Estimating and Forecasting Volatility using Leverage Effect *

Christina Dan Wang
Columbia University

Per A. Mykland
University of Chicago

Lan Zhang
University of Illinois at Chicago

November 30, 2017

Abstract

This research provides a theoretical foundation for our previous empirical finding that leverage effect has a role in estimating and forecasting volatility. This empirics is also related to earlier econometric studies of news impact curves (Engle and Ng, Chen and Ghysels). Our new theoretical development is based on the concept of projection on stable subspaces of semi-martingales. We show that this projection provides a framework for forecasting (across time periods) that is internally consistent with the semi-martingale model which is used for the intra-day high frequency asymptotics. The paper shows that the approach provides improved estimation and forecasting both theoretically, in simulation, and in data.

KEYWORDS: integrated volatility, leverage effect, realized volatility, discrete observation, Itô process, microstructure noise, pre-averaging, stable convergence,.

*Financial support from the National Science Foundation under grants DMS 17-13129 and DMS 17-13118 is gratefully acknowledged.

1 Introduction

The estimation of volatility from high frequency data is a well researched topic. In a continuous semimartingale (Itô process)

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

the object is to estimate integrated volatility for time period (day, or other) $\#i$

$$IV_i = \int_{T_{i-1}}^{T_i} \sigma_t^2 dt. \tag{2}$$

In the absence of microstructure, the basic estimator is *realized volatility* (RV),[1] given by

$$RV_i = \sum_{T_{i-1} < t_j \leq T_i} (X_{t_j} - X_{t_{j-1}})^2 \tag{3}$$

When microstructure is present, more complex estimators are in order, and we shall return to this matter below. To cover both scenarios, we write $\widehat{IV}_i$ to denote an estimator of IV_i based on intraday data.

Under certain conditions, the RV is a nonparametric maximum likelihood estimator of integrated volatility,[2] and hence efficient. One of these conditions, however, is the absence of leverage effect, *i.e.*, that W_t and σ_t^2 are independent. However, this is typically not the case.[3] This raises the question of whether an estimator of the leverage effect can be parlayed into a better estimator of integrated volatility.

Our purpose in this paper is to show that this is indeed the case.

Specifically, we shall argue that an improved estimator of volatility for time period $\#i$ is obtained by using an optimal linear combinaion of $\widehat{IV}_i$ and the alternative estimator

$$\widetilde{IV}_i = \widehat{IV}_{i-1} + \text{ an estimator of } (C_i + P_i) \tag{4}$$

[1]See, in particular, Andersen and Bollerslev (1998a,b), Andersen, Bollerslev, Diebold, and Labys (2001, 2003), Barndorff-Nielsen and Shephard (2001, 2002); Barndorff-Nielsen (2004), Jacod and Protter (1998), Zhang (2001), Mykland and Zhang (2006), and the group at Olsen and Associates (Dacorogna, Gençay, Müller, Olsen, and Pictet (2001)), An instantaneous version of RV was earlier proposed by Foster and Nelson (1996) and Comte and Renault (1998).

[2]Section 2 of Mykland (2012).

[3]For a review of the leverage effect, see Wang and Mykland (2014), as well as the discussion of news inpact curves below.

where

$$C_i = \int_{T_i}^{T_{i+1}} (T_{i+1} - t) \frac{d}{dt} \langle \log \sigma, X \rangle_t dX_t \quad \text{and}$$
$$P_i = \int_{T_{i-1}}^{T_i} (t - T_{i-1}) d \frac{d}{dt} \langle \log \sigma, X \rangle_t dX_t. \tag{5}$$

Similarly, a close to optimal forecast of IV_i is given by

$$PIV_i = \widehat{IV}_{i-1} + \text{ an estimator of } P_i. \tag{6}$$

The estimator of $d\langle \log \sigma, X \rangle_t / dt$ is based on the estimator of leverage effect developed in (Wang and Mykland (2014)). Meanwhile, dX_t is replaced by ΔX_{t_j} when there is no microstructure, and by a preaveraged $\overline{\Delta X}_{\tau_j}$ when microstructure is present.[4] The estimators are described in detail in Section 3.

The approach in (4)-(6) has two parts. On the one hand, one can draw on the past to better estimate or predict current volatility. On the other hand, and this is new, we have a specific formula for how to draw on the past.

The paper is closely related to earlier work by Meddahi (2002) and Ghysels, Mykland, and Renault (2012), The premise in these papers is that one can reinforce the current day's estimate of volatility with the previous day's estimate.[5]. Our approach follows this tradition, but with the variation that one can to some extent estimate the quantities P_i and C_i from high frequency data, without relying on stationarity assumptions. We here use high frequency regression, as discussed in Section 2.

In particular, if the quantities IV_{i-1} and P_i were observed, the prediction equation (6) can be used without further estimation. In our data analysis, however, we shall rely on stationarity-based

[4]Pre-averaging with flat weights (see Jacod, Li, Mykland, Podolskij, and Vetter (2009a), Podolskij and Vetter (2009a,b), Jacod, Podolskij, and Vetter (2009b), Mykland and Zhang (2016)). Other approcahes to estimation under microstructure include are *Two and Multi Scales Realized Volatility* (Zhang, Mykland, and Aït-Sahalia (2005), Zhang (2006)), *Realized Kernel Volatility*, which uses weighted autocovariances (Barndorff-Nielsen, Hansen, Lunde, and Shephard (2008)), *Quasi-likelihood* (Xiu (2010)), and the *spectral* approach of Bibinger and Reiß (2014) and Altmeyer and Bibinger (2015).

[5]There is, of course, also a lively literature on forecasting, see Andersen, Bollerslev, and Meddahi (2005) and the references therein

estimation to obtain the coefficients of $\widehat{IV}_{i-1}$ and $\hat{P}_i$ (for forecasting) and also of $\hat{C}_i$ and $\widehat{IV}_i$ (for estimation) because the relevant quantities are observed with error (*i.e.*, estimated), so the theoretical coefficients may not be optimal. The linear combination of $\widehat{IV}_i$ and $\widetilde{IV}_i$ also needs to be determined empirically.

We shall see in Remark 1 below that the quantities C_i and P_i are closely related to leverage effect and skewness. In continuous processes, these two concepts are the same object up to a constant (Section 7 of Wang and Mykland (2014)). The current paper therefore provides a way for skewness to predict volatility. This has earlier been studied in the conext of *news impact curves* (Engle and Ng (1993), Chen and Ghysels (2011)).

The plan of the paper is as follows. In Section 2 we review high frequency regression and the notion of projection on a stable subspace. This provides the theoretical foundation for formulae (4)-(6). Section 3 describes how to estimate the relevant quantities, while consistency and asymptotic normality is shown in Section 4. The theoretical theoretical asymptotic normality is further corroborated in simulation in Section 4.1. Theoretical results under market microstructure noise are provided in Section 5. We explore the method in estimating and forecasting volatilities from the standpoint of simulation in Section 6, and provide a data example in Section 7.

2 High frequency regression

Consider two continuous semimartingales X_t and Y_t. In the hypothetical case of continuous observation, the regression of Y on X is given by

$$dY_t = f_t dX_t + dZ_t \tag{7}$$

where

$$f_t = \frac{d\langle Y, X\rangle_t}{d\langle X, X\rangle_t} \text{ and } \langle Z, X\rangle_t \equiv 0. \tag{8}$$

The integral $\int f_t dX_t$ is the projection of Y on the *stable subspace* generated by X (Jacod (1979)). For the standard case of two general processes Y and X, estimation of f_t is similar to the estimation of spot volatility.[6]

[6] See Zhang (2001, 2012) as well as Mykland and Zhang (2008, 2009).

In our case, a special form of regression emerges. Write

$$IV_i = IV_{i-1} + Y^{(i)}_{T_{i+1}}, \tag{9}$$

THEOREM 1. (PROJECTION RESULT.) *Assume that σ^2 is a continuous seminartingale. Also assume that the T_i are equidistant. Write*

$$dY^{(i)}_t = \begin{cases} (t - T_{i-1})d\sigma^2_t, \text{ for } T_{i-1} < t \leq T_i \\ (T_{i+1} - t)d\sigma^2_t, \text{ for } T_i \leq t < T_{i+1} \end{cases} \tag{10}$$

with $Y^{(i)}_t = 0$ for $t \leq T_{i-1}$ and $Y^{(i)}_t = 0$ for $t > T_{i+1}$. Then

$$IV_i = IV_{i-1} + Y^{(i)}_{T_{i+1}}, \tag{11}$$

and the regression (7) of $Y^{(i)}$ on X has regression coefficient $f^{(i)}$ given by

$$f^{(i)}_t = \begin{cases} (t - T_{i-1})\frac{d}{dt}\langle \log\sigma^2, X\rangle_t, \text{ for } T_{i-1} < t \leq T_i \\ (T_{i+1} - t)\frac{d}{dt}\langle \log\sigma^2, X\rangle_t, \text{ for } T_i \leq t < T_{i+1} \end{cases} \tag{12}$$

with $f^{(i)}_t = 0$ for $t \leq T_{i-1}$ and $t > T_{i+1}$.

Proof of Theorem 1. The form (11) follows from Theorem 2 (p. 206) of Mykland and Zhang (2017). The high frequency projection is then obtained using (13) below. ∎

The theorem yields the formulae (4)-(6) as a corollary.

REMARK 1. (THREE INTERPRETATIONS OF THE REGRESSION COEFFICIENT.) A crucial ingredient in this result is that[7]

$$\frac{d\langle\sigma^2, X\rangle_t}{d\langle X, X\rangle_t} = \frac{d}{dt}\langle \log\sigma^2, X\rangle_t. \tag{13}$$

Thus, on the one hand, $d\langle \log\sigma^2, X\rangle_t/dt$ is the regression coefficient from the projection of σ^2 on X.[8] On the other hand, it is itself a form of leverage effect, which can be estimated using the methods of Wang and Mykland (2014). From the right hand side of (13), we also note that the regression coefficient is a normalized high frequency skewness (*ibid, Section 7).* □

[7] Cf. Section 8 of Wang and Mykland (2014)

[8] And so the regression is not the same for instantaneous and interated volatility.

3 Estimators

We shall start from the equidistant case when microstructure noise is not present. In addition to the log price process, we will assume the volatility process is another semi-martingale.

ASSUMPTION (H): The log-price and volatility processes are both Itô semimartingales:

$$X_t = X_0 + \int_0^t \mu_s ds + \int_0^t \sigma_s dW_s, \tag{14}$$

$$\sigma_t = \sigma_0 + \int_0^t \tilde{\mu}_s ds + \int_0^t a_s dW_s + \int_0^t b_s dB_s, \tag{15}$$

where W and B are independent standard Brownian motions; Additionally,

(a) The processes $\mu_t(\omega)$, $\tilde{\mu}_t(\omega)$, $a_t(\omega)$, and $b_t(\omega)$ are locally bounded;

(b) All paths $t \mapsto \mu_t(\omega)$, $t \mapsto \tilde{\mu}_t(\omega)$, $t \mapsto a_t(\omega)$, $t \mapsto b_t(\omega)$ are càglàd (left-continuous with right limits), and a_+ (right limit of a) is continuous and smooth.

(c) The processes σ^2 and σ^2_- (left limit of σ^2) are bounded away from zero.

Under Assumption (H), the regression coefficient can be written as

$$\frac{d}{dt}\langle \log \sigma^2, X\rangle_t = 2a_t. \tag{16}$$

Suppose the data are observed every $\Delta_n = T/n$ units of time without any measurement error. The full grid containing all of the observation points is given by:

$$\mathcal{U} = \{0 = t_0^n < t_1^n < t_2^n < \cdots < t_n^n = \lfloor T/\Delta_n \rfloor \Delta_n\}, \tag{17}$$

where $t_j^n = j\Delta_n$ for each j. Then the increment (log-return) of log-prices over the j-th interval is $\Delta_j^n X := X_{t_j^n} - X_{t_{j-1}^n}$.

We will take a local window of $m_n = \lfloor c\sqrt{n} \rfloor$ time intervals for estimating the spot volatility, since it provides the optimal convergence rate while estimating leverage effect (see Wang and Mykland (2014)). Let $I_n^-(i) = \{i - m_n, \ldots, i-1\}$ (if $i > m_n$) and $I_n^+(i) = \{i+1, \ldots, i+m_n\}$ define two local windows in time of length $m_n\Delta_n$, just before and after time $i\Delta_n$. Furthermore, we need to take another window of k_n time intervals for estimating the spot leverage effect. k_n will be an integer number taking the value $\lfloor c_1 m_n n^b \rfloor$. Then if we can define

$$
\begin{aligned}
\widehat{C_i} &= \sum_{\lambda=m_n+1}^{n-k_n-m_n} (T_{i+1} - t_\lambda)\, \widehat{a}_{t_\lambda} \Delta^n_{\lambda+1} X, \quad \text{and} \\
\widehat{P}_{i+1} &= \sum_{\lambda=m_n+1}^{n-k_n-m_n} (t_\lambda - T_i)\, \widehat{a}_{t_\lambda} \Delta^n_{\lambda+1} X \quad \text{with} \\
\widehat{a}_{t_\lambda} &= \frac{1}{k_n \Delta_n} \sum_{j=\lambda+1}^{\lambda+k_n} \Delta^n_j X\, (\log(\widehat{\sigma}^2_{j+}) - \log(\widehat{\sigma}^2_{j-})), \\
\widehat{\sigma}^2_{j+} &= \frac{1}{m_n \Delta_n} \sum_{k \in I^+_n(j)} (\Delta^n_k X)^2, \quad \widehat{\sigma}^2_{j-} = \frac{1}{m_n \Delta_n} \sum_{k \in I^-_n(j)} (\Delta^n_k X)^2.
\end{aligned}
\tag{18}
$$

REMARK 2. If $b \geq 1/4$, the above estimators need to be modified by tail estimators, as follows.

Denote $t' = T_{i+1} - (k_n + m_n)\Delta_n$

$$
\begin{aligned}
\widehat{C}^e_i &= \sum_{\lambda=n-k_n-m_n+1}^{n} (T_{i+1} - t_\lambda)\, \widehat{a}_{t'} \Delta^n_{\lambda+1} X, \quad \text{and} \\
\widehat{P}^e_{i+1} &= \sum_{\lambda=n-k_n-m_n+1}^{n} (t_\lambda - T_i)\, \widehat{a}_{t'} \Delta^n_{\lambda+1} X \quad \text{with} \\
\widehat{a}_{t'} &= \frac{1}{k_n \Delta_n} \sum_{j=n-k_n-m_n+1}^{n-m_n} \Delta^n_j X\, (\log(\widehat{\sigma}^2_{j+}) - \log(\widehat{\sigma}^2_{j-})), \\
\widehat{\sigma}^2_{j+} &= \frac{1}{m_n \Delta_n} \sum_{k \in I^+_n(j)} (\Delta^n_k X)^2, \quad \widehat{\sigma}^2_{j-} = \frac{1}{m_n \Delta_n} \sum_{k \in I^-_n(j)} (\Delta^n_k X)^2.
\end{aligned}
\tag{19}
$$

The modified estimators become $\widehat{C}_i + \widehat{C}^e_i$ and $\widehat{P}_{i+1} + \widehat{P}^e_{i+1}$, and for $b \geq 1/4$, these are the estimators for which limit behavior is presented below. □

4 Consistency and CLT of the estimator

Theorem 2. *For some positive constant c and c_1, let $m_n = \lfloor c\sqrt{n} \rfloor$ and $k_n = \lfloor c_1 m_n n^b \rfloor$ with $0 < b < \frac{1}{2}$. Under Assumption (H) in Section 3, we have*[9]

$$\widehat{C}_i \xrightarrow{u.c.p.} C_i, \text{ and } \quad \widehat{P}_i \xrightarrow{u.c.p.} P_i, \text{ as } n \to \infty. \tag{20}$$

Theorem 3. *For some positive constant c and c_1, let $m_n = \lfloor c\sqrt{n} \rfloor$ and $k_n = \lfloor c_1 m_n n^b \rfloor$ with $0 < b < \frac{1}{2}$. Then, under Assumption (H), $\widehat{C}_i$ converges stably in law to a limiting random variable defined on an extension of the original probability space. That is,*

$$\sqrt{n^b}\left(\widehat{C}_i - C_i\right) \xrightarrow{\mathscr{L}_{st}} \int_{T_i}^{T_{i+1}} \eta_t \, dB_t, \tag{21}$$

where B is a standard Wiener process independent of $\mathcal{F}$, and η_t satisfies:

$$\int_{T_i}^{T_{i+1}} \eta_t^2 dt = \frac{4}{c^2 c_1 T} \int_{T_i}^{T_{i+1}} (T_{i+1} - t)^2 \sigma_t^4 \, dt + \frac{8}{3c_1} \int_{T_i}^{T_{i+1}} (T_{i+1} - t)^2 \sigma_t^2 \, d\langle \sigma, \sigma \rangle_t. \tag{22}$$

Then, under Assumption (H), $\widehat{C}_i$ converges stably in law to a limiting random variable defined on an extension of the original probability space. That is,

$$\sqrt{n^b}\left(\widehat{P}_i - P_i\right) \xrightarrow{\mathscr{L}_{st}} \int_{T_{i-1}}^{T_i} \eta'_t \, dB_t, \tag{23}$$

where B is a standard Wiener process independent of $\mathcal{F}$, and η'_t satisfies:

$$\int_{T_{i-1}}^{T_i} \eta_t'^2 dt = \frac{4}{c^2 c_1 T} \int_{T_{i-1}}^{T_i} (t - T_{i-1})^2 \sigma_t^4 \, dt + \frac{8}{3c_1} \int_{T_{i-1}}^{T_i} (t - T_{i-1})^2 \sigma_t^2 \, d\langle \sigma, \sigma \rangle_t. \tag{24}$$

4.1 Normality demonstration

In the simulation, we use the Heston model to generate the log-price process X_t and the volatility process σ_t^2:

$$\begin{cases} dX_t = (\mu - \sigma_t^2/2)dt + \sigma_t dW_t \\ d\sigma_t^2 = \kappa(\theta - \sigma_t^2)dt + \eta\sigma_t(\rho dW_t + \sqrt{1-\rho^2} dV_t), \end{cases} \tag{25}$$

[9]Here, $Z_T^n \xrightarrow{u.c.p.} Z_T$ means that the sequence of stochastic processes Z_T^n converges in probability, locally uniformly in time, to a limit Z_T, that is, $\sup_{s \leq T} \mid Z_s^n - Z_s \mid \xrightarrow{\mathbb{P}} 0$ for all finite T.

where W and V are independent standard Brownian motions. We take the following parameter values: $\theta = 0.1, \eta = 0.5, \kappa = 5, \rho = -0.8$ and $\mu = 0.05$.

We simulate data using daily 23400 observations. We repeat each simulation run 2000 times. We take k_n in 3 different settings: $m_n n^{1/4}$, $0.1 m_n n^{0.4}$ and $0.1n$. The last setting is stretching the theorem to the boundary case.

Figure 1 presents the simulation results. The first and third rows present the densities of the estimated (black dashed line) P_i and C_i and the standard normal r.v. (red solid line). The corresponding Q-Q plots are provided underneath the density plots, respectively. In all three scenarios, both the density curves and the Q-Q plots demonstrate close fit to the standard normal distribution.

5 Market Microstructure Noise

In high-frequency financial applications, the presence of market microstructure noise in asset prices can be non-negligible. To deal with market microstructure noise, we employ pre-averaging. The contaminated log return process Y_t is observed every $\Delta t_{n,i} = T/n$ units of time, at times $0 = t_{n,0} < t_{n,1} < t_{n,2} < \cdots < t_{n,n} = T$. The noise term has the following structure:

Assumption 1.

$$Y_t = X_t + \epsilon_t, \textit{ where the } \epsilon_t\textit{'s are i.i.d. } N(0, a^2) \textit{ and } \epsilon_t \perp\!\!\!\perp W_t \textit{ and } B_t, \textit{ for all } t \geq 0. \tag{26}$$

Blocks are defined on a much less dense grid of $\tau_{n,i}$'s, also spanning $[0, T]$, so that

$$\text{block } i = \{t_{n,j} : \tau_{n,i} \leq t_{n,j} < \tau_{n,i+1}\} \tag{27}$$

(the last block, however, includes T). We define the block size, $M_{n,i}$, by

$$L_{n,i} = \#\{j : \tau_{n,i} \leq t_{n,j} < \tau_{n,i+1}\}. \tag{28}$$

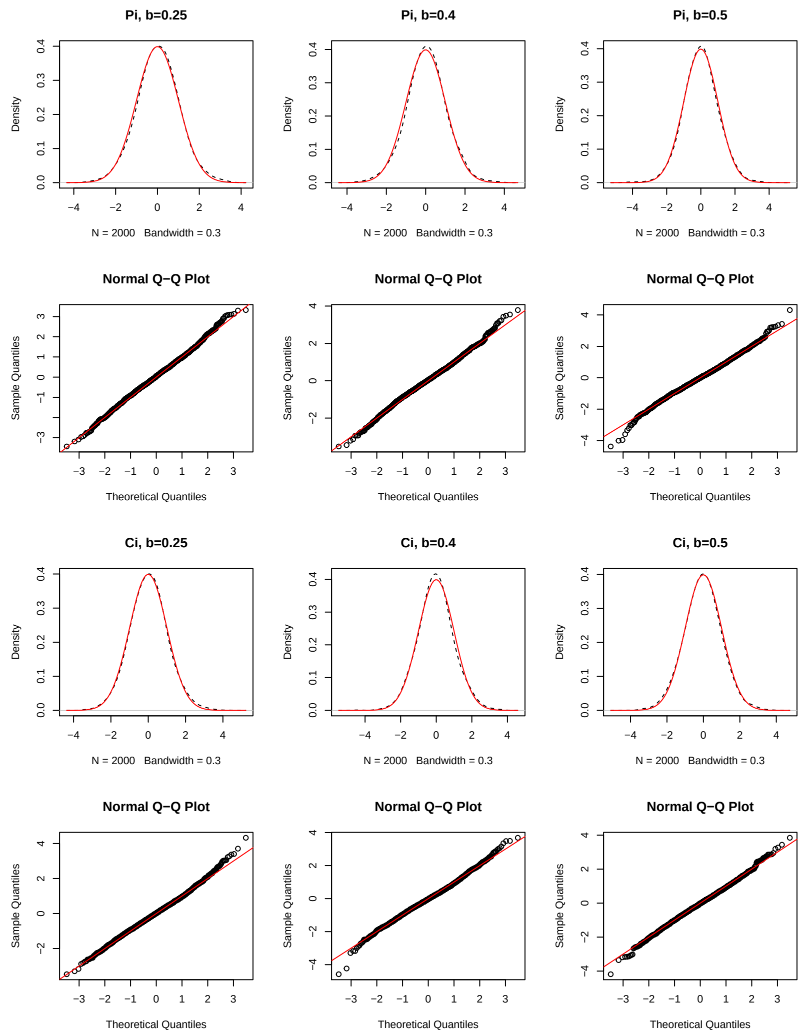

Figure 1: The density of the estimated P_i and C_i for different choice of window sizes. The first column provides the results corresponding to $k_n = m_n n^{0.25}$. The second column corresponds to $k_n = 0.25 m_n n^{0.4}$. The third column corresponds to $k_n = 0.1 m_n n^{0.5}$.

In principle, the block size can vary across the trading period $[0, T]$, but for this development we take $L_{n,i} = L$: 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 $[\tau_{n,i}, \tau_{n,i+1})$:

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

Let $J_n^-(j) = \{j-(m_n-1)L, j-m_nL, \dots, j-L\}$ if $j > m_nL$ and $J_n^+(j) = \{j+L, j+2L, \dots, j+$

$(m_n+1)L\}$ define two local windows in time of length $m_nL\Delta_n$ just before and after time $jL\Delta_n$. Then if $b<1/4$, we can define

$$
\begin{aligned}
\widehat{C}_i &= \sum_{\lambda=m_n+1}^{\lfloor n/L\rfloor-k_n-m_n} (T_{i+1}-t_{\lambda L})\,\widehat{a}_{t_{\lambda L}}\Delta^n_{\lambda L+1}\hat{X}, \quad \text{and} \\
\widehat{P}_{i+1} &= \sum_{\lambda=m_n+1}^{\lfloor n/L\rfloor-k_n-m_n} (t_{\lambda L}-T_i)\,\widehat{a}_{t_{\lambda L}}\Delta^n_{\lambda L+1}\hat{X} \quad \text{with} \\
\widehat{a}_{t_{\lambda L}} &= \frac{1}{k_nL\Delta_n}\sum_{j=\lambda+1}^{\lambda+k_n} \Delta^n_{jL}\hat{X}\,(\log(\widehat{\sigma}^2_{jL+})-\log(\widehat{\sigma}^2_{jL-})), \\
\widehat{\sigma}^2_{jL+} &= \frac{3}{2m_nL\Delta_n}\sum_{k\in J^+_n(jL)}(\Delta^n_k\hat{X})^2, \quad \widehat{\sigma}^2_{jL-} = \frac{3}{2m_nL\Delta_n}\sum_{k\in J^-_n(jL)}(\Delta^n_k\hat{X})^2.
\end{aligned}
\tag{29}
$$

REMARK 3. For $b\geq 1/4$, the estimators will be modified in the similar way as mentioned in Remark 2, by adding tail estimators. Denote $t'=T_{i+1}-(k_n+m_n)L\Delta_n$ and the tail estimators are defined by

$$
\begin{aligned}
\widehat{C}^e_i &= \sum_{\lambda=\lfloor n/L\rfloor-k_n-m_n+1}^{\lfloor n/L\rfloor} (T_{i+1}-t_{\lambda L})\,\widehat{a}_{t'}\Delta^n_{\lambda L+1}\hat{X}, \quad \text{and} \\
\widehat{P}^e_{i+1} &= \sum_{\lambda=\lfloor n/L\rfloor-k_n-m_n+1}^{\lfloor n/L\rfloor} (t_{\lambda L}-T_i)\,\widehat{a}_{t'}\Delta^n_{\lambda L+1}\hat{X} \quad \text{with} \\
\widehat{a}_{t'} &= \frac{1}{k_n\Delta_n}\sum_{j=\lfloor n/L\rfloor-k_n-m_n+1}^{\lfloor n/L\rfloor-m_n} \Delta^n_{jL}\hat{X}\,(\log(\widehat{\sigma}^2_{jL+})-\log(\widehat{\sigma}^2_{jL-})), \\
\widehat{\sigma}^2_{j+} &= \frac{3}{2m_nL\Delta_n}\sum_{k\in J^+_n(jL)}(\Delta^n_k\hat{X})^2, \quad \widehat{\sigma}^2_{j-} = \frac{3}{2m_nL\Delta_n}\sum_{k\in J^-_n(jL)}(\Delta^n_k\hat{X})^2.
\end{aligned}
\tag{30}
$$

The modified estimators again become $\widehat{C}_i+\widehat{C}^e_i$ and $\widehat{P}_{i+1}+\widehat{P}^e_{i+1}$. □

THEOREM 4. *For some positive constant c and c_1, let $L=\lfloor c_2\sqrt{n}\rfloor$, $m_n=\lfloor c\sqrt{n/L}\rfloor$ and $k_n=\lfloor c_1m_nn^{b/2}\rfloor$ with $0<b<\frac{1}{2}$. Under Assumption (H) in Section 3, we have*[10]

$$
\widehat{C}_i \xrightarrow{u.c.p.} C_i, \;\; \textit{and} \quad \widehat{P}_i \xrightarrow{u.c.p.} P_i, \;\; \textit{as } n\to\infty. \tag{31}
$$

[10]Here, $Z^n_T \xrightarrow{u.c.p.} Z_T$ means that the sequence of stochastic processes Z^n_T converges in probability, locally uniformly in time, to a limit Z_T, that is, $\sup_{s\leq T}\mid Z^n_s-Z_s\mid \xrightarrow{\mathbb{P}} 0$ for all finite T.

THEOREM 5. *For some positive constant c and c_1, let $L = \lfloor c_2\sqrt{n}\rfloor$, $m_n = \lfloor c\sqrt{n/L}\rfloor$ and $k_n = \lfloor c_1 m_n n^{b/2}\rfloor$ with $0 < b < \frac{1}{2}$. Then, under Assumption (H), $\widehat{C}_i$ converges stably in law to a limiting random variable defined on an extension of the original probability space. That is,*

$$\sqrt{n^{b/2}}\left(\widehat{C}_i - C_i\right) \xrightarrow{\mathscr{L}_{st}} \int_{T_i}^{T_{i+1}} \eta_t\, dB_t, \tag{32}$$

where B is a standard Wiener process independent of $\mathcal{F}$, and η_t satisfies:

$$\begin{aligned}\int_{T_i}^{T_{i+1}} \eta_t^2 dt =& \frac{4}{c^2 c_1 T}\int_{T_i}^{T_{i+1}} (T_{i+1}-t)^2\sigma_t^4\, dt + \frac{8}{3c_1}\int_{T_i}^{T_{i+1}} (T_{i+1}-t)^2\sigma_t^2\, d\langle\sigma,\sigma\rangle_t \\ &+ \frac{18a^2}{c^2 c_1 c_2^2 T^2}\int_{T_i}^{T_{i+1}} (T_{i+1}-t)^2\sigma_t^2 dt + \frac{18a^4}{c^2 c_1 c_2^4}.\end{aligned} \tag{33}$$

Then, under Assumption (H), $\widehat{C}_i$ converges stably in law to a limiting random variable defined on an extension of the original probability space. That is,

$$\sqrt{n^{b/2}}\left(\widehat{P}_i - P_i\right) \xrightarrow{\mathscr{L}_{st}} \int_{T_{i-1}}^{T_i} \eta_t'\, dB_t, \tag{34}$$

where B is a standard Wiener process independent of $\mathcal{F}$, and η_t' satisfies:

$$\begin{aligned}\int_{T_{i-1}}^{T_i} \eta_t'^2 dt =& \frac{4}{c^2 c_1 T}\int_{T_{i-1}}^{T_i} (t-T_{i-1})^2\sigma_t^4\, dt + \frac{8}{3c_1}\int_{T_{i-1}}^{T_i} (t-T_{i-1})^2\sigma_t^2\, d\langle\sigma,\sigma\rangle_t. \\ &+ \frac{18a^2}{c^2 c_1 c_2^2 T^2}\int_{T_{i-1}}^{T_i} (t-T_{i-1})^2\sigma_t^2 dt + \frac{18a^4}{c^2 c_1 c_2^4}.\end{aligned} \tag{35}$$

5.1 Normality demonstration

In the simulation, the Heston model 25 is adopted and the parameterization remains unchanged. The additional noise term is assumed to be an independent normal random variable, $N(0, 0.005^2)$. Due to the pre-averaging step, a larger sample size 10^6 is taken to show the behavior of the estimators. For brevity, in the case with microstructure noise, b is chosen as 0.25. For other choice of b, the simulation results are similar and thus omitted due to issue of space. The simulation results are presented in Figure 2, and the estimators exhibit a close fit to normal random variables.

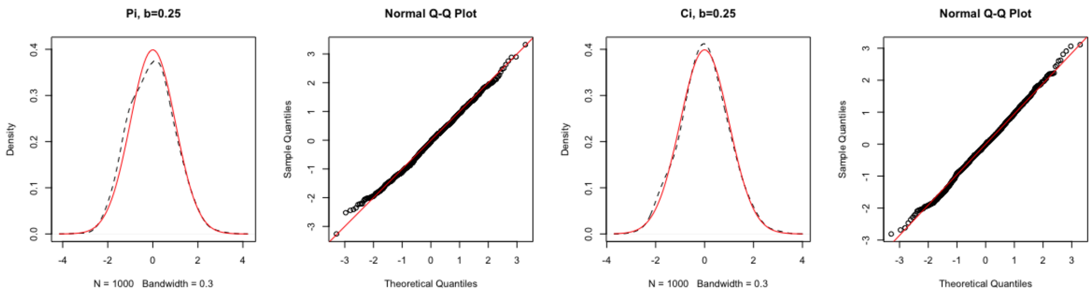

Figure 2: The density and QQ plots for the estimated P_i and C_i. The left two plots demonstrate the behavior of $\widehat{P}_i$. The remaining two plots demonstrate the behavior of $\widehat{C_i}$. Both are simulated under $b = 0.25$.

6 Simulation experiments for volatility estimation and prediction

The Heston model as in equation 25 is adopted in the simulation. To mimic real data, microstructure noise is included into the Heston model. The parameterization is the same as that in Section 4.1 and Section 5.1. The simulation is based on 1000 sample price path. Each price path generates 93600 data, corresponding to 4 day's returns observed at every second frequency. The estimators in the simulation are constructed based on equation 29.

In the first step, we study the projection of IV_i on to $\widetilde{IV}$ in fomula 4 as an alternative estimator of integrated volatility. Before studying the estimation performance, we starts from illustrating the theoretical results in formula 11 and 12 with the true values of all the covariates in Table 1, the true values of integrated volatility, C_i and P_i. Table 1 shows that all the coefficients are around 1 and significantly different from zeros, which is consistent with the theories of high frequency projection. Then we examine the estimation of integrated volatilities by substituting the estimated covariates as in formula 4. Table 2 demonstrates the regression results. As seen from this table, all estimated covariates are significant and have strong predictive powers for the true IV_i. In addition, the sum of squares contributed by the estimated covariates are comparable to those contributed by the true covariates.

Next, we move onto investigating the prediction of integrated volatility IV_i by PIV as in formula

	Estimate	P(> $\|t\|$)	SS explained	vif
$Intercept_i$	5.011×10^{-5}	8.52×10^{-15}		
IV_{i-1}	0.932	$< 2 \times 10^{-16}$	2.954×10^{-4}	1.019474
C_i	1.059	$< 2 \times 10^{-16}$	8.4686×10^{-5}	1.008998
P_i	1.028	$< 2 \times 10^{-16}$	3.7258×10^{-5}	1.024653

Table 1: Volatility prediction results of IV_i by IV_{i-1}, C_i and P_i with the true values of all the covariates. The first column provides the estimated coefficient for each covariate. The second column presents the p-value. The third column shows the sum of squares explained by each covariate. The last column checks the collinearity by vif.

	Estimate	P(> $\|t\|$)	SS explained	vif
$Intercept_i$	2.977×10^{-5}	0.000309		
$\widehat{IV}_{i-1}$	0.9624	$< 2 \times 10^{-16}$	2.9363×10^{-4}	1.023253
$\widehat{C}_i$	0.2442	8.47×10^{-7}	9.4166×10^{-7}	1.028287
$\widehat{P}_i$	0.09548	1.12×10^{-11}	1.3520×10^{-6}	1.021508

Table 2: Volatility prediction results of IV_i by $\widehat{IV}_{i-1}$, $\widehat{C}_i$ and $\widehat{P}_i$.

6. Similar to the previous study, we provide the prediction results in two scenarios, with true values of IV_{i-1} and P_i and the estimated values of IV_{i-1} and P_i. The results with all true covariates are presented in Table 3. Clearly, besides the well acknowledged predictor IV_{i-1}, P_i is a significant additional predictor for integrated volatilities. This significance can be seen by both the small p -value and the sum of square. Table 4 provides the results when the covariates are substituted by the estimated ones The coefficients are still significant and the sum of squares are of the similar size as those of the true covariates.

	Estimate	P(> $\|t\|$)	SS explained	vif
$Intercept_i$	4.379×10^{-5}	1.31×10^{-8}		
IV_{i-1}	0.9379	$< 2 \times 10^{-16}$	2.954×10^{-4}	1.017984
P_i	1.146	$< 2 \times 10^{-16}$	4.66×10^{-5}	1.017984

Table 3: Volatility prediction results of IV_i by IV_{i-1} and P_i with the true values of all the covariates. This should be compared to formula (6).

In addition to the prediction of true IV_i, the prediction of $\widehat{IV}_i$ by $\widehat{IV}_{i-1}$ and $\widehat{P}_i$ is also studied, since in application, true IV_i is not available and the meaningful prediction of integrated volatility

	Estimate	P(> $\|t\|$)	SS explained	vif
$Intercept_i$	3.235×10^{-5}	0.000104		
$\widehat{IV}_{i-1}$	0.957	$< 2 \times 10^{-16}$	2.9363×10^{-4}	1.009484
$\widehat{P}_i$	0.1029	3.66×10^{-13}	1.591×10^{-6}	1.009484

Table 4: Volatility prediction results of IV_i by $\widehat{IV}_{i-1}$ and $\widehat{P}_i$.

is the prediction of $\widehat{IV}_i$. Clearly, $\widehat{P}_i$ still contributes significantly to $\widehat{IV}_i$ on top of $\widehat{IV}_{i-1}$.

	Estimate	P(> $\|t\|$)	SS explained	vif
$Intercept_i$	3.405×10^{-5}	6.85×10^{-5}		
$\widehat{IV}_{i-1}$	0.9553	$< 2 \times 10^{-16}$	2.9306×10^{-4}	1.009484
$\widehat{P}_i$	0.09205	2.07×10^{-10}	1.2720×10^{-6}	1.009484

Table 5: Volatility prediction results of $\widehat{IV}_i$ by $\widehat{IV}_{i-1}$ and $\widehat{P}_i$.

The third step of the simulation illustrates how the combination of $\widehat{IV}_i$ and $\widetilde{IV}_i$ will improve the estimation of integrated volatility. For the sake of space, on the results with estimated covariates are presented in Table 2. With no argument, IV_i is the most significant predictor. But it is surprising to find that $\widetilde{IV}_i$ also explains quite a lot of variation in integrated volatilities.

	Estimate	P(> $\|t\|$)	SS explained	vif
$Intercept_i$	-3.152×10^{-9}	0.86271		
$\widehat{IV}_i$	0.9485	$< 2 \cdot 10^{-16}$	3.2293×10^{-4}	10.786885
$\widehat{IV}_{i-1}$	5.165×10^{-2}	9.81×10^{-14}	5.2942×10^{-8}	10.975811
$\widehat{C}_i$	3.552	0.00121	1.8293×10^{-8}	1.047763
$\widehat{P}_i$	1.455×10^{-2}	3.35×10^{-6}	3.0307×10^{-8}	1.058112

Table 6: Volatility prediction of true IV_i by the estimated $\widehat{IV}$, $\widehat{C}_i$ and $\widehat{P}_i$.

To further examine the performance of $\widetilde{IV}_i$, we also consider a time series model to incorporate dependence structure within sequence and extra noise terms. Results from VARMA(1,1) model are shown in Table 7. Even with extra MA noise, the prediction of IV_i still relies on both IV_{i-1} and P_i significantly. And with the extra MA part, C_i is not as significant as in the previous regression. But this is alright since in prediction, C_i is not observable.

Equation	Parameter	Estimate	Error	t Value	$Pr > \|t\|$
vol	CONST1	5.143×10^{-5}	1.044×10^{-5}	4.925	1.85×10^{-6}
	IV_{i-1}	0.9255	1.192×10^{-2}	77.634	$< 2 \times 10^{-16}$
	C_i	-0.2770	0.8676	-0.319	0.7495
	P_i	1.227	0.1632	7.518	5.55×10^{-14}
	MA1-1-1	0.218	0.03515	6.203	5.55×10^{-10}
	$MA1-1-2$	1.339	0.8683	1.542	0.1230
	$MA1-1-3$	-0.2736	0.1784	-1.533	0.1252
crl	CONST2	-5.666×10^{-6}	2.985×10^{-6}	-1.898	0.0577
	IV_{i-1}	6.247×10^{-3}	3.632×10^{-3}	1.720	0.0854
	C_i	0.5484	0.1267	4.329	1.50×10^{-5}
	P_i	-0.1769	0.1075	-1.644	0.1001
	$MA1-2-1$	-0.01535	0.02629	-0.584	0.5593
	$MA1-2-2$	0.5172	0.1269	-4.076	4.58×10^{-5}
	$MA1-2-3$	0.1327	0.1392	0.953	0.3406
prl	CONST3	-1.199×10^{-6}	3.594×10^{-6}	-0.334	0.7386
	IV_{i-1}	4.094×10^{-4}	3.684×10^{-6}	-0.111	0.9115
	C_i	0.2533	0.4419	0.573	0.5666
	P_i	1.624×10^{-2}	4.942×10^{-2}	0.329	0.7424
	$MA1-3-1$	0.01750	0.01257	1.393	0.1637
	$MA1-3-2$	0.1759	0.4448	0.395	0.6926
	$MA1-3-3$	0.04996	0.05763	0.867	0.3860

Table 7: VARMA(1,1) results for volatility estimation of IV_i with the true values of all the covariates.

7 A data illustration

In the empirical study, we employ Microsoft stock trades data from the New York Stock Exchange (NYSE TAQ). The years under study are 2007 through 2010. 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 the 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. To get rid of the impact of microstructure noise, we pre-average the log price over every 5-min period. Due to the limited number of observations after pre-averaging, the regression are conducted for the weekly integrated volatility. The regression results are provided in Table 8 and Table 9.

Table 8 presents the results for volatility prediction by the observed data in the past. It shows that in most of the years under study, both $\widehat{IV}_{i-1}$ and $\widehat{P}_i$ are significant in predicting next period's volatility. $\widehat{P}_i$ can explain comparable amount of variation according to the sum of squares in ANOVA, and sometimes even more as shown in year 2007 and 2010. In 2010, when $\widehat{IV}_{i-1}$ and $\widehat{P}_i$ are not as significant, the variability explained by $\widehat{P}_i$ is much higher than that explained by $\widehat{IV}_{i-1}$. The results suggest that $\widehat{P}_i$ is at least as powerful as $\widehat{IV}_{i-1}$ in forecasting volatilities.

		Microsoft 2007		
	Estimate	P($>\lvert t\rvert$)	SS explained	vif
$Intercept_i$	6.147×10^{-5}	0.796723		
$\widehat{IV}_{i-1}$	1.178	0.000243	4.2278×10^{-6}	6.673157
$\widehat{P}_i$	-0.271	0.001364	1.0953×10^{-5}	6.673157
		Microsoft 2008		
	Estimate	P($>\lvert t\rvert$)	SS explained	vif
$Intercept_i$	0.0009726	0.005037		
$\widehat{IV}_{i-1}$	0.6266049	7.13×10^{-8}	1.1555×10^{-4}	1.051665
$\widehat{P}_i$	-0.4755511	0.000731	4.8481×10^{-5}	1.051665
		Microsoft 2009		
	Estimate	P($>\lvert t\rvert$)	SS explained	vif
$Intercept_i$	0.0006008	0.13082		
$\widehat{IV}_{i-1}$	0.7594722	4.33×10^{-5}	3.2118×10^{-5}	1.855224
$\widehat{P}_i$	-0.2364149	0.00395	2.6384×10^{-5}	1.855224
		Microsoft 2010		
	Estimate	P($>\lvert t\rvert$)	SS explained	vif
$Intercept_i$	0.0004508	0.00383		
$\widehat{IV}_{i-1}$	0.3859716	0.10958	1.9382×10^{-7}	1.385599
$\widehat{P}_i$	-0.1290625	0.14372	8.6756×10^{-7}	1.398536

Table 8: Volatility prediction results of $\widehat{IV}_i$ by $\widehat{IV_{i-1}}$ and $\widehat{P}_i$ in empirical study .

Table 9 provides the results when $\widehat{C}_i$ is added into the estimation, so that it gives an alternative estimators of integrated volatilities by $\widetilde{IV}$ from equation 4. The empirical results show that almost

Microsoft 2007				
	Estimate	P(> $\|t\|$)	SS explained	vif
$Intercept_i$	4.548×10^{-4}	1.85×10^{-5}		
$\widehat{IV}_{i-1}$	0.2674	0.0119	5.8×10^{-8}	3.033585
$\widehat{C}_i$	0.1499	$< 2e - 16$	2.4562×10^{-4}	1.005309
$\widehat{P}_i$	-0.1494	0.0199	1.1930×10^{-6}	3.030491
Microsoft 2008				
	Estimate	P(> $\|t\|$)	SS explained	vif
$Intercept_i$	0.0011450	0.019964		
$\widehat{IV}_{i-1}$	0.55987711	2.83×10^{-6}	1.6637×10^{-4}	1.017429
$\widehat{C}_i$	0.2723415	0.000181	8.7191×10^{-5}	1.018063
$\widehat{P}_i$	-0.3065788	0.030991	2.9448×10^{-5}	1.032872
Microsoft 2009				
	Estimate	P(> $\|t\|$)	SS explained	vif
$Intercept_i$	0.0007874	0.016751		
$\widehat{IV}_{i-1}$	0.5158307	0.000801	3.1989×10^{-5}	2.088067
$\widehat{C}_i$	0.3472574	6.57×10^{-6}	7.0843×10^{-5}	1.194996
$\widehat{P}_i$	-0.0977630	0.159443	3.7930×10^{-6}	2.206918
Microsoft 2010				
	Estimate	P(> $\|t\|$)	SS explained	vif
$Intercept_i$	0.0006369	6.53×10^{-7}		
$\widehat{IV}_{i-1}$	0.0489809	0.5339	1.276×10^{-6}	1.385599
$\widehat{C}_i$	0.2508148	$< 2 \times 10^{-16}$	4.5199×10^{-5}	1.010995
$\widehat{P}_i$	0.0747220	0.0815	8.2400×10^{-7}	1.398536

Table 9: Volatility prediction results $\widehat{IV}$, $\widehat{C}_i$ and $\widehat{P}_i$ in empirical study .

over every year, the $\widehat{C}_i$ and $\widehat{P}_i$ have significant contribution to the volatility estimation. There are significant extra powers in addition to the ones from the previous period's integrated volatility. When the previous period's integrated volatility is not significant in the prediction, the $\widehat{C}_i$ and $\widehat{P}_i$ still contribute considerably in explaining the variation in the integrated volatility, such as in year 2010.

8 Conclusion

In this paper, we have shown that (1) the leverage effect has a rôle to play in estimating and forecasting volatility, and (2) that it permits the extraction of information from previous days' data.

This is the first theoretical result to this effect that does not rely on stationarity or other additional conditions. We have also seen that the resulting method improves estimation and prediction of volatility, both in simulation and in data. The improvement for real data is particularly striking, as we have seen in Table 9 in Section 7.

REFERENCES

ALTMEYER, R. AND M. BIBINGER (2015): "Functional stable limit theorems for quasi-efficient spectral covolatility estimators," *Stochastic Processes and Their Applications*, 125, 4556–4600.

ANDERSEN, T., T. BOLLERSLEV, AND N. MEDDAHI (2005): "Analytic Evaluation of Volatility Forecasts," *Econometrica*, 73, 279–296.

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

——— (1998b): "Deutsche Mark-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements, and Longer Run Dependencies," *Journal of Finance*, 53, 219–265.

ANDERSEN, T. G., T. BOLLERSLEV, F. X. DIEBOLD, AND P. LABYS (2001): "The Distribution of Realized Exchange Rate Volatility," *Journal of the American Statistical Association*, 96, 42–55.

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

BARNDORFF-NIELSEN, O. E. (2004): "Power and Bipower Variation with Stochastic Volatility and Jumps," *Journal of Financial Econometrics*, 2, 1–37.

BARNDORFF-NIELSEN, O. E., P. R. HANSEN, A. LUNDE, AND N. SHEPHARD (2008): "Designing realised kernels to measure ex-post variation of equity prices in the presence of noise," *Econometrica*, 76(6), 1481–1536.

BARNDORFF-NIELSEN, O. E. AND N. SHEPHARD (2001): "Non-Gaussian Ornstein-Uhlenbeck-Based Models And Some Of Their Uses In Financial Economics," *Journal of the Royal Statistical Society, B*, 63, 167–241.

——— (2002): "Econometric Analysis of Realized Volatility and Its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, B*, 64, 253–280.

Bibinger, M. and M. Reiss (2014): "Spectral Estimation of Covolatility from Noisy Observations Using Local Weights," *Scandinavian Journal of Statistics*, 41(1), 23–50.

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.

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

Dacorogna, M. M., R. Gençay, U. Müller, R. B. Olsen, and O. V. Pictet (2001): *An Introduction to High-Frequency Finance*, San Diego: Academic Press.

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

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

Ghysels, E., P. A. Mykland, and E. Renault (2012): "In-sample Asymptotics and Across-sample Efficiency Gains for High Frequency Data Statistics," Tech. rep., University of Chicago.

Jacod, J. (1979): *Calcul stochastique et problèmes de martingales*, Lecture Notes in Mathematics 714, Springer-Verlag.

——— (2012): "Statistics and high frequency data," in *Statistical Methods for Stochastic Differential Equations*, ed. by A. L. Matthieu Kessler and M. Sorensen, Taylor and Francis, 191–310.

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

Jacod, J., M. Podolskij, and M. Vetter (2009b): "Limit theorems for moving averages of discretized processes plus noise," *Annals of Statistics*, 38, 1478–1545.

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

——— (2012): *Discretization of Processes*, New York: Springer-Verlag, first ed.

Meddahi, N. (2002): "A theoretical comparison between integrated and realized volatility," *Journal of Applied Econometrics*, 17, 479–508.

Mykland, P. and L. Zhang (2016): "Between data cleaning and Inference: Pre-Averaging and other robust estimators of the efficient price," *Journal of Econometrics*, 194, 242–262.

Mykland, P. A. (2012): "A Gaussian Calculus for Inference from High Frequency Data," *Annals of Finance*, 8, 235–258.

Mykland, P. A. and L. Zhang (2006): "ANOVA for Diffusions and Itô Processes," *Annals of Statistics*, 34, 1931–1963.

——— (2008): "Inference for Volatility Type Objects and Implications for Hedging," *Statistics and its Interface*, 1, 255–278.

——— (2009): "Inference for continuous semimartingales observed at high frequency," *Econometrica*, 77, 1403–1455.

——— (2017): "Assessment of Uncertainty in High Frequency Data: The Observed Asymptotic Variance," *Econometrica*, 85, 197231.

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

——— (2009b): "Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps," *Bernoulli*, 15, 634–658.

Wang, C. D. and P. A. Mykland (2014): "The Estimation of Leverage Effect with High Frequency Data," *Journal of the American Statistical Association*, 109, 197–215.

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.

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

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

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

——— (2012): "Implied and realized volatility: Empirical model selection," *Annals of Finance*, 8, 259–275.

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 Preliminary results

A.1 Localization

As shown in, for example, Jacod and Protter (2012), localization is a simple but very powerful standard procedure to prove limit theorems for discretized processes over a finite time interval. Adopting the localization procedure, it is sufficient to prove our results only under a stronger version of Assumption (H), and the same results will remain valid under the original Assumption (H). In particular, we can strengthen Assumption (H) by replacing the locally boundedness conditions by boundedness, and we only need to prove our results under the boundedness condition. More precisely, we set

Assumption (SH): We have (H) and, for some constant Λ and all (ω, t, x),

$$\left.\begin{array}{lll} |\mu_t(\omega)| \leq \Lambda, & |\sigma_t(\omega)| \leq \Lambda, & |X_t(\omega)| \leq \Lambda; \\ |\tilde{\mu}_t(\omega)| \leq \Lambda, & |a_t(\omega)| \leq \Lambda, & |b_t(\omega)| \leq \Lambda; \\ |\delta(\omega,t,x)| \leq \Lambda(\gamma(x) \wedge 1), & \multicolumn{2}{l}{|\tilde{\delta}(\omega,t,x)| \leq \Lambda(\tilde{\gamma}(x) \wedge 1);} \\ \multicolumn{3}{l}{\text{the coefficients of } a_t \text{ are also bounded by } \Lambda.} \end{array}\right\} \tag{A.1}$$

To prove the consistency and CLT, we will consider a decomposition of the difference between the estimators and the parameters. First, we will define two intermedia processes $\widetilde{C}_i$ and $\widetilde{P}_i$ as follows:

$$\begin{aligned} \widetilde{C}_i &= \sum_{\lambda=m_n+1}^{n-k_n-m_n} (T_{i+1} - t_\lambda)\, \widetilde{a}_{t_\lambda} \Delta_{\lambda+1}^n X \text{ and} \\ \widetilde{P}_{i+1} &= \sum_{\lambda=m_n+1}^{n-k_n-m_n} (t_\lambda - T_i)\, \widetilde{a}_{t_\lambda} \Delta_{\lambda+1}^n X, \text{ with} \\ \widetilde{a}_{t_\lambda} &= \frac{1}{k_n \Delta_n} \sum_{j=\lambda+1}^{\lambda+k_n} \Delta_j^n X\, \Delta_j^n \log \sigma^2. \end{aligned} \tag{A.2}$$

From the definition, $\widetilde{C_i}$ and $\widetilde{P}_i$ can be viewed as intermedia estimators of C_i and P_i by substituting true spot volatilities for the estimated ones. With $\widetilde{C_i}$ and $\widetilde{P}_i$, we have the following decomposition:

$$\begin{aligned} \widehat{C_i} - C_i &= (\widehat{C_i} - \widetilde{C_i}) + (\widetilde{C_i} - C_i) \text{ and} \\ \widehat{P}_{i+1} - P_{i+1} &= (\widehat{P}_{i+1} - \widetilde{P}_{i+1}) + (\widetilde{P}_{i+1} - P_{i+1}) \end{aligned} \tag{A.3}$$

We shall prove that the two terms on the right hand side will converge to zero in probability for the consistency. In addition, they converge stably to asymptotic random variables up to proper scaling with the correct convergence rate. In the proof of consistency and CLT, we will mainly discuss $\widehat{P}_i$ since the corresponding proofs for $\widehat{C_i}$ can be argued from analogy.

$$\widehat{P}_{i+1} - \widetilde{P}_{i+1} = \sum_{\lambda=m_n+1}^{n-k_n-m_n} (t_\lambda - T_i)\Delta_{\lambda+1}^n X \frac{1}{k_n\Delta_n} \sum_{j=\lambda+1}^{\lambda+k_n} \Delta_j^n X \, \left(\log(\widehat{\sigma}_{j+}^2) - \log(\widehat{\sigma}_{j-}^2) - \Delta_j^n \log\sigma^2\right)$$

For simplicity, the second summation will be denoted as

$$[\widehat{X, \log\sigma}^2]_\lambda - [\widetilde{X, \log\sigma}^2]_\lambda = \sum_{j=\lambda+1}^{\lambda+k_n} \Delta_j^n X \, \left(\log(\widehat{\sigma}_{j+}^2) - \log(\widehat{\sigma}_{j-}^2) - \Delta_j^n \log\sigma^2\right), \tag{A.4}$$

A.2 Auxiliary Results

LEMMA 1. *Let $u_n = n^{b'\wedge(1-b')}$, where b satisfies condition*

$$\frac{1}{K} \le k_n \Delta_n^{b'} \le K, \quad \textit{with } 0 < b' < 1, \tag{A.5}$$

for some positive constant K. At any time t_i we have

$$\sqrt{u_n}\left(\log(\widehat{\sigma}_{i+}^2) - \log(\sigma_{i+}^2), \log(\widehat{\sigma}_{i-}^2) - \log(\sigma_{i-}^2)\right) \xrightarrow{\mathscr{L}_{st}} (V_i^+, V_i^-), \tag{A.6}$$

where (V_i^+, V_i^-) is a vector of normal random variables independent of $\mathcal{F}$. They have zero $\mathcal{F}$-conditional covariance and

$$\mathbb{E}((V_i^{\pm})^2|\mathcal{F}) = \frac{2}{c}1_{\{b'\in(0,1/2]\}} + \frac{c}{3\sigma_{i\pm}^4}\Big(\frac{d\langle\sigma^2,\sigma^2\rangle_t}{dt}\Big|_{t=t_i^{\pm}}\Big)1_{\{b'\in[1/2,1)\}}. \tag{A.7}$$

B Proof of Theorem 3

$$\widehat{P}_{i+1} - P_{i+1} = (\widehat{P}_{i+1} - \widetilde{P}_{i+1}) + (\widetilde{P}_{i+1} - P_{i+1}) = V_t^n + D(1)_t^n + D(2)_t^n + D(3)_t^n, \tag{B.8}$$

where

$$V_t^n = \sum_{\lambda=m_n+1}^{n-k_n-m_n} (t_\lambda - T_i)\Delta_{\lambda+1}^n X \frac{1}{k_n\Delta_n} \sum_{j=\lambda+1}^{\lambda+k_n} \Delta_j^n X \left(\log(\widehat{\sigma}_{j+}^2) - \log(\widehat{\sigma}_{j-}^2) - \Delta_j^n \log(\sigma^2)\right),$$

$$D(1)_t^n = -\sum_{\lambda=0}^{m_n} (t_\lambda - T_i)\Delta_{\lambda+1}^n X \frac{1}{k_n\Delta_n} \sum_{j=\lambda+1}^{\lambda+k_n} \Delta_j^n X \Delta_j^n \log\sigma^2$$

$$D(2)_t^n = \sum_{\lambda=1}^{n-k_n-m_n} (t_\lambda - T_i)\Delta_{\lambda+1}^n X \frac{1}{k_n\Delta_n} \sum_{j=\lambda+1}^{\lambda+k_n} \Delta_j^n X \Delta_j^n \log\sigma^2 - \int_{T_i}^{T_{i+1}-(k_n+m_n)\Delta_n} (t-T_i)\, 2a_t\, dX_t$$

$$D(3)_t^n = \widehat{P}_{i+1}^e \cdot 1_{b\geq 1/4} - \int_{T_{i+1}-(k_n+m_n)\Delta_n}^{T_{i+1}} (t-T_i)\, 2a_t\, dX_t$$

Step 1. Let $u_n = n^b$ and $b < 1/4$. In this step we analyze $\sqrt{u_n}\, V_t^n$. Define

$$\begin{aligned}\xi_\lambda^n =& \sqrt{u_n}(t_\lambda - T_i)\Delta_{\lambda+1}^n X \frac{1}{k_n\Delta_n} \sum_{j=\lambda+1}^{\lambda+k_n} \Delta_j^n X \left(\log(\widehat{\sigma}_{j+}^2) - \log(\widehat{\sigma}_{j-}^2) - \Delta_j^n \log(\sigma^2)\right) \\ =& \sqrt{u_n}(t_\lambda - T_i)\Delta_{\lambda+1}^n X \frac{[\widehat{X, \log\sigma^2}]_\lambda - [\widetilde{X, \log\sigma^2}]_\lambda}{k_n\Delta_n}\end{aligned}$$

The variable ξ_i^n has a vanishing $\mathcal{F}_{(i-1)\Delta_n}$-conditional expectation, but it is not $\mathcal{F}_{i\Delta_n}$-measurable. To induce "some conditional independence" of the successive summands, we split the sum over i into big blocks of size $\tilde{m}k_n$ ($\tilde{m}$ will eventually go to infinity, to ensure that the summation over these big blocks is asymptotically equivalent to the summation over all blocks), separated by small blocks of size $2k_n$; cf. Section 12.2.4 of Jacod and Protter (2012). The condition on $\tilde{m}$ is $\tilde{m} \to \infty$.

More specifically, define $I(\tilde{m}, n, l) = (l-1)(\tilde{m}+2)k_n + 1$. Then the l-th big block contains ξ_λ^n for all i between $I(\tilde{m}, n, l) + k_n + 1$ and $I(\tilde{m}, n, l) + (\tilde{m}+1)k_n$, and the total number of such blocks is $l_n(\tilde{m}, t) = \lfloor \frac{\lfloor t/\Delta_n \rfloor - 1}{(\tilde{m}+2)k_n} \rfloor$. Let

$$\xi(\tilde{m})_\lambda^n = \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \xi_{I(\tilde{m},n,\lambda)+r}, \qquad Z(\tilde{m})_t^n = \sum_{i=1}^{l_n(\tilde{m},t)} \xi(\tilde{m})_\lambda^n,$$

$$\widetilde{\xi}(\tilde{m})_\lambda^n = \sum_{r=-k_n}^{k_n} \xi_{I(\tilde{m},n,\lambda)+r}, \qquad \widetilde{Z}(\tilde{m})_t^n = \sum_{i=2}^{l_n(\tilde{m},t)} \widetilde{\xi}(\tilde{m})_\lambda^n.$$

So $\sqrt{u_n}\, V_t^n = Z(\tilde{m})_t^n + \widetilde{Z}(\tilde{m})_t^n$. We are going to show that $\widetilde{Z}(\tilde{m})_t^n$ is asymptotically negligible first. By successive conditioning, we get

$$\begin{aligned}
\mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}\big(\xi_{I(\tilde{m},n,\lambda)+r}\big) &= \mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}\left(\sqrt{u_n}(t_\lambda - T_i)\Delta_{\lambda+1}^n X\frac{\widehat{[X,\log\sigma^2]}_\lambda - \widetilde{[X,\log\sigma^2]}_\lambda}{k_n\Delta_n}\right)\\
&= \mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}\big((t_\lambda - T_i)\Delta_{\lambda+1}^n X\cdot O_p(\psi_n)\big)\\
&= O_p(\psi_n\Delta_n),\\
\mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}\big((\xi_{I(\tilde{m},n,\lambda)+r})^2\big) &= \mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}\left(u_n(t_\lambda - T_i)^2\Delta_{\lambda+1}^n X^2\big(\frac{\widehat{[X,\log\sigma^2]}_\lambda - \widetilde{[X,\log\sigma^2]}_\lambda}{k_n\Delta_n}\big)^2\right)\\
&= \rho_{I(\tilde{m},n,\lambda)-k_n-1}\Delta_n + o_p(\Delta_n).
\end{aligned}$$

As before, $\psi_n \to 0$, and it may change from line to line. Observe that, for any given n and $\tilde{m}$, there is no overlap among the sequence $\widetilde{\xi}(\tilde{m})_i^n$. Then it is easy to verify that

$$\mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}(\widetilde{\xi}(\tilde{m})_i^n) = O_p(k_n\Delta_n\psi_n),$$

and

$$\begin{aligned}
\mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}(\widetilde{\xi}(\tilde{m})_i^n)^2 &= \mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}\left(\sum_{r=-k_n}^{k_n}(\xi_{I(\tilde{m},n,\lambda)+r})^2\right)\\
&\quad + \mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}\left(\sum_{r,j=-k_n}^{k_n} 1_{\{j\neq r\}}\xi_{I(\tilde{m},n,\lambda)+r}\xi_{I(\tilde{m},n,\lambda)+j}\right)\\
&= O_p(k_n\Delta_n) + O_p(k_n^2\Delta_n^2).
\end{aligned}$$

The first term on the right-hand side of the expression above is $O_p(k_n\Delta_n)$. For the second term, when $j > r$, by successive conditioning, we get

$$\begin{aligned}
&\mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}\big(\xi_{I(\tilde{m},n,\lambda)+r}\xi_{I(\tilde{m},n,\lambda)+j}\big)\\
&= \mathbb{E}_{I(\tilde{m},n,\lambda)-k_n-1}\big(\Delta_r^n X O_p(1)\,\mathbb{E}_{I(\tilde{m},n,\lambda)+j-1}(\Delta_j^n X O_p(1))\big),
\end{aligned}$$

where $O_p(1)$ comes from the standardized estimation error of spot volatility. Irrespective of whether $\Delta_j^n X'$ is correlated with its associated $O_p(1)$ or not, the conditional expectation of their product is $O_p(\Delta_n)$. The same argument applies to $\Delta_r^n X'$. Hence, the above result readily follows. Then, as long as $\tilde{m}$ goes to infinity, Lemma 4.1 in Jacod (2012) yields that $\widetilde{Z}(\tilde{m})_t^n \xrightarrow{u.c.p.} 0$.

Next, notice that the variable $\xi(\tilde{m})^n_i$ has vanishing $\mathcal{F}_{I(\tilde{m},n,i)}$-conditional expectation, and is $\mathcal{F}_{I(\tilde{m},n,i+1)}$-measurable. That is, it behaves like a martingale difference. We are going to prove that

$$\begin{cases} \sum_{\lambda=1}^{l_n(\tilde{m},t)} \mathbb{E}\Big(\big(\xi(\tilde{m})^n_\lambda\big)^2 \Big| \mathcal{F}_{I(\tilde{m},n,\lambda)}\Big) \xrightarrow{\mathbb{P}} \int_{T_i}^{T_{i+1}} \eta_s^2 ds, \\ \sum_{\lambda=1}^{l_n(\tilde{m},t)} \mathbb{E}\Big(\big(\xi(\tilde{m})^n_\lambda\big)^4 \Big| \mathcal{F}_{I(\tilde{m},n,\lambda)}\Big) \xrightarrow{\mathbb{P}} 0, \\ \sum_{\lambda=1}^{l_n(\tilde{m},t)} \mathbb{E}\Big(\xi(\tilde{m})^n_\lambda \; \Delta^n_{i,\tilde{m}} M \Big| \mathcal{F}_{I(\tilde{m},n,\lambda)}\Big) \xrightarrow{\mathbb{P}} 0, \end{cases} \tag{B.9}$$

where $\Delta^n_{i,\tilde{m}} M = M_{(I(\tilde{m},n,\lambda)+(\tilde{m}+1)k_n)\Delta_n} - M_{(I(\tilde{m},n,\lambda)+k_n+1)\Delta_n}$. The remaining difficulty is that the ξ^n_λ may have overlaps within the big block. To deal with this, recall that $I^-_n(i) = \{i-k_n, \ldots, i-1\}$ if $i > k_n$ and $I^+_n(i) = \{i+1, \ldots, i+k_n\}$, which define two local windows of length $k_n\Delta_n$ just before and after the time point $i\Delta_n$. Let $I^{\pm}_n(i)$ be the union of them. Furthermore, let

$$J(\tilde{m}, n, \lambda, j) = \{I(n, \tilde{m}, \lambda) + k_n + 1, \ldots, I(\tilde{m}, n, \lambda) + (\tilde{m}+1)k_n\} \setminus (I^{\pm}_n(j) \cup j).$$

With these notations, we can decompose the conditional second moment of $\xi(\tilde{m})^n_\lambda$, as follows:

$$\begin{aligned} &\mathbb{E}\Big(\big(\xi(\tilde{m})^n_\lambda\big)^2 \Big| \mathcal{F}_{I(\tilde{m},n,\lambda)}\Big) \\ =& \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \sum_{j=k_n+1}^{(\tilde{m}+1)k_n} \mathbb{E}\big(\xi_{I(\tilde{m},n,\lambda)+r} \; \xi_{I(\tilde{m},n,\lambda)+j} | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) \\ =& \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \sum_{j=r} \mathbb{E}\big(\xi_{I(\tilde{m},n,\lambda)+r} \; \xi_{I(\tilde{m},n,\lambda)+j} | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) + \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \sum_{j\in I^{\pm}_n(r)} \mathbb{E}\big(\xi_{I(\tilde{m},n,\lambda)+r} \; \xi_{I(\tilde{m},n,\lambda)+j} | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) \\ &+ \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \sum_{j\in J(\tilde{m},n,\lambda,r)} \mathbb{E}\big(\xi_{I(\tilde{m},n,\lambda)+r} \; \xi_{I(\tilde{m},n,\lambda)+j} | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) \\ =:& \; H(\tilde{m},1)^n_\lambda + H(\tilde{m},2)^n_\lambda + H(\tilde{m},3)^n_\lambda. \end{aligned}$$

From Lemmas 1 , we have

$$\sum_{\lambda=1}^{l_n(\tilde{m},t)} H(\tilde{m},1)_\lambda^n = \sum_{\lambda=1}^{l_n(\tilde{m},t)} \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \mathbb{E}\big((\xi_{I(\tilde{m},n,\lambda)+r})^2 | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) \xrightarrow{\mathbb{P}} \int_{T_i}^{T_{i+1}} \eta_s^2 ds,$$

$$\begin{aligned}\sum_{\lambda=1}^{l_n(\tilde{m},t)} H(\tilde{m},2)_\lambda^n &= \sum_{\lambda=1}^{l_n(\tilde{m},t)} \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \sum_{j\in I_n(r)} \mathbb{E}\big(\xi_{I(\tilde{m},n,\lambda)+r}\ \xi_{I(\tilde{m},n,\lambda)+j} | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) \\ &\leq \sum_{\lambda=1}^{l_n(\tilde{m},t)} K\tilde{m}k_n^2\Delta_n^2 \to 0.\end{aligned}$$

Next, notice that, when $j \in J(\tilde{m},n,i,r)$, there is no overlap between $\xi_{I(\tilde{m},n,i)+r}$ and $\xi_{I(\tilde{m},n,i)+j}$. Hence, by successive conditioning, we obtain

$$\begin{aligned}\sum_{\lambda=1}^{l_n(\tilde{m},t)} H(\tilde{m},3)_i^n &= \sum_{\lambda=1}^{l_n(\tilde{m},t)} \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \sum_{j\in J(\tilde{m},n,\lambda,r)} \mathbb{E}\big(\xi_{I(\tilde{m},n,\lambda)+r}\ \xi_{I(\tilde{m},n,\lambda)+j} | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) \\ &\leq \sum_{\lambda=1}^{l_n(\tilde{m},t)} K\tilde{m}^2k_n^2\Delta_n^2\psi_n \to 0.\end{aligned}$$

The calculation of the fourth moments is even more tedious; we present partial results and omit the remainder of the calculations for brevity:

$$\begin{aligned}&\sum_{\lambda=1}^{l_n(\tilde{m},t)} \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \mathbb{E}\big((\xi_{I(\tilde{m},n,\lambda)+r})^4 | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) \\ =& \sum_{\lambda=1}^{l_n(\tilde{m},t)} \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \mathbb{E}\big((\Delta^n_{I(\tilde{m},n,\lambda)+r}X)^4 O_p(1) | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) \\ \leq& \sum_{\lambda=1}^{l_n(\tilde{m},t)} \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} K\Delta_n^2 = Kt\Delta_n \to 0, \\ &\sum_{\lambda=1}^{l_n(\tilde{m},t)} \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \sum_{j\neq r} \mathbb{E}\big((\xi_{I(\tilde{m},n,\lambda)+r})^2\ (\xi_{I(\tilde{m},n,\lambda)+j})^2 | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) \\ =& \sum_{\lambda=1}^{l_n(\tilde{m},t)} \sum_{r=k_n+1}^{(\tilde{m}+1)k_n} \sum_{j\neq r} \mathbb{E}\big((\Delta^n_{I(\tilde{m},n,\lambda)+r}X)^2 (\Delta^n_{I(\tilde{m},n,\lambda)+j}X)^2 O_p(1) | \mathcal{F}_{I(\tilde{m},n,\lambda)}\big) \\ \leq& \sum_{\lambda=1}^{l_n(\tilde{m},t)} K\tilde{m}^2k_n^2\Delta_n^2 = Kt\tilde{m}k_n\Delta_n \to 0.\end{aligned}$$

As for the last equation in (B.10), we first note that it holds when M is orthogonal to W and B. Besides, when $M = W$ or $M = B$, according to the proof of Lemma 1, one can verify by successive conditioning that

$$\sum_{\lambda=1}^{l_n(\tilde{m},t)} \mathbb{E}\Big(\xi(\tilde{m})^n_\lambda\ \Delta^n_{i,\tilde{m}} M \Big| \mathcal{F}_{I(\tilde{m},n,\lambda)}\Big) \leq O_p(\sqrt{u_n \Delta_n}) \xrightarrow{\mathbb{P}} 0.$$

In any other case, M can be decomposed into the sum of two components, one driven by W and B and the other orthogonal to B and W. Thus the result readily follows.

Step 2. In this step, we are going to prove the following result for $j = 1, 2, 3$:

$$\sqrt{u_n}\ D(j)^n_t \xrightarrow{u.c.p.} 0.$$

For $D(2)^n_t$, we have by Itô's formula:

$$\Delta^n_j X \Delta^n_j \log \sigma^2 = \int_j^{j+1} (X_t - X_j)\, d \log \sigma^2_t + \int_j^{j+1} (\log \sigma^2_t - \log \sigma^2_j)\, dX_t + \int_j^{j+1} 2a_t\, dt$$

Consequently, we obtain

$$\mathbb{E}_\lambda\big(\sum_{j=\lambda+1}^{\lambda+k_n} \Delta^n_j X \Delta^n_j \log \sigma - 2a_\lambda k_n \Delta t\big) = K k_n \Delta_n^{3/2},$$
$$\mathbb{E}_\lambda\big((\sum_{j=\lambda+1}^{\lambda+k_n} \Delta^n_j X \Delta^n_j \log \sigma - 2a_\lambda k_n \Delta t)^2\big) \leq K k_n \Delta_n^2.$$

Let $\zeta^n_\lambda = \sqrt{u_n}(t_\lambda - T_i)\Delta^n_{\lambda+1} X \frac{1}{k_n \Delta_n}(\sum_{j=\lambda+1}^{\lambda+k_n} \Delta^n_j X \Delta^n_j \log \sigma - 2a_\lambda k_n \Delta t)$. Note that ζ^n_λ is $\mathcal{F}_\lambda$-measurable. Then the above equations yield

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{\lambda-1}\big(\zeta^n_\lambda\big) = K\sqrt{u_n \Delta_n} \longrightarrow 0,$$
$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}_{\lambda-1}\big((\zeta^n_\lambda)^2\big) \leq K\ u_n/k_n \longrightarrow 0.$$

Hence, Lemma 4.1 in Jacod (2012) implies that $\sqrt{u_n}\ D(2)^n_t \xrightarrow{u.c.p.} 0$.

For $\sqrt{u_n}\ D(1)^n_t$, by the Cauchy-Schwarz inequality, we obtain

$$\mathbb{E}\big(|\Delta^n_i X \Delta^n_i \sigma^2|\big) \leq \sqrt{\mathbb{E}\big((\Delta^n_i X)^2\big)\ \mathbb{E}\big((\Delta^n_i \sigma^2)^2\big)} \leq K\Delta_n,$$

for some constant K. Then the first term $\sqrt{u_n}\ D(1)_t^n \xrightarrow{u.c.p.} 0$ readily follows from the fact that

$$\limsup_{n\to\infty} \sqrt{u_n}\ \mathbb{E}(|D(1)_t^n|) \le \limsup_{n\to\infty} K\sqrt{k_n\Delta_n} = 0.$$

Since $D(3)_t^n$ is a Itô integral, its quadratic variation can be calculated as:

$$\begin{aligned} u_n[D(3)]_t^n &= \int_{T_{i+1}-(k_n+m_n)\Delta_n}^{T_{i+1}} (t-T_i)^2\, 4a_t^2\, d[X,X]_t \\ &\le K u_n(k_n+m_n)\Delta_n T^2 \to 0. \end{aligned}$$

Then $\sqrt{u_n}\ D(3)_t^n \xrightarrow{u.c.p.} 0$.

When $b \ge 1/4$, the proof will go through the similar analysis from step 1 to step 2. But in the first step, besides considering V_t^n, $D(3)_t^n$ should be added because of the tail estimator. Therefore, in step 1, the asymptotic variance will be derived from $\sqrt{u_n}(V_t^n + D(3)_t^n)$. We will only provide some details in step 1. For step 2, all the convergence can be verified in the similar manner as in the proof under $b < 1/4$.

The big block and small block technique will still be applied and the last big block will be the block $[n-k_n-m_n, n]$. On this block, the asymptotic variance of $D(3)_t^n$ will be calculated and added to the variance of $\sqrt{u_n}V_t^n$

$$\begin{aligned} D(3)_t^n =& \widehat{P}_{i+1}^e \cdot 1_{b\ge 1/4} - \int_{T_{i+1}-(k_n+m_n)\Delta_n}^{T_{i+1}} (t-T_i)\, 2a_t\, dX_t \\ =& \sum_{\lambda=n-k_n-m_n+1}^{n} (t_\lambda - T_i)\, (\hat{a}_{t'} - \tilde{a}_t)\Delta_{\lambda+1}^n X \\ &+ \sum_{\lambda=n-k_n-m_n+1}^{n} (t_\lambda - T_i)\, \tilde{a}_t \Delta_{\lambda+1}^n X - \int_{T_{i+1}-(k_n+m_n)\Delta_n}^{T_{i+1}} (t-T_i)\, 2a_t\, dX_t \end{aligned}$$

Set $\xi_e = \sqrt{u_n}D(3)_t^n$

$$\begin{cases} \displaystyle\sum_{\lambda=1}^{l_n(\tilde{m},t)} \mathbb{E}\Big(\big(\xi(\tilde{m})_\lambda^n\big)^2 \Big| \mathcal{F}_{I(\tilde{m},n,\lambda)} \Big) + \mathbb{E}\Big(\xi_e^2 \Big| \mathcal{F}_{t'}\Big) \xrightarrow{\mathbb{P}} \int_{T_i}^{T_{i+1}} \eta_s^2 ds, \\ \displaystyle\sum_{\lambda=1}^{l_n(\tilde{m},t)} \mathbb{E}\Big(\big(\xi(\tilde{m})_\lambda^n\big)^4 \Big| \mathcal{F}_{I(\tilde{m},n,\lambda)} \Big) + \mathbb{E}\Big(\xi_e^4 \Big| \mathcal{F}_{t'}\Big) \xrightarrow{\mathbb{P}} 0, \\ \displaystyle\sum_{\lambda=1}^{l_n(\tilde{m},t)} \mathbb{E}\Big(\xi(\tilde{m})_\lambda^n\ \Delta_{i,\tilde{m}}^n M \Big| \mathcal{F}_{I(\tilde{m},n,\lambda)} \Big) + \mathbb{E}\Big(\xi_e\ \Delta_e M \Big| \mathcal{F}_{t'}\Big) \xrightarrow{\mathbb{P}} 0, \end{cases} \tag{B.10}$$

C Proof of Theorem 5

For this proof, we decompose $\hat{X}$ into $\bar{X} + \bar{\epsilon}$ (with obvious definitions in view of Assumption 1). Then the estimator becomes:

$$
\begin{aligned}
&\Delta_i^n \hat{X}(\log \widehat{\sigma}_{i+}^2 - \log \sigma_{i+}^2) \\
&= \Delta_i^n \bar{X} \frac{1}{m_n L \Delta_n \sigma_{i+}^2} \Big(\sum_{j \in J_n^+(i)} (\Delta_j^n \bar{X})^2 - m_n L \Delta_n \sigma_{i+}^2 \\
&\quad + \sum_{j \in J_n^+(i)} 2(\Delta_j^n \bar{X})(\Delta_j^n \bar{\epsilon}) + \sum_{j \in J_n^+(i)} (\Delta_j^n \bar{\epsilon})^2 \Big) + O_p(m_n L \Delta_n).
\end{aligned}
$$

The $O_p(m_n L \Delta_n)$ terms can be easily verified, due to the independence between $\bar{X}$ and $\bar{\epsilon}$, and the fact that $\bar{\epsilon}^2 = O_p(\frac{1}{L})$. To prove Theorems 4 and 5, we will go through very similar steps as in the proofs of Theorems 2 and 3. So we will only point out the similarities and differences and omit the full derivations for brevity.

Furthermore, define $\widehat{\sigma}_{i+}'^2$ in the same way as $\widehat{\sigma}_{i+}^2$ except that the noise terms are not included and only the $\bar{X}_i$ is considered.

$$
\begin{aligned}
\widehat{\sigma}_{i+}'^2 - \sigma_{i+}^2 &= \frac{1}{m_n L \Delta_n} \sum_{j \in J_n^+(i)} \frac{3}{2L} \sum_{t_l^n \in (\tau_{j-1}^n, \tau_j^n]} \Big(2 \int_{t_l^n}^{t_l^n + L\Delta_n} (X_s' - X_{t_{j-1}^n}') dX_s' \\
&\quad + \int_{t_l^n}^{t_l^n + L\Delta_n} (\sigma_s^2 - \sigma_{i+}^2) ds \Big) \\
&\quad + \frac{1}{m_n L \Delta_n} \sum_{j \in J_n^+(i)} \frac{3}{L} \sum_{\substack{t_l^n \in (\tau_{j-1}^n, \tau_j^n] \\ t_k^n \in (\tau_{j-1}^n, \tau_j^n]}} \Big(2 \int_{t_k^n}^{t_l^n + L\Delta_n} (X_s' - X_{t_{j-1}^n}') dX_s' \\
&\quad + \int_{t_k^n}^{t_l^n + L\Delta_n} (\sigma_s^2 - \sigma_{i+}^2) ds \Big).
\end{aligned} \tag{C.11}
$$

For Theorem 5, we can still apply the same decomposition as in equation (C.12), except that we replace $\Delta_i^n X$ by $\Delta_i^n \bar{X}$ and Δ_n by $\Delta_n' = L\Delta_n$

$$
\widehat{P}_{i+1} - P_{i+1} = (\widehat{P}_{i+1} - \widetilde{P}_{i+1}) + (\widetilde{P}_{i+1} - P_{i+1}) = V_t^n + D(1)_t^n + D(2)_t^n + D(3)_t^n, \tag{C.12}
$$

where

$$V_t^n = \sum_{\lambda=m_n+1}^{\lfloor n/L\rfloor-k_n-m_n} (t_{\lambda L}-T_i)\Delta_{\lambda L+1}^n \bar{X}\frac{1}{k_n\Delta_n'}\sum_{j=\lambda+1}^{\lambda+k_n}\Delta_{jL}^n\bar{X}\,\left(\log(\widehat{\sigma}_{jL+}^2)-\log(\widehat{\sigma}_{jL-}^2)-\Delta_{jL}^n\log(\sigma^2)\right),$$

$$D(1)_t^n = -\sum_{\lambda=0}^{m_n}(t_{\lambda L}-T_i)\Delta_{\lambda L+1}^n\bar{X}\frac{1}{k_n\Delta_n'}\sum_{j=\lambda+1}^{\lambda+k_n}\Delta_{jL}^n\bar{X}\Delta_{jL}^n\log\sigma^2$$

$$D(2)_t^n = \sum_{\lambda=1}^{\lfloor t/\Delta_n'\rfloor}(t_{\lambda L}-T_i)\Delta_{\lambda L+1}^n X\frac{1}{k_n\Delta_n'}\sum_{j=\lambda+1}^{\lambda+k_n}\Delta_{jL}^n\bar{X}\Delta_{jL}^n\log\sigma^2-\int_{T_i}^{T_{i+1}}(t-T_i)\,2a_t\,dX_t$$

$$D(3)_t^n = \hat{P}_{i+1}^e\cdot 1_{b\geq 1/4}-\int_{T_{i+1}-(k_n+m_n)\Delta_n'}^{T_{i+1}}(t-T_i)\,2a_t\,dX_t$$

Then we can deploy essentially the same ideas as in the proof of Theorem 3. The last three terms in the decomposition will still converge to 0 if $b<1/4$. If $b\geq 1/4$, $D(3)_t^n$ will be added into V_t^n as the last piece of the big block.

For the asymptotic variance, if we consider Δ_n' as the unit of time change, then we can keep the same notation as in the proof for the case without microstructure noise. There is only one difference that we need to point out: $\Delta_{jL}^n\bar{X}$ and $\Delta_{(j+1)L}^n\bar{X}$ are not conditionally independent while $\Delta_j^n X$ and $\Delta_{j+1}^n X$ are. Therefore, except for the conditional second moments, there will be an extra cross product term contributing to the asymptotic variance. In particular,

$$\sum_{i=1}^{l_n(\tilde{m},t)}\sum_{r=k_n+1}^{(\tilde{m}+1)k_n}\mathbb{E}\big((\xi_{I(\tilde{m},n,i)+r})^2|\mathcal{F}_{I(\tilde{m},n,i)}\big)\xrightarrow{\mathbb{P}}\frac{2}{3}\int_{T_i}^{T_{i+1}}\tilde{\eta}_s^2ds, \tag{C.13}$$

$$\sum_{i=1}^{l_n(\tilde{m},t)}\sum_{r=k_n+1}^{(\tilde{m}+1)k_n-1}2\mathbb{E}\big(\xi_{I(\tilde{m},n,i)+r}\,\xi_{I(\tilde{m},n,i)+r+1}|\mathcal{F}_{I(\tilde{m},n,i)}\big)\xrightarrow{\mathbb{P}}\frac{1}{3}\int_{T_i}^{T_{i+1}}\tilde{\eta}_s^2ds. \tag{C.14}$$

By similar arguments as in the proof of Theorem 3, one can then prove the following equations to establish the CLT:

$$\begin{cases} \sum_{i=1}^{l_n(\tilde{m},t)} \mathbb{E}\Big(\big(\xi(\tilde{m})_i^n\big)^2 \Big| \mathcal{F}_{I(\tilde{m},n,i)} \Big) \xrightarrow{\mathbb{P}} \int_0^t \tilde{\eta}_s^2 ds, \\ \sum_{i=1}^{l_n(\tilde{m},t)} \mathbb{E}\Big(\big(\xi(\tilde{m})_i^n\big)^4 \Big| \mathcal{F}_{I(\tilde{m},n,i)} \Big) \xrightarrow{\mathbb{P}} 0, \\ \sum_{i=1}^{l_n(\tilde{m},t)} \mathbb{E}\Big(\xi(\tilde{m})_i^n \; \Delta_{i,\tilde{m}}^n M \Big| \mathcal{F}_{I(\tilde{m},n,i)} \Big) \xrightarrow{\mathbb{P}} 0. \end{cases}$$

The first equation is proven by the summation of (C.13) and (C.14). The remaining two equations can be proven through tedious calculations, analogous to the ones in the proof of Theorem 3. As a result, we complete the proof of Theorem 3.