Unified Discrete-Time and Continuous-Time Models and Statistical Inferences for Merged Low-Frequency and High-Frequency Financial Data

Donggyu Kim and Yazhen Wang
University of Wisconsin-Madison
December 24, 2014

Abstract

This paper introduces a unified model, which can accommodate both a continuous-time Itô process used to model high-frequency stock prices and a GARCH process employed to model low-frequency stock prices, by embedding a discrete-time GARCH volatility in its continuous-time instantaneous volatility. This model is called a unified GARCH-Itô model. We adopt realized volatility estimators based on high-frequency financial data and the quasi-likelihood function for the low-frequency GARCH structure to develop parameter estimation methods for the combined high-frequency and low-frequency data. We establish asymptotic theory for the proposed estimators and conduct a simulation study to check finite sample performances of the estimators. We apply the proposed estimation approach to Bank of America stock price data.

Keywords GARCH, high-frequency data, low-frequency data, Itô process, quasi-maximum likelihood estimator, realized volatility, stochastic differential equation.

Running title Unified Models and Inferences for Low- and High-Frequency Data
1 Introduction

Since the seminal papers by Engle (1982) and Bollerslev (1986), generalized autoregressive conditional heteroskedastic (GARCH) models have been often used in volatility analysis of low-frequency financial data. On the other hand, as high-frequency financial data are now widely available, it becomes popular to model high-frequency financial data by continuous-time Itô processes and develop nonparametric realized volatility estimators from the high-frequency data (Zhang et al. (2005), Zhang (2006), Fan and Wang (2007), Barndorff-Nielsen et al. (2008), Jacod et al. (2009), and Xiu (2010)). Here low-frequency financial data are referred to as observed price data on financial assets at daily or longer time horizons, with high-frequency financial data for intra-day data observed on the prices of financial assets. Financial models and econometric analysis for low-frequency and high-frequency data are developed quite independently. As high-frequency and low-frequency are just different time scales, the models and data at these different time scales must be inter-related. However, typical GARCH models for low-frequency data and stochastic processes for high-frequency data are often separately treated as unrelated. It is scientifically interesting and financially important to develop unified model and statistical inference approaches for combining both high-frequency and low-frequency financial data. There are some attempts to bridge the gap between discrete-time and continuous-time modeling and reconcile volatility analysis at high-frequency and low-frequency levels. On the model side, Nelson (1990) established the continuous-time diffusion limit for the discrete-time GARCH model by showing that GARCH processes weakly converge to some bivariate diffusions, as the length of the discrete time intervals goes to zero, Wang (2002) studied the statistical relationship between the GARCH and diffusion models, and Kallsen and Taqqu (1998) embedded a discrete-time ARCH model in a continuous-time diffusion model and investigated the option pricing relationship between the GARCH and diffusion models. For volatility analysis, combined inferences include the realized GARCH model (Hansen et al. (2012)), the high-frequency-based volatility model (Shephard and Shephard (2010)), the multiplicative error model
(Engle and Gallo (2006)), and the realized volatility and factor model approach (Tao et al. (2011)).

In this paper we introduce a unified model at all time scales for modeling both low-frequency and high-frequency financial data and drawing inferences for the combined high-frequency and low-frequency data. The unified model is a continuous-time Itô process with a GARCH volatility embedded in its continuous instantaneous volatility at integer time points. As a result, the instantaneous volatility at integer time points is a GARCH volatility, and thus the model at the discrete time points can be viewed as a volatility model for low-frequency log price data, while the model between the integer time points is a continuous-time Itô process often used as a log price model for high-frequency financial data. It provides a unified price model for both low-frequency and high-frequency financial data. The model is called a unified GARCH-Itô model. For the unified model, its low-frequency volatility obeys a parametric GARCH volatility structure, we may estimate the model parameters by the following combing high-frequency and low-frequency approach. First, as the high-frequency volatility is nonparametric, we may use high-frequency data to construct nonparametric realized volatility estimators of integrated volatility over consecutive integer time points. Second, we treat the realized volatility estimators as proxy for low-frequency volatilities in the GARCH volatility structure of the unified model and develop a quasi-maximum likelihood estimation for the GARCH structure to estimate the model parameters. As a comparison, we also employ only the GARCH volatility structure at integer time points to develop a GARCH quasi-maximum likelihood estimation for estimating model parameters based on low-frequency data alone. We establish asymptotic theory for the proposed estimators. A simulation study is conducted to check finite sample performances of the estimation methods. Both asymptotic theory and simulation study show that while both estimators have good performances, the estimator based on the combined data is much better than the estimator using only low-frequency data. We illustrate the estimation methods through an application to Bank of America stock price data.
The rest of the paper is organized as follows. Section 2 introduces a unified GARCH-Itô model and illustrates volatility properties for the model. Section 3 proposes quasi-maximum likelihood estimation methods based on combined high-frequency and low-frequency data and based on low-frequency data alone and develops asymptotic theory for the proposed estimators. Section 4 presents a simulation study to check the finite sample performances of the proposed estimators and an application of the estimation methods to Bank of America stock price data. All the proofs are collected in Section 5.

2 Unified discrete-time and continuous-time models

2.1 Discrete-time and continuous-time models

Discrete-time GARCH processes and continuous-time Itô processes provide common stochastic models in finance. A discrete-time GARCH model may be described as follows,

\[ X_t - X_{t-1} = \mu + \xi_t, \quad t = 1, \ldots, n, \]

where we denote by \( X_t \) the true log price at time \( t \), random errors \( \xi_t \) satisfy \( E[\xi_t | \mathcal{F}_{t-1}] = 0 \) a.s., \( \mathcal{F}_{t-1} = \sigma(X_{t-1}, X_{t-1}, \ldots) \), and their conditional variances obey

\[ E[\xi_t^2 | \mathcal{F}_{t-1}] = \omega + \gamma E[\xi_{t-1}^2 | \mathcal{F}_{t-2}] + \beta \xi_{t-1}^2. \]

Because of simplicity in statistical inferences, GARCH models are often employed in empirical financial studies.

Continuous-time models play a pivotal role in theoretical finance particularly in asset pricing. The models are often defined by stochastic differential equations. Although the two kinds of models are developed independently and have very different characteristics, there are deep relationships between GARCH and diffusion models (Nelson (1990), Wang (2002), Duan et al. (2009), and Kallsen and Taqqu (1998)). For example, Kallsen and Taqqu (1998) embedded a special ARCH model in an Itô process to study GARCH option pricing. Discrete GARCH models are often used to model low-frequency financial data,
and continuous-time Itô processes are commonly employed in the study of high-frequency financial data. It is desirable to combine both low-frequency and high-frequency financial data by developing unified financial models. Some recent attempts in such studies may include Engle and Gallo (2006), Shephard and Shephard (2010), Hansen et al. (2012), and Tao et al. (2011).

2.2 Instantaneous volatility process

Set $\mathbb{R}_+ = [0, \infty)$ and denote by $\mathbb{N}$ the set of all non-negative integers. We define a unified modeling approach by embedding a standard GARCH(1,1) model into an Itô process with an instantaneous volatility as follows.

**Definition 1** We call a log stock price $X_t$, $t \in \mathbb{R}_+$, to follow a unified GARCH-Itô model if it satisfies

$$dX_t = \mu dt + \sigma_t dB_t, \quad (2.1)$$

$$\sigma_t^2 = \sigma_{[t]}^2 + (t - [t])\{\omega + (\gamma - 1)\sigma_{[t]}^2\} + \beta \left(\int_{[t]}^t \sigma_s dB_s\right)^2, \quad (2.2)$$

where $\mu$ is a drift, $[t]$ denotes the integer part of $t$, $B_t$ is a standard Brownian motion with respect to a filtration $\mathcal{F}_t$, $\sigma_t$ is a volatility process adapted to $\mathcal{F}_t$, and $\theta = (\omega, \beta, \gamma)$ are model parameters.

For a unified GARCH-Itô model, instantaneous volatility $\sigma_t^2$ is a continuous-time process defined at all times $t \in \mathbb{R}_+$, and when restricted to integer times $t \in \mathbb{N}$ it obeys a GARCH structure

$$\sigma_t^2 = \omega + \gamma \sigma_{t-1}^2 + \beta Z_t^2, \quad Z_t = \int_{t-1}^t \sigma_s dB_s, \quad t \in \mathbb{N}. \quad (2.3)$$

We may interpret the unified GARCH-Itô model as a unified model for both high-frequency and low-frequency financial data as follows. The process $X_t$ and volatility $\sigma_t^2$ at integer time points $t \in \mathbb{N}$ are treated as low-frequency the daily log price and daily GARCH volatility, respectively. For time $t$ between integers, we may view $X_t$ as the high-frequency log price, $\sigma_t^2$.
as high-frequency instantaneous volatility, and integral of $\sigma^2_t$ over two consecutive integers as integrated volatility. The continuous-time GARCH model proposed by Kallsen and Taqqu (1998) is a special case of the unified GARCH-Itô model.

2.3 Integrated volatility for the unified model

High-frequency volatility analysis often estimates integrated volatility $\int_{n-1}^n \sigma^2_t \, dt$, while the low-frequency volatilities $\sigma^2_n$ obey a GARCH structure used for parameter estimation and volatility prediction. We study integrated volatilities over consecutive integers, which will be employed for statistical inferences later.

Proposition 1 For a unified GARCH-Itô model we have the iterative relationships for integrated volatilities.

(a) For $0 < \beta < 1$ and $k, n \in \mathbb{N}$,

$$R(k) = \int_{n-1}^n \frac{(n-t)^k}{k!} \sigma^2_t \, dt = \omega + (\gamma + k + 1)\sigma^2_{n-1} + 2\beta \int_{n-1}^n \frac{(n-t)^{k+1}}{(k+1)!} \int_{n-1}^t \sigma_s dB_s \sigma_t dB_t + \beta R(k+1).$$

In particular,

$$\int_{n-1}^n \sigma^2_t \, dt = \sum_{k=0}^{\infty} \frac{\beta^k (\omega + (\gamma + k + 1)\sigma^2_{n-1})}{(k+2)!} + 2 \int_{n-1}^n \frac{(\beta(t-n))^{k+1}}{(k+1)!} \int_{n-1}^t \sigma_s dB_s \sigma_t dB_t$$

$$= h_n(\theta) + D_n,$$

where

$$h_n(\theta) = \beta^{-2} (e^\beta - 1 - \beta) \omega + [(\gamma - 1) \beta^{-2} (e^\beta - 1 - \beta) + \beta^{-1} (e^\beta - 1)] \sigma^2_{n-1}$$

$$= \omega^g + \gamma h_{n-1}(\theta) + \beta^g Z^2_{n-1},$$

$$\omega^g = \beta^{-1} (e^\beta - 1) \omega, \quad \beta^g = \beta^{-1} (\gamma - 1) (e^\beta - 1 - \beta) + e^\beta - 1, \quad \tau(\theta) = (\omega^g, \beta^g, \gamma),$$

and

$$D_n = 2 \int_{n-1}^n (e^{(n-t)\beta} - 1) \int_{n-1}^t \sigma_s dB_s \sigma_t dB_t$$

is a martingale difference.
(b) For $0 < \beta < 1$ and $k, n \in \mathbb{N}$,
\[
E \left[ \int_{n-1}^{n} \frac{(n-t)^k}{k!} \sigma_t^2 dt \middle| \mathcal{F}_{n-1} \right] = \omega + (\gamma + k + 1) \sigma_{n-1}^2 + \beta E \left[ \int_{n-1}^{n} \frac{(n-t)^{k+1}}{(k+1)!} \sigma_t^2 dt \middle| \mathcal{F}_{n-1} \right].
\]
In particular,
\[
E \left[ \int_{n-1}^{n} \sigma_t^2 dt \middle| \mathcal{F}_{n-1} \right] = h_n(\theta),
\]
where $h_n$ is defined in (2.5).

(c) For $0 < \beta^g + \gamma < 1$ and $n \in \mathbb{N}$,
\[
E [h_n(\theta)] = \frac{\omega^g}{1 - \beta^g - \gamma} \quad \text{and} \quad E [\sigma_n^2] = \frac{\omega (1 - \beta^g - \gamma) + \beta \omega^g}{(1 - \beta^g - \gamma)(1 - \gamma)},
\]
where $\omega^g$ and $\beta^g$ are defined in (2.6).

Proposition 1 (a) shows that the integrated volatilities can be decomposed into $h_n(\theta)$ and $D_n$. $h_n(\theta)$ is adapted to the low-frequency filtration, $\mathcal{F}_{n-1}^{LF} = \sigma (\{X_i : i \in \{n-1, n-2, \ldots\}\})$, which is a sigma field generated by daily stock prices and may be viewed as low-frequency information based on daily stock prices. Note that the filtration $\mathcal{F}_{n-1}$ at integers $n-1$ is generated by stock prices at all past times and interpreted as the high-frequency filtration, and of course $\mathcal{F}_{n-1} \supset \mathcal{F}_{n-1}^{LF}$. Both $\sigma_n^2$ and $h_n(\theta)$ obey linear GARCH(1,1) volatility structures with model parameter $\theta = (\omega, \beta, \gamma)$ for $\sigma_n^2$ and model parameter $\tau(\theta)$ in (2.6) for $h_n(\theta)$. The decomposition in (2.4) shows that the integrated volatilities consist of a GARCH volatility $h_n(\theta)$ and a martingale difference $D_n$. The proof of Lemma 2 (b) further shows that $D_n$ can be expressed as
\[
D_n = \beta \int_{n-1}^{n} e^{(n-t)\beta} \left( \int_{n-1}^{t} \sigma_s dB_s \right)^2 dt - \int_{n-1}^{n} (e^{(n-t)\beta} - 1) \sigma_t^2 dt.
\]

3 Parameter estimation for the unified model

3.1 Low-frequency and high-frequency financial data

Discrete-time GARCH models are often used to fit low-frequency financial data, and continuous-time Itô processes are commonly employed to model log stock prices in high-
frequency financial data. It is desirable to unify both high-frequency and low-frequency financial models and fit the unified models to the combined high-frequency and low-frequency financial data. For low-frequency data, GARCH models are parametric models, and model parameters are estimated by (quasi) maximum likelihood estimation. For high-frequency data, Itô processes are non-parametric models, and realized volatility methods are adopted to estimate integrated volatility for the Itô processes. Our proposal for the unified GARCH-Itô model is the quasi-maximum likelihood estimation (QMLE) approach, where we substitute volatilities in a quasi-likelihood function by the realized volatility estimators obtained from high-frequency data and estimate GARCH parameters by maximizing the quasi-likelihood function.

We assume that the underlying log price process follows a unified GARCH-Itô model. The low-frequency data are observed log prices at integer times, namely, \( X_t, t = 0, 1, 2, \ldots, n \). The high-frequency data are observed log prices at time points between integers, that is, let \( t_{i,j} \) be the high-frequency observed time points during the \( i \)-th period that satisfy \( i - 1 = t_{i,0} < t_{i,1} < \ldots < t_{i,m_i} < t_{i,m_i+1} = t_{i+1,0} = i \), and observed log prices \( Y_{t_{i,j}} \) obey

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

(3.1)

where \( \epsilon_{t_{i,j}} \) are micro-structure noises independent of price and volatility processes, and for each \( i, \epsilon_{t_{i,j}}, j = 1, \ldots, m_i \), are i.i.d. with \( E\epsilon_{t_{i,j}} = 0 \). The low-frequency drift \( \mu \) over consecutive integers can be easily estimated by the sample mean of low-frequency returns, while the drift \( \mu \) has asymptotically negligible effects on high-frequency realized volatility estimators. Thus, to focus on volatility analysis we may take \( \mu = 0 \) in (2.1) for simplicity.

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

We can nonparametrically estimate integrated volatility by using high-frequency financial data in the \( i \)-th period. In the low-frequency view, we may treat the estimated integrated
volatilities as "observations" and define a quasi-likelihood function $L_{m,n}^{GH}(\theta)$ for the low-frequency GARCH structure as follows,

$$\hat{L}_{n,m}^{GH}(\theta) = -\frac{1}{2n} \sum_{i=1}^{n} \log(h_i(\theta)) + \frac{RV_i}{h_i(\theta)},$$

where $RV_i$ is the multi-scale realized volatility estimator (kernel realized volatility estimator or pre-averaging volatility estimator) based on $m_i$ high-frequency data during the $i$-th period. As an estimator of the integrated volatility over the $i$-th period, $RV_i$ may have convergence rate $m_i^{-1/4}$. From Proposition 1, the integrated volatility over the $i$-th period is equal to $h_i(\theta_0)$ plus a martingale difference $D_i$, where $\theta_0 = (\omega_0, \beta_0, \gamma_0)$ denotes the true parameter value. In the quasi-likelihood function we drop the martingale differences from the integrated volatilities, because the effects of martingale differences are asymptotically negligible.

To evaluate the quasi-likelihood function we need initial volatility $\sigma_0^2$ to obtain $h_i(\theta)$. Taking $Z_i^2 = (X_i - X_0)^2$ as an initial value of $\sigma_0^2$, we compute the quasi-likelihood function $\hat{L}_{n,m}^{GH}(\theta)$ and maximize it over parameter space $\Theta$. Denote the maximizer by $\hat{\theta}^{GH}$, that is,

$$\hat{\theta}^{GH} = \arg\max_{\theta \in \Theta} \hat{L}_{n,m}^{GH}(\theta).$$

We estimate $\theta_0$ by $\hat{\theta}^{GH}$ and model parameter $\tau(\theta_0)$ in (2.6) by $\tau(\hat{\theta}^{GH})$.

### 3.2.1 Asymptotic theory

This section establishes consistency and asymptotic distribution for the proposed estimator $\hat{\theta}^{GH}$. First we fix notations. For a matrix $A = (A_{i,j})_{i,j=1,\ldots,k}$, and a vector $a = (a_1, \ldots, a_k)$, define $\|A\|_{\text{max}} = \max_{i,j} |A_{i,j}|$ and $\|a\|_{\text{max}} = \max_i |a_i|$. Given a random variable $X$ and $p \geq 1$, let $\|X\|_{L_p} = \{E[|X|^p]\}^{1/p}$. Let $C$'s be positive generic constants whose values are free of $\theta$, $n$ and $m_i$, and may change from appearance to appearance. We need the following assumptions.

**Assumption 1**
Let
\[
\Theta = \{\theta = (\omega, \beta, \gamma) : \omega_l < \omega < \omega_u, \beta_l < \beta < \beta_u, \gamma_l < \gamma < \gamma_u, \gamma + \beta g < 1\},
\]
where \(\omega_l, \omega_u, \gamma_l, \gamma_u, \beta_l, \beta_u\) are known positive constants, and \(\beta g\) is defined in (2.6).

(b) \(|D_i| : i \in \mathbb{N}\) is uniformly integrable.

(c) One of the following conditions is satisfied.

(c1) \(E\left[\frac{Z_i^4}{h_i(\theta_0)}\right] \leq C \) a.s. for any \(i \in \mathbb{N}\).

(c2) There exists a positive constant \(\delta\) such that \(E\left[\left(\frac{Z_i^2}{h_i(\theta_0)}\right)^{2+\delta}\right] \leq C\) for any \(i \in \mathbb{N}\).

(d) \((D_i, Z_i^2)\) is a stationary ergodic process.

(e) Let \(m = \sum_{i=1}^{n} m_i/n\). We have \(C_1 m \leq m_i \leq C_2 m\), 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 \(m, n \to \infty\).

(f) \(\sup_{i \in \mathbb{N}} \left\|RV_i - \int_{i-1}^{i} \sigma_t^2 dt\right\|_{L_{1+\delta}} \leq C \cdot m^{-1/4}\) for some \(\delta > 0\).

(g) For any \(i \in \mathbb{N}\), \(E[RV_i|\mathcal{F}_{i-1}] \leq C \cdot E\left[\int_{i-1}^{i} \sigma_t^2 dt\right] + C\) a.s.

Assumption 1 (a)-(d) are for the low-frequency part of the model, with Assumption 1 (e)-(g) for the high-frequency part. A sufficient condition of Assumption 1 (b) is \(\sup_{i \in \mathbb{N}} E(|D_i|^{1+\delta}) < \infty\) for some \(\delta > 0\). Assumption 1 (c) is required moment conditions where (c1) is similar to the moment condition in Lee and Hansen (1992) and (c2) is relatively easy to check. Since \(m\) is the average number of high-frequency data, with \(n\) for low-frequency sample size, Assumption 1 (e) is typical conditions in high-frequency volatility analysis. We may take \(RV_i\) as one of multi-scale realized volatility (Zhang (2006) and Fan and Wang (2007)), kernel realized volatility (Barndorff-Nielsen et al. (2008)), and pre-averaging realized volatility (Jacod et al. (2009)), and Tao et al. (2013, theorem 1) and Kim and Wang (2014, theorems 1 and 3) indicate that Assumption 1 (f)-(g) are reasonable.

The following theorems establish consistency and convergence rate for \(\hat{\theta}^{GH}\).
**Theorem 1** Under Assumption 1 (a)-(b), (e)-(f), there is a unique maximizer of $L^G_H(\theta)$ and as $m, n \to \infty$, $\hat{\theta}^G_H \to \theta_0$ in probability.

**Theorem 2** Under Assumption 1 (a)-(c), (e)-(g), we have

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

Theorem 2 shows that the convergence rate of $\hat{\theta}^G_H$ has two components $m^{-1/4}$ and $n^{-1/2}$. The rate $m^{-1/4}$ is due to the high-frequency volatility estimation, which is the optimal convergence rate for estimating integrated volatilities based on high-frequency data with market microstructure noise, and $n^{-1/2}$ is from the usual parametric convergence rate based on low-frequency data.

The stationary ergodic condition Assumption 1 (d) is used to obtain asymptotic normality.

**Theorem 3** Under Assumption 1, we have as $m, n \to \infty$,

$$\sqrt{n}(\hat{\theta}^G_H - \theta_0) \xrightarrow{d} N(0, B^{-1}A^G_H B^{-1}),$$

where

$$A^G_H = E \left[ \frac{\partial h_1(\theta)}{\partial \theta} \frac{\partial h_1(\theta)}{\partial \theta^T} \right]_{\theta=\theta_0} h_1^{-4}(\theta_0) \int_0^1 (e^{\beta_0(1-t)} - 1)^2 (X_t - X_0)^2 \sigma_t^2 dt$$

and

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

Theorem 3 provides an asymptotic normal distribution for $\hat{\theta}^G_H$. Though the asymptotic variance is quite complicated, we will demonstrate in the next section via asymptotic analysis that the estimation based on combined high-frequency and low-frequency approach is better than that based on low-frequency alone.
3.3 Quasi-maximum likelihood estimation based on only low-frequency data

This section considers the estimation of \( \theta_0 \) using only low-frequency data \( X_0, X_1, \cdots, X_n \) and compare it with combined low-frequency and high-frequency approach. As we have seen in Section 3.2, \( \theta_0 \) and \( \tau(\theta_0) \) are model parameters for \( \sigma^2_i \) and \( h_i(\theta_0) \), respectively. Due to the continuous nature of the unified GARCH-Itô model, low-frequency return \( Z_i = X_i - X_{i-1} \) corresponds to integrated volatility of \( \sigma^2_t \) over \( (i-1, i] \) but not \( \sigma^2_i \). Thus we can not directly apply standard GARCH likelihood or quasi-likelihood estimation methods to return data \( Z_i \) and estimate \( \theta_0 \), although (2.3) may lead to think such direct applications.

We estimate \( \theta_0 \) using the GARCH structure between low-frequency return squares \( Z^2_i \) and GARCH volatilities \( h_i(\theta_0) \). Specifically, by Itô’s lemma and Proposition 1 (a), we have

\[
Z^2_i = (X_i - X_{i-1})^2 = 2 \int_{i-1}^i (X_t - X_{i-1})dX_t + \int_{i-1}^i \sigma^2_t dt
\]

\[
= h_i(\theta_0) + 2 \int_{i-1}^i e^{(n-t)\beta_0}(X_t - X_{i-1})dX_t = h_i(\theta_0) + D^L_i,
\]

where \( D^L_i = 2 \int_{i-1}^i e^{(n-t)\beta_0}(X_t - X_{i-1})dX_t \). We define the quasi-likelihood function

\[
\hat{L}^{GL}_n(\theta) = -\frac{1}{2n} \sum_{i=1}^n \log(h_i(\theta)) + \frac{Z^2_i}{h_i(\theta)},
\]

and maximizes \( \hat{L}^{GL}_n(\theta) \) over the parameter space \( \Theta \). Denote the maximizer by \( \hat{\theta}^{GL} \), that is,

\[
\hat{\theta}^{GL} = \arg\max_{\theta \in \Theta} \hat{L}^{GL}_n(\theta),
\]

where we use \( Z^2_1 \) as an initial value of \( \sigma^2_0 \) in the expressions of \( h_i(\theta) \) and \( \hat{L}^{GL}_n(\theta) \), as in Section 3.2. We estimate \( \theta_0 \) by \( \hat{\theta}^{GL} \) and \( \tau(\theta_0) \) in (2.6) by \( \tau(\hat{\theta}^{GL}) \).

The following theorems establish consistency and convergence rate for \( \hat{\theta}^{GL} \).

**Theorem 4** Under Assumption 1 (a)-(b), there exists a unique maximizer of \( L^{GL}_n(\theta) \) and as \( n \to \infty \), \( \hat{\theta}^{GL} \to \theta_0 \) in probability.
Theorem 5 Under Assumption 1 (a)-(c), we have
\[ \left\| \hat{\theta}_{GL} - \theta_0 \right\|_{\text{max}} = O_p(n^{-1/2}). \]

Theorem 6 Under Assumption 1 (a)-(d), we have as \( n \to \infty \),
\[ \sqrt{n}(\hat{\theta}_{GL} - \theta_0) \xrightarrow{d} N(0, A_{GL}^{-1}B^{-1}), \]
where
\[ A_{GL} = E\left[ \frac{\partial h_1(\theta) \partial h_1(\theta)}{\partial \theta \partial \theta^T} \bigg|_{\theta = \theta_0} h_1^{-4}(\theta_0) \int_0^1 e^{2\beta_0(1-t)}(X_t - X_0)^2 \sigma_t^2 dt \right], \]
and \( B \) is defined in Theorem 3.

Theorems 3 and 6 establish asymptotic normality for \( \hat{\theta}_{GH} \) and \( \hat{\theta}_{GL} \), respectively. Their asymptotic variances have the same structure, but \( A_{GL} \) is larger than \( A_{GH} \). It may be intuitively explained by the fact that \( \hat{\theta}_{GH} \) is constructed by combining both high-frequency and low-frequency data, while \( \hat{\theta}_{GH} \) is obtained by using only low-frequency data. Technically it is due to the difference between \( D_t \) and \( D_t^L \), where \( E[(D_t^L)^2] = E[\int_0^1 e^{2\beta_0(1-t)}(X_t - X_0)^2 \sigma_t^2 dt] \) and \( E[D_t^2] = E[\int_0^1 (e^{\beta_0(1-t)} - 1)^2(X_t - X_0)^2 \sigma_t^2 dt] \), and the integrated volatilities \( \int_{i-1}^i \sigma_t^2 dt \) are closer to the GARCH volatilities \( h_i(\theta_0) \) than the low-frequency return squares \( Z_t^2 \). Thus, \( \hat{\theta}_{GL} \) has a bigger asymptotic variance than \( \hat{\theta}_{GH} \), and \( \hat{\theta}_{GH} \) is asymptotically more efficient.

4 Numerical Analysis

4.1 A simulation study

We conducted simulations to check the finite sample performances of the proposed estimators \( \hat{\theta}_{GH} \) and \( \hat{\theta}_{GL} \) and compare their performances. We generated the log prices \( X_{t,i,j} \), \( t_{i,j} = i - 1 + j/m, i = 1, \cdots, n, j = 1, \cdots, m \), from the unified GARCH-Itô model in (2.1) and (2.2) with the following form,
\[ dX_t = \sigma_t dB_t, \quad d\sigma_t^2 = (\omega_0 + (\gamma_0 - 1)\sigma_t^2 + \beta_0 \sigma_t^2) dt + 2\beta_0 (X_t - X_0) \sigma_t dB_t, \]
where \( \theta_0 = (\omega_0, \beta_0, \gamma_0) = (0.2, 0.3, 0.4) \), and \( B_t \) is a standard Brownian motion. We took \( n = 250 \), \( m = 2160 \), and initial values \( X_0 = 10 \) and \( \sigma_0^2 = E(\sigma_1^2) = 0.5198 \) in (2.8). The values for \( m \) and \( n \) were chosen so that low-frequency data are daily observations and high-frequency data correspond to stock prices observed every ten seconds. The Euler scheme was used to discretize the continuous-time Itô process driving by \( B_t \). The low-frequency data were taken to be \( X_i, i = 0, \ldots, n \). The high-frequency data \( Y_{t_{i,j}} \) were simulated from model (3.1) where \( X_{t_{i,j}} \) were taken from the generated log prices from the unified GARCH-Itô model, and market microstructure noise \( \epsilon_{t_{i,j}} \) were simulated from i.i.d. normal distribution with mean zero and standard deviation \( \sigma_\epsilon = 0.05 \). In the simulation, we estimated integrated volatilities by the multi-scale realized volatility estimator (Zhang (2006) and Fan and Wang (2007)). Specifically, the \( i \)-th period integrated volatility estimator is

\[
RV_i = \sum_{k=1}^{\sqrt{m_i}} a_k \cdot RV_i^{K_k} + \zeta \left( RV_i^{K_1} - RV_i^{K_\sqrt{m_i}} \right),
\]

where

\[
RV_i^{K} = \frac{1}{K} \sum_{r=K}^{m_i} [Y(t_{i,r}) - Y(t_{i,r-K})]^2, \quad a_k = \frac{12(k + \sqrt{m_i})(k - \sqrt{m_i}/2 - 1/2)}{\sqrt{m_i}(m_i - 1)},
\]

\[K_k = \sqrt{m_i} + k, \quad \text{and} \quad \zeta = \frac{2\sqrt{m_i} (\sqrt{m_i} + 1)}{(m_i + 1)(\sqrt{m_i} - 1)}.\]

We estimated the model parameters by the two estimators \( \hat{\theta}^{GH} \) and \( \hat{\theta}^{GL} \) proposed in Section 3. The whole simulation procedure was repeated 1000 times.

Table 1: The squared biases, variances, and MSEs of \( \hat{\theta}^{GH} \) and \( \hat{\theta}^{GL} \) for estimating \( \theta_0 = (\omega_0, \beta_0, \gamma_0) \) based on data simulated from the unified GARCH-Itô model with \( m = 2160 \), \( n = 250 \) and \( \sigma_\epsilon = 0.05 \). The squared biases, variances, and MSEs reported are based on 1000 repetitions and multiplied by 100.

<table>
<thead>
<tr>
<th></th>
<th>( \omega )</th>
<th></th>
<th>( \beta )</th>
<th></th>
<th>( \gamma )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias(^2)</td>
<td>Variance</td>
<td>MSE</td>
<td>Bias(^2)</td>
<td>Variance</td>
</tr>
<tr>
<td>( \hat{\theta}^{GH} )</td>
<td>0.0040</td>
<td>0.0286</td>
<td>0.0326</td>
<td>0.0022</td>
<td>0.0783</td>
</tr>
<tr>
<td>( \hat{\theta}^{GL} )</td>
<td>0.0555</td>
<td>1.2004</td>
<td>1.2559</td>
<td>0.0459</td>
<td>2.7784</td>
</tr>
</tbody>
</table>

The simulation results are reported in Table 1 and displayed in Figure 1. Table 1 provides the squared biases, variances, and mean squared errors (MSEs) of the proposed
estimators $\hat{\theta}^{GH}$ based on combined high-frequency and low-frequency data and $\hat{\theta}^{GL}$ based on low-frequency data alone, and Figure 1 shows their box plots. The simulation results demonstrate that the proposed estimators have good finite sample performances, and as we expected, $\hat{\theta}^{GL}$ has much larger bias and variance than $\hat{\theta}^{GH}$. The simulations confirm the theoretical findings in Sections 2 and 3 that the bias and variance components for constructing $\hat{\theta}^{GH}$ based on the combined data are smaller than those for $\hat{\theta}^{GL}$ based on only low-frequency data.

![Boxplot](image)

Figure 1: Boxplots of $\hat{\theta}^{GH}$ and $\hat{\theta}^{GL}$ for estimating $\theta_0 = (\omega_0, \beta_0, \gamma_0)$ based on data simulated from the unified GARCH-Itô model with $m = 2160$, $n = 250$ and $\sigma_\epsilon = 0.05$, where GH and GL represent $\hat{\theta}^{GH}$ and $\hat{\theta}^{GL}$, respectively.

4.2 An empirical study

We applied the unified GARCH-Itô model to high-frequency trading data of Bank of America Corporation (BAC) over a period of 146 days in 2012 with $n = 146$ and $m = 175,745$. The high-frequency data were available between ‘open’ and ‘close’, and low-frequency data were obtained by computing open-to-close returns. The multi-scale realized volatility estimators $RV_i$ in (4.1) were employed to estimate daily integrated volatilities based on the high-frequency data, and according to the description in Section 3, we calculated $\hat{\theta}^{GH}$ based on combined high-frequency and low-frequency data (all data)
Volatility for BAC with open–to–close returns

Figure 2: Plots of daily volatility estimates $RV_i$, $h_i(\hat{GH})$ and $h_i(\hat{GL})$, where RV, GH and GL stand for realized volatility estimates $RV_i$, GARCH volatility estimates $h_i(\hat{GH})$ based on all data, and GARCH volatility estimates $h_i(\hat{GL})$ based on low-frequency data alone.

and $\hat{\theta}^{GL}$ based on only the low-frequency data to estimate $\theta$. The estimated values were $\hat{\theta}^{GH} = (0.0002319, 0.2162, 0.1685)$ and $\hat{\theta}^{GL} = (0.0001745, 0.04851, 0.6384)$. In Figure 2 we displayed the estimated daily volatilities by realized volatility estimators $RV_i$, daily volatility estimators $h_i(\hat{\theta}^{GH})$ based on combined high-frequency and low-frequency data and daily volatility estimators $h_i(\hat{\theta}^{GL})$ using low-frequency data alone. From the plot we can see that while the low-frequency estimators $h_i(\hat{\theta}^{GL})$ were quite flat, $RV_i$ are very rough and volatile, and the combined estimators $h_i(\hat{\theta}^{GH})$ were somewhat between the two types of the estimators and followed the realized volatility estimators more than the low-frequency ones. The phenomenon may be due to the facts that $RV_i$ are nonparametric estimators and free of any parametric assumption, $h_i(\hat{\theta}^{GH})$ and $h_i(\hat{\theta}^{GL})$ were obtained under the unified GARCH-Itô model, and $h_i(\hat{\theta}^{GL})$ more or less rely on low-frequency data and the parametric GARCH structure embedded in the unified GARCH-Itô model, while $h_i(\hat{\theta}^{GH})$ pool the strength of the parametric GARCH structure embedded in the unified GARCH-Itô model for the low-frequency data and the nonparametric realized volatility approach for the high-frequency data.
5 Proofs

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

5.1 Proof of Proposition 1

Proof of Proposition 1. (a). By Itô’s lemma we have

$$R(k) = \sigma^2 \int_{n-1}^{n} \frac{(n-t)^k}{k!} dt + (\omega + (\gamma - 1)\sigma^2_{n-1}) \int_{n-1}^{n} t_r \frac{(n-t)^k}{k!} dt$$

$$\quad + \beta \int_{n-1}^{n} \int_{s}^{n} \frac{(n-t)^k}{k!} \sigma^2_s ds dt + 2\beta \int_{n-1}^{n} \int_{s}^{n} \frac{(n-t)^k}{k!} dt \left( \int_{s}^{t} \sigma_h dB_h \right) \sigma_s dB_s$$

$$= \frac{\omega + (\gamma + k + 1)\sigma^2_{n-1}}{(k+2)!} + 2\beta \int_{n-1}^{n} \frac{(n-t)^{k+1}}{(k+1)!} \int_{n-1}^{t} \sigma_s dB_s \sigma_t dB_t + \beta R(k + 1).$$

Using the iterative relationship we can easily prove (2.4).

To show (2.5), since $\sigma^2_n$ has a standard GARCH(1,1) form, we have

$$h_n(\theta) = \beta^{-2}(e^\beta - 1 - \beta)\omega$$

$$+ [(\gamma - 1)\beta^{-2}(e^\beta - 1 - \beta) + \beta^{-1}(e^\beta - 1)] \left( \frac{\omega}{1 - \gamma} + \beta \sum_{i=0}^{\infty} \gamma^i Z_{n-1-i}^2 \right)$$

$$= \frac{\omega^g}{1 - \gamma} + \beta^g \sum_{i=0}^{\infty} \gamma^i Z_{n-1-i}^2 = \omega^g + \gamma h_{n-1}(\theta) + \beta^g Z^2_{n-1}.$$  

As the integrand of $D_n$ is predictable, $D_n$ is a martingale difference.

(b). It is an immediate consequence of $E[D_n|F_{n-1}] = 0$.

(c). By Itô’s lemma we conclude

$$E[Z^2_n] = 2E \left[ \int_{n-1}^{n} \int_{n-1}^{t} \sigma_s dB_s \sigma_t dB_t \right] + E \left[ \int_{n-1}^{n} \sigma^2_t dt \right]$$

$$= E[h_n(\theta)],$$

where the last equality is due to the fact that $\int_{n-1}^{n} \int_{n-1}^{t} \sigma_s dB_s \sigma_t dB_t$ is a martingale difference. Then, we have

$$E[h_n(\theta)] = \omega^g + (\gamma + \beta^g)E[h_{n-1}(\theta)] = \frac{\omega^g}{1 - \gamma - \beta^g}.$$
Similarly, we can show
\[ E[\sigma^2_n] = \omega + \gamma E[\sigma^2_{n-1}] + \beta \frac{\omega^g}{1 - \gamma - \beta^g} = \omega(1 - \beta^g - \gamma) + \beta \omega^g. \]
\[
\sigma^2_n \cdot \omega = \gamma \sigma^2_{n-1} + \beta \sigma^2_{n-2} + \omega^g \frac{1}{1 - \beta^g - \gamma}. \]

\[ 5.2 \text{ Initial value} \]

We use \( Z_1^2 \) as the initial value of \( \sigma^2_0 \) to evaluate \( h_i(\theta) \) and the quasi-likelihood. The following lemma shows that the impact of the initial value is asymptotically negligible.

**Lemma 1** As \( h_i(\theta) \) depends on the initial value \( \sigma^2_0 \), write \( h_i(\theta) \) as \( h_i(\theta, \sigma^2_0) \) to explicitly express its dependence on the initial value \( \sigma^2_0 \). Under Assumption 1 (a), we have for any \( \xi = O_p(1) \) and \( i \in \mathbb{N} \),
\[ |h_i(\theta, \xi) - h_i(\theta, \sigma^2_0)| = O_p(\gamma^{i-1}). \]

*Proof:* Simple algebra shows
\[ h_i(\theta, \xi) - h_i(\theta, \sigma^2_0) = \beta^{-1} \beta^g (\xi - \sigma^2_0) = O_p(\gamma^{i-1}). \]

Lemma 1 shows that the dependence of \( h_i(\theta) \) on the initial value is exponentially decaying. Thus, as \( n \to \infty \), the difference between the quasi-likelihood functions with the true initial value \( \sigma^2_0 \) and initial value \( Z_1^2 \) decays faster than \( O_p(n^{-1}) \). Thus we may use the true initial value \( \sigma^2_0 \) in the rest of proofs.

\[ 5.3 \text{ Proof of Theorem 1} \]

Note the notation \( \tau(\theta_0) = (\omega^g_0, \beta^g_0, \gamma_0) \). Let
\[ \hat{L}_{n,m}^{GH}(\theta) = -\frac{1}{2n} \sum_{i=1}^{n} \log(h_i(\theta)) + \frac{RV_i}{h_i(\theta)} = -\frac{1}{2n} \sum_{i=1}^{n} \hat{l}^{GH}_i(\theta) \quad \text{and} \quad \hat{\psi}_{n,m}^{GH}(\theta) = \frac{\partial \hat{L}_{n,m}^{GH}(\theta)}{\partial \theta}; \]
Let \((\beta^g_0, \beta^u_0)\) and \((\omega^g_0, \omega^u_0)\) be the lower bound and the upper bound of \(\beta^g\) and \(\omega^g\), respectively.

To ease notations, we denote derivatives of any function \(f\) at \(x_0\) by

\[
\frac{\partial f(x_0)}{\partial x} = \frac{\partial f(x)}{\partial x} \bigg|_{x=x_0}.
\]

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

(a) \(E(Z^2_0) = E \left( \int_{i-1}^{i} \sigma^2_t dt \right) = E(h_i(\theta_0))\), \(\sup_{i \in \mathbb{N}} E(Z^2_i) \leq \frac{\omega^g_0}{1 - \gamma_0 - \beta^g_0} + E(h_1(\theta_0)) < \infty\), and \(\sup_{i \in \mathbb{N}} E(h_i(\theta_0)) < \infty\);

(b) \(D_i = \beta_0 \int_{i-1}^{i} (e^{(n-t)\beta_0} - 1) \sigma^2_t dt\) a.s. for any \(i \in \mathbb{N}\);

(c) if \(\gamma \geq \gamma_0\), \(\sup_{i \in \mathbb{N}} \frac{h_i(\theta_0)}{h_i(\theta)} < \infty\) a.s.;

(d) there exists a neighborhood \(B(\theta_0)\) of \(\theta_0\) such that for any \(p \geq 1\), \(\sup_{i \in \mathbb{N}} \left\| \sup_{\theta \in B(\theta_0)} \frac{h_i(\theta_0)}{h_i(\theta)} \right\|_{L^p} < \infty\) and \(B(\theta_0) \subset \Theta\);

(e) for any \(p \geq 1\), \(\sup_{i \in \mathbb{N}} \left\| \sup_{\theta \in \Theta} \frac{\partial^2 h_i(\theta)}{\partial \theta_j \partial \theta_k} \right\|_{L^p} \leq \tilde{C}\) and \(\sup_{i \in \mathbb{N}} \left\| \sup_{\theta \in \Theta} \frac{\partial^3 h_i(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} \right\|_{L^p} \leq \tilde{C}\) for any \(j, k, l \in \{1, 2, 3\}\), where \(\theta = (\theta_1, \theta_2, \theta_3) = (\omega, \beta, \gamma)\).

**Proof:** (a). By Itô’s lemma and Proposition 1 (b) we have

\[
E(Z^2_0) = E \left( \int_{i-1}^{i} \sigma^2_t dt \right) = E(h_i(\theta_0)).
\]

Simple algebra shows

\[
E(h_i(\theta_0)) = \omega^g_0 + \gamma_0 E(h_{i-1}(\theta_0)) + \beta^g_0 E(Z^2_{i-1}) = \omega^g_0 + (\beta^g_0 + \gamma_0) E(h_{i-1}(\theta_0)).
\]

Using this iterative equation and the fact that \(\beta^g_0 + \gamma_0 < 1\) we can easily prove

\[
E(h_i(\theta_0)) = \frac{\omega^g_0 (1 - (\beta^g_0 + \gamma_0)^{i-1})}{1 - \beta^g_0 - \gamma_0} + (\beta^g_0 + \gamma_0)^{i-1} E(h_1(\theta_0)) \leq \frac{\omega^g_0}{1 - \gamma_0 - \beta^g_0} + E(h_1(\theta_0)) < \infty.
\]
Similarly, we can show
\[ \sup_{i \in \mathbb{N}} E \left( \sup_{\theta \in \Theta} h_i(\theta) \right) < \infty. \]

(b). Let \( f(t, X_t) = \left( e^{(i-t)\beta_0} - 1 \right) (X_t - X_{t-1})^2 \). By Itô’s lemma we have
\[ df(t, X_t) = \left[ -\beta_0 e^{(i-t)\beta_0} (X_t - X_{t-1})^2 + (e^{(i-t)\beta_0} - 1) \sigma_i^2 \right] dt \\
+ 2 \left( e^{(i-t)\beta_0} - 1 \right) (X_t - X_{t-1}) dX_t. \]

Then
\[ 0 = \int_{t-1}^{t} \left[ -\beta_0 e^{(i-t)\beta_0} (X_t - X_{t-1})^2 + (e^{(i-t)\beta_0} - 1) \sigma_i^2 \right] dt + D_i. \]

(c). The statement follows from Lemma 4 (4) in Lee and Hansen (1994).

(d). For any \( \delta > 0 \), there exists a neighborhood \( B(\theta_0) \subset \Theta \) such that \( \gamma_0 \leq (1 + \delta) \gamma \) for any \( \theta \in B(\theta_0) \). By the fact that \( x/(1 + x) \leq x^s \) for all \( x \geq 0 \) and any \( s \in [0, 1] \) we have
\[ \sup_{\theta \in B(\theta_0)} \frac{h_i(\theta_0)}{h_i(\theta)} \leq C + \frac{\beta_0^2}{\beta_1^3} \sum_{k=0}^{i-2} \sup_{\theta \in B(\theta_0)} \left( \frac{\gamma_0}{\gamma} \right)^k \frac{\gamma^k \beta^g Z_{i-1-k}^2}{\omega^g + \gamma^k \beta^g Z_{i-1-k}^2} \]
\[ \leq C + C \sum_{k=0}^{i-2} \sup_{\theta \in B(\theta_0)} \left( \frac{\gamma_0}{\gamma} \right)^k \gamma^s \beta^g Z_{i-1-k}^2 \]
\[ \leq C + C \sum_{k=0}^{i-2} (1 + \delta)^k \gamma^s \beta^g Z_{i-1-k}^2. \]

Take \( 0 < \delta < \frac{1-\gamma_s^s}{\gamma_s^s} \). Then \((1 + \delta) \gamma_s^s < 1\). Choose \( s \in [0, 1] \) such that \( \sup_{i \in \mathbb{N}} E \left( Z_{i-1-k}^{2\rho_s} \right) < \infty \), and by Minkowski’s inequality we have
\[ \left\| \sup_{\theta \in B(\theta_0)} \frac{h_i(\theta_0)}{h_i(\theta)} \right\|_{L_p} \leq C + C \sum_{k=0}^{i-2} \rho^k \| Z_{i-1-k}^{2\rho_s} \|_{L_p} < \infty. \]

Since \( |\rho| < 1 \),
\[ \sup_{i \in \mathbb{N}} \left\| \sup_{\theta \in B(\theta_0)} \frac{h_i(\theta_0)}{h_i(\theta)} \right\|_{L_p} < \infty. \]

(e). First, we investigate the first derivatives. Since \( \sigma_i^2 \) is the linear function of \( \beta \) and \( \omega \), we can show
\[ h_i(\theta)^{-1} \frac{\partial h_i(\theta)}{\partial \theta_j} \leq C \text{ a.s. for } j = 1, 2. \]  

(5.1)
Using the fact that \( x/(1 + x) \leq x^s \) for any \( s \in [0, 1] \) and all \( x \geq 0 \) we can show

\[
0 \leq h_i(\theta)^{-1} \frac{\partial h_i(\theta)}{\partial \gamma} = h_i(\theta)^{-1} \left\{ \sum_{k=1}^{i-2} \left[ k \gamma^{k-1}(\omega^g + \beta^g Z_{i-1-k}^2) + \gamma^{k-1}(e^\beta - 1 - \beta)Z_{i-1-k}^2 \right] \\
+ (i-1) \gamma^{i-2} \frac{\partial h_i(\theta)}{\partial \gamma} \right\} \leq C \sum_{k=1}^{i-2} k \frac{\gamma^{k}(\omega^g + \beta^g Z_{i-1-k}^2)}{\omega^g + \gamma^{k}(\omega^g + \beta^g Z_{i-1-k}^2)} + C
\]

\[
\leq C \sum_{k=1}^{i-2} k \rho^{ks}(\omega^g + \beta^g Z_{i-1-k}^2)^s + C.
\]

Choose \( s \in [0, 1] \) such that \( E \left( \omega^g + \beta^g Z_{i-1-k}^2 \right)^s p < \infty \), and by the fact that \( |\rho| < 1 \) we have

\[
\sup_{i \in \mathbb{N}} \left\| \sup_{\theta \in \Theta} \left\| \frac{\partial h_i(\theta)}{\partial \gamma} \right\|_{L^p} \right\| \leq C.
\]

Similarly, we can show the bound for the second and third derivatives. ■

**Lemma 3** Under Assumption 1 (a), (b), (e), and (f), we have

\[
\sup_{\theta \in \Theta} \left| \hat{L}_{n,m}^{GH}(\theta) - L_n^{GH}(\theta) \right| = O_p(m^{-1/4}) + o_p(1). \tag{5.2}
\]

**Proof**: By the triangular inequality we have

\[
\left| \hat{L}_{n,m}^{GH}(\theta) - L_n^{GH}(\theta) \right| \leq \left| \hat{L}_{n,m}^{GH}(\theta) - \hat{L}_n^{GH}(\theta) \right| + \left| \hat{L}_n^{GH}(\theta) - L_n^{GH}(\theta) \right|. \tag{5.3}
\]

For the first term on the right hand side of (5.3), since \( h_i(\theta) \) stays away from zero, we have

\[
E \left[ \sup_{\theta \in \Theta} \left| \hat{L}_{n,m}^{GH}(\theta) - \hat{L}_n^{GH}(\theta) \right| \right] \leq C \frac{1}{n} \sum_{i=1}^{n} E \left[ \left| RV_i - \int_{i-1}^{i} \sigma_t^2 dt \right| \right] \leq C m^{-1/4},
\]

where the last inequality is due to Assumption 1 (f). Thus,

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

For the second term on the right hand side of (5.3), simple algebra shows

\[
\hat{L}_n^{GH}(\theta) - L_n^{GH}(\theta) = \frac{1}{2n} \sum_{i=1}^{n} \frac{D_i}{h_i(\theta)}.
\]
Since \( h_i(\theta) \) is adapted to \( F_{i-1} \), \( \frac{D_n}{h_i(\theta)} \) is also a martingale difference. The fact that \( \left| \frac{D_n}{h_i(\theta)} \right| \leq \frac{1}{\omega^2} |D_i| \) implies the uniform integrability of \( \left| \frac{D_n}{h_i(\theta)} \right| \). By Theorem 2.22 in Hall and Heyde (1980), we have

\[
\left| L_n^{GH}(\theta) - L_n^{GH}(\theta) \right| \to 0 \text{ in probability.}
\]

Now we need to establish the uniform convergence. Let \( G_n(\theta) = \hat{L}_n^{GH}(\theta) - L_n^{GH}(\theta) \). If \( G_n(\theta) \) is stochastic equicontinuous, the uniform convergence can be obtained. By Taylor expansion and the mean value theorem there exists \( \theta^* \) between \( \theta \) and \( \theta' \) such that

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

In case of \( \beta \) and \( \omega \), by (5.1) we have for \( j = 1, 2 \),

\[
\left\| \sup_{\theta \in \Theta} \left| \frac{\partial h_i(\theta^*)}{\partial \theta_j} \frac{D_i}{h_i^2(\theta^*)} \right| \right\|_{L_1} \leq C \| D_i \|_{L_1} \leq C < \infty.
\]

In case of \( \gamma \), by Lemma 2 (b) and (e) we have

\[
\left\| \sup_{\theta^* \in \Theta} \left| \frac{\partial h_i(\theta^*)}{\partial \gamma} \frac{D_i}{h_i^2(\theta^*)} \right| \right\|_{L_1} \leq C \left\| \sup_{\theta^* \in \Theta} \left| \frac{\partial h_i(\theta^*)}{\partial \gamma} \frac{\int_{i-1}^{i} (X_t - X_{i-1})^2 dt + \int_{i-1}^{i} \sigma_t^2 dt}{h_i^2(\theta^*)} \right| \right\|_{L_1}
\leq C \left\| \sup_{\theta^* \in \Theta} \left| \frac{\partial h_i(\theta^*)}{\partial \gamma} \frac{h_i(\theta_0)}{h_i^{\prime}(\theta^*)} \right| \right\|_{L_1},
\]

(5.4)

where the second inequality is due to Itô's Lemma and the tower property. If \( \gamma \geq \gamma_0 \), Lemma 2 (c) and (e) imply that (5.4) is bounded. Thus we just need to work for the case of \( \gamma \leq \gamma_0 \). For any \( 0 \leq k \leq i - 1 \), we have

\[
\frac{h_i(\theta_0)}{h_i(\theta)} \leq C + C \sum_{l=0}^{k-1} \frac{Z_i^{\prime} Z_{i-1}^{\prime l}}{\gamma \gamma_i Z_{i-1}^{\prime l}} + \frac{\gamma \gamma_i}{h_i(\theta_0)} \leq C + C \left[ k \left( \frac{\gamma}{\gamma_0} \right)^k \right] + \frac{\gamma \gamma_i}{h_i(\theta_0)}.
\]

Thus, by Lemma 2 (a) we obtain that the right hand side of (5.4) is bounded by

\[
C \left\| \sup_{\theta^* \in \Theta} \left\{ \sum_{k=0}^{i-2} \frac{\gamma^k h_i^{\prime - k}(\theta)}{h_i(\theta)} \left\{ C + C \left[ k \left( \frac{\gamma}{\gamma_0} \right)^k \right] + \gamma \gamma_i \gamma_i^{\prime - k}(\theta) \right\} \right\} \right\|_{L_1}
\leq C \left\| \sup_{\theta^* \in \Theta} \left\{ \frac{\gamma^k h_i^{\prime - k}(\theta)}{h_i(\theta)} + \frac{k \gamma \gamma_i}{h_i(\theta)} + \frac{\gamma \gamma_i}{h_i(\theta)} \right\} \right\|_{L_1}.
\]

22
\[
\begin{align*}
\leq C \sum_{k=0}^{i-2} \left\{ \gamma_u \left\| \sup_{\theta \in \Theta} h_{i-1-k}(\theta) \right\|_{L_1} + k \gamma_0 \left\| \sup_{\theta \in \Theta} h_{i-1-k}(\theta) \right\|_{L_1} + \gamma_k \left\| h_{i-k}(\theta_0) \right\|_{L_1} \right\} \\
\leq C < \infty,
\end{align*}
\]

where the first inequality is due to \( \frac{\gamma_k h_{i-1-k}(\theta)}{h_i(\theta)} \leq 1 \). Therefore, \( G_n(\theta) \) satisfies a weak Lipschitz condition, and Theorem 3 in Andrews (1992) implies that it uniformly converges to zero. ■

**Proof of Theorem 1.** First, we show that there is a unique maximizer of \( L_n^{GH}(\theta) \).

From the definition of \( L_n^{GH}(\theta) \), simple algebra shows

\[
\max_{\theta \in \Theta} L_n^{GH}(\theta) \leq -\frac{1}{2n} \sum_{i=1}^{n} \min_{\theta_i \in \Theta} \log(h_i(\theta_i)) + \frac{h_i(\theta_0)}{h_i(\theta_i)}.
\]

\( \theta_{0i} \) is the minimizer of the \( i \)-th summand on the right hand side of above inequality, if \( \theta_{0i} \) satisfies \( h_i(\theta_{0i}) = h_i(\theta_0) \). Thus, if there exists \( \theta^* \in \Theta \) such that \( h_i(\theta^*) = h_i(\theta_0) \) for all \( i = 1, 2, \ldots, n \), then \( \theta^* \) is the maximizer. Note that \( \theta_0 \) is one of the candidates \( \theta^* \). Below we will show that in fact such \( \theta^* \) must be equal to \( \theta_0 \) a.s. Since

\[
h_i(\theta) = \omega^g + \gamma h_i-1(\theta) + \beta^g Z_{i-1}^2,
\]

\( \theta^* \) and \( \theta_0 \) satisfies the following equation,

\[
\begin{pmatrix}
1 & h_1(\theta_0) & Z_1^2 \\
1 & h_2(\theta_0) & Z_2^2 \\
\vdots & \vdots & \vdots \\
1 & h_{n-1}(\theta_0) & Z_{n-1}^2
\end{pmatrix}
\begin{pmatrix}
\omega^g - \omega_0^g \\
\gamma^* - \gamma_0 \\
\beta^* - \beta_0^g
\end{pmatrix}
\equiv M
\begin{pmatrix}
\omega^g - \omega_0^g \\
\gamma^* - \gamma_0 \\
\beta^* - \beta_0^g
\end{pmatrix}
= 0 \text{ a.s.,}
\]

where \( \omega^g = \beta^* - 1)(\omega^* - 1) \omega^* \) and \( \beta^* = \beta^* - 1)(\gamma^* - 1)(e^{\beta^*} - 1 - \beta^*) + e^{\beta^*} - 1 \). Since \( Z_i \)'s are nondegenerate, \( M \) is of full rank. Then, \( M^T M \) is invertible, and

\[
\begin{pmatrix}
\omega^g - \omega_0^g \\
\gamma^* - \gamma_0 \\
\beta^* - \beta_0^g
\end{pmatrix}
= 0 \text{ a.s.}
\]

Since given \( \gamma, \beta^g \) is a strictly increasing function with respect to \( \beta \), \( \theta^* = \theta_0 \), and there is a unique maximizer. Then, since \( L_n^{GH}(\theta) \) is a continuous function, for any \( \varepsilon > 0 \), there is a

23
constant \( \vartheta > 0 \) such that

\[
L_n^{GH}(\theta_0) - \max_{\theta \in \mathcal{B}} L_n^{GH}(\theta) > \vartheta \text{ a.s.}
\] (5.5)

Now the theorem is a consequence of Theorem 1 in Xiu (2010), Lemma 3 and (5.5).

### 5.4 Proof of Theorem 2

**Lemma 4** Under Assumption 1 (a) and (g), we have

(a) there exists a neighborhood \( B(\theta_0) \) of \( \theta_0 \) such that \( \sup_{i \in \mathbb{N}} \left\| \sup_{\theta \in B(\theta_0)} \frac{\partial^3 \psi_{n}^{GH}(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} \right\|_{L_1} < \infty \)

for any \( j, k, l \in \{1, 2, 3\} \), where \( \theta = (\theta_1, \theta_2, \theta_3) = (\omega, \beta, \gamma) \);

(b) \(-\nabla \psi_{n}^{GH}(\theta_0)\) is a positive definite matrix for \( n \geq 3 \).

**Proof:** (a). Simple algebraic manipulations show

\[
\frac{\partial^3 \psi_{n}^{GH}(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} = \left\{ 1 - \frac{RV_i}{h_i(\theta)} \right\} \left\{ 1 \frac{\partial^2 h_i(\theta)}{h_i(\theta) \partial \theta_j \partial \theta_k \partial \theta_l} \right\} + \left\{ \frac{2}{h_i(\theta)} - 1 \right\} \left\{ \frac{1}{h_i(\theta) \partial \theta_j} \left\{ \frac{1}{h_i(\theta) \partial \theta_k} \left\{ \frac{1}{h_i(\theta) \partial \theta_l} \right\} \right\} \right\} + \left\{ \frac{2}{h_i(\theta)} - 1 \right\} \left\{ \frac{1}{h_i(\theta) \partial \theta_k} \left\{ \frac{1}{h_i(\theta) \partial \theta_l} \right\} \right\} + \left\{ \frac{2}{h_i(\theta)} - 1 \right\} \left\{ \frac{1}{h_i(\theta) \partial \theta_l} \left\{ \frac{1}{h_i(\theta) \partial \theta_k} \right\} \right\} + \left\{ \frac{2}{h_i(\theta)} - 1 \right\} \left\{ \frac{1}{h_i(\theta) \partial \theta_k} \left\{ \frac{1}{h_i(\theta) \partial \theta_l} \right\} \right\} + \left\{ \frac{2}{h_i(\theta)} - 1 \right\} \left\{ \frac{1}{h_i(\theta) \partial \theta_l} \left\{ \frac{1}{h_i(\theta) \partial \theta_k} \right\} \right\}. \]

By Assumption 1 (g) we get

\[
E \left[ RV_i | \mathcal{F}_{i-1} \right] \leq CE \left[ \int_{i-1}^{i} \sigma_t^2 dt \right] + C = C \cdot h_i(\theta_0) + C \text{ a.s.}
\]

Then, by Lemma 2 (d) and (e) we have

\[
E \left[ \sup_{\theta \in B(\theta_0)} \left| \frac{RV_i}{h_i(\theta)} \left\{ \frac{1}{h_i(\theta) \partial \theta_j \partial \theta_k \partial \theta_l} \right\} \right| \right] \leq CE \left[ \sup_{\theta \in B(\theta_0)} \left| \frac{h_i(\theta)}{h_i(\theta)} \left\{ \frac{1}{h_i(\theta) \partial \theta_j \partial \theta_k \partial \theta_l} \right\} \right| \right] + C
\]

\[
\leq C \left\| \sup_{\theta \in B(\theta_0)} \left| \frac{h_i(\theta)}{h_i(\theta)} \right| \right\|_{L_p} \left\| \sup_{\theta \in B(\theta_0)} \left| \frac{1}{h_i(\theta) \partial \theta_j \partial \theta_k \partial \theta_l} \right| \right\|_{L_q} + C \leq C < \infty,
\]

24
Since

\[
\text{convergence rate of } \hat{h}_\theta \\text{ some }
\]

If

\[-\lambda \text{ the left hand side is of full rank a.s. Thus,}
\]

Then, there exists \( \lambda \neq 0 \) such that \( \frac{1}{2n} \sum_{i=1}^{n} \lambda^T h_{\theta,i} h_{\theta,i}^T \lambda = 0 \). This implies

\[ h_{\theta,i}^T \lambda = 0 \text{ a.s. for all } i = 1, \ldots, n. \]

Since \( h_i(\theta_0) \) stays away from zero, we can rewrite this as

\[
\left( \begin{array}{ccc}
\frac{\partial h_1(\theta_0)}{\partial \omega} & \frac{\partial h_1(\theta_0)}{\partial \beta} & \frac{\partial h_1(\theta_0)}{\partial \gamma} \\
\frac{\partial \omega_0}{\partial \omega} + \gamma \frac{\partial h_1(\theta_0)}{\partial \omega} & \frac{\partial \omega_0}{\partial \beta} + \frac{\partial \omega_0}{\partial \beta} Z_1^2 + \gamma \frac{\partial h_1(\theta_0)}{\partial \beta} & \frac{\partial h_1(\theta_0)}{\partial \gamma} \\
\vdots & \vdots & \vdots \\
\frac{\partial \omega_n}{\partial \omega} + \gamma \frac{\partial h_{n-1}(\theta_0)}{\partial \omega} & \frac{\partial \omega_n}{\partial \beta} + \frac{\partial \omega_n}{\partial \beta} Z_{n-1}^2 + \gamma \frac{\partial h_{n-1}(\theta_0)}{\partial \beta} & \frac{\partial h_{n-1}(\theta_0)}{\partial \gamma}
\end{array} \right) \lambda = 0 \text{ a.s.,}
\]

where \( \frac{\partial \omega_n}{\partial \omega} = \beta_0^{-1}(e^{\beta_0} - 1), \frac{\partial \omega_n}{\partial \beta} = \beta_0^{-2}(1 - e^{\beta_0}) + e^{\beta_0} \), \( \frac{\partial \omega_n}{\partial \gamma} = \beta_0^{-1}(e^{\beta_0} - 1 - \beta_0) \), and

\[ \frac{\partial \omega_n}{\partial \eta} = (\gamma_0 - 1)(e^{\beta_0} - \beta_0^{-2}e^{\beta_0} + \beta_0^{-2}) + \beta_0 e^{\beta_0} \].

Since \( Z_i \)'s are nondegenerate, the matrix on the left hand side is of full rank a.s. Thus, \( \lambda = 0 \) a.s., which is a contradiction. \( \blacksquare \)

**Proof of Theorem 2.** By Taylor expansion and the mean value theorem we have for some \( \theta^* \) between \( \theta_0 \) and \( \hat{\theta}^{GH} \):

\[
\hat{\psi}_{n,m}^{GH}(\hat{\theta}^{GH}) - \hat{\psi}_{n,m}^{GH}(\theta_0) = -\hat{\psi}_{n,m}^{GH}(\theta_0) = \nabla \hat{\psi}_{n,m}(\theta^*)(\hat{\theta}^{GH} - \theta_0).
\]

If \( -\nabla \hat{\psi}_{n,m}(\theta^*) \stackrel{p}{\to} -\nabla \hat{\psi}_{n,m}^{GH}(\theta_0) \) which is a positive definite matrix by Lemma 4 (b), then the convergence rate of \( \hat{\theta}^{GH} - \theta_0 \) is the same as that of \( \hat{\psi}_{n,m}(\theta_0) \). So it is enough to show

\[
\left\| \nabla \hat{\psi}_{n,m}(\theta^*) - \nabla \hat{\psi}_{n}^{GH}(\theta_0) \right\|_{\text{max}} = o_p(1) \quad \text{and} \quad \hat{\psi}_{n,m}(\theta_0) = O_p(m^{-1/4}) + O_p(n^{-1/2}).
\]

To prove the second result, we first show

\[
\left\| \hat{\psi}_{n,m}(\theta_0) - \hat{\psi}_{n}^{GH}(\theta_0) \right\|_{\text{max}} = O_p(m^{-1/4}).
\]

25
For any \( j \in \{1, 2, 3\} \), by Hölder’s inequality and Lemma 2 (e) we have

\[
\left\| \frac{1}{2n} \sum_{i=1}^{n} \frac{\partial h_i(\theta)}{\partial \theta_j} h_i(\theta)^{-2} \left( RV_i - \int_{i-1}^{i} \sigma_i^2 dt \right) \right\|_{L_1} \leq C \frac{1}{n} \sum_{i=1}^{n} \left\| \frac{\partial h_i(\theta)}{\partial \theta_j} h_i(\theta)^{-1} \right\|_{L_q} \left\| RV_i - \int_{i-1}^{i} \sigma_i^2 dt \right\|_{L_p} \leq C m^{-1/4}, \quad (5.6)
\]

where \( 1 < p \leq 1 + \delta \) and \( 1/p + 1/q = 1 \), and the last inequality is due to Assumption 1 (f).

Then, we have

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

\[
= \frac{1}{2n} \sum_{i=1}^{n} \frac{\partial h_i(\theta_0)}{\partial \theta} h_i(\theta_0)^{-1} D_i \frac{h_i(\theta_0)}{h_i(\theta_0)} + O_p(m^{-1/4}).
\]

By Itô’s lemma and Itô isometry we have for any \( j \in \{1, 2, 3\} \),

\[
E \left[ \left( \frac{1}{2n} \sum_{i=1}^{n} \frac{\partial h_i(\theta_0)}{\partial \theta_j} h_i(\theta_0)^{-1} D_i \frac{h_i(\theta_0)}{h_i(\theta_0)} \right)^2 \right] = \frac{1}{4n^2} \sum_{i=1}^{n} E \left( \left( \frac{\partial h_i(\theta_0)}{\partial \theta_j} \right)^2 h_i(\theta_0)^{-2} \frac{D_i^2}{h_i^2(\theta_0)} \right)
\]

\[
= \frac{1}{4n^2} \sum_{i=1}^{n} E \left( \left( \frac{\partial h_i(\theta_0)}{\partial \theta_j} \right)^2 h_i(\theta_0)^{-2} \frac{E[D_i^2 | F_{i-1}]}{h_i^2(\theta_0)} \right)
\]

\[
= \frac{1}{4n^2} \sum_{i=1}^{n} E \left( \left( \frac{\partial h_i(\theta_0)}{\partial \theta_j} \right)^2 h_i(\theta_0)^{-2} \frac{\int_{i-1}^{i} (e^{(i-j)} - 1)^2 (X_i - X_{i-1})^2 \sigma_i^2 dt | F_{i-1}}{h_i^2(\theta_0)} \right)
\]

\[
\leq C \frac{1}{n^2} \sum_{i=1}^{n} E \left( \left( \frac{\partial h_i(\theta_0)}{\partial \theta_j} \right)^2 h_i(\theta_0)^{-2} \frac{Z_{i-1}^4 | F_{i-1}}{h_i^2(\theta_0)} \right). \quad (5.7)
\]

If the condition (c1) in Assumption 1 (c) is satisfied, we immediately conclude from Lemma 2 (e) that (5.7) is of order \( n^{-1} \). If the condition (c2) in Assumption 1 (c) is satisfied, by Hölder’s inequality and Lemma 2 (e) we have that (5.7) is of order \( n^{-1} \). Thus,

\[
\hat{\psi}_{n,m}^GH(\theta_0) = O_p(m^{-1/4}) + O_p(n^{-1/2}).
\]

In the case of \( \left\| \nabla \hat{\psi}_{n,m}^GH(\theta^*) - \nabla \hat{\psi}_n^GH(\theta_0) \right\|_{max} = o_p(1) \), by the triangular inequality we have

\[
\left\| \nabla \hat{\psi}_{n,m}^GH(\theta^*) - \nabla \hat{\psi}_n^GH(\theta_0) \right\|_{max} \leq \left\| \nabla \hat{\psi}_{n,m}^GH(\theta^*) - \nabla \hat{\psi}_{n,m}^GH(\theta_0) \right\|_{max} + \left\| \nabla \hat{\psi}_{n,m}^GH(\theta_0) - \nabla \hat{\psi}_n^GH(\theta_0) \right\|_{max}. \quad (5.8)
\]

26
For the first term on the right hand side of (5.8), let \( M_i = \max_{j,k,l \in \{1,2,3\}} \sup_{\theta \in B(\theta_0)} \left| \frac{\partial^3 \hat{\psi}_{GH}^n(\theta)}{\partial \theta_j \partial \theta_k \partial \theta_l} \right| \).

For large enough \( n \) and \( m \), by Taylor expansion and the mean value theorem we have

\[
\left\| \nabla \hat{\psi}_{n,m}^G(\theta^*) - \nabla \hat{\psi}_{n,m}^G(\theta_0) \right\|_{\max} \leq \frac{C}{n} \sum_{i=1}^{n} M_i \| \theta^* - \theta_0 \|_{\max} = o_p(1),
\]

where the last equality is derived by Theorem 1 and the fact that \( \frac{1}{n} \sum_{i=1}^{n} M_i = O_p(1) \) implied by Lemma 4 (a).

For the second term on the right hand side of (5.8), similar to the proof of (5.6), we can apply Hölder’s inequality and Lemma 2 (e) to show

\[
\left\| \nabla \hat{\psi}_{n,m}^G(\theta_0) - \nabla \hat{\psi}_{n,m}^G(\theta_0) \right\|_{\max} = O_p(m^{-1/4}).
\]

On the other hand, we have

\[
\nabla \hat{\psi}_{n,m}^G(\theta_0) = \nabla \hat{\psi}_{n}^G(\theta_0) + O_p(m^{-1/4}) = \nabla \hat{\psi}_{n}^G(\theta_0) + \eta_n + O_p(m^{-1/4}) \text{ a.s.,}
\]

where \( \eta_n = \frac{1}{2n} \sum_{i=1}^{n} \frac{\partial^2 h_i(\theta_0)}{\partial \theta^2} h_i(\theta_0)^{-1} \left( - \frac{D_i}{h_i(\theta_0)} \right) + \left( \frac{\partial h_i(\theta_0)}{\partial \theta} \right) \left( \frac{\partial h_i(\theta_0)}{\partial \theta} \right)^T h_i(\theta_0)^{-1} \frac{2D_i}{h_i(\theta_0)} \). Note that \( \eta_n \) is a martingale. Then we can show \( \| \eta_n \|_{\max} = O_p(n^{-1/2}) \) similar to the proof of (5.7). Finally we establish

\[
\nabla \hat{\psi}_{n,m}^G(\theta_0) = \nabla \hat{\psi}_{n}^G(\theta_0) + O_p(n^{-1/2}) + O_p(m^{-1/4}).
\]

\( \blacksquare \)

5.5 Proof of Theorem 3

**Proof of Theorem 3.** For any \( \lambda \in \mathbb{R}^3 \), let

\[
d_i = \lambda^T \frac{\partial h_i(\theta_0)}{\partial \theta} h_i(\theta_0)^{-1} \frac{D_i}{h_i(\theta_0)}.
\]

Then \( d_i \) is a martingale difference, and similar to the proof of (5.7) we can show \( E(d_i^2) < \infty \). \((D_i, Z_i^2)\)'s are stationary and ergodic processes, thus \( d_i \) is stationary and ergodic. Applying the martingale central limit theorem we obtain

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} d_i \overset{d}{\to} N(0, E(d_i^2)).
\]
Using Cramér-Wold device we can show

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

By the ergodic theorem and the result in the proof of Theorem 2 we have

$$-\nabla \hat{\psi}_{n,m}^{GH}(\theta_0) \to B \text{ in probability},$$

and similar to the proof of Lemma 4 (b) we can show that $B$ is a positive definite matrix.

Finally we conclude

$$\sqrt{n}(\hat{\theta}^{GH} - \theta_0) = -\sqrt{n}B^{-1}\hat{\psi}_{n}^{GH}(\theta_0) + O_p(n^{1/2}m^{-1/4}) + o_p(1) \to N(0, B^{-1}A^{GH}B^{-1}).$$


5.6 Proofs of Theorems 4-6

The arguments to prove Theorems 4-6 are similar to the proofs of Theorems 1-3.

Acknowledgement. Yazhen Wang’s research was supported in part by the NSF Grants DMS-10-5635 and DMS-12-65203.

References


Donggyu Kim, DEPARTMENT OF STATISTICS, UNIVERSITY OF WISCONSIN-MADISON, 1300 UNIVERSITY AVENUE, MADISON, WI 53706.

E-mail address: kimd@stat.wisc.edu

Yazhen Wang, DEPARTMENT OF STATISTICS, UNIVERSITY OF WISCONSIN-MADISON, 1300 UNIVERSITY AVENUE, MADISON, WI 53706.

E-mail address: yzwang@stat.wisc.edu