Volatility Analysis with Realized GARCH-Itô Models

Xinyu Song[1], Donggyu Kim[2], Huiling Yuan[3], Xiangyu Cui[1], Zhiping Lu[4], Young Zhou[4&5], Yazhen Wang[4&6]*

[1] Shanghai University of Finance and Economics
[2] Korea Advanced Institute of Science and Technology
[3] Shanghai University of International Business and Economics
[4] East China Normal University
[5] Chinese Academy of Sciences
[6] University of Wisconsin-Madison

July 3, 2019

Abstract

This paper introduces a unified approach for modeling high-frequency financial data that can accommodate both the continuous-time jump-diffusion and discrete-time realized GARCH model by embedding the discrete realized GARCH structure in the continuous instantaneous volatility process. The key feature of the proposed model is that the corresponding conditional daily integrated volatility adopts an autoregressive structure where both integrated volatility and jump variation serve as innovations. We name it as the realized GARCH-Itô model. Given the autoregressive structure in the conditional daily integrated volatility, we propose a quasi-likelihood function for parameter estimation and establish its asymptotic properties. To improve the parameter estimation, we propose a joint quasi-likelihood function that is built on the marriage of daily integrated volatility estimated by high-frequency data and nonparametric volatility estimator obtained from option data. We conduct a simulation study to check the finite sample performance of the proposed methodologies and an empirical study with the S&P500 stock index and option data.

JEL classification: C10, C22, C58

Keywords: High-frequency financial data, option data, quasi-maximum likelihood estimation, stochastic differential equation, volatility estimation and prediction.

*Corresponding author: Yazhen Wang. Address: 1220 Medical Science Center, 1300 University Avenue , Madison, WI 53706. Phone: 6082626399. Fax: 6082620032. E-mail: yzwang@stat.wisc.edu.

1 Introduction

In modern financial markets, volatility measures the degree of dispersion for assets and plays a crucial role in portfolio allocation, performance evaluation, and risk management. Low-frequency and high-frequency stock data are widely adopted to model the dynamic evolution of daily volatilities. Option data provide one more natural source for the more precise forecast of volatilities and have been investigated thoroughly since the seminal work of Black and Scholes (1973). In traditional volatility analysis, researchers employ discrete parametric econometric models and low-frequency data. Examples include the generalized autoregressive conditional heteroskedasticity (GARCH) models (Bollerslev, 1986; Engle, 1982) which adopt squared daily log returns as innovations in the conditional volatilities. However, when the volatility changes rapidly to a new level, it is often difficult to catch up with the new level immediately using only the daily log returns as the innovations (Andersen et al., 2003b). On the other hand, high-frequency financial data that refer to intra-daily observations such as tick-by-tick stock prices became available thanks to advances in information technology. Major challenges in estimating volatilities with high-frequency data are the market microstructure noises and price jumps. Without the presence of price jumps, Zhang et al. (2005) proposed two-time scale realized volatility (TSRV) which is a consistent estimator for daily variation while Zhang (2006) further improved the TSRV to multi-scale realized volatility (MSRV) so that it can achieve the optimal convergence rate. Other forms of estimators that can achieve the optimal convergence rate only in the presence of market microstructure noises are kernel realized volatility (KRV) (Barndorff-Nielsen et al., 2008), quasi-maximum likelihood estimator (QMLE) (Aït-Sahalia et al., 2010; Xiu, 2010), and pre-averaging realized volatility (PRV) (Jacod et al., 2009). Empirical studies support the existence of price jumps, and decomposition of daily variation into its continuous and jump components can improve volatility forecasts (Aït-Sahalia et al., 2012; Andersen et al., 2003a, 2007; Barndorff-Nielsen and Shephard, 2006; Corsi et al., 2010). For example, Mancini (2004) studied a threshold method for jump-detection and presented the order of an optimal threshold, and Davies and Tauchen (2018) further examined a data-driven type threshold method. Also Fan and Wang (2007) and Zhang et al. (2016) employed wavelet method to identify the jumps given noisy high-frequency data. We refer to the estimators of daily variation based on high-frequency data as the realized volatility estimators. Such estimators are more informative compared to simple squared daily log returns as the innovations, which may help to catch up with rapid changes in the volatility process better.

Efforts made for volatility estimation usually employ low- and high-frequency data independently. However, the inter-correlation between low- and high-frequency data gathered at the two different time scales cannot be ignored as low-frequency data present high-frequency data in an aggregated form. There are several attempts to bridge the gap between the two types of data. For example, multiple studies proposed new GARCH type models, which include realized volatilities as innovations in the conditional volatilities (Engle and Gallo, 2006; Shephard and Sheppard, 2010; Hansen et al., 2012). On the other hand, Wang (2002) showed that the standard GARCH model and its diffusion limit are nonequivalent asymptotically, which discredits the direct application of statistical inferences derived for the GARCH model to its diffusion limit. Thus, Kim and Wang (2016) introduced the unified GARCH-Itô model by embedding the standard GARCH volatility structure in the instantaneous volatilities of an Itô diffusion process. The unified GARCH-Itô model is a continuous-time process at the high-frequency timescale and when restricted to the low-frequency timescale, retains the standard GARCH structure.

In this paper, we expand the unified GARCH-Itô model (Kim and Wang, 2016) so that features of financial data at both frequencies can be better captured as follows. First, price jumps that are well-documented in empirical studies are allowed. Second, we embed the realized GARCH volatility structure (Hansen et al., 2012) in the instantaneous volatilities of a jump-diffusion process, which employs the more informative high-frequency data-based innovations. Third, the well-known intra-day U-shape volatility pattern is accounted for (Admati and Pfleiderer, 1988; Andersen et al., 1997, 2018; Hong and Wang, 2000). We name the proposed model as the realized GARCH-Itô model. The key feature of the proposed model is that its conditional volatility has integrated volatility and jump variation as innovations. Based on the structure of the conditional volatility process, we propose a quasi-likelihood function for estimating model parameters. Specifically, the quasi-likelihood function that is usually adopted in the standard GARCH type models is employed, and the realized volatility estimators are used as the proxy for conditional volatilities. We call the proposed estimator the quasi-maximum likelihood estimator based on high-frequency data and low-frequency structure (QMLE-HL). The proposed model and this estimating approach are constructed purely based on stock data. We as well harness option data to improve the model parameter estimation. In specific, Todorov (2019) developed nonparametric volatility estimator based on a portfolio of short-dated option contracts given a general setting where jumps are present. As stated in Todorov (2019), the estimator can be viewed as the option counterpart of high-frequency data-based volatility estimators. To incorporate the option-based nonparametric volatility estimator, we construct a joint quasi-likelihood function. We call the proposed estimator the quasi-maximum likelihood estimator based on high-frequency data, low-frequency structure and additional option data (QMLE-HLO). Both the QMLE-HL and the QMLE-HLO present good consistency and asymptotic properties. In numerical analysis, we further demonstrate that the joint estimation method QMLE-HLO performs better in estimation and prediction than the QMLE-HL.

This paper is organized as follows. Section 2 introduces the realized GARCH-Itô model. We demonstrate its connection with the realized GARCH model and discuss its advantages comparing to the unified GARCH-Itô model. Section 3 introduces quasi-likelihood estimation methods and investigates their asymptotic behaviors. Section 4 conducts a simulation study to check the finite sample performance for the proposed estimators. Section 5 carries out an empirical analysis with S&P500 stock and option data to demonstrate the advantage of the proposed model in volatility analysis. We collect all the proofs in the Appendix.

2 Realized GARCH-Itô model

The realized GARCH-Itô model is an innovated jump-diffusion process that can incorporate high-frequency based volatility model (Shephard and Sheppard, 2010) and realized GARCH model (Hansen et al., 2012) structures. Let $\mathbb{R}_+ = [0, \infty)$ and $\mathbb{N}$ be the set of all non-negative integers. Our proposed model is formulated as follows.

Definition 1. Log stock price X_t, $t \in \mathbb{R}_+$, obeys a realized GARCH-Itô model if it satisfies

$$dX_t = \mu_t dt + \sigma_t(\theta) dB_t + L_t d\Lambda_t, \tag{2.1}$$

$$\begin{aligned} \sigma_t^2(\theta) &= \sigma_{\lceil t-1 \rceil}^2(\theta) + \gamma(t - \lceil t-1 \rceil)^2 \left\{\omega_1 + \sigma_{\lceil t-1 \rceil}^2(\theta)\right\} - (t - \lceil t-1 \rceil) \left\{\omega_2 + \sigma_{\lceil t-1 \rceil}^2(\theta)\right\} \\ &\quad + \alpha \int_{\lceil t-1 \rceil}^{t} \sigma_s^2(\theta) ds + \beta \int_{\lceil t-1 \rceil}^{t} L_s^2 d\Lambda_s + \nu \left(\lceil t-1 \rceil + 1 - t\right) Z_t^2, \end{aligned} \tag{2.2}$$

where $\lceil t-1 \rceil$ denotes the ceiling of $t-1$, $Z_t = \int_{\lceil t-1 \rceil}^{t} dW_t$, where B_t and W_t are standard Brownian motions with respect to filtration $\mathcal{F}_t$ with $dW_t dB_t = \rho dt$ a.s., μ_t is a predictable process that is known as the drift, and $\sigma_t(\theta)$ is the volatility process that is adapted to $\mathcal{F}_t$. For the jump part, Λ_t is the standard Poisson process with constant intensity λ and L_t denotes the i.i.d. jump sizes which are independent of the Poisson and continuous diffusion processes.

Remark 1. The i.i.d. assumption on jump sizes can be rewritten as

$$L_t^2 = \omega_L + M_t, \tag{2.3}$$

where M_t's are i.i.d. random variables with mean zero and variance ζ^2, $\omega_L + M_t$ is restricted to be positive. For instance, if the jump sizes L_t's obey the Normal distribution with mean δ and variance η, then the corresponding ω_L takes value $\delta^2 + \eta$ while M_t has mean zero and variance $4\delta^2\eta + 2\eta^2$.

The instantaneous volatility $\sigma_t^2(\theta)$ in (2.2) is defined at all times for $t \in \mathbb{R}_+$ and also retains some U-shape pattern within the intra-day. Specifically, when considering the deterministic process part of the instantaneous volatility, it is convex with respect to time t and for an appropriate parameter, it has the smallest value in the middle section of the day. This U-shape instantaneous volatility pattern is often observed in empirical data and supported by financial market (Admati and Pfleiderer, 1988; Andersen et al., 1997, 2018; Hong and Wang, 2000). Moreover, random fluctuations are accounted for in the instantaneous volatility process. We note that when the process is restricted to integer times, it employs the realized GARCH model type structure (Hansen et al., 2012) with an additional jump innovation term as follows:

$$\sigma_n^2(\theta) = \omega + \gamma \sigma_{n-1}^2(\theta) + \alpha \int_{n-1}^{n} \sigma_s^2(\theta) ds + \beta \int_{n-1}^{n} L_s^2 d\Lambda_s, \tag{2.4}$$

where $\omega = \gamma\omega_1 - \omega_2$ and $n \in \mathbb{N}$. Therefore, the instantaneous volatility process is affected by both the integrated volatilities and the jump variations of the stock price process. In comparison to the unified GARCH-Itô model (Kim and Wang, 2016), the realized GARCH-Itô model considers price jumps, accounts for intra-day U-shape volatility pattern, and adopts a richer volatility dynamics with random fluctuations.

For statistical inferences, we study the integrated volatilities obtained from the realized GARCH-Itô model over consecutive integers, that is, $\int_{n-1}^{n} \sigma_t^2(\theta) dt$.

Proposition 1. *Iterative relationship exists in integrated volatilities for the realized GARCH-Itô model defined in Definition 1 and when condition* (2.3) *is met.*

(a) For $0 < \alpha < 1$ *and* $n \in \mathbb{N}$*, the realized GARCH-Itô model implies that*

$$\int_{n-1}^{n} \sigma_t^2(\theta) dt = h_n(\theta) + D_n \quad a.s., \tag{2.5}$$

where

$$h_n(\theta) = \omega^g + \gamma h_{n-1}(\theta) + \alpha^g \int_{n-2}^{n-1} \sigma_s^2(\theta) ds + \beta^g \int_{n-2}^{n-1} L_t^2 d\Lambda_t, \tag{2.6}$$

$$\begin{aligned} \omega^g &= \gamma(\rho_1 - \varrho_2 + 2\varrho_3)\omega_1 - (\varrho_1 - \gamma\varrho_2 + 2\gamma\varrho_3)\omega_2 + (1-\gamma)\{(\varrho_2 - 2\varrho_3)\nu + \varrho_2\beta\lambda\omega_L\}, \\ \alpha^g &= (\rho_1 - \rho_2 + 2\gamma\varrho_3)\,\alpha, \quad \beta^g = (\rho_1 - \rho_2 + 2\gamma\varrho_3)\,\beta, \quad \theta = (\omega^g, \alpha^g, \beta^g, \gamma)\,, \\ \rho_1 &= \alpha^{-1}(e^{\alpha} - 1), \quad \rho_2 = \alpha^{-2}(e^{\alpha} - 1 - \alpha), \quad \rho_3 = \alpha^{-3}(e^{\alpha} - 1 - \alpha - \frac{\alpha^2}{2}), \end{aligned} \tag{2.7}$$

and

$$\begin{gathered} D_n = D_n^c + D_n^J, \\ D_n^c = 2\nu\alpha^{-2} \int_{n-1}^{n} \left\{\alpha(n - t - \alpha^{-1})e^{\alpha(n-t)} + 1\right\} Z_t dZ_t, \\ D_n^J = \beta\alpha^{-1} \left\{ \int_{n-1}^{n} \left(e^{\alpha(n-t)} - 1\right) M_t d\Lambda_t + \omega_L \int_{n-1}^{n} \left(e^{\alpha(n-t)} - 1\right)(d\Lambda_t - \lambda dt) \right\} \end{gathered}$$

are all martingale differences.

(b) For $0 < \alpha < 1$ *and* $n \in \mathbb{N}$*,*

$$E\left[\int_{n-1}^{n} \sigma_t^2(\theta) dt \middle| \mathcal{F}_{n-1}\right] = h_n(\theta) \quad a.s., \tag{2.8}$$

where $h_n(\theta)$ *is defined in (2.6).*

(c) For $0 < \alpha^g + \gamma < 1$ *and* $n \in \mathbb{N}$*,*

$$E[h_n(\theta)] = \frac{\omega^g + \beta^g\lambda\omega_L}{1 - \alpha^g - \gamma}, \quad E[\sigma_n^2] = \frac{(\omega + \beta\lambda\omega_L)(1 - \alpha^g - \gamma) + \alpha(\omega^g + \beta^g\lambda\omega_L)}{(1 - \alpha^g - \gamma)(1 - \gamma)}, \tag{2.9}$$

where ω^g*,* α^g *and* β^g *are defined in* (2.7).

Proposition 1 (a) indicates that the daily integrated volatility can be decomposed into the realized GARCH volatility $h_n(\theta)$ and the martingale difference D_n, where the GARCH volatility $h_n(\theta)$ can be further explained by historical integrated volatilities and jump variations. We utilize this model feature to build up parameter estimation methods in Section 3.

3 Parameter estimation

In this section, we first discuss the model set-up and review nonparametric estimation methods for the integrated volatility in the presence of market microstructure noises given the jump-diffusion process. With the well-performing realized volatility and jump variation estimators, we construct quasi-maximum likelihood estimation procedures and investigate their asymptotic behaviors.

3.1 The model set-up and the realized volatility estimators

Let n be the total number of low-frequency observations and m_i be the total number of high-frequency observations during the ith low-frequency period, for example, the ith day. We further denote $m = \sum_{i=1}^{n} m_i/n$. The underlying log price process is assumed to obey the realized GARCH-Itô model as described in Definition 1. The low-frequency data are the true log prices at integer times, $X_i, i = 0, 1, \ldots, n$. The high-frequency data are observations between integer times and are contaminated by market microstructure noises. Major sources for the market microstructure noises are bid-ask bounce, discreteness of price change, and infrequent trading that only play a role in high-frequency trading (Ait-Sahalia and Yu, 2008). We let $t_{i,j}$ be the high-frequency observed time points during the ith low-frequency period such that $i-1 = t_{i,0} < t_{i,1} < \cdots < t_{i,m_i} = t_{i+1,0} = i$. In this regard, we take the well-agreed assumption in high-frequency literature such that

$$Y_{t_{i,j}} = X_{t_{i,j}} + \epsilon_{t_{i,j}},$$

where $\epsilon_{t_{i,j}}$ are market microstructure noises. Moreover, we note that the effect of the drift term μ_t on high-frequency data based volatility estimators is negligible asymptotically and since the term can be estimated well with the sample mean of intra-day high-frequency log returns, we take $\mu_t = 0$ to highlight on modeling the volatility and jump processes.

Without the presence of price jumps, researchers have constructed nonparametric realized volatility estimators that take advantage of sub-sampling and re-averaging techniques to remove the effect of market microstructure noises so that the integrated volatility can be estimated consistently and efficiently. Such estimators include the multi-scale realized volatility estimator (Zhang, 2006, 2011), the pre-averaging realized volatility estimator (Christensen et al., 2010; Jacod et al., 2009), and the kernel realized volatility estimator (Barndorff-Nielsen et al., 2008). To identify the jump locations given noisy high-frequency data, Fan and Wang (2007) and Zhang et al. (2016) proposed wavelet methods to detect jumps and applied the

MSRV method to jump-adjusted data. They demonstrated that the estimator of jump variation has the convergence rate of $m^{-1/4}$, which further helps the estimator of integrated volatility to achieve the optimal convergence rate of $m^{-1/4}$. In this paper, we let JV_i to be the estimator of jump variation for the ith day and RV_i to be the corresponding estimator of daily integrated volatility that is robust to microstructure noises and price jumps, where both estimators can achieve the convergence rate $m^{-1/4}$.

3.2 Quasi-maximum likelihood estimation based on high-frequency data and low-frequency structure

3.2.1 Estimation procedure

Recall that the integrated volatility over the ith period can be decomposed into the realized GARCH volatility $h_i(\theta)$ and martingale difference D_i as described in Proposition 1 (a). We harness this information for making inferences on the true parameter $\theta_0 = (\omega_0^g, \alpha_0^g, \beta_0^g, \gamma_0)$. Specifically, using the likelihood of the standard GARCH model and the low-frequency structure of the realized GARCH-Itô model, we define the following quasi-likelihood function

$$L_{n,m}^{GH}(\theta) = -\sum_{i=1}^{n} \left[\log(h_i(\theta)) + \frac{RV_i}{h_i(\theta)} \right].$$

Under some technical conditions, the impact of the martingale difference term D_i is negligible in the asymptotic sense. Therefore, the realized volatility estimators RV_i's can be considered as the observed value for $h_i(\theta)$'s and are employed as the proxy. To harness the proposed quasi-likelihood function, we first need to evaluate the realized GARCH term $h_i(\theta)$. Recall the iterative relationship in the realized GARCH term $h_i(\theta)$ as described in Proposition 1 (a):

$$\begin{aligned} h_i(\theta) =& \omega^g + \gamma h_{i-1}(\theta) + \alpha^g \int_{i-2}^{i-1} \sigma_t^2(\theta) dt + \beta^g \int_{i-2}^{i-1} L_t^2 d\Lambda_t \\ =& \sum_{l=1}^{i-1} \gamma^{l-1} \left\{ \omega^g + \alpha^g \int_{i-l-1}^{i-l} \sigma_t^2(\theta) dt + \beta^g \int_{i-2}^{i-1} L_t^2 d\Lambda_t \right\} + \gamma^{i-1} h_1(\theta), \quad i = 2, \ldots, n. \end{aligned}$$

The initial $h_1(\theta)$ is selected to be $E[h_1(\theta)]$ that is given in Proposition 1 (c). Specifically, we take

$$h_1(\theta) = \frac{\omega^g + \beta^g \lambda \omega_L}{1 - \alpha^g - \gamma}.$$

The true integrated volatilities and jump variations are not observed so that we adopt their estimators RV_i and JV_i, respectively. Specifically, let

$$\widehat{h}_i(\theta) = \sum_{l=1}^{i-1} \gamma^{l-1} \left\{ \omega^g + \alpha^g RV_{i-l} + \beta^g JV_{i-l} \right\} + \gamma^{i-1} h_1(\theta), \quad i = 2, \ldots, n.$$

With the realized GARCH volatility estimator $\widehat{h}_i(\theta)$, the quasi-likelihood function is updated to the following:

$$\widehat{L}_{n,m}^{GH}(\theta) = -\sum_{i=1}^{n}\left[\log(\widehat{h}_i(\theta)) + \frac{RV_i}{\widehat{h}_i(\theta)}\right].$$

We estimate the parameter θ_0 by maximizing the quasi-likelihood function $\widehat{L}_{n,m}^{GH}(\theta)$,

$$\widehat{\theta}^{GH} = \underset{\theta\in\Theta}{\operatorname{argmax}}\ \widehat{L}_{n,m}^{GH}(\theta),$$

and name the maximizer $\widehat{\theta}^{GH}$ the quasi-maximum likelihood estimator based on high-frequency data and low-frequency structure combined (QMLE-HL), because the data at the high-frequencies and the dynamics at the low-frequencies are crucial to our model proposal and estimation procedure.

3.2.2 Asymptotic theory

This section establishes the consistency and asymptotic distribution for the proposed estimator $\widehat{\theta}^{GH}$. We first define some notations. For any given random variable X and $p \geq 1$, define $\|X\|_{L_p} = \{E[|X|^p]\}^{1/p}$. For a matrix $A = (A_{i,j})_{1\leq i\leq k', 1\leq j\leq k}$, let $\|A\|_{max} = \max_{i,j}|A_{i,j}|$. Let C's be positive generic constants whose values are free of θ, n, and m_i, and may change from occurrence to occurrence. To investigate the asymptotic behaviors of proposed estimation method, we require the following technical assumptions.

Assumption 1.

(a) *Let*

$$\Theta = \{(\omega^g, \alpha^g, \beta^g, \gamma) : \omega_l^g < \omega^g < \omega_u^g, \alpha_l^g < \alpha^g < \alpha_u^g, \beta_l^g < \beta^g < \beta_u^g, \gamma_l < \gamma < \gamma_u, \alpha^g + \gamma < 1\},$$

where $\omega_l^g, \omega_u^g, \alpha_l^g, \alpha_u^g, \beta_l^g, \beta_u^g, \gamma_l, \gamma_u$ *are known positive constants.*

(b) *We have* $\max_{t\in\mathbb{R}_+} E\left\{\sigma_t^4(\theta_0)\right\} < \infty$ *and* $E(\epsilon_{t_{i,j}}^4) < \infty$.

(c) *There exist some fixed constants* C_1 *and* C_2 *such that* $C_1 m \leq m_i \leq C_2 m$, *and* $\sup_{1\leq j\leq m_i}|t_{i,j} - t_{i,j-1}| = O(m^{-1})$ *and* $n^2 m^{-1} \to 0$ *as* $m, n \to \infty$.

(d) *One of the following conditions is satisfied.*

(d1) *There exists a positive constant* δ *such that* $E\left[\left(\frac{R_i^2}{h_i(\theta_0)}\right)^{2+\delta}\right] \leq C$ *for any* $i \in \mathbb{N}$, *where* $R_i = \int_{i-1}^{i}\sigma_t(\theta_0)dB_t$.

(d2) $\frac{E[R_i^4|\mathcal{F}_{i-1}]}{h_i^2(\theta_0)} \leq C$ *a.s. for any* $i \in \mathbb{N}$.

(e) $\sup_{i\in\mathbb{N}} \left\| RV_i - \int_{i-1}^{i} \sigma_s^2(\theta_0) ds \right\|_{L_2} \leq Cm^{-1/4}$ *and* $\sup_{i\in\mathbb{N}} \left\| JV_i - \int_{i-1}^{i} L_s^2 d\Lambda_s \right\|_{L_2} \leq Cm^{-1/4}$.

(f) For any $i \in \mathbb{N}$, $E\left[RV_i | \mathcal{F}_{i-1}\right] \leq C\, E\left[\int_{i-1}^{i} \sigma_s^2 ds | \mathcal{F}_{i-1}\right] + C$ *a.s.*

(g) $\left(D_i, \int_{i-1}^{i} \sigma_t^2(\theta_0) dt, R_i^2\right)$ *is a stationary ergodic process.*

Remark 2. The parameters of interests are related to volatilities (the 2nd moment), and thus, to study their asymptotic behaviors, we require some finite 4th moment conditions such as Assumption 1 (b) and (d). Therefore, these conditions are not restrictive at all. Assumption 1 (c) is a well-known key condition in high-frequency data based volatility analysis. Under the finite 4th moment condition, Kim et al. (2016) showed that the realized volatility estimators satisfy Assumption 1 (e). Finally, the stationary ergodic condition Assumption 1 (g) is used to obtain asymptotic normality for the QMLE-HL.

The following theorems establish the convergence rate and asymptotic normality for the QMLE-HL $\widehat{\theta}^{GH}$.

Theorem 1. *Under Assumption 1 (a)-(f) (except for* $n^2 m^{-1} \to 0$ *in Assumption 1 (c)), we have*

$$\left\| \widehat{\theta}^{GH} - \theta_0 \right\|_{max} = O_p\left(m^{-1/4} + n^{-1/2}\right).$$

Theorem 2. *Under Assumption 1, we have as* $m, n \to \infty$,

$$\sqrt{n}\left(\widehat{\theta}^{GH} - \theta_0\right) \xrightarrow{d} N\left(0, B^{-1} A^{GH} B^{-1}\right),$$

where

$$\begin{aligned} A^{GH} &= E\Bigg[\Bigg\{\alpha_0^{-4}\nu_0^2 \int_0^1 \left\{\alpha_0(1-t-\alpha_0^{-1})e^{\alpha_0(1-t)} + 1\right\}^2 t dt \\ &\qquad + \frac{\lambda_0\beta_0^2}{4\alpha_0^2} \int_{n-1}^{n} \left(e^{\alpha_0(n-t)} - 1\right)^2 (M_t^2 + \omega_{L0}^2) dt \Bigg\} \frac{\partial h_1(\theta)}{\partial \theta} \frac{\partial h_1(\theta)}{\partial \theta^T}\bigg|_{\theta=\theta_0} h_1^{-4}(\theta_0)\Bigg] \end{aligned}$$

and

$$B = \frac{1}{2} E\left[\frac{\partial h_1(\theta)}{\partial \theta} \frac{\partial h_1(\theta)}{\partial \theta^T}\bigg|_{\theta=\theta_0} h_1^{-2}(\theta_0)\right].$$

Remark 3. Theorem 1 shows that the convergence rate of $\widehat{\theta}^{GH}$ is $m^{-1/4} + n^{-1/2}$. The rate $n^{-1/2}$ is coming from the usual parametric convergence rate based on the low-frequency structure while the rate $m^{-1/4}$ is due to the high-frequency volatility and jump variation estimations and is known as the optimal convergence rate for estimating integrated volatilities with the presence of market microstructure noises and price jumps. Theorem 2 provides the asymptotic normal distribution for $\widehat{\theta}^{GH}$.

Remark 4. We note that when replacing $m^{-1/4}$ in Assumption 1 (e) by $m^{-\xi}$ for some positive constant $0 < \xi < 1/4$, the convergence rate in Theorem 1 will change to $m^{-\xi} + n^{-1/2}$. On the other hand, the condition $n^2 m^{-1} \to 0$ in Assumption 1 (c) is required to derive the asymptotic normality in Theorem 2. Specifically, there exist two types of estimation errors: the errors coming from estimating the realized volatility and jump variation with high-frequency data and have the rate of $m^{-1/4}$; the errors coming from estimating model parameters with the quasi-likelihood function that is constructed on the low-frequency structure and have the rate of $n^{-1/2}$. This condition $n^2 m^{-1} \to 0$ is needed to make sure that the high-frequency estimation errors are negligible when deriving the asymptotic normality in Theorem 2. When the condition $n^2 m^{-1} \to 0$ is not satisfied, the asymptotic normality may depend on $m^{1/4}(RV_i - \int_{i-1}^{i} \sigma_s^2(\theta_0) ds)$, which is the quantity related to high-frequency estimation. For example, if $m^{1/4}(RV_i - \int_{i-1}^{i} \sigma_s^2(\theta_0) ds)$ is some martingale difference sequence, we can relax the condition $n^2 m^{-1} \to 0$ to $n m^{-1} \to 0$.

3.3 Quasi-maximum likelihood estimation based on based on high-frequency data, low-frequency structure, and additional option data

3.3.1 Estimation procedure

In this section, we discuss how to incorporate additional option data information in parameter estimation. The famous Black-Scholes model indicates that option prices are determined by several factors such as time to expiration, strike price, underline asset price, and its volatility, and so one can deduce the volatility from option data. For example, the VIX presents the stock market's general expectation of volatility. However, we usually find that the VIX is different from the historical nonparametric realized volatility. This may be because of the jumps in stock prices and the wedge between the risk-neutral and statistical probabilities. Recently, Todorov (2019) proposed a nonparametric volatility estimator based on a portfolio of noisy short-dated option contracts with different strike prices. This estimator is robust to price jumps and does not require any assumption on the wedge between risk-neutral and statistical probabilities. Specifically, let T be the time to expiration for an option contract, k_ℓ be the ℓth log strike price, where $k_1 < k_2 < \cdots < k_N$ and $\Delta_\ell = k_\ell - k_{\ell-1}$ for $\ell = 2, \ldots, N$. Let $\kappa_T(k_\ell)$ be the true option price given expiration T and log-strike k_ℓ. Due to observation errors in empirical derivatives pricing, the observed option price $\widehat{\kappa}_T(k_\ell)$ obeys

$$\widehat{\kappa}_T(k_\ell) = \kappa_T(k_\ell) + \varepsilon_\ell,$$

where the noises ε_ℓ's are random variables with mean zero. Given this set-up, Todorov (2019) proposed the following nonparametric volatility estimator

$$NV_i = \frac{-2}{Tu} \mathcal{R} \left(\log \left(\widehat{f}_i(u) \wedge T \right) \right),$$

where

$$\widehat{f}_i(u) = 1 - (u^2 + \sqrt{-1}u) \sum_{\ell=2}^{N} e^{(\sqrt{-1}u-1)k_{\ell-1} - \sqrt{-1}uX_i} \widehat{\kappa}_T(k_{\ell-1})\Delta_\ell,$$

$\mathcal{R}(A)$ is the real part of a complex number A, and u is a tuning parameter.

Under some technical conditions, as T goes to zero, this nonparametric volatility estimator NV_i converges to the true spot volatility $\sigma_i^2(\theta_0)$ (Todorov, 2019). However, empirically, option contracts are quoted at the market open or close on each trading day so that the minimum choice of T is 1 business day. In this sense, NV_i may contain integrated volatility for the remaining period from time i. Also Todorov (2019) showed that the estimates NV_i's hold a close relationship with the jump-robust realized type volatility estimates RV_i's in his empirical study. Based on his results, we assume that the nonparametric volatility estimator NV_{i-1} and the conditional daily integrated volatility $h_i(\theta)$ have the following linear relationship:

$$NV_{i-1} = b + ah_i(\theta) + e_i, \quad i = 1, \dots, n, \tag{3.1}$$

where b and a are the intercept and slope coefficients, respectively. Moreover, e_i's are martingale differences with mean zero and variance σ_e^2, and they are independent of the price process and the microstructure component.

Let $\varphi = (\omega^g, \alpha^g, \beta^g, \gamma, a, b)$ and $\phi = (\omega^g, \alpha^g, \beta^g, \gamma, a, b, \sigma_e^2)$. Note that θ is a subset of φ and ϕ. We propose the following joint quasi-likelihood function based on high-frequency and option data for estimating the true parameter $\phi_0 = (\omega_0^g, \alpha_0^g, \beta_0^g, \gamma_0, a_0, b_0, \sigma_{e0}^2)$

$$\widehat{L}_{n,m}^{GHO}(\phi) = -\sum_{i=1}^{n} \left[\log(\widehat{h}_i(\theta)) + \frac{RV_i}{\widehat{h}_i(\theta)} \right] - \sum_{i=1}^{n} \left[\log(\sigma_e^2) + \frac{(NV_{i-1} - b - a\widehat{h}_i(\theta))^2}{\sigma_e^2} \right].$$

We maximize $\widehat{L}_{n,m}^{GHO}(\phi)$ to obtain parameter estimators, that is,

$$\widehat{\phi}^{GHO} = \operatorname*{argmax}_{\phi \in \Phi} \widehat{L}_{n,m}^{GHO}(\phi),$$

where Φ is the parameter space of ϕ. We call the proposed estimator the quasi-maximum likelihood estimator based on high-frequency data, low-frequency structure, and additional option data combined (QMLE-HLO).

Remark 5. We connect the nonparametric volatility estimator proposed by Todorov (2019) with the conditional integrated volatility using the linear structure assumption in (3.1). This is because practically, T ranges from 1 to 5 business days so that NV_i contains volatility information for the remaining period. Another approach is to connect NV_i with the spot volatility at integer time i, that is, $\sigma_i^2(\theta_0)$, from the realized GARCH-Itô model. In fact, Theorems 1 and 3 in Todorov (2019) show that as T goes to 0, the nonparametric volatility estimator NV_i converges to the spot volatility $\sigma_i^2(\theta_0)$. Similar to $\widehat{L}_{n,m}^{GHO}(\phi)$, we could construct a joint quasi-likelihood function by taking advantage of the asymptotic limits provided in Todorov (2019). In this case, the convergence rate of the corresponding estimator will depend on the convergence rate of NV_i, which may not be optimal. Specifically, Theorem 1 in Todorov (2019) shows that NV_i converges to $\sigma_i^2(\theta_0)$ with the convergence rate T^τ, where τ is a positive number that is determined by the range of log-strikes and the choice of u. In this case, T^τ plays the dominant role in the convergence rate of the quasi-maximum likelihood estimator, which is rather slow due to the fixed remaining period of T in practice. Also the numerical study shows that the QMLE-HLO method has better performance and thus, we present the QMLE-HLO method in this paper.

3.3.2 Asymptotic theory

To establish the asymptotic behaviors of the proposed estimation method, we require the following additional assumptions.

Assumption 2.

(a) *Let*

$$\Phi = \{(\omega^g, \alpha^g, \beta^g, \gamma, a, b, \sigma_e^2) : (\omega^g, \alpha^g, \beta^g, \gamma) \in \Theta, a_l < a < a_u, b_l < b < b_u, \sigma_{e_l}^2 < \sigma_e^2 < \sigma_{e_u}^2\},$$

where $a_l, a_u, b_l, b_u, \sigma_{e_l}^2, \sigma_{e_u}^2$ *are known positive constants.*

(b) $\sup_{i\in\mathbb{N}} E\left[e_i^4\right] < \infty$.

(c) $\left(D_i, \int_{i-1}^i \sigma_t^2(\phi_0)dt, R_i^2, e_i\right)$ *is a stationary ergodic process.*

The following theorems establish the convergence rate and asymptotic normality for the QMLE-HLO $\widehat{\phi}^{GHO}$.

Theorem 3. *Under Assumption 1 (a)–(f) (except for* $n^2m^{-1} \to 0$ *in Assumption 1 (c)) and Assumption 2 (a)–(b), we have*

$$\left\|\widehat{\phi}^{GHO} - \phi_0\right\|_{max} = O_p\left(n^{-1/2} + m^{-1/4}\right).$$

Theorem 4. *Under Assumption 1 and Assumption 2, we have as* $m, n \to \infty$,

$$\sqrt{n}\left(\widehat{\phi}^{GHO} - \phi_0\right) \xrightarrow{d} N\left(0, \left(B^{GHO}\right)^{-1} A^{GHO} \left(B^{GHO}\right)^{-1}\right),$$

where

$$A^{GHO} = \begin{pmatrix} A^{GH} & \mathbf{0}_{4\times3} \\ \mathbf{0}_{4\times3}^T & \mathbf{0}_{3\times3} \end{pmatrix} + A^O, \qquad B^{GHO} = \begin{pmatrix} B^{\varphi} & \mathbf{0}_{6\times1} \\ \mathbf{0}_{6\times1}^T & \frac{1}{2}\sigma_{e0}^{-4} \end{pmatrix},$$

$$A^O = E\left[\begin{pmatrix} \frac{\partial f_1(\varphi)}{\partial\varphi}\frac{\partial f_1(\varphi)}{\partial\varphi^T}\Big|_{\varphi=\varphi_0}\frac{1}{\sigma_{e0}^2} & \frac{\partial f_1(\varphi)}{\partial\varphi}\Big|_{\varphi=\varphi_0}\frac{e_1^3}{2\sigma_{e0}^6} \\ \frac{\partial f_1(\varphi)}{\partial\varphi^T}\Big|_{\varphi=\varphi_0}\frac{e_1^3}{2\sigma_{e0}^6} & \frac{(e_1^2-\sigma_{e0}^2)^2}{4\sigma_{e0}^8} \end{pmatrix}\right],$$

$$B^{\varphi} = \frac{1}{2} E \left[\frac{\partial h_1(\theta)}{\partial \varphi} \frac{\partial h_1(\theta)}{\partial \varphi^T} \bigg|_{\varphi=\varphi_0} h_1^{-2}(\theta_0) + \frac{\partial f_1(\varphi)}{\partial \varphi} \frac{\partial f_1(\varphi)}{\partial \varphi^T} \bigg|_{\varphi=\varphi_0} \frac{2}{\sigma_{e0}^2} \right],$$

and $f_i(\varphi) = b + a h_i(\theta)$ *for* $i = 1, \dots, n$*. Here* $\mathbf{0}_{i \times j}$ *denotes an i-by-j matrix of zeros.*

Remark 6. Theorem 3 shows that the convergence rate for the QMLE-HLO is the same as the QMLE-HL. Theorem 4 provides the asymptotic normal distribution for the QMLE-HLO.

4 Simulation study

In this section, we conducted a simulation study to check the finite sample performance of the estimators $\widehat{\theta}^{GH}$ and $\widehat{\theta}^{GHO}$, as well as to investigate the prediction performance of the realized GARCH volatilities $\widehat{h}_i(\widehat{\theta}^{GH})$ and $\widehat{h}_i(\widehat{\theta}^{GHO})$, which was also compared with the performance of the GARCH volatilities used in Kim and Wang (2016). The true log prices $X_{t_{i,j}}$, $t_{i,j} = i-1+j/m$, $i = 1, \dots, n$, $j = 1, \dots, m$, were generated based on the proposed realized GARCH-Itô model defined in (2.1) and (2.2) with the following set of parameters $\omega_1 = 2.755$, $\omega_2 = 0.898$, $\alpha = 0.480$, $\beta = 0.126$, $\nu = 0.8$, $\gamma = 0.366$, and $\rho = -0.6$. For the jump process, we took the intensity λ to be 27 and generated L_t^2 such that $L_t^2 = \omega_L + M_t$, where $\omega_L = 0.005$ and M_t follows the normal distribution with mean zero and standard deviation 0.001. Each jump L_t was further assigned to be either positive or negative randomly. The chosen parameters resulted in the following target parameter $\theta = (\omega^g, \alpha^g, \beta^g, \gamma) = (0.111, 0.398, 0.105, 0.366)$ for modeling the dynamics in conditional integrated volatilities. We note that the parameter ω^g was scaled by 50000 times compared to its empirical counterpart while the rest parameters remained the same. Scaling in this simulation study was done in order to avoid the generation of any negative value for the instantaneous volatilities due to the U-shape intra-day pattern. Initial values for the simulation were chosen to be $X_0 = 10$ and $\sigma_0^2 = E(\sigma_1^2) = 0.601$. For the high-frequency data, market microstructure noises were added to simulated log prices $X_{t_{i,j}}$'s between integer times, and the noises were modeled by i.i.d normal random variables with mean 0 and standard deviation 0.005. For the option model described in (3.1), we took $a = 0.768$, $b = 0.1$, $\sigma_e = 0.08$, where the intercept b and standard deviation σ_e were scaled by roughly 50000 times comparing to their empirical estimates. We took $n = 125, 250, 500, 1000$ and $m = 390, 780, 2340, 23400$. For each combination of n and m, we repeated the simulation procedure for 2000 times. We followed the procedure as described in Fan and Wang (2007) to detect the jump locations, estimate the jump variations, and compute the jump-adjusted MSRV estimators. Model parameter estimators were obtained by maximizing the proposed quasi-likelihood functions $\widehat{L}_{n,m}^{GH}(\theta)$ and $\widehat{L}_{n,m}^{GHO}(\phi)$.

Table 1 reports the mean squared errors (MSEs) for the jump parameters ω_L and λ. We find that the MSEs decrease as the number of high-frequency observations increases for each n, and larger n often helps to locate the jumps and to estimate the parameters ω_L and λ better. Table 2 presents the MSEs for the QMLE-HL and QMLE-HLO. The proposed estimating procedures present good finite sample performances and support the theoretical results derived in Section 3. For each estimation method, as the number of low-frequency or high-frequency observations increases, the MSEs decrease. When comparing the two methods, the QMLE-HLO has smaller MSE than the QMLE-HL. Thus, it is reasonable to conclude that additional option data help to enhance the estimation of model parameters.

The major motivation of our model proposal is to predict future volatilities by taking advantage of the imposed autoregressive type of model structure at the low-frequency. So we examined the finite sample performance of the proposed predictors $\widehat{h}_i(\widehat{\theta}^{GH})$ and $\widehat{h}_i(\widehat{\theta}^{GHO})$. For comparison purpose, we as well investigated the prediction performance of the unified GARCH-Itô model proposed by Kim and Wang (2016), and denote the predictor by $\widehat{h}_{i0}(\widehat{\theta}_0^{GH})$. Specifically, we evaluated the mean squared prediction errors (MSPEs) by

$$\frac{1}{n-h}\sum_{i=h+1}^{n}\left(\widehat{H}_i - h_i(\theta)\right)^2,$$

where $\widehat{H}_i$ is one of the followings: $\widehat{h}_i(\widehat{\theta}^{GH})$, $\widehat{h}_i(\widehat{\theta}^{GHO})$, or $\widehat{h}_{i0}(\widehat{\theta}_0^{GH})$. As a benchmark, we as well considered the prediction of $h_i(\theta)$ using RV_{i-1}. We let the initial forecast origin to be $h = n - 20$ and expanded the observation window by one low-frequency period at a time. Each time, the model parameters were estimated and the predictors were obtained.

Table 3 summarizes the MSPEs and Figure 1 presents the log MSPEs against the number of high-frequency observations. Overall, the MSPE for the realized GARCH-Itô approach decreases as the number of low-frequency or high-frequency observations increases. Moreover, the QMLE-HLO method presents the best performance regarding the MSPE. That is, the numerical results indicate that utilizing information contained in an additional data source can improve both the estimation and prediction performance of the proposed methodology. On the other hand, the unified GARCH-Itô model is not capable of explaining the rich dynamics in order to predict the conditional integrated volatilities. This may be because it takes into account neither the realized volatility nor the jump variation as an innovation. The benchmark method does not perform well because the realized GARCH-Itô model has rich dynamics that cannot be fully captured by the jump-adjusted MSRV method.

5 Empirical analysis

In this section, we illustrate the proposed estimation methods with trading data in second for S&P500 stock index and option data quoted at the market opening on each trading day where S&P500 stock index is the underline asset. The data sets were obtained from the TAQ and the CBOE database, respectively. We examined the period from January 3rd, 2017 to November 30th, 2017 so that the number of low-frequency periods is $n = 231$. The high-frequency data are available between open and close of the market so that the number of high-frequency observations for a full trading day is $m = 23400$. We followed the procedure given in Fan and Wang (2007) to detect jumps, as well as to compute the jump variation estimates JV_i's and jump-adjusted MSRV estimates RV_i's. We estimated the intensity λ by the daily averaged number of price jumps, and the parameter ω_L by the sample median of all squared price jumps because the sample median better described the center of the distribution formed by squared jumps. The estimated values are $\widehat{\lambda} = 26.926$ and $\widehat{\omega}_L = 2.320 \times 10^{-8}$. For the option data, we followed the procedure presented in Todorov (2019) as their empirical study covered the period studied in this paper and considered the S&P500 index as well. Specifically, we took the option contracts where the time to expiration ranges from 1 to 2 business days and skipped the contracts that were settled on a holiday. The average number of strikes per date was 51.549 and the values of the tuning parameters were set to be the same as in Todorov (2019). Denote the option-based nonparametric volatility estimates by NV_i's. Figure 3 displays the auto-correlation functions for the RV_i's, JV_i's, and NV_i's, which provides promising evidence for explaining the rich dynamics with these innovations. The QMLE-HL estimates are $\widehat{\omega}^g = 2.222 \times 10^{-6}, \widehat{\alpha}^g = 0.398, \widehat{\beta}^g = 0.105$, and $\widehat{\gamma} = 0.366$, and the QMLE-HLO estimates are $\widehat{\omega}^g = 1.226 \times 10^{-6}, \widehat{\alpha}^g = 0.284, \widehat{\beta}^g = 1.382, \widehat{\gamma} = 0.449, \widehat{a} = 0.768, \widehat{b} = 2.024 \times 10^{-6}, \widehat{\sigma}_e = 1.510 \times 10^{-6}$. The parameter ω^g denotes the intercept term in the realized GARCH volatility dynamics while the parameter b denotes the intercept term in model (3.1). Their small estimated values reflect the overall level of daily volatilities that can be seen in Figure 3.

Figure 3 displays the jump-adjusted MSRV estimates, the option-based nonparametric volatility estimates, the realized GARCH volatility estimates from the QMLE-HL and the QMLE-HLO. For comparison purpose, we as well present the GARCH volatilities adopted in the unified GARCH-Itô model (Kim and Wang, 2016). Figure 3 shows that the nonparametric jump-adjusted MSRV and the option-based nonparametric volatility estimates are both volatile, and the realized GARCH volatility estimates from the QMLE-HL and QMLE-HLO methods can account for these dynamics well. Moreover, when comparing with the unified GARCH-Itô estimates, the proposed realized GARCH-Itô estimates are closer to the jump-adjusted MSRV estimates. This may be because the realized GARCH-Itô model includes realized volatilities and jump variations as innovations while the unified GARCH-Itô model comprises squared daily log returns as innovations. That is, the proposed structure in the realized GARCH-Itô model helps to capture the market dynamics promptly.

To investigate the prediction performance of the proposed methodologies, we employed the MSPE criteria again. Denote the forecast origin by h. To further examine the dependency of split points, we took $h = 168, 188, 210$, where each value corresponds to the last trading day of August, September, and October. Since the exact conditional daily integrated volatilities are unknown for empirical data, we used the jump-adjusted MSRV estimates instead and evaluated the following MSPE:

$$\frac{1}{n-h} \sum_{i=h+1}^{n} \left(\widehat{H}_i - RV_i \right)^2,$$

where $\widehat{H}_i$ is one of the followings: $\widehat{h}_i(\widehat{\theta}^{GH})$, $\widehat{h}_i(\widehat{\theta}^{GHO})$, $\widehat{h}_{i0}(\widehat{\theta}_0^{GH})$, or RV_{i-1}.

Table 4 summarizes the MSPEs from the realized GARCH-Itô, the unified GARCH-Itô, and the jump-adjusted MSRV estimates. Overall, the proposed realized GARCH-Itô estimates outperform the other methods in terms of the MSPE across various split points. When comparing the realized GARCH-Itô estimates, the QMLE-HLO presents smaller MSPE than the QMLE-HL. The empirical results indicate that the realized GARCH-Itô model holds advantages in predicting future volatilities as it utilizes the autoregressive structure in daily integrated volatilities and emphasizes high-frequency based information by using both realized volatilities and jump variations as innovations. Moreover, incorporating option-based nonparametric volatility estimates could help to predict future volatilities.

6 Conclusion

In this paper, we introduce a novel realized GARCH-Itô model based on a jump-diffusion process which embeds the discrete realized GARCH model structure (Hansen et al., 2012) in its instantaneous volatility process. When the model is restricted to the low-frequency period, it employs an autoregressive type structure to explain the co-dynamics in the integrated volatilities and jump variations. Model parameters in the realized GARCH-Itô model are estimated by maximizing a quasi-likelihood function. To improve the statistical performance of the proposed estimating approach and to incorporate additional information from option data, we as well connect the nonparametric volatility estimator proposed by Todorov (2019) with the conditional integrated volatility from the proposed model. A joint quasi-likelihood function is then adopted and we show that this method helps to improve accounting for the market dynamics in the numerical analysis.

We also leave some open issues for future study. For example, we may observe some heterogeneous variance in model (3.1). One possible approach is to generalize the homogeneous variance in (3.1) to heterogeneous variance such as replacing σ_e^2 by $\sigma_e^2 h_i^{\zeta}(\theta)$, where parameter $\zeta > 0$ is used to adjust the level of heteroscedasticity with $\zeta = 0$ corresponding to the homogeneous case. We replace σ_e^2 by $\sigma_e^2 \widehat{h}_i^{\zeta}(\theta)$ in $\widehat{L}_{n,m}^{GHO}(\phi)$ and then estimate ζ jointly with the other parameters by maximizing $\widehat{L}_{n,m}^{GHO}(\phi)$. Moreover, we may consider more complex relationships than the linear structure in model (3.1) and use estimators other than the NV_i's.

A Appendix

Let $C > 0$ and $0 < \rho < 1$ be generic constants whose values are free of θ, ϕ, n, and m and may change from occurrence to occurrence.

A.1 Proof of Proposition 1

Proof of Proposition 1. For $k, n \in \mathbb{N}$, let

$$R(k) \equiv \int_{n-1}^{n} \frac{(n-t)^k}{k!} \sigma_t^2(\theta) dt.$$

By the Itô's Lemma, we have

$$\begin{aligned} R(k) =& \frac{(k+1)\nu}{(k+3)!} + \frac{\gamma\omega_1 - \omega_2 + \gamma\sigma_{n-1}^2(\theta)}{(k+1)!} + \frac{\omega_2 - 2\gamma\omega_1 + (1-2\gamma)\sigma_{n-1}^2(\theta)}{(k+2)k!} \\ &+ \frac{\gamma\omega_1 + \gamma\sigma_{n-1}^2(\theta)}{(k+3)k!} + \frac{\beta\lambda\omega_L}{(k+2)!} + \alpha R(k+1) \\ &+ \beta \int_{n-1}^{n} \frac{(n-t)^{k+1}}{(k+1)!} M_t d\Lambda_t + \beta \int_{n-1}^{n} \frac{(n-t)^{k+1}}{(k+1)!} \omega_L (d\Lambda_t - \lambda dt) \\ &+ 2\nu \int_{n-1}^{n} \left(\frac{(n-t)^{k+2}}{(k+1)!} - \frac{(n-t)^{k+2}}{(k+2)!} \right) Z_t dZ_t. \end{aligned}$$

Then simple algebraic manipulations show

$$\begin{aligned} \int_{n-1}^{n} \sigma_t(\theta)^2 dt =& R(0) \\ =& (\varrho_2 - 2\varrho_3)\nu + \varrho_2\beta\lambda\omega_L + \varrho_1 \left\{ \gamma\omega_1 - \omega_2 + \gamma\sigma_{n-1}^2(\theta) \right\} \\ &+ (\varrho_1 - \varrho_2) \left\{ \omega_2 - 2\gamma\omega_1 + (1-2\gamma)\sigma_{n-1}^2(\theta) \right\} \\ &+ (\varrho_1 - 2\varrho_2 + 2\varrho_3) \left\{ \gamma\omega_1 + \gamma\sigma_{n-1}^2(\theta) \right\} + D_n^J + D_n^c \\ =& (\varrho_2 - 2\varrho_3)\nu + \varrho_2\beta\lambda\omega_L + 2\varrho_3\gamma\omega_1 - \varrho_2\omega_2 + (\varrho_1 - \varrho_2 + 2\gamma\varrho_3)\sigma_{n-1}^2(\theta) \\ &+ D_n^J + D_n^c \quad \text{a.s.} \end{aligned}$$

Since

$$\sigma_n^2(\theta) = \omega + \gamma\sigma_{n-1}^2(\theta) + \alpha \int_{n-1}^{n} \sigma_s^2(\theta) ds + \beta \int_{n-1}^{n} L_s^2 d\Lambda_s,$$

we have

$$\begin{aligned} h_n(\theta) &= (\varrho_2 - 2\varrho_3)\nu + \varrho_2\beta\lambda\omega_L + 2\varrho_3\gamma\omega_1 - \varrho_2\omega_2 + (\varrho_1 - \varrho_2 + 2\gamma\varrho_3)\sigma_{n-1}^2(\theta) \\ &= (\varrho_2 - 2\varrho_3)\nu + \varrho_2\beta\lambda\omega_L + 2\varrho_3\gamma\omega_1 - \varrho_2\omega_2 \\ &\quad + (\varrho_1 - \varrho_2 + 2\gamma\varrho_3) \left(\omega + \gamma\sigma_{n-2}^2(\theta) + \alpha \int_{n-2}^{n-1} \sigma_s^2(\theta) ds + \beta \int_{n-2}^{n-1} L_s^2 d\Lambda_s \right) \end{aligned}$$

$$= \omega^g + \gamma h_{n-1}(\theta) + \alpha^g \int_{n-2}^{n-1} \sigma_s^2(\theta) ds + \beta^g \int_{n-2}^{n-1} L_t^2 d\Lambda_t,$$

where ω^g, α^g and β^g are defined in (2.7). Thus, we have

$$\int_{n-1}^{n} \sigma_t(\theta)^2 dt = h_n(\theta) + D_n,$$

where $D_n = D_n^c + D_n^J$. Since the integrand of D_n^c is predictable, D_n is a martingale difference. Proposition 1 (b) and (c) can be showed immediately following the results of Proposition 1 (a). ■

A.2 Proof of Theorem 1

Maximizing $\widehat{L}_{n,m}^{GH}$ proposed in Section 3.2 is equivalent to maximizing

$$\widehat{L}_{n,m}^{GH} = -\frac{1}{2n} \sum_{i=1}^{n} \left[\log(\widehat{h}_i(\theta)) + \frac{RV_i}{\widehat{h}_i(\theta)} \right].$$

We focus on $\widehat{L}_{n,m}^{GH}$ defined above in this proof. Define

$$\widehat{L}_{n,m}^{GH}(\theta) = -\frac{1}{2n} \sum_{i=1}^{n} \left[\log(\widehat{h}_i(\theta)) + \frac{RV_i}{\widehat{h}_i(\theta)} \right] = -\frac{1}{2n} \sum_{i=1}^{n} \widehat{l}_i^{GH}(\theta) \quad \text{and} \quad \widehat{\psi}_{n,m}^{GH}(\theta) = \frac{\partial \widehat{L}_{n,m}^{GH}(\theta)}{\partial \theta};$$
$$\widehat{L}_{n}^{GH}(\theta) = -\frac{1}{2n} \sum_{i=1}^{n} \left[\log(h_i(\theta)) + \frac{\int_{i-1}^{i} \sigma_t^2(\theta_0) dt}{h_i(\theta)} \right] \quad \text{and} \quad \widehat{\psi}_{n}^{GH}(\theta) = \frac{\partial \widehat{L}_{n}^{GH}(\theta)}{\partial \theta};$$
$$L_{n}^{GH}(\theta) = -\frac{1}{2n} \sum_{i=1}^{n} \left[\log(h_i(\theta)) + \frac{h_i(\theta_0)}{h_i(\theta)} \right] \quad \text{and} \quad \psi_{n}^{GH}(\theta) = \frac{\partial L_{n}^{GH}(\theta)}{\partial \theta}.$$

To ease notations, we denote derivatives of any given function g at x_0 by

$$\frac{\partial g(x_0)}{\partial x} = \left. \frac{\partial g(x)}{\partial x} \right|_{x=x_0}.$$

Lemma 1 in Kim and Wang (2016) shows that the dependence of $h_i(\theta)$ on the initial value decays exponentially. Thus, we may use the true initial value $\sigma_0^2(\theta_0)$ during the rest of the proofs.

Lemma 1. *Under Assumption 1 (a)-(f), we have*

(a) $E\left(R_i^2\right) = E\left(\int_{i-1}^{i} \sigma_t^2(\theta_0) dt\right) = E\left\{h_i(\theta_0)\right\}$, $\sup_{i\in\mathbb{N}} E(R_i^2) \le \frac{\omega_0^g + \beta_0^g \lambda \omega_L}{1-\alpha_0^g - \gamma_0} + E(h_1(\theta_0)) < \infty$, *and* $\sup_{i\in\mathbb{N}} E(\sup_{\theta\in\Theta} h_i(\theta)) < \infty$;

(b) *for any* $p \geq 1$,

$$\sup_{i\in\mathbb{N}} \left\| \sup_{\theta\in\Theta} \widehat{h}_i^{-1}(\theta) \frac{\partial \widehat{h}_i(\theta)}{\partial\theta_j} \right\|_{L_p} \leq C, \quad \sup_{i\in\mathbb{N}} \left\| \sup_{\theta\in\Theta} \widehat{h}_i^{-1}(\theta) \frac{\partial^2 \widehat{h}_i(\theta)}{\partial\theta_j\partial\theta_k} \right\|_{L_p} \leq C,$$
$$\textit{and} \ \sup_{i\in\mathbb{N}} \left\| \sup_{\theta\in\Theta} \widehat{h}_i^{-1}(\theta) \frac{\partial^3 \widehat{h}_i(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l} \right\|_{L_p} \leq C$$

for any $j,k,l \in \{1,2,3,4\}$*, where* $\theta = (\theta_1,\theta_2,\theta_3,\theta_4) = (\omega^g, \alpha^g, \beta^g, \gamma)$.

Proof of Lemma 1. The statements can be showed similar to the proofs of Lemma 2 (Kim and Wang, 2016). ■

Lemma 2. *Under Assumption 1 (a)-(d), we have*

(a) *there exists a neighborhood* $B(\theta_0)$ *of* θ_0 *such that*

$$\sup_{i\in\mathbb{N}} \left\| \sup_{\theta\in B(\theta_0)} \frac{\partial^3 \widehat{l}_i^{GH}(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l} \right\|_{L_1} < \infty$$

for any $j,k,l \in \{1,2,3,4\}$ *where* $\theta = (\theta_1,\theta_2,\theta_3,\theta_4) = (\omega^g, \alpha^g, \beta^g, \gamma)$;

(b) $-\nabla\psi_n^{GH}(\theta_0)$ *is a positive definite matrix for* $n \geq 5$.

Proof of Lemma 2. (a). We have

$$\begin{aligned}
\frac{\partial^3 \widehat{l}_i^{GH}(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l} &= \left\{1 - \frac{RV_i}{\widehat{h}_i(\theta)}\right\}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial^3 \widehat{h}_i(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l}\right\} \\
&+ \left\{2\frac{RV_i}{\widehat{h}_i(\theta)} - 1\right\}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial \widehat{h}_i(\theta)}{\partial\theta_j}\right\}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial^2 \widehat{h}_i(\theta)}{\partial\theta_k\partial\theta_l}\right\} \\
&+ \left\{2\frac{RV_i}{\widehat{h}_i(\theta)} - 1\right\}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial \widehat{h}_i(\theta)}{\partial\theta_k}\right\}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial^2 \widehat{h}_i(\theta)}{\partial\theta_j\partial\theta_l}\right\} \\
&+ \left\{2\frac{RV_i}{\widehat{h}_i(\theta)} - 1\right\}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial \widehat{h}_i(\theta)}{\partial\theta_l}\right\}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial^2 \widehat{h}_i(\theta)}{\partial\theta_j\partial\theta_k}\right\} \\
&+ \left\{2 - 6\frac{RV_i}{\widehat{h}_i(\theta)}\right\}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial \widehat{h}_i(\theta)}{\partial\theta_j}\right\}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial \widehat{h}_i(\theta)}{\partial\theta_k}\right\}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial \widehat{h}_i(\theta)}{\partial\theta_l}\right\}.
\end{aligned}$$

Since $\widehat{h}_i(\theta)$ stays away from 0, we have

$$E\left[\sup_{\theta\in B(\theta_0)} \left|\frac{RV_i}{\widehat{h}_i(\theta)}\left\{\frac{1}{\widehat{h}_i(\theta)}\frac{\partial \widehat{h}_i(\theta)}{\partial\theta_j\partial\theta_k\partial\theta_l}\right\}\right|\right]$$

$$
\begin{aligned}
&\leq C \left\| RV_i \right\|_{L_2} \left\| \sup_{\theta \in B(\theta_0)} \left| \frac{1}{\widehat{h}_i(\theta)} \frac{\partial \widehat{h}_i(\theta)}{\partial\theta_j \partial\theta_k \partial\theta_l} \right| \right\|_{L_2} \\
&\leq C,
\end{aligned}
$$

where the first and second inequalities are due to Hölder's inequality and Lemma 1 (b), respectively. Similarly, we can bound the rest of terms.

(b). The statement can be showed similar to the proofs of Lemma 4 (b) in Kim and Wang (2016). ■

Lemma 3. *Under Assumption 1 (a)-(f), we have*

$$
\sup_{\theta \in \Theta} \left| \widehat{L}_{n,m}^{GH}(\theta) - \widehat{L}_n^{GH}(\theta) \right| = O_p(m^{-1/4}), \tag{A.1}
$$

$$
\sup_{\theta \in \Theta} \left| \widehat{L}_n^{GH}(\theta) - L_n^{GH}(\theta) \right| = o_p(1), \tag{A.2}
$$

$$
\sup_{\theta \in \Theta} \left| \widehat{L}_{n,m}^{GH}(\theta) - L_n^{GH}(\theta) \right| = O_p(m^{-1/4}) + o_p(1). \tag{A.3}
$$

Proof of Lemma 3. First consider (A.1). We have

$$
\begin{aligned}
&\left| \widehat{L}_{n,m}^{GH}(\theta) - \widehat{L}_n^{GH}(\theta) \right| \\
&\leq \frac{1}{2n} \sum_{i=1}^{n} \left| \log \left(\widehat{h}_i(\theta)/h_i(\theta) \right) \right| + \frac{1}{2n} \sum_{i=1}^{n} \left| \frac{RV_i - \int_{i-1}^{i} \sigma_t^2(\theta_0) dt}{\widehat{h}_i(\theta)} \right| \\
&\quad + \frac{1}{2n} \sum_{i=1}^{n} \int_{i-1}^{i} \sigma_t^2(\theta_0) dt \left| \frac{\widehat{h}_i(\theta) - h_i(\theta)}{h_i(\theta) \widehat{h}_i(\theta)} \right| \\
&= (I) + (II) + (III).
\end{aligned}
$$

For (I), since $\log(1+x) \leq x$ for all $x > -1$, we have

$$
\begin{aligned}
&E \left\{ \sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \left| \log \left(\widehat{h}_i(\theta)/h_i(\theta) \right) \right| \right\} \\
&\leq \frac{1}{n} \sum_{i=2}^{n} E \left\{ \sup_{\theta \in \Theta} \left| \frac{\widehat{h}_i(\theta) - h_i(\theta)}{h_i(\theta)} \right| \right\} \\
&\leq \frac{C}{n} \sum_{i=2}^{n} \sum_{j=1}^{i-1} \gamma_u^{j-1} \left\{ E \left(\left| RV_{i-j} - \int_{i-j-1}^{i-j} \sigma_t^2(\theta_0) dt \right| \right) + E \left(\left| JV_{i-j} - \int_{i-j-1}^{i-j} L_t^2 d\Lambda_t \right| \right) \right\} \\
&\leq C m^{-1/4},
\end{aligned}
$$

where the second and last inequalities are due to Assumption 1 (a) and (e), respectively. For (II), since $\widehat{h}_i(\theta)$ stays away from zero, we have

$$
E \left[\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} \left| \frac{RV_i - \int_{i-1}^{i} \sigma_t^2(\theta_0) dt}{\widehat{h}_i(\theta)} \right| \right]
$$

$$
\begin{aligned}
&\leq C\frac{1}{n}\sum_{i=1}^{n} E\left[\left|RV_i - \int_{i-1}^{i} \sigma_t^2(\theta_0)dt\right|\right] \\
&\leq Cm^{-1/4},
\end{aligned}
$$

where the last inequality is due to Assumption 1 (e). For (III), since $\widehat{h}_i(\theta)$ and $h_i(\theta)$ stay away from zero, we have

$$
\begin{aligned}
&E\left[\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\int_{i-1}^{i}\sigma_t^2(\theta_0)dt\left|\frac{\widehat{h}_i(\theta)-h_i(\theta)}{h_i(\theta)\widehat{h}_i(\theta)}\right|\right] \\
&\leq \frac{C}{n}\sum_{i=1}^{n} E\left[\int_{i-1}^{i}\sigma_t^2(\theta_0)dt\sum_{j=1}^{i-1}\gamma_u^{j-1}\left\{\left|RV_{i-j}-\int_{i-j-1}^{i-j}\sigma_t^2(\theta_0)dt\right|+\left|JV_{i-j}-\int_{i-j-1}^{i-j}L_t^2 d\Lambda_t\right|\right\}\right] \\
&\leq \frac{C}{n}\sum_{i=1}^{n}\sum_{j=1}^{i-1}\gamma_u^{j-1}E\left[\left(\int_{i-1}^{i}\sigma_t^2(\theta_0)dt\right)^2\right]^{1/2} \\
&\qquad\times\left\{E\left(\left|RV_{i-j}-\int_{i-j-1}^{i-j}\sigma_t^2(\theta_0)dt\right|^2\right)^{1/2}+E\left(\left|JV_{i-j}-\int_{i-j-1}^{i-j}L_t^2 d\Lambda_t\right|^2\right)^{1/2}\right\} \\
&\leq Cm^{-1/4},
\end{aligned}
$$

where the last inequality is due to Assumption 1 (e). Thus, we have

$$
\sup_{\theta\in\Theta}\left|\widehat{L}_{n,m}^{GH}(\theta)-\widehat{L}_n^{GH}(\theta)\right| = O_p(m^{-1/4}).
$$

Consider (A.2). Simple algebra shows

$$
\widehat{L}_n^{GH}(\theta)-L_n^{GH}(\theta)=\frac{1}{2n}\sum_{i=1}^{n}\frac{D_i}{h_i(\theta)}.
$$

Since $h_i(\theta)$ is adapted to $\mathcal{F}_{i-1}$, $\frac{D_i}{h_i(\theta)}$ is also a martingale difference. Also, $\left|\frac{D_i}{h_i(\theta)}\right| \leq \frac{1}{\omega^g}\left|D_i\right|$, which implies the uniform integrability of $\left|\frac{D_i}{h_i(\theta)}\right|$. By the application of Theorem 2.22 in Hall and Heyde (2014), we can show

$$
\left|\widehat{L}_n^{GH}(\theta)-L_n^{GH}(\theta)\right| \to 0 \text{ in probability.}
$$

Now we will show its uniform convergence. Define

$$
G_n(\theta)=\widehat{L}_n^{GH}(\theta)-L_n^{GH}(\theta).
$$

If $G_n(\theta)$ is stochastic equicontinuous, Theorem 3 in Andrews (1992) implies that it uniformly converges to zero. So it is enough to show that $G_n(\theta)$ is stochastic equicontinuous. By the mean value theorem and Taylor expansion, there exists θ^* between θ and θ' such that

$$
|G_n(\theta)-G_n(\theta')| = \left|\frac{1}{2n}\sum_{i=1}^{n}\frac{\partial h_i(\theta^*)}{\partial\theta}\frac{D_i}{h_i^2(\theta^*)}(\theta-\theta')\right|
$$

$$\leq \quad C\frac{1}{2n}\sum_{i=1}^{n}\left\|\sup_{\theta^*\in\Theta}\frac{\partial h_i(\theta^*)}{\partial\theta}\frac{D_i}{h_i^2(\theta^*)}\right\|_{max}\|\theta-\theta'\|_{max}.$$

Similar to the proofs of Lemma 2 (e) in Kim and Wang (2016), we can show that

$$\left\|\sup_{\theta^*\in\Theta} h_i(\theta)^{-1}\frac{\partial h_i(\theta)}{\partial\theta_j}\right\|_{L_2}\leq C \quad \text{for } j=1,2,3,4.$$

By the Itô's Isometric, we have

$$\begin{aligned}
E\left(D_i^2\middle|\mathcal{F}_{i-1}\right) &= E\left(4\nu_0^2\alpha_0^{-4}\int_{i-1}^{i}\{\alpha_0(i-t-\alpha_0^{-1})e^{\alpha_0(i-t)}+1\}^2 Z_t^2 dt\middle|\mathcal{F}_{i-1}\right)+Var\left(D_n^J\right)\\
&= 4\nu_0^2\alpha_0^{-4}\int_{i-1}^{i}\{\alpha_0(i-t-\alpha_0^{-1})e^{\alpha_0(i-t)}+1\}^2(t-i+1)dt+Var\left(D_n^J\right)\\
&\leq C \text{ a.s.}
\end{aligned} \tag{A.4}$$

Thus,

$$E\left\{\left\|\sup_{\theta^*\in\Theta}\frac{\partial h_i(\theta^*)}{\partial\theta}\frac{D_i}{h_i^2(\theta^*)}\right\|_{max}\right\}\leq C,$$

which implies that $G_n(\theta)$ is stochastic equicontinuous.

Finally, the triangular inequality shows (A.3). ■

Proposition 2. *Under Assumption 1 (a)-(d), there is a unique maximizer of $L_n^{GH}(\theta)$ and as $m,n\to\infty$, $\widehat{\theta}^{GH}\to\theta_0$ in probability.*

Proof of Proposition 2. The statement can be showed similar to the proofs of Theorem 1 (Kim and Wang, 2016) together with the result of Lemma 3. ■

Proof of Theorem 1. By the mean value theorem and Taylor expansion, there exists θ^* between θ_0 and $\widehat{\theta}^{GH}$ such that

$$\widehat{\psi}_{n,m}^{GH}(\theta_0)-\widehat{\psi}_{n,m}^{GH}(\widehat{\theta}^{GH})=\widehat{\psi}_{n,m}^{GH}(\theta_0)=-\nabla\widehat{\psi}_{n,m}^{GH}(\theta^*)(\widehat{\theta}^{GH}-\theta_0).$$

If $-\nabla\widehat{\psi}_{n,m}^{GH}(\theta^*)\xrightarrow{p}-\nabla\psi_n^{GH}(\theta_0)$ which is a positive definite matrix by Lemma 2 (b), the convergence rate of $\|\widehat{\theta}^{GH}-\theta_0\|_{max}$ is the same as that of $\widehat{\psi}_{n,m}^{GH}(\theta_0)$. Thus, it is enough to show

$$\widehat{\psi}_{n,m}^{GH}(\theta_0)=O_p(m^{-1/4})+O_p(n^{-1/2})$$

and

$$\left\|\nabla\widehat{\psi}_{n,m}^{GH}(\theta^*)-\nabla\psi_n^{GH}(\theta_0)\right\|_{max}=o_p(1).$$

First consider $\widehat{\psi}_{n,m}^{GH}(\theta_0)=O_p(m^{-1/4})+O_p(n^{-1/2})$. Similar to the proofs of Theorem 2 (Kim and Wang, 2016), we can show that

$$\widehat{\psi}_{n,m}^{GH}(\theta_0) \quad = \quad \psi_n^{GH}(\theta_0)+\frac{1}{2n}\sum_{i=1}^{n}\frac{\partial h_i(\theta_0)}{\partial\theta}h_i(\theta_0)^{-1}\frac{D_i}{h_i(\theta_0)}+O_p(m^{-1/4})$$

$$= \quad \frac{1}{2n}\sum_{i=1}^{n}\frac{\partial h_i(\theta_0)}{\partial \theta}h_i(\theta_0)^{-1}\frac{D_i}{h_i(\theta_0)} + O_p(m^{-1/4}). \tag{A.5}$$

By the application of the Itô's lemma and Itô's isometry, we can show for any $j \in \{1,2,3,4\}$,

$$\begin{aligned}
& E\left[\left(\frac{1}{2n}\sum_{i=1}^{n}\frac{\partial h_i(\theta_0)}{\partial \theta_j}h_i(\theta_0)^{-1}\frac{D_i}{h_i(\theta_0)}\right)^2\right] \\
&= \frac{1}{4n^2}\sum_{i=1}^{n}E\left[\left(\frac{\partial h_i(\theta_0)}{\partial \theta_j}\right)^2 h_i(\theta_0)^{-2}\frac{E\left[D_i^2|\mathcal{F}_{i-1}\right]}{h_i^2(\theta_0)}\right] \\
&\le \frac{C}{n^2}\sum_{i=1}^{n}E\left[\left(\frac{\partial h_i(\theta_0)}{\partial \theta_j}\right)^2 h_i(\theta_0)^{-2}\frac{1}{h_i^2(\theta_0)}\right] \\
&\le \frac{C}{n^2}\sum_{i=1}^{n}E\left[\left(\frac{\partial h_i(\theta_0)}{\partial \theta_j}\right)^2 h_i(\theta_0)^{-2}\right] \\
&\le Cn^{-1},
\end{aligned} \tag{A.6}$$

where the first and last inequalities are due to (A.4) and Lemma 1 (b), respectively. Similar to the proofs of Theorem 2 (Kim and Wang, 2016) together with the results of Lemma 2 and Proposition 2, we can show that

$$\left\|\nabla\widehat{\psi}_{n,m}^{GH}(\theta^*) - \nabla\psi_n^{GH}(\theta_0)\right\|_{max} = o_p(1).$$

■

A.3 Proof of Theorem 2

Proof of Theorem 2. By the mean value theorem and Taylor expansion, we have for some θ^* between θ_0 and $\widehat{\theta}^{GH}$,

$$\begin{aligned}
& -\nabla\widehat{\psi}_{n,m}^{GH}(\theta^*)(\widehat{\theta}^{GH} - \theta_0) \\
&= \widehat{\psi}_n^{GH}(\theta_0) + \left\{\widehat{\psi}_{n,m}^{GH}(\theta_0) - \widehat{\psi}_n^{GH}(\theta_0)\right\} \\
&= \frac{1}{2n}\sum_{i=1}^{n}\frac{\partial h_i(\theta_0)}{\partial \theta}h_i(\theta_0)^{-1}\frac{D_i}{h_i(\theta_0)} + O_p(m^{-1/4}),
\end{aligned}$$

where the second equality is due to (A.5). By the ergodic theorem and the result in the proof of Theorem 1, we have

$$-\nabla\widehat{\psi}_{n,m}^{GH}(\theta^*) \to B \text{ in probability,}$$

and B is a positive definite matrix. For any $f \in \mathbb{R}^4$, let

$$d_i = f^T\frac{\partial h_i(\theta_0)}{\partial \theta}h_i(\theta_0)^{-1}\frac{D_i}{h_i(\theta_0)}.$$

Then d_i is a martingale difference with $E(d_i^2) < \infty$.

Since $\left(D_i, \int_{i-1}^{i} \sigma_t^2(\theta_0)dt, R_i^2\right)$'s are stationary and ergodic processes, d_i is also stationary and ergodic. By the martingale central limit theorem and Cramér-Wold device, we have

$$-\sqrt{n}\widehat{\psi}_n^{GH}(\theta_0) = \sqrt{n}\frac{1}{2n}\sum_{i=1}^{n}\frac{\partial h_i(\theta_0)}{\partial\theta}h_i(\theta_0)^{-1}\frac{D_i}{h_i(\theta_0)} \xrightarrow{d} N(0, A^{GH}).$$

Therefore, by Slutsky's theorem, we conclude that

$$\sqrt{n}(\widehat{\theta}^{GH} - \theta_0) \quad \xrightarrow{d} \quad N(0, B^{-1}A^{GH}B^{-1}).$$

■

A.4 Proof of Theorem 3

Maximizing $\widehat{L}_{n,m}^{GHO}$ is equivalent to maximizing

$$\widehat{L}_{n,m}^{GHO}(\phi) = -\frac{1}{2n}\sum_{i=1}^{n}\left[\log(\widehat{h}_i(\theta)) + \frac{RV_i}{\widehat{h}_i(\theta)}\right] - \frac{1}{2n}\sum_{i=1}^{n}\left[\log\sigma_e^2 + \frac{(NV_{i-1} - \widehat{f}_i(\varphi))^2}{\sigma_e^2}\right],$$

where $\widehat{f}_i(\varphi) = b + a\widehat{h}_i(\theta)$. We focus on $\widehat{L}_{n,m}^{GHO}$ defined above in this proof. Define

$$\begin{aligned}\widehat{L}_{n,m}^{GHO}(\phi) &= -\frac{1}{2n}\sum_{i=1}^{n}\left[\log(\widehat{h}_i(\theta)) + \frac{RV_i}{\widehat{h}_i(\theta)}\right] - \frac{1}{2n}\sum_{i=1}^{n}\left[\log\sigma_e^2 + \frac{(NV_{i-1} - \widehat{f}_i(\varphi))^2}{\sigma_e^2}\right] \\ &= -\frac{1}{2n}\sum_{i=1}^{n}\widehat{l}_i^{GH}(\theta) - \frac{1}{2n}\sum_{i=1}^{n}\widehat{l}_i^{GO}(\phi)\end{aligned}$$

and

$$\widehat{\psi}_{n,m}^{GHO}(\phi) = \frac{\partial\widehat{L}_{n,m}^{GHO}(\phi)}{\partial\phi};$$

$$\widehat{L}_n^{GHO}(\phi) = -\frac{1}{2n}\sum_{i=1}^{n}\left[\log(h_i(\theta)) + \frac{\int_{i-1}^{i}\sigma_t^2(\theta_0)dt}{h_i(\theta)}\right] - \frac{1}{2n}\sum_{i=1}^{n}\left[\log\sigma_e^2 + \frac{(NV_{i-1} - f_i(\varphi))^2}{\sigma_e^2}\right]$$

and

$$\widehat{\psi}_n^{GHO}(\phi) = \frac{\partial\widehat{L}_n^{GHO}(\phi)}{\partial\phi};$$

$$L_n^{GHO}(\phi) = -\frac{1}{2n}\sum_{i=1}^{n}\left(\log h_i(\theta) + \frac{h_i(\theta_0)}{h_i(\theta)}\right) - \frac{1}{2n}\sum_{i=1}^{n}\left\{\log\sigma_e^2 + \frac{[f_i(\varphi) - f_i(\varphi_0)]^2 + \sigma_{e0}^2}{\sigma_e^2}\right\}$$

and

$$\psi_n^{GHO}(\phi) = \frac{\partial L_n^{GHO}(\phi)}{\partial\phi}.$$

Lemma 4. *Under Assumption 1 (a)–(f) and Assumption 2 (a)–(b),*

(a) there exists a neighborhood $B(\phi_0)$ around ϕ_0 such that

$$\sup_{i\in\mathbb{N}} \left\| \sup_{\phi\in B(\phi_0)} \frac{\partial^3 \widehat{l}_i^{GO}(\phi)}{\partial\phi_j\partial\phi_k\partial\phi_l} \right\|_{L_1} < \infty$$

for any $j,k,l \in \{1,2,\ldots,7\}$, where $\phi = (\phi_1,\phi_2,\phi_3,\phi_4,\phi_5,\phi_6,\phi_7) = (\omega^g, \alpha^g, \beta^g, \gamma, a, b, \sigma_e^2)$;

(b) $-\nabla\psi_n^{GHO}(\theta_0)$ is a positive definite matrix for $n \geq 7$.

Proof of Lemma 4. (a). Similar to the proofs of Lemma 2 (Kim and Wang, 2016), we can show the statement.

(b). We have

$$\begin{aligned}
& -\nabla\psi_n^{GHO}(\phi_0) \\
=& \frac{1}{2n}\begin{pmatrix} \sum_{i=1}^n \frac{\partial h_i(\theta_0)}{\partial\varphi}\frac{\partial h_i(\theta_0)}{\partial\varphi^T} h_i(\theta_0)^{-2} & \mathbf{0}_{6\times1} \\ \mathbf{0}_{6\times1}^T & 0 \end{pmatrix} + \frac{1}{2n}\begin{pmatrix} \frac{2}{\sigma_{e0}^2}\sum_{i=1}^n \frac{\partial f_i(\varphi_0)}{\partial\varphi}\frac{\partial f_i(\varphi_0)}{\partial\varphi^T} & \mathbf{0}_{6\times1} \\ \mathbf{0}_{6\times1}^T & \frac{n}{\sigma_{e0}^4} \end{pmatrix},
\end{aligned}$$

where

$$\begin{aligned}
\frac{\partial h_i(\theta_0)}{\partial\varphi} &= \left(\frac{\partial h_i(\theta_0)}{\partial\omega^g}, \frac{\partial h_i(\theta_0)}{\partial\alpha^g}, \frac{\partial h_i(\theta_0)}{\partial\beta^g}, \frac{\partial h_i(\theta_0)}{\partial\gamma}, 0, 0\right)^T, \\
\frac{\partial f_j(\varphi_0)}{\partial\varphi} &= \left(\frac{\partial f_j(\varphi_0)}{\partial\omega^g}, \frac{\partial f_j(\varphi_0)}{\partial\alpha^g}, \frac{\partial f_j(\varphi_0)}{\partial\beta^g}, \frac{\partial f_j(\varphi_0)}{\partial\gamma}, \frac{\partial f_j(\varphi_0)}{\partial a}, \frac{\partial f_j(\varphi_0)}{\partial b}\right)^T.
\end{aligned}$$

Then, similar to the proofs of Lemma 4 (b) in Kim and Wang (2016), we can show the statement. ■

Lemma 5. *Under Assumption 1 (a)-(f) and Assumption 2 (a)–(b), we have*

$$\begin{aligned}
\sup_{\phi\in\Phi}\left|\widehat{L}_{n,m}^{GHO}(\phi) - \widehat{L}_n^{GHO}(\phi)\right| &= O_p(m^{-1/4}), \\
\sup_{\phi\in\Phi}\left|\widehat{L}_n^{GHO}(\phi) - L_n^{GHO}(\phi)\right| &= o_p(1), \\
\sup_{\phi\in\Phi}\left|\widehat{L}_{n,m}^{GHO}(\phi) - L_n^{GHO}(\phi)\right| &= O_p(m^{-1/4}) + o_p(1).
\end{aligned}$$

Proof of Lemma 5. First consider $\left|\widehat{L}_{n,m}^{GHO}(\phi) - \widehat{L}_n^{GHO}(\phi)\right|$. Since

$$\sup_{\varphi}|\widehat{f}_i(\varphi) - f_i(\varphi)| \leq b_u \sup_{\theta\in\Theta}|\widehat{h}_i(\theta) - h_i(\theta)| = O_p(m^{-1/4}),$$

with the result in Lemma 3, we have

$$\sup_{\phi} \left| \widehat{L}_{n,m}^{GHO}(\phi) - \widehat{L}_{n}^{GHO}(\phi) \right| = O_p(m^{-1/4}).$$

Now consider $\left| \widehat{L}_{n}^{GHO}(\phi) - L_{n}^{GHO}(\phi) \right|$. Simple algebra shows

$$\begin{aligned}
& \widehat{L}_{n}^{GHO}(\phi) - L_{n}^{GHO}(\phi) \\
= & \ \frac{1}{2n} \sum_{i=1}^{n} \frac{-D_i}{h_i(\theta)} - \frac{1}{2n} \sum_{i=1}^{n} \frac{[NV_{i-1} - f_i(\varphi)]^2 - [f_i(\varphi) - f_i(\varphi_0)]^2 - \sigma_{e0}^2}{\sigma_e^2} \\
= & \ \frac{1}{2n} \sum_{i=1}^{n} \frac{-D_i}{h_i(\theta)} - \frac{1}{2n} \sum_{i=1}^{n} \frac{e_i^2 - \sigma_{e0}^2 - 2e_i\left[f_i(\varphi) - f_i(\varphi_0)\right]}{\sigma_e^2}.
\end{aligned}$$

Since σ_e^2 is bounded and e_i is independent of $f_i(\varphi) - f_i(\varphi_0)$, the second term of the right hand side uniformly converges to zero with the convergence rate $n^{-1/2}$. Lemma 3 shows the first term of the right hand side uniformly converges to zero. Therefore, we have

$$\sup_{\phi \in \Phi} \left| \widehat{L}_{n}^{GHO}(\phi) - L_{n}^{GHO}(\phi) \right| = o_p(1).$$

Finally, by the triangular inequality, we have

$$\sup_{\phi \in \Phi} \left| \widehat{L}_{n,m}^{GHO}(\phi) - L_{n}^{GHO}(\phi) \right| = O_p(m^{-1/4}) + o_p(1).$$

■

Proposition 3. *Under Assumption 1 (a)-(f) and Assumption 2 (a)–(b), there exists a unique maximizer for $L_n^{GHO}(\phi)$. As $m, n \to \infty$, $\widehat{\phi}^{GHO} \to \phi_0$ in probability, where ϕ_0 is a vector of true parameters.*

Proof of Proposition 3. According to the definition of $L_n^{GHO}(\phi)$, we have

$$\begin{aligned}
\max_{\phi \in \Phi} \ L_n^{GHO}(\phi) \leq & - \frac{1}{2n} \sum_{i=1}^{n} \min_{\theta_i \in \Theta} \ \left[\log(h_i(\theta_i)) + \frac{h_i(\theta_0)}{h_i(\theta_i)} \right] \\
& - \frac{1}{2n} \sum_{i=1}^{n} \min_{\phi_i \in \Phi} \ \left[\log \sigma_{ei}^2 + \frac{(f_i(\varphi_i) - f_i(\varphi_0))^2 + \sigma_{e0}^2}{\sigma_{ei}^2} \right].
\end{aligned}$$

Then, similar to the proofs in Theorem 1 of Kim and Wang (2016), we can show the uniqueness of the solution of $L_n^{GHO}(\phi)$, which together with Lemma 5 implies Proposition 3. ■

Proof of Theorem 3. By the mean value theorem and Taylor expansion, we have

$$\widehat{\psi}_{n,m}^{GHO}(\widehat{\phi}^{GHO}) - \widehat{\psi}_{n,m}^{GHO}(\phi_0) = -\widehat{\psi}_{n,m}^{GHO}(\phi_0) = \nabla \widehat{\psi}_{n,m}^{GHO}(\phi^*)(\widehat{\phi}^{GHO} - \phi_0),$$

where ϕ^* is between ϕ_0 and $\widetilde{\phi}^{GHO}$. According to Lemma 4 (b), $-\nabla\psi_n^{GHO}(\phi_0)$ is a positive definite matrix. If $-\nabla\widehat{\psi}_{n,m}^{GHO}(\phi^*) \xrightarrow{p} -\nabla\psi_n^{GHO}(\phi_0)$, then the convergence rate of $\widehat{\phi}^{GHO} - \phi_0$ is the same as the convergence rate of $\widehat{\psi}_{n,m}^{GHO}(\phi_0)$.

By the similar arguments in the proof of Theorem 1, we can show

$$\left\|\widehat{\psi}_{n,m}^{GHO}(\phi_0) - \widehat{\psi}_n^{GHO}(\phi_0)\right\|_{L_1} \leq Cm^{-1/4}.$$

We have

$$\widehat{\psi}_n^{GHO}(\phi_0) = \frac{1}{2n}\sum_{i=1}^n \begin{pmatrix} \frac{D_i}{h_i^2(\theta_0)}\frac{\partial h_i(\theta_0)}{\partial\varphi} \\ 0 \end{pmatrix} - \frac{1}{2n}\sum_{i=1}^n \begin{pmatrix} \frac{-2e_i}{\sigma_{e0}^2}\frac{\partial f_i(\varphi_0)}{\partial\varphi} \\ \frac{1}{\sigma_{e0}^2} - \frac{e_i^2}{\sigma_{e0}^4} \end{pmatrix}. \tag{A.7}$$

The arguments in the proof of Theorem 1 shows that the first term of the right side of (A.7) is $O_p(n^{-1/2})$. Since e_i is independent of $\frac{\partial f_i(\varphi_0)}{\partial\varphi}$, the second term of the right side of (A.7) is also $O_p(n^{-1/2})$. Thus, the convergence rate of $\widehat{\psi}_{n,m}^{GHO}(\phi_0)$ is $n^{-1/2} + m^{-1/4}$.

Similar to the proof of Theorem 1, we can show

$$\left\|\nabla\widehat{\psi}_{n,m}^{GHO}(\phi^*) - \nabla\psi_n^{GHO}(\phi_0)\right\|_{\max} = o_p(1).$$

Therefore, the statement is proved. ■

A.5 Proof of Theorem 4

Proof of Theorem 4. Since the mean value theorem and Taylor expansion provides

$$\widehat{\psi}_{n,m}^{GHO}(\widehat{\phi}^{GHO}) - \widehat{\psi}_{n,m}^{GHO}(\phi_0) = -\widehat{\psi}_{n,m}^{GHO}(\phi_0) = \nabla\widehat{\psi}_{n,m}^{GHO}(\phi^*)(\widehat{\phi}^{GHO} - \phi_0),$$

where ϕ^* is between ϕ_0 and $\widehat{\phi}^{GHO}$, we have

$$\sqrt{n}(\widehat{\phi}^{GHO} - \phi_0) = -\sqrt{n}\left(\nabla\psi_n^{GHO}(\phi_0) + o_p(1)\right)^{-1}\widehat{\psi}_n^{GHO}(\phi_0) + o_p(1),$$

where the equality can be showed similar to the proof of Theorem 1. Since e_i is independent of D_i and (D_i, e_i, Z_i^2) is stationary and ergodic, by the Cramér-Wold device and the martingale central limit theorem, we have

$$\sqrt{n}\widehat{\psi}_n^{GHO}(\phi_0) = \frac{\sqrt{n}}{2n}\sum_{i=1}^n \begin{pmatrix} \frac{D_i}{h_i^2(\theta_0)}\frac{\partial h_i(\theta_0)}{\partial\varphi} \\ 0 \end{pmatrix} - \frac{\sqrt{n}}{2n}\sum_{i=1}^n \begin{pmatrix} \frac{-2e_i}{\sigma_{e0}^2}\frac{\partial f_i(\varphi_0)}{\partial\varphi} \\ \frac{1}{\sigma_{e0}^2} - \frac{e_i^2}{\sigma_{e0}^4} \end{pmatrix} \xrightarrow{d} N(0, A^{GHO}).$$

On the other hand,

$$\begin{aligned} &-\nabla\psi_n^{GHO}(\phi_0) \\ &= \frac{1}{2n}\begin{pmatrix} \sum_{i=1}^n \frac{\partial h_i(\theta_0)}{\partial\varphi}\frac{\partial h_i(\theta_0)}{\partial\varphi^T}h_i(\theta_0)^{-2} & \mathbf{0}_{6\times 1} \\ \mathbf{0}_{6\times 1}^T & 0 \end{pmatrix} + \frac{1}{2n}\begin{pmatrix} \frac{2}{\sigma_{e0}^2}\sum_{i=1}^n \frac{\partial f_i(\varphi_0)}{\partial\varphi}\frac{\partial f_i(\varphi_0)}{\partial\varphi^T} & \mathbf{0}_{6\times 1} \\ \mathbf{0}_{6\times 1}^T & \frac{n}{\sigma_{e0}^4} \end{pmatrix} \end{aligned}$$

$$\xrightarrow{p} B^{GHO}.$$

Therefore, by the Slutsky's theorem,

$$\sqrt{n}(\widehat{\phi}^{GHO} - \phi_0) \xrightarrow{d} N(0, (B^{GHO})^{-1} A^{GHO} (B^{GHO})^{-1}).$$

■

Acknowledgements

The research of Xinyu Song was supported by the Fundamental Research Funds for the Central Universities (2018110128) and China Scholarship Council (201806485017). The research of Donggyu Kim was supported in part by KAIST Settlement/Research Subsidies for Newly-hired Faculty grant G04170049 and KAIST Basic Research Funds by Faculty (A0601003029). The research of Huiling Yuan was supported by the State Scholarship Fund. The research of Xiangyu Cui was supported by National Natural Science Foundation of China (71671106). The research of Zhiping Lu was supported by Natural Science Foundation of Shanghai (17ZR1409000) and the 111 Project (B14019). The research of Young Zhou was supported by the State Key Program of National Natural Science Foundation of China (71331006), the State Key Program in the Major Research Plan of National Natural Science Foundation of China (91546202). The research of Yazhen Wang was supported in part by NSF Grants DMS-12-65203, DMS-15-28375, and DMS-17-07605.

We thank the Associate Editor, Viktor Todorov, and two anonymous referees for many constructive suggestions that have significantly improved the paper.

This research was performed using the compute resources and assistance of the UW-Madison Center For High Throughput Computing (CHTC) in the Department of Computer Sciences. The CHTC is supported by UW-Madison, the Advanced Computing Initiative, the Wisconsin Alumni Research Foundation, the Wisconsin Institutes for Discovery, and the National Science Foundation, and is an active member of the Open Science Grid, which is supported by the National Science Foundation and the U.S. Department of Energy's Office of Science.

References

Admati, A. R. and Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies*, 1(1):3–40.

Aït-Sahalia, Y., Fan, J., and Xiu, D. (2010). High-frequency covariance estimates with noisy and asynchronous financial data. *Journal of the American Statistical Association*, 105(492):1504–1517.

Aït-Sahalia, Y., Jacod, J., and Li, J. (2012). Testing for jumps in noisy high frequency data. *Journal of Econometrics*, 168(2):207–222.

Ait-Sahalia, Y. and Yu, J. (2008). High frequency market microstructure noise estimates and liquidity measures. Technical report, National Bureau of Economic Research.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2003a). Some like it smooth, and some like it rough: Untangling continuous and jump components in measuring, modeling, and forecasting asset return volatility.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The review of economics and statistics*, 89(4):701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003b). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Andersen, T. G., Bollerslev, T., et al. (1997). Intraday periodicity and volatility persistence in financial markets. *Journal of empirical finance*, 4(2-3):115–158.

Andersen, T. G., Thyrsgaard, M., Todorov, V., et al. (2018). Time-varying periodicity in intraday volatility. Technical report.

Andrews, D. W. (1992). Generic uniform convergence. *Econometric theory*, 8(2):241–257.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2008). Designing realized kernels to measure the ex post variation of equity prices in the presence of noise. *Econometrica*, 76(6):1481–1536.

Barndorff-Nielsen, O. E. and Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of financial Econometrics*, 4(1):1–30.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 31(3):307–327.

Christensen, K., Kinnebrock, S., and Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159(1):116–133.

Corsi, F., Pirino, D., and Reno, R. (2010). Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics*, 159(2):276–288.

Davies, R. and Tauchen, G. (2018). Data-driven jump detection thresholds for application in jump regressions. *Econometrics*, 6(2):16.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

Engle, R. F. and Gallo, G. M. (2006). A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1):3–27.

Fan, J. and Wang, Y. (2007). Multi-scale jump and volatility analysis for high-frequency financial data. *Journal of the American Statistical Association*, 102(480):1349–1362.

Hall, P. and Heyde, C. C. (2014). *Martingale limit theory and its application*. Academic press.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized garch: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

Hong, H. and Wang, J. (2000). Trading and returns under periodic market closures. *The Journal of Finance*, 55(1):297–354.

Jacod, J., Li, Y., Mykland, P. A., Podolskij, M., and Vetter, M. (2009). Microstructure noise in the continuous case: the pre-averaging approach. *Stochastic processes and their applications*, 119(7):2249–2276.

Kim, D. and Wang, Y. (2016). Unified discrete-time and continuous-time models and statistical inferences for merged low-frequency and high-frequency financial data. *Journal of Econometrics*, 194(2):220–230.

Kim, D., Wang, Y., and Zou, J. (2016). Asymptotic theory for large volatility matrix estimation based on high-frequency financial data. *Stochastic Processes and their Applications*, 126(11):3527–3577.

Mancini, C. (2004). Estimation of the characteristics of the jumps of a general poisson-diffusion model. *Scandinavian Actuarial Journal*, 2004(1):42–52.

Shephard, N. and Sheppard, K. (2010). Realising the future: forecasting with high-frequency-based volatility (heavy) models. *Journal of Applied Econometrics*, 25(2):197–231.

Todorov, V. (2019). Nonparametric spot volatility from options. *Manuscript*.

Wang, Y. (2002). Asymptotic nonequivalence of garch models and diffusions. *The Annals of Statistics*, 30(3):754–783.

Xiu, D. (2010). Quasi-maximum likelihood estimation of volatility with high frequency data. *Journal of Econometrics*, 159(1):235–250.

Zhang, L. (2006). Efficient estimation of stochastic volatility using noisy observations: A multi-scale approach. *Bernoulli*, 12(6):1019–1043.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411.

Zhang, X., Kim, D., and Wang, Y. (2016). Jump variation estimation with noisy high frequency financial data via wavelets. *Econometrics*, 4(3):34.

Table 1: The mean squared errors (MSEs) for the jump process parameters ω_L and λ given $n = 125, 250, 500, 1000$ and $m = 390, 780, 2340, 23400$.

	MSE							
	ω_L				λ			
$n \setminus m$	390	780	2340	23400	390	780	2340	23400
125	1.772×10^{-4}	3.398×10^{-5}	1.703×10^{-6}	8.315×10^{-10}	570.841	399.781	63.034	0.278
250	1.732×10^{-4}	3.311×10^{-5}	1.654×10^{-6}	4.604×10^{-10}	573.157	398.359	61.935	0.123
500	1.726×10^{-4}	3.267×10^{-5}	1.645×10^{-6}	2.254×10^{-10}	575.315	398.013	61.862	0.034
1000	1.720×10^{-4}	3.248×10^{-5}	1.634×10^{-6}	1.229×10^{-10}	575.976	398.531	61.876	0.002

Table 2: The mean squared errors (MSEs) for the QMLE-HL and QMLE-HLO methods on estimating realized GARCH volatility parameters for $n = 125, 250, 500, 1000$ and $m = 390, 780, 2340, 23400$.

		MSE $\times 10^3$										
		QMLE-HL				**QMLE-HLO**						
n	m	ω^g	α^g	β^g	γ	ω^g	α^g	β^g	γ	a	b	σ_e
125	390	16.742	26.239	358.192	49.352	6.473	18.909	132.689	25.984	77.966	23.278	0.069
	780	15.589	19.289	341.371	38.736	6.263	15.415	129.867	17.473	59.975	17.768	0.045
	2340	11.813	14.378	331.959	33.394	4.700	11.600	114.979	11.894	48.852	12.523	0.030
	23400	11.057	10.994	280.268	32.078	4.518	10.267	97.788	9.328	41.762	10.750	0.026
250	390	8.577	19.426	207.105	31.688	3.308	15.345	78.090	20.216	41.400	11.941	0.069
	780	7.898	12.899	198.560	22.871	3.249	10.720	73.681	10.795	28.111	7.598	0.042
	2340	5.706	8.621	189.687	18.175	2.590	7.272	54.402	6.648	24.265	6.169	0.020
	23400	5.272	6.463	139.513	15.775	2.242	5.352	45.373	4.786	21.990	5.495	0.013
500	390	4.327	16.835	125.098	24.255	1.832	14.185	48.383	17.562	18.030	5.530	0.068
	780	4.128	10.187	113.344	13.848	1.796	8.523	45.854	8.414	14.791	4.153	0.038
	2340	2.828	5.814	98.984	10.152	1.287	5.261	27.692	4.612	12.496	3.144	0.015
	23400	2.161	3.456	66.545	7.939	1.020	3.050	19.562	2.456	10.993	2.593	0.007
1000	390	2.299	15.110	72.092	20.232	1.238	13.620	34.918	16.413	8.015	2.922	0.065
	780	2.198	8.834	64.902	10.826	1.055	7.572	32.384	7.184	6.122	1.444	0.036
	2340	1.355	4.185	49.286	6.278	0.636	3.752	13.372	3.380	5.044	1.314	0.014
	23400	0.985	1.812	33.705	4.011	0.433	1.769	9.740	1.414	4.913	1.166	0.004

Table 3: The mean squared prediction errors (MSPEs) of the realized GARCH volatility predictors $h_i(\theta)$ proposed in realized GARCH-Itô model with the QMLE-HL and the QMLE-HLO methods, the GARCH volatility predictor $h_{i0}(\theta_0)$ proposed in unified GARCH-Itô model (Kim and Wang, 2016), and the benchmark jump-adjusted MSRV method for $n = 125, 250, 500, 1000$ and $m = 390, 780, 2340, 23400$.

		MSPE $\times 10^3$			
		Realized GARCH-Itô		**Unified GARCH-Itô**	**Jump-adjusted**
n	m	**QMLE-HL**	**QMLE-HLO**	**QMLE-HL**	**MSRV**
125	390	4.859	4.721	14.931	26.944
	780	4.558	4.452	19.979	27.642
	2340	2.825	2.765	22.607	21.764
	23400	1.923	1.859	26.744	15.221
250	390	4.223	3.721	14.283	27.608
	780	3.745	3.601	19.379	29.218
	2340	2.227	1.899	21.835	21.825
	23400	1.370	1.086	26.407	14.580
500	390	3.390	3.085	13.614	27.951
	780	3.212	3.078	18.987	28.942
	2340	1.668	1.413	21.657	21.755
	23400	0.948	0.726	25.878	15.401
1000	390	2.922	2.909	13.423	27.152
	780	2.915	2.851	19.036	28.375
	2340	1.220	1.197	21.761	22.001
	23400	0.630	0.547	25.697	15.170

Table 4: Mean squared prediction errors (MSPEs) of the realized GARCH-Itô estimates with the QMLE-HL and the QMLE-HLO, the unified GARCH-Itô estimates with the QMLE-HL, and the jump-adjusted MSRV estimates.

Forecast Origin	**MSPE $\times 10^{11}$**			
	Realized GARCH-Itô		**Unified GARCH-Itô**	**Jump-adjusted**
	QMLE-HL	**QMLE-HLO**	**QMLE-HL**	**MSRV**
$h = 168$	2.458	2.201	3.894	3.166
$h = 188$	2.279	2.141	3.527	2.885
$h = 210$	3.297	3.262	3.511	4.657

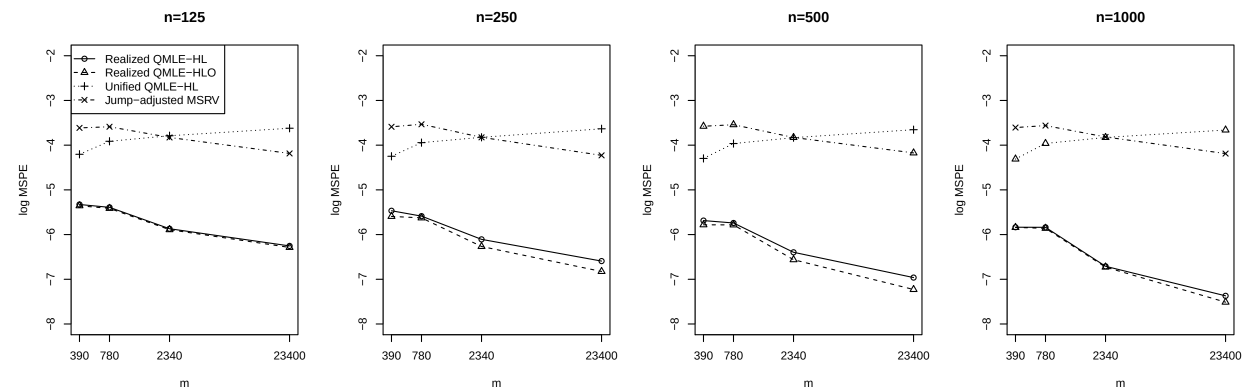

Figure 1: The log mean squared prediction errors (MSPEs) of the realized GARCH volatility predictors $h_i(\theta)$, the GARCH volatility predictors $h_{i0}(\theta_0)$ and the benchmark jump-adjusted MSRV predictors RV_{i-1} against m for the different n choices.

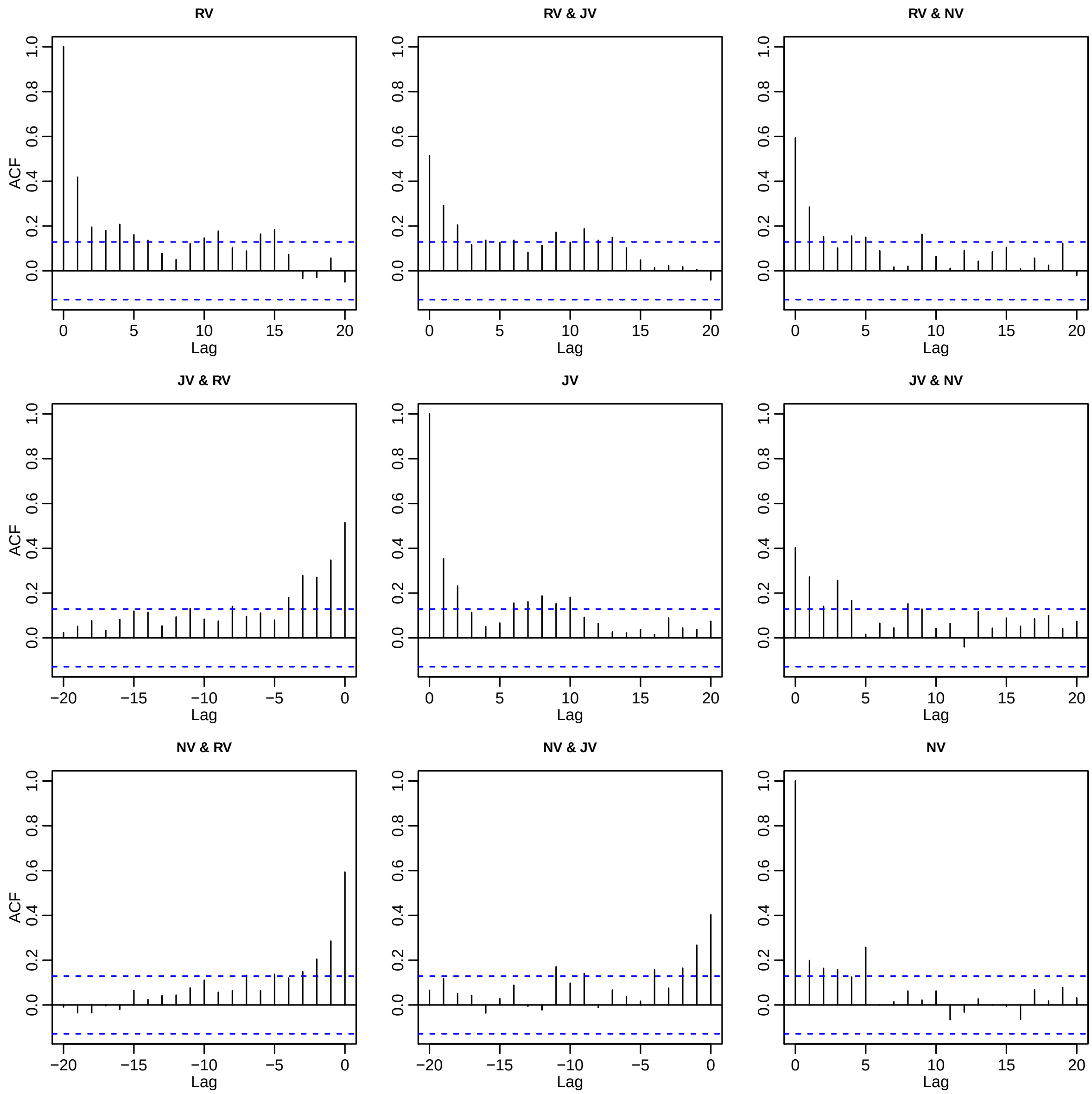

Figure 2: Auto-correlation functions (ACFs) plots for the time series of the daily jump-adjusted MSRV (RV) estimators, the daily jump variation (JV) estimators and the daily nonparametric volatility (NV) estimators with option data.

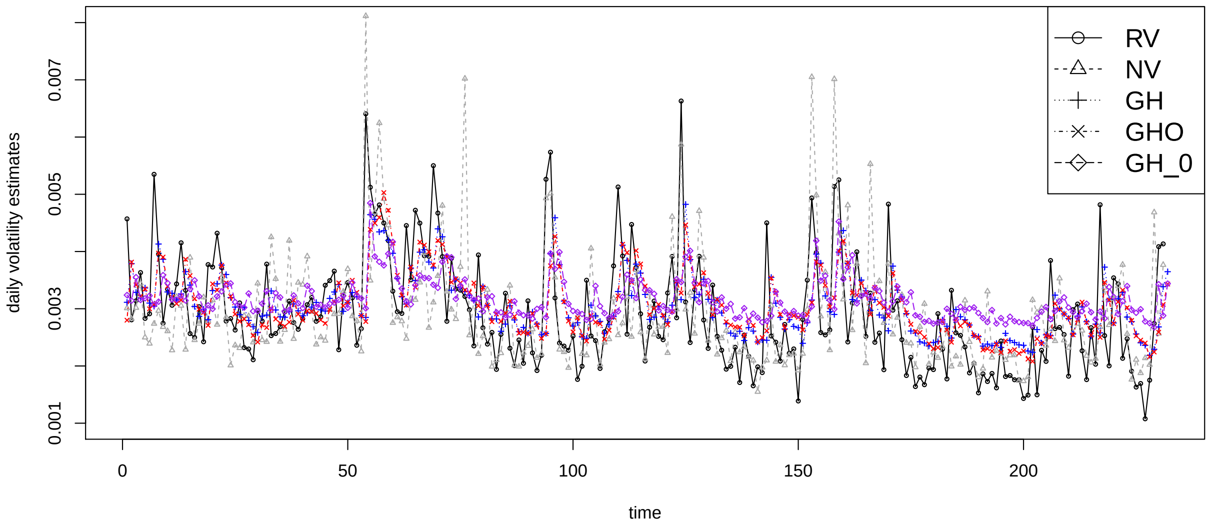

Figure 3: Daily volatility estimates with 1) RV: jump-adjusted MSRV estimates $\sqrt{RV_i}$; 2) NV: option-based nonparametric volatility estimates: $\sqrt{NV_i}$; 3) GH: realized GARCH volatility estimates $\sqrt{\widehat{h}_i(\widehat{\theta}^{GH})}$ with the QMLE-HL; 4) GHO: realized GARCH volatility estimates $\sqrt{\widehat{h}_i(\widehat{\theta}^{GHO})}$ with the QMLE-HLO; 5) GH_0: GARCH volatility estimates $\sqrt{h_{i0}(\widehat{\theta}_0^{GH})}$ given the unified GARCH-Itô model.