i)^2\\
=&O_p(M_n^2\Delta t^2). \qquad \text{(C.20)}
\end{aligned}
$$

This proves $G_n^1 \xrightarrow{p} \frac{8}{c}\int_0^T (F'(\sigma_t^2))^2\sigma_t^6 dt + cT\int_0^T (F'(\sigma_t^2))^2\sigma_t^4(\frac{28}{3}f_t^2 + \frac{16}{3}g_t^2)dt$.
By similar argument, one can prove $G_n^2 \xrightarrow{p} \frac{8}{c}\int_0^T (F'(\sigma_t^2))^2\sigma_t^6 dt + cT\int_0^T (F'(\sigma_t^2))^2\sigma_t^4\frac{16}{3}(f_t^2 + g_t^2)dt$.
And the stable convergence in Theorem 3 is the direct results of convergence in probability of the asymptotic variance and the stable convergence in Theorem 1.

D Proof of Theorem 4

The proof of Theorem 4 will be done similarly to that of Theorem 1 in a manner to compare the difference between $(X_{t_{n,j+1}} - X_{t_{n,j}})$ and $(\bar{X}_{t_{n,j+1}} - \bar{X}_{t_{n,j}})$ in each step. As seen in the case without microstructure noise, it is enough to prove the case $F(x) = x$. For the convenience of later calculations, we always assume $\tau_{n,j+1} \in (\lambda_{n,i}, \lambda_{n,i+1}]$, $\tau'_{n,j+1} \in (\lambda_{n,i+1}, \lambda_{n,i+2}]$ and $\tau_{n,j+1}$ and $\tau'_{n,j+1}$ are corresponding $(j+1)$th observation time in the consecutive two big λ-blocks.

D.1 Aggregate conditional expectation of the estimator

According to the contiguity to Gaussian noise [Mykland and Zhang (2011a)], the process can be simplified as follows:

$$\hat{X}_{\tau_{n,i}} = \frac{1}{M}\sum_{t_{n,j}\in[\tau_i,\tau_{i+1})} Y_{t_{n,j}} = \frac{1}{M}\sum_{t_{n,j}\in[\tau_i,\tau_{i+1})} X_{t_{n,j}} + M^{-1/2}Z_{\tau_{n,i}} = \bar{X}_{\tau_{n,i}} + M^{-1/2}Z_{\tau_{n,i}},$$

$$\begin{aligned}\hat{\sigma}^2_{\lambda_{n,i+1}} &= \frac{1}{LM\Delta t}\sum_{\tau_{n,j+1}\in(\lambda_{n,i},\lambda_{n,i+1}]} (\hat{X}_{\tau_{n,j+1}} - \hat{X}_{\tau_{n,j}})^2 \\ &= \frac{1}{LM\Delta t}\sum_{\tau_{n,j+1}\in(\lambda_{n,i},\lambda_{n,i+1}]} (\Delta\bar{X}_{\tau_{n,j+1}} + \Delta Z_{\tau_{n,j+1}}M^{-1/2})^2,\end{aligned}$$

$$\begin{aligned}\text{and } \widehat{\langle X, \sigma^2\rangle}_T &= 3\sum_{i=0}^{K_n-2}(\hat{X}_{\lambda_{n,i+1}} - \hat{X}_{\lambda_{n,i}})(\hat{\sigma}^2_{\lambda_{n,i+1}} - \hat{\sigma}^2_{\lambda_{n,i}}) \\ &= \frac{3}{LM\Delta t}\sum_{i=0}^{K_n-2}(\Delta\bar{X}_{\lambda_{n,i+1}} + \Delta Z_{\lambda_{n,i+1}}M^{-1/2})\Big(\sum_{\tau'_{n,j+1}}(\Delta\bar{X}_{\tau'_{n,j+1}} + \Delta Z_{\tau'_{n,j+1}}M^{-1/2})^2 \\ &\qquad - \sum_{\tau_{n,j+1}}(\Delta\bar{X}_{\tau_{n,j+1}} + \Delta Z_{\tau_{n,j+1}}M^{-1/2})^2\Big),\end{aligned} \tag{D.21}$$

where the $Z'_{\tau_{n,i}}s$ are i.i.d $N(0, a^2)$. This iid normality leads to the following aggregate conditional expectation of the estimator:

$$
\begin{aligned}
& E\Big(\frac{3}{LM\Delta t}\sum_{i=0}^{K_n-2}\Delta\bar{X}_{\lambda_{n,i+1}}\Big(\sum_{\tau'_{n,j+1}\in(\lambda_{n,i+1},\lambda_{n,i+2}]}\Delta\bar{X}^2_{\tau'_{n,j+1}}-\sum_{\tau_{n,j+1}\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta\bar{X}^2_{\tau_{n,j+1}}\Big)|\mathcal{F}_i\Big)+O_p(LM\Delta t)\\
= & \frac{3}{LM\Delta t}\sum_{i=0}^{K_n-2}E\Big(\frac{1}{M^3}\sum_{\substack{\tau'_{n,j+1}\in(\lambda_{n,i+1},\lambda_{n,i+2}]\\ t_{n,l}\in[\lambda_{n,i},\tau_{n,1})\\ t'_{n,l}\in[\lambda_{n,i+1},\tau'_{n,1})\\ t_{n,k}\in[\tau'_{n,j},\tau'_{n,j+1})\\ t_{n,k}\in[\tau'_{n,j+1},\tau'_{n,j+2})}}(X_{t'_{n,l}}-X_{t_{n,l}})(X_{t'_{n,k}}-X_{t_{n,k}})^2\\
& -\frac{1}{M^3}\sum_{\substack{t_{n,l}\in[\lambda_{n,i},\tau_{n,1})\\ t'_{n,l}\in[\lambda_{n,i+1},\tau'_{n,1})}}(X_{t'_{n,l}}-X_{t_{n,l}})\Big(\sum_{\substack{\tau_{n,j+1}\in(\lambda_{n,i},\lambda_{n,i+1}]\\ t_{n,k}\in[\tau_{n,j},\tau_{n,j+1})\\ t_{n,k}\in[\tau_{n,j+1},\tau_{n,j+2})}}(X_{t'_{n,k}}-X_{t_{n,k}})\Big)^2|\mathcal{F}_i\Big)+O_p(LM\Delta t)\\
= & \sum_{i=0}^{K_n-2}2\sigma^2_{\lambda_{n,i}}f_{\lambda_{n,i}}LM\Delta t+O_p(LM\Delta t).
\end{aligned}
$$

The last step applies Lemma 1 (A.3) and (A.4), and we conclude that:

$$\widehat{\langle X,\sigma^2\rangle}_T\xrightarrow{p}\int_0^T 2\sigma_t^2 f_t dt=\langle X,\sigma^2\rangle_T. \tag{D.22}$$

There is another way to look at $\Delta\bar{X}_{\tau_{n,j+1}}$, which leads to a continuous approximation and provides a simplified version to prove the results.

$$
\begin{aligned}
& \bar{X}_{\tau_{n,j+1}}-\bar{X}_{\tau_{n,j}}\\
= & X_{\tau_{n,j+1}}-X_{\tau_{n,j}}+\sum_{\substack{k=1\\ t'_{n,k}\in[\tau_{n,j+1},\tau_{n,j+2})}}^{M}\Big(\frac{M-k}{M}\Big)(X_{t'_{n,k+1}}-X_{t'_{n,k}})\\
& -\sum_{\substack{k=1\\ t_{n,k}\in[\tau_{n,j},\tau_{n,j+1})}}^{M}\Big(\frac{M-k}{M}\Big)(X_{t_{n,k+1}}-X_{t_{n,k}})\\
\simeq & \Delta X_{\tau_{n,j+1}}+\frac{1}{M\Delta t}\int_{\tau_{n,j+1}}^{\tau_{n,j+2}}(\tau_{n,j+2}-t)dX_t-\frac{1}{M\Delta t}\int_{\tau_{n,j}}^{\tau_{n,j+1}}(\tau_{n,j+1}-t)dX_t,
\end{aligned}
\tag{D.23}
$$

and

$$
\begin{aligned}
& \bar{X}_{\lambda_{n,i+1}}-\bar{X}_{\lambda_{n,i}}\\
= & X_{\lambda_{n,i+1}}-X_{\lambda_{n,i}}+\sum_{\substack{k=1\\ t'_{n,k}\in[\lambda_{n,i+1},\tau'_{n,1})}}^{M}\Big(\frac{M-k}{M}\Big)(X_{t'_{n,k+1}}-X_{t'_{n,k}})-\sum_{\substack{k=1\\ t_{n,k}\in[\lambda_{n,i},\tau_{n,1})}}^{M}\Big(\frac{M-k}{M}\Big)(X_{t_{n,k+1}}-X_{t_{n,k}})\\
\simeq & X_{\lambda_{n,i+1}}-X_{\lambda_{n,i}}+\frac{1}{M\Delta t}\int_{\lambda_{n,i+1}}^{\tau'_{n,1}}(\tau'_{n,1}-t)dX_t-\frac{1}{M\Delta t}\int_{\lambda_{n,i}}^{\tau_{n,1}}(\tau_{n,1}-t)dX_t.
\end{aligned}
\tag{D.24}
$$

One can verify that, to the relevant order, the aggregate conditional expectation by this continuous approximation gives the same result but less complicated calculations.

D.2 Conditional variance of the approximate martingale

Similarly as in the proof of Theorem 1,the martingale is constructed as follows.
Up to time t,

$$M_t^n = \sum_{\lambda_{n,i+1}\leq t} \{\Delta \widehat{V}^n_{\lambda_{n,i+1}} - E(\Delta \widehat{V}^n_{\lambda_{n,i+1}}|\mathcal{F}_i)\}, \tag{D.25}$$

where, except a few terms at the edge,

$$\begin{aligned}\Delta \widehat{V}^n_{\lambda_{n,i+1}} = \frac{3}{\sqrt{LM\Delta t}}\{&(\widehat{X}_{\lambda_{n,i+1}} - \widehat{X}_{\lambda_{n,i}})(\sigma^2_{\lambda_{n,i+1}} - \sigma^2_{\lambda_{n,i}}) + (\widehat{X}_{\lambda_{n,i}} - \widehat{X}_{\lambda_{n,i-1}})(\frac{1}{LM\Delta t}\sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta\bar{X}^2_{\tau_i} - \sigma^2_{\lambda_{n,i}})\\
&- (\widehat{X}_{\lambda_{n,i+1}} - \widehat{X}_{\lambda_{n,i}})(\frac{1}{LM\Delta t}\sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta\bar{X}^2_{\tau_i} - \sigma^2_{\lambda_{n,i}}) - \int_{\lambda_i}^{\lambda_{i+1}}\frac{2}{3}\sigma_t^2 f_t\,dt\\
&+ (\widehat{X}_{\lambda_{n,i+1}} - \widehat{X}_{\lambda_{n,i}})\frac{1}{LM\Delta t}(\sum_{\tau_i\in(\lambda_{n,i+1},\lambda_{n,i+2}]}\Delta Z^2_{\tau_i}M^{-1} - \sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta Z^2_{\tau_i}M^{-1})\\
&+ (\widehat{X}_{\lambda_{n,i}} - \widehat{X}_{\lambda_{n,i-1}})(\frac{1}{LM\Delta t}\sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta\bar{X}_{\tau_i}\Delta Z_{\tau_i}M^{-1/2})\\
&- (\widehat{X}_{\lambda_{n,i+1}} - \widehat{X}_{\lambda_{n,i}})(\frac{1}{LM\Delta t}\sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta\bar{X}_{\tau_i}\Delta Z_{\tau_i}M^{-1/2})\}.\end{aligned}$$

Again, we will interpolate this martingale into a continuous martingale up to any time $t \in (0, T]$ as in the proof of Theorem 1.

From this point on, the continuous approximation of $\bar{X}$ with integral will be adopted for simplicity and transparency. However, all the calculation can be checked by applying Lemma 1 to prove directly without continuous approximation

Lemma 3.

$$\begin{aligned}&E((\frac{1}{M\Delta t}\int_{\tau_{n,j}}^{\tau_{n,j+1}}(\tau_{n,j+1}-t)dX_t)^2\sigma^p_{\tau_{n,m}}|\mathcal{F}_j)\\
&= \frac{1}{3}\sigma^{p+2}_{\tau_{n,j}}(\tau_{n,j+1}-\tau_{n,j}) + [\frac{p(p-1)}{6}(\tau_{n,m}-\tau_{n,j})(\tau_{n,j+1}-\tau_{n,j}) + \frac{p^2}{4}\sigma^p_{\tau_{n,j}}f^2_{\tau_{n,j}}(\tau_{n,j+1}-\tau_{n,j})^2\\
&\quad+\frac{2p+1}{12}(\tau_{n,j+1}-\tau_{n,j})^2]\sigma^p_{\tau_{n,j}}(f_{\tau_{n,j}}+g^2_{\tau_{n,j}}) + \frac{p}{4}\sigma^p_{\tau_{n,j}}\langle X,f\rangle_{\tau_{n,j}}(\tau_{n,j+1}-\tau_{n,j})^2 + O_p(\Delta t^{\frac{3}{2}}),\end{aligned} \tag{D.26}$$

$$\begin{aligned}&E((X_{\tau_{n,j}} - X_{\tau_{n,k}})\frac{1}{M\Delta t}\int_{\tau_{n,k}}^{\tau_{n,k+1}}(\tau_{n,k+1}-t)dX_t\sigma^p_{\tau_{n,m}}|\mathcal{F}_k)\\
&= \frac{1}{2}\sigma^{p+2}_{\tau_{n,k}}(\tau_{n,k+1}-\tau_{n,k}) + \sigma^p_{\tau_{n,k}}(\frac{p^2}{2}f^2_{\tau_{n,k}} + \frac{p}{2}\langle X,f\rangle_{\tau_{n,k}})[(\tau_{n,k+1}-\tau_{n,k})^2 + (\tau_{n,k+1}-\tau_{n,k})(\tau_{n,j}-\tau_{n,k+1})]\\
&\quad+\frac{p(p-1)}{4}\sigma^p_{\tau_{n,k}}(f^2_{\tau_{n,k}}+g^2_{\tau_{n,k}})(\tau_{n,k+1}-\tau_{n,k})(\tau_{n,m}-\tau_{n,k})\\
&\quad+\frac{2p+1}{6}\sigma^p_{\tau_{n,k}}(f^2_{\tau_{n,k}}+g^2_{\tau_{n,k}})(\tau_{n,k+1}-\tau_{n,k})^2 + O_p(\Delta t^{\frac{3}{2}}),\end{aligned} \tag{D.27}$$

$$\begin{aligned}&E((X_{\tau_{n,j}} - X_{\tau_{n,k}})\frac{1}{M\Delta t}\int_{\tau_{n,j}}^{\tau_{n,j+1}}(\tau_{n,j+1}-t)dX_t\sigma^p_{\tau_{n,m}}|\mathcal{F}_k)\\
&= \sigma^p_{\tau_{n,k}}(\frac{p^2}{2}f^2_{\tau_{n,k}} + \frac{p}{2}\langle X,f\rangle_{\tau_{n,k}})(\tau_{n,j+1}-\tau_{n,j})(\tau_{n,j}-\tau_{n,k}) + O_p(\Delta t^{\frac{3}{2}}), \textit{ and}\end{aligned} \tag{D.28}$$

$$E(\frac{1}{M\Delta t}\int_{\tau_{n,k}}^{\tau_{n,k+1}}(\tau_{n,k+1}-t)dX_t\frac{1}{M\Delta t}\int_{\tau_{n,j}}^{\tau_{n,j+1}}(\tau_{n,j+1}-t)dX_t\sigma^p_{\tau_{n,m}}|\mathcal{F}_k)$$
$$=\sigma^p_{\tau_{n,k}}(\frac{p^2}{4}f^2_{\tau_{n,k}}+\frac{p}{4}\langle X,f\rangle_{\tau_{n,k}})(\tau_{n,j+1}-\tau_{n,j})(\tau_{n,k+1}-\tau_{n,k})+O_p(\Delta t^{\frac{3}{2}}), \tag{D.29}$$

where $\tau_{n,k}<\tau_{n,j}\leq\tau_{n,m}$, $\mathcal{F}_k=\mathcal{F}_{n,k}$.

The proof of this Lemma is similar to that of Lemma 1.

It is easy to see that the aggregate conditional expectation part in M_t^n is of order $O_p(LM\Delta t)$ from (D.22). So the aggregate conditional variance will be the same as the second moment of the part before the aggregate conditional expectation. Let $K_n=[\frac{\sqrt{n}}{c_1}]$:

$$\begin{aligned}
&\sum_{\lambda_{n,i+1}\leq t}\mathrm{Var}(\Delta M^n_{\lambda_{n,i+1}}|\mathcal{F}_i)\\
=\;&\frac{9}{LM\Delta t}\sum_{\lambda_{n,i+1}\leq t}E((\widehat{X}_{\lambda_{n,i+1}}-\widehat{X}_{\lambda_{n,i}})^2(\sigma^2_{\lambda_{n,i+1}}-\sigma^2_{\lambda_{n,i}})^2+(\hat{X}_{\lambda_{n,i}}-\hat{X}_{\lambda_{n,i-1}})^2(\hat{\sigma}^2_{\lambda_{n,i}}-\sigma^2_{\lambda_{n,i}})^2\\
&+(\hat{X}_{\lambda_{n,i+1}}-\hat{X}_{\lambda_{n,i}})^2(\hat{\sigma}^2_{\lambda_{n,i}}-\sigma^2_{\lambda_{n,i}})^2-2(\hat{X}_{\lambda_{n,i+1}}-\hat{X}_{\lambda_{n,i}})^2(\hat{\sigma}^2_{\lambda_{n,i}}-\sigma^2_{\lambda_{n,i}})(\sigma^2_{\lambda_{n,i}}-\sigma^2_{\lambda_{n,i}})\\
&+(\int_{\lambda_i}^{\lambda_{i+1}}\frac{2}{3}\sigma^2_t f_t\,dt)^2-\frac{4}{3}(\int_{\lambda_i}^{\lambda_{i+1}}\sigma^2_t f_t\,dt)(\widehat{X}_{\lambda_{n,i+1}}-\widehat{X}_{\lambda_{n,i}})(\sigma^2_{\lambda_{n,i+1}}-\sigma^2_{\lambda_{n,i}})\\
&+(\widehat{X}_{\lambda_{n,i+1}}-\widehat{X}_{\lambda_{n,i}})^2\frac{1}{L^2M^2\Delta t^2}(\sum_{\tau_i\in(\lambda_{n,i+1},\lambda_{n,i+2}]}\Delta Z^2_{\tau_i}M^{-1}-\sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta Z^2_{\tau_i}M^{-1})^2\\
&+(\widehat{X}_{\lambda_{n,i}}-\widehat{X}_{\lambda_{n,i-1}})^2(\frac{1}{LM\Delta t}\sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta\bar{X}_{\tau_i}\Delta Z_{\tau_i}M^{-1/2})^2\\
&+(\widehat{X}_{\lambda_{n,i+1}}-\widehat{X}_{\lambda_{n,i}})^2(\frac{1}{LM\Delta t}\sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta\bar{X}_{\tau_i}\Delta Z_{\tau_i}M^{-1/2})^2|\mathcal{F}_i)+O_p(L^2M^2\Delta t^2).
\end{aligned}$$

We will separately calculate the conditional variance involving the microstructure noise $\Delta ZM^{-1/2}$ and the variance involving only the semimartingale process, denoted accordingly as $\mathrm{Var}(\Delta M^{n,noise}_{\lambda_{n,i+1}}|\mathcal{F}_i)$ and $\mathrm{Var}(\Delta M^{n,process}_{\lambda_{n,i+1}}|\mathcal{F}_i)$. Applying Lemma 1 and Lemma 3, we can derive:

$$\begin{aligned}
&\sum\mathrm{Var}(\Delta M^{n,\text{ noise}}_{\lambda_{n,i+1}}|\mathcal{F}_i)\\
=&\frac{9}{L^3M^3\Delta t^3}\sum_{\lambda_{n,i+1}\leq t}E((\widehat{X}_{\lambda_{n,i+1}}-\widehat{X}_{\lambda_{n,i}})^2(\sum_{\tau_i\in(\lambda_{n,i+1},\lambda_{n,i+2}]}\Delta Z^2_{\tau_i}M^{-1}-\sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta Z^2_{\tau_i}M^{-1})^2\\
&\qquad+(\widehat{X}_{\lambda_{n,i}}-\widehat{X}_{\lambda_{n,i-1}})^2(\sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta\bar{X}_{\tau_i}\Delta Z_{\tau_i}M^{-1/2})^2\\
&\qquad+(\widehat{X}_{\lambda_{n,i+1}}-\widehat{X}_{\lambda_{n,i}})^2(\sum_{\tau_i\in(\lambda_{n,i},\lambda_{n,i+1}]}\Delta\bar{X}_{\tau_i}\Delta Z_{\tau_i}M^{-1/2})^2|\mathcal{F}_i)+O_p(L^2M^2\Delta t^2)\\
=&\sum_{\lambda_{n,i}\leq t}\frac{96a^2}{L^3M^4\Delta t^3}\sigma^4_{\lambda_{n,i}}(LM\Delta t)^2+\frac{216a^4}{L^2M^5\Delta t^3}\sum_{\lambda_{n,i+1}\leq t}E(\Delta\bar{X}^2_{\lambda_{n,i+1}}|\mathcal{F}_i)+O_p(L^2M^2\Delta t^2)\\
\xrightarrow{p}&\frac{96a^2}{3c^2c_1^2T^2}\int_0^T\sigma^4_t dt+\frac{216a^4}{c^2c_1^4T^3}\int_0^T\sigma^2_t dt,
\end{aligned} \tag{D.30}$$

and

$$
\begin{aligned}
&\sum \mathrm{Var}(\Delta M^{n,\text{ process}}_{\lambda_{n,i+1}}|\mathcal{F}_i) \\
=&\frac{9}{LM\Delta t}\sum_{\lambda_{n,i+1}\le t} E((\widehat{X}_{\lambda_{n,i+1}}-\widehat{X}_{\lambda_{n,i}})^2(\sigma^2_{\lambda_{n,i+1}}-\sigma^2_{\lambda_{n,i}})^2+(\hat{X}_{\lambda_{n,i}}-\hat{X}_{\lambda_{n,i-1}})^2(\hat{\sigma}^2_{\lambda_{n,i}}-\sigma^2_{\lambda_{n,i}})^2 \\
&+(\hat{X}_{\lambda_{n,i+1}}-\hat{X}_{\lambda_{n,i}})^2(\hat{\sigma}^2_{\lambda_{n,i}}-\sigma^2_{\lambda_{n,i}})^2-2(\hat{X}_{\lambda_{n,i+1}}-\hat{X}_{\lambda_{n,i}})^2(\hat{\sigma}^2_{\lambda_{n,i}}-\sigma^2_{\lambda_{n,i}})(\sigma^2_{\lambda_{n,i}}-\sigma^2_{\lambda_{n,i}}) \\
&+(\int_{\lambda_i}^{\lambda_{i+1}}\frac{2}{3}\sigma_t^2 f_t\,dt)^2-\frac{4}{3}(\int_{\lambda_i}^{\lambda_{i+1}}\sigma_t^2 f_t\,dt)(\widehat{X}_{\lambda_{n,i+1}}-\widehat{X}_{\lambda_{n,i}})(\sigma^2_{\lambda_{n,i+1}}-\sigma^2_{\lambda_{n,i}})|\mathcal{F}_i)+O_p(L^2M^2\Delta t^2) \\
=&\sum_{\lambda_{n,i}\le t}(\frac{16}{L^2M\Delta t}\sigma^6_{\lambda_{n,i}}LM\Delta t+\frac{44}{3}\sigma^4_{\lambda_{n,i}}f^2_{\lambda_{n,i}}LM\Delta t+\frac{32}{3}\sigma^4_{\lambda_{n,i}}g^2_{\lambda_{n,i}}LM\Delta t)+O_p(LM\Delta t).
\end{aligned}
\tag{D.31}
$$

These results prove that, for any $t\in(0,T]$:

$$
[M,M]_t\xrightarrow{p}\frac{16}{c^2t}\int_0^t\sigma_s^6ds+\int_0^t\sigma_s^4(\frac{44}{3}f_s^2+\frac{32}{3}g_s^2)ds+\frac{96a^2}{c^2c_1^2t^2}\int_0^t\sigma_s^4ds+\frac{216a^4}{c^2c_1^4t^3}\int_0^t\sigma_s^2ds.
\tag{D.32}
$$

By similar methods, we can derive that the bias term also converges to zero (here we omit the lengthy proof):

$$
\frac{1}{\sqrt{LM\Delta t}}\sum_i \mathrm{Cov}(\Delta M^n_{\lambda_{n,i+1}},\Delta W_{\lambda_{n,i+1}}|\mathcal{F}_i)=O_p((LM\Delta t)^{3/2}).
\tag{D.33}
$$

Thus, $[M^n,W]_t=O_p((LM\Delta t)^{3/2})$ for any $t\in(0,T]$. This completes the proof of Theorem 4.

E Proof of Theorem 5

Theorem 5 can be easily proved by the consistency of asymptotic variance (27) and stable convergence in Theorem 4. One only needs to establish the convergence in probability of equation (27). And this proof adopts the similar martingale technique as in the proof of Theorme 3, by applying Lemma 1, Lemma 2 and Lemma 3. One can prove: $E(\frac{81}{4L^2M^2\Delta t^2}\sum_i(\hat{X}_{\lambda_{n,i+1}}-\hat{X}_{\lambda_{n,i}})^4(F(\hat{\sigma}^2_{\lambda_{n,i+1}})-F(\hat{\sigma}^2_{\lambda_{n,i}}))^4|\mathcal{F}_i)-(E(\frac{9}{2LM\Delta t}\sum_i(\hat{X}_{\lambda_{n,i+1}}-\hat{X}_{\lambda_{n,i}})^2(F(\hat{\sigma}^2_{\lambda_{n,i+1}})-F(\hat{\sigma}^2_{\lambda_{n,i}}))^2|\mathcal{F}_i))^2=O_p(LM\Delta t)$.
This gives the consistency of $G_{n,1}$ in equation 27. By similar argument, the consistency of $G_{n,2}$ can be proved and consequently completing the proof of the convergence of equation 27 and Theorem 5.

F Proof of Proposition 2

Assume equivalent measure where both X_t and σ_t are martingales. This can be done in analogy with the development in [Mykland and Zhang (2009)], Section 2.2, by a stopping argument and so long as the instantaneous correlation between X_t and σ_t is not ±1. (In the latter case, the proof is different but straightforward.)
By the third Bartlett identity [Mykland (1994)]

$$
E(\Delta X^3_{t_{n,i+1}}|\mathcal{F}_{t_i}) \quad = \quad 3E(\Delta X_{t_{n,i+1}}\Delta\langle X,X\rangle|\mathcal{F}_{t_i})
$$

$$\begin{aligned} &= 3E(\Delta X_{t_{n,i+1}} \int_{t_{n,i}}^{t_{n,i+1}} \sigma_u^2 du | \mathcal{F}_{t_i}) \\ &= 3E(\Delta X_{t_{n,i+1}} \int_{t_{n,i}}^{t_{n,i+1}} (\sigma_u^2 - \sigma_{t_{n,i}}^2) du | \mathcal{F}_{t_i}) \text{ By MG property of } X_t. \quad \text{(F.34)} \end{aligned}$$

By the Itô formula, $d(t_{n,i+1} - t)(\sigma_u^2 - \sigma_{t_{n,i}}^2) = -(\sigma_u^2 - \sigma_{t_{n,i}}^2)dt + (t_{n,i+1} - t)d\sigma_u^2$, we obtain:

$$\begin{aligned} \text{(F.34)} &= 3E(\Delta X_{t_{n,i+1}} \int_{t_{n,i}}^{t_{n,i+1}} (t_{n,i+1} - t) d\sigma_u^2 | \mathcal{F}_{t_i}) \\ &= 3E(\int_{t_{n,i}}^{t_{n,i+1}} (t_{n,i+1} - t) d\langle X, \sigma_u^2 \rangle | \mathcal{F}_{t_i}). \end{aligned}$$

Hence,

$$\begin{aligned} \frac{n}{T} \sum_{t_{n,i+1} \leq T} E(\Delta X_{t_{n,i+1}}^3 | \mathcal{F}_{t_i}) &= \frac{n}{T} \frac{3}{2} \sum_{t_{n,i+1} \leq T} \Delta t_{n,i+1}^2 \langle X, \sigma^2 \rangle'_{t_{n,i}} + \text{higher order terms} \\ &\xrightarrow{p} \frac{3}{2} \int_0^T \langle X, \sigma^2 \rangle'_u du \\ &= \frac{3}{2} \langle X, \sigma^2 \rangle_T. \end{aligned}$$

We have here used that $\Delta t_i = \frac{T}{n}$ (equidistant spacing); for the more general case, the expression will involve the quadratic variation of time.