# Limit theorems for a class of martingale arrays with applications in nonlinear cointegrating regression

Qiying Wang

The University of Sydney *

August 22, 2026

**Abstract**

This paper develops a new asymptotic theory for a broad class of martingales, establishing convergence to limiting distributions that involve a functional of stochastic integrals. The proposed limit theorem substantially extends existing martingale asymptotic theory by accommodating a wider class of dependence structures. As a primary application, the theory is applied to nonlinear regression models with nonstationary time series, yielding a rigorous framework for asymptotic inference on nonlinear least square estimators.

## 1 Introduction

Let $\{\epsilon_{nj}, \mathcal{F}_{nj}\}_{j\geq 1, n\geq 1}$ be a martingale difference array defined on a common probability space $(\Omega, \mathcal{F}, P)$ with $\mathbb{E}\big(\epsilon_{nj}^2|\mathcal{F}_{n,j-1}\big)=1, a.s.$. For each $n\geq 1$, let $\{u_{nj}\}_{j\geq 1}$ and $\{x_{nj}\}_{j\geq 1}$ be arbitrary random sequences so that both are adapted to $\mathcal{F}_n=\{\mathcal{F}_{n,j-1}\}_{j\geq 1}$ and, on $D_{\mathbb{R}^2}[0,\infty)$,

$$\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\epsilon_{nj},\ x_{n,[nt]}\Big) \ \Rightarrow\ (B_t, X_t),$$

*Address correspondence to: School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia. Email: qiying.wang@sydney.edu.au.

where $B = \{B_t\}_{t\geq 0}$ is a Brownian motion and $X = \{X_t\}_{t\geq 0}$ is a continuous process. Define a martingale array by $\{S_{nk}, \mathcal{F}_{nk}\}_{k\geq 2, n\geq 1}$, where

$$S_{nk} = \frac{1}{\sqrt{n}} \sum_{j=1}^{k} H(x_{nj}, u_{nj})\, \epsilon_{nj}, \quad k = 2, 3, ...$$

and $H$ is a continuous function of its components. In this paper, we concern with the asymptotics of a statistic $S_n$ defined by $S_n := S_{nn}$, as $n \to \infty$ and their applications in nonlinear cointegrating regression.

The limit theory for martingales and/or statistics like $S_n$ derived from martingale arrays is a celebrated achievement in the development of Probability and Statistics. For various martingales with different structures, extensive articles have established a number of fundamental results on central limit theorems, convergence to a mixture of normal distributions and convergence to stochastic integrals. The earlier contributions to these topics were summarized in excellent books by **?**, **?** and **?**.

To the best of our knowledge, however, the asymptotic behavior of $S_n$ has not been investigated in the literature, because of the presence of both stationary (e.g.,$u_{nj}$) and non-statinnary (e.g.,$x_{nj}$) variables. First of all, when $u_{nj}$ is stationary, no limitation exists for $H(x_{n,[nt]}, u_{n,[nt]})$ as $n \to \infty$ and therefore the classical theory on the convergence of stochastic integrals (e.g., **?**) is unlikely to be used for such a martingale array $\{S_{nk}, \mathcal{F}_{nk}\}_{k\geq 2, n\geq 1}$. On the other hand, by writing the conditional variance $V_n^2$ of $S_n$ by $V_n^2 = \frac{1}{n}\sum_{j=1}^{n} H^2(x_{nj}, u_{nj})$, we may prove that $V_n^2 \to_D \int_0^1 \widehat{H}^2(X_t)dt$ as in Remark 2.3, where $\widehat{H}^2(x)$ is defined as in Asssumption 2.3 below, but it is usually impossible to show that $V_n^2 \to_P \int_0^1 \widehat{H}^2(X_t)dt$ under natural settings on $H(x, y)$ and $u_{nj}$[1]. This claim suggests that the classical martingale limit theorems are unable to establish the asymptotics of $S_n$. Indeed, many earlier results on the convergence to a mixture of normal distributions and on the convergence to a semimartingale were established under the assumption that $V_n^2 \to_P V^2$, where $V^2$ is an almost finite random variable (e.g.,**?**; **?**). More recently, in investigating the convergence to a mixture

[1]For a simple illustration, assume that $H(x, y)$ is a continuous function on $\mathbb{R}^2$ and $x_{nj} = \frac{1}{\sqrt{n}}\sum_{j=1}^{j} \epsilon_i$, where $\{\epsilon_j\}_{j\geq 1}$ is a sequence of iid random variables with $E\epsilon_1 = 0$ and $E\epsilon_1^2 = 1$. We further assume that $u_{nj} = \epsilon_j$ and $\epsilon_{nj} = \epsilon_j$ so that $V_n^2 = \frac{1}{n}\sum_{j=1}^{n} H^2(x_{nj}, \epsilon_j)$. Note that $x_{n,[nt]} \Rightarrow B_t$ on $D[0, 1]$, but the convergence in distribution can not be enhanced to the convergence in probability. We may prove that $V_n^2 \to_D \int_0^1 \hat{H}(B_t)dt$, where $\hat{H}(x) = EH^2(x, \epsilon_1)$, but it is incorrect to say that $V_n^2 \to_P \int_0^1 \hat{H}(B_t)dt$.

of normal distributions for a martingale array, **?**, **??** and **???** successfully relaxed the requirement $V_n^2 \to_P V^2$ to the weaker condition $V_n^2 \to_D V^2$. However, when their results are applied to the present martingale array $\{S_{nk}, \mathcal{F}_{nk}\}_{k\geq 2, n\geq 1}$, a crucial additional assumption imposed in the cited papers, namely, $\frac{1}{n}\sum_{j=1}^{n} H(x_{nj}, u_{nj}) \to_P 0$, is not satisfied.

The main aim of this paper is to remove this restriction. Under certain regular conditions on $x_{nj}$ and $u_{nj}$, ensuring that $\frac{1}{n}\sum_{j=1}^{n} H(x_{nj}, u_{nj}) \to_D \int_0^1 \widetilde{H}(X_t)dt$, our main result Theorem 2.1 establishes new asymptotics involving a functional of stochastic integrals for $S_{nn}$, providing a framework in an extension to the existing martingale limit theorem. In the remainder of this paper, our new result is applied to nonlinear regression with nonstationary time series in Section 3, allowing for the models to include an integrated regressor with certain lags, which are most commonly used in practice. Using Theorem 2.1 as a main tool, Theorem 3.1 investigate the asymptotics for the conventional least square estimators in a nonlinear cointegrating regression model. These results improve those in the existing literature. All technical proofs are given in Section 4.

Throughout the paper, we denote by $C, C_1, ...$ constants, which may change at each appearance. The notation $\to_D$ ($\to_P$ respectively) denotes convergence in distribution (in probability respectively) for a sequence of random variables (or vectors). The notation $=_D$ denotes equality in distribution. If $\alpha_n^{(1)}$, $\alpha_n^{(2)}$,..., $\alpha_n^{(k)}$ $(1 \leq n \leq \infty)$ are random elements on $D[0,1]$ or $D[0,\infty)$ or $\mathbf{R}$, we will understand the condition

$$(\alpha_n^{(1)}, \alpha_n^{(2)}, ..., \alpha_n^{(k)}) \Rightarrow (\alpha_\infty^{(1)}, \alpha_\infty^{(2)}, ..., \alpha_\infty^{(k)})$$

to mean that for all $\alpha_\infty^{(1)}$, $\alpha_\infty^{(2)}$,..., $\alpha_\infty^{(k)}$-continuity sets $A_1$, $A_2$,...,$A_k$,

$$P\big(\alpha_n^{(1)} \in A_1, \alpha_n^{(2)} \in A_2, ..., \alpha_n^{(k)} \in A_k\big) \to P\big(\alpha_\infty^{(1)} \in A_1, \alpha_\infty^{(2)} \in A_2, ..., \alpha_\infty^{(k)} \in A_k\big).$$

[see Theorem 3.1 of **?**]. As usual, $D[0,1]$ ($D[0,\infty)$ respectively) denotes the space of cadlag functions on $[0,1]$ ($[0,\infty)$ respectively), which is equipped with Skorohod topology.

## 2 Main results

We recall that $\{\epsilon_{nj}, \mathcal{F}_{nj}\}_{j\geq 1, n\geq 1}$ is a martingale difference array defined on a common probability space $(\Omega, \mathcal{F}, P)$ with $\mathbb{E}\big(\epsilon_{nj}^2 | \mathcal{F}_{n,j-1}\big) = 1, a.s.$ for all $n, j \geq 1$. For each

$n \geq 1$, we always assume that $\mathcal{F}_{n0}$ is a trivial $\sigma$-field, $\mathcal{F}_{nj} \subseteq \mathcal{F}_{n,j+1}, j \geq 1$, and both random vector sequences $\{x_{nj}\}_{j\geq 1}$ (with degree $d_1$) and $\{u_{nj}\}_{j\geq 1}$ (with degree $d_2$) are adapted to the filtration $\mathcal{F}_n = \{\mathcal{F}_{n,j-1}\}_{j\geq 1}$. In particular, we allow for $d_2 = \infty$, i.e., we may have $u_{nj} = (u_{nj,1}, u_{nj,2}, ...)$. This section investigates the asymptotics of $S_n$ defined by

$$S_n \;=\; \frac{1}{\sqrt{n}} \sum_{j=1}^{n} H(x_{nj}, u_{nj})\, \epsilon_{nj}.$$

To this end, we make use of the following additional assumptions.

**Assumption 2.1.** *There exists a d-dimensional martingale difference array* $\{(\eta_{nj,1}, ..., \eta_{nj,d}), \mathcal{F}_{nj}\}_{j\geq 1, n\geq 1}$ *on* $(\Omega, \mathcal{F}, P)$ *so that* $\mathbb{E}\big(\epsilon_{nj}\eta_{nj,1} \mid \mathcal{F}_{n,j-1}\big) = ... = \mathbb{E}\big(\epsilon_{nj}\eta_{nj,d} \mid \mathcal{F}_{n,j-1}\big) = 0$ *for all* $n, j \geq 1$ *and, on* $D_{\mathbb{R}^{1+d+d_1}}[0,\infty)$,

$$\Big(\frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \epsilon_{nj},\ \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \eta_{nj},\ x_{n,[nt]}\Big) \;\Rightarrow\; \big(B_t, B_{1t}, X_t\big), \tag{2.1}$$

*where* $\eta_{nj} = (\eta_{nj,1}, ..., \eta_{nj,d})$, $(B_t, B_{1t}), t \geq 0$, *is a* $(1+d)-$*dimensional standard Brownian motion and* $X = \{X_t\}_{t\geq 0}$ *is a continuous process adapted to the filtration* $\mathcal{F} = \{\sigma(B_s, B_{1s}, s \leq t)\}_{t\geq 0}$.

**Assumption 2.2.** $H(x, y)$ *is a real function satisfying the conditions:*

*(a)* $H(x, y)$ *is piecewise continuous with respect to* $x \in \mathbb{R}^{d_1}$*;*

*(b) for each* $A > 0$, $\{g_A^2(u_{nj})\}_{n\geq 1, j\geq 1}$ *is uniformly integrable, where* $g_A(y) = \sup_{||x||\leq A} |H(x, y)|$.

**Assumption 2.3.** *There exist piecewise continuous functions* $\widetilde{H}(x)$ *and* $\widehat{H}(x)$ *on* $\mathbb{R}^{d_1}$ *and for some sequence* $0 < m := m_n \to \infty$ *with* $n/m \to \infty$ *so that, for each fixed* $x$,

$$\max_{m\leq k\leq n-m} \mathbb{E}\Big|\frac{1}{m} \sum_{j=km+1}^{(k+1)m} H(x, u_{nj}) - \widetilde{H}(x)\Big| \;\to\; 0, \tag{2.2}$$

$$\max_{m\leq k\leq n-m} \mathbb{E}\Big|\frac{1}{m} \sum_{j=km+1}^{(k+1)m} H^2(x, u_{nj}) - \widehat{H}^2(x)\Big| \;\to\; 0, \tag{2.3}$$

*as* $m \to \infty$.

We have the following main result.

**Theorem 2.1.** *Suppose that Assumptions 2.1 - 2.3 hold. Then, as* $n \to \infty$, *we have*

$$S_n \;\to_D\; S := \int_0^1 \widetilde{H}(X_t) dB_t + \sigma(X)\, \mathcal{N}, \tag{2.4}$$

*where* $\sigma^2(X) = \int_0^1 \big[\widehat{H}^2(X_t) - \widetilde{H}^2(X_t)\big]dt$ *and* $\mathcal{N}$ *is a standard normal variate independent of* $B = \{B_t\}_{t\ge 0}$ *and* $X = \{X_t\}_{t\ge 0}$.

**Remark 2.1.** *The condition (2.1) in Assumption 2.1 is standard in investigating general limit theorems for martingale arrays. See, for instance, ??. The martingale difference array* $\{\eta_{nj}, \mathcal{F}_{nj}\}_{j\ge 1, n\ge 1}$ *is usually involved in exploring the limit law of* $\{x_{nj}\}_{j\ge 1}$. *Its existence is obvious in many applications as seen in the following (3.12). If* $X_t$ *is measurable with respect to* $\sigma(B_s : s \le t)$, *we may take* $\eta_{nj} \equiv 0$. *Note that, if* $X$ *is independent of* $B$, *we may rewrite (2.4) as* $S_n \to_D \sigma_1(X)\mathcal{N}$, *where* $\sigma_1^2(X) = \int_0^1 \widehat{H}^2(X_t)dt$. *Assumption 2.2 is weak, where the piecewise continuity of* $H(x, y)$ *in Assumption 2.2(a) is enough for many empiriacal applications. Assumption 2.2(b) suggests that (2.2) and (2.3) only involve the stationary components of* $S_n$, *which are easy to be verified in general. In particular, when* $u_{nj} \equiv u_j$ *is (strictly) stationary satisfying* $Eg_A^2(u_1) < \infty$ *for each* $A > 0$, *the conditions (2.2) and (2.3) hold with* $\widetilde{H}(x) = \mathbb{E}H(x, u_1)$ *and* $\widehat{H}^2(x) = \mathbb{E}H^2(x, u_1)$, *respectively.*

**Remark 2.2.** *For any square integrable martingale difference array* $\{X_{nj}, \mathcal{F}_{nj}\}_{j\ge 1, n\ge 1}$, *we may write* $X_{nj} = \sigma_{nj}\,\epsilon_{nj}$, *where* $\sigma_{nj}^2 = \mathbb{E}\big(X_{nj}^2|\mathcal{F}_{n,j-1}\big)$ *and* $\{\epsilon_{nj}, \mathcal{F}_{nj}\}_{j\ge 1, n\ge 1}$ *is a martingale difference array with* $\mathbb{E}\big(\epsilon_{nj}^2|\mathcal{F}_{n,j-1}\big) = 1, a.s.$ *for all* $n, j \ge 1$. *Theorem 2.1 imposes the structure* $\sigma_{nj} = H(x_{nj}, u_{nj})$ *with that* $x_{nj}$ *is a sequence of random vectors satisfying* $x_{n,[nt]} \Rightarrow X_t$ *on* $D_{\mathbb{R}^{d_1}}[0, \infty)$. *In earlier works, a similar structure* $\sigma_{nj} = H(r_n\, x_{nj}, u_{nj})$, *where* $r_n \to \infty$ *is a sequence of constants, was considered in ?. We mention that the asymptotics of* $\sum_{j=1}^n H(x_{nj}, u_{nj})\epsilon_{nj}$ *are essentially different from those of* $\sum_{j=1}^n H(r_n x_{nj}, u_{nj})\epsilon_{nj}$ *with* $r_n \to \infty$. *In latter case, the limiting distribution is usually a mixture of normal distributions rather than a function of stochastic integrals. Indeed, under certain regular conditions, when* $r_n/n \to 0$ *and* $r_n \to \infty$, *we may have*

$$\Big(\frac{r_n}{n}\Big)^{1/2} \sum_{j=1}^n H(r_n x_{nj}, u_{nj})\epsilon_{nj} \to_D c_0\, L_X^{1/2}(1, 0)\mathcal{N}$$

*where* $c_0$ *is a constant,* $L_X^{1/2}(1, 0)$ *is a local time process of* $X$ *and* $\mathcal{N}$ *is a normal variate independent of* $X$. *For more details in related topics, together with the applications in financial econometrics, we refer to ?, ???, ? and references therein.*

**Remark 2.3.** *The limiting result (2.4) of* $S_n$ *depends essentially on its conditional variance* $V_n^2 := V_{nn}^2$, *where, for* $k \ge 1$,

$$V_{nk}^2 \;=\; \frac{1}{n}\sum_{j=1}^k H^2(x_{nj}, u_{nj})$$

*and the conditional co-variance $R_{nn}$ between $S_n$ and $\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\epsilon_{nj}$, where, for $k\geq 1$,*

$$R_{nk} = \frac{1}{n}\sum_{j=1}^{k} H(x_{nj}, u_{nj}).$$

*Under (2.2) and (2.3), Lemma 4.2 in Section 4 yields the results:*

$$\max_{1\leq k\leq n}\big|R_{nk}-\frac{1}{n}\sum_{j=1}^{k}\widetilde{H}(x_{nj})\big| = o_P(1),$$

$$\max_{1\leq k\leq n}\big|V_{nk}-\frac{1}{n}\sum_{j=1}^{k}\widehat{H}^2(x_{nj})\big| = o_P(1),$$

*indicating that $R_{nn}\to_D \int_0^1 \widetilde{H}(X_t)dt$ and $V_n^2 \to_D \int_0^1 \widehat{H}^2(X_t)dt$. These facts play a central role in the proof of Theorem 2.1, where we have in fact established the following joint convergence: on $D_{\mathbb{R}^{d+d_1+3}}[0,\infty)$,*

$$\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\epsilon_{nj},\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\eta_{nj},\ x_{n,[nt]}, V_n^2, S_n\Big) \Rightarrow \Big(B_t, B_{1t}, X_t, \int_0^1 \widehat{H}^2(X_t)dt, S\Big). \tag{2.5}$$

We next consider a simple corollary of Theorem 2.1, illustrating the difference from previous works. More general application will be investigated in Section 3.

Let $\epsilon_i$ be iid random variables with $E\epsilon_1=0$ and $E\epsilon_1^2=1$. Let $x_{nj}=\frac{1}{\sqrt{n}}\sum_{k=1}^{j-1}\epsilon_k$ and write

$$S_{n1} = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} x_{nj}K(\epsilon_{j-1})\epsilon_j, \quad S_{n2} = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\big[x_{nj}+K(\epsilon_{j-1})\big]\epsilon_j,$$

where $K(y)$ is a real function satisfying that $|K(y)|\leq C\,|y|^\alpha$ for some $\alpha\geq 0$. By Theorem 2.1 with $u_{nj}=K(\epsilon_{j-1})$, $=1$ and $H(x,y)=xK(y)$ or $H(x,y)=x+K(y)$ respectively, we have the following corollary.

**Corollary 2.1.** *Suppose that $\mathbb{E}|\epsilon_1|^{\max\{2\alpha,2\}}<\infty$. Then, as $n\to\infty$, we have*

$$S_{n1} \to_D \mathbb{E}K(\epsilon_1)\int_0^1 B_t dB_t + \sigma_K\sqrt{\int_0^1 B_t^2 dt}\,\mathcal{N}, \tag{2.6}$$

$$S_{n2} \to_D \int_0^1 \big[B_t+\mathbb{E}K(\epsilon_1)\big]dB_t + \sigma_K\mathcal{N}, \tag{2.7}$$

*where $\sigma_K^2 = var\big[K(\epsilon_1)\big]$, $B_t$ is a standard Brownian motion and $\mathcal{N}$ is a standard normal variate independent of $B=\{B_t\}_{t\geq 0}$.*

If $\mathbb{E}K(\epsilon_1) = 0$, the limiting distribution of $S_{n1}$ is mixture normal, which is similar to that provided in recent works such as **?**, **??** and **???**. When $\mathbb{E}K(\epsilon_1) \neq 0$, the limiting distributions of $S_{n1}$ and $S_{n2}$ involve a functional of stochastic integrals, which seems new to the literature.

# 3 A regression application

This section considers nonlinear cointegrating regression model having the following form:

$$y_t = f(x_t, \Delta x_t, ..., \Delta x_{t-d},\ \theta) + u_t, \quad t = d, d+1, ..., \tag{3.8}$$

for some $d \geq 0$, where $f(.; ..., \theta)$ is a given real function indexed by $\theta = (\theta_1, ..., \theta_q)$, a vector of unknown parameters, $x_t$ is an integrated regressor, $\Delta x_k = x_k - x_{k-1}, k = t, ..., t-d$ denotes the lags of the integrated regressor $x_t$, and $(u_t, \mathcal{F}_t)_{t\geq 1}$ is a sequence of martingale differences such that $x_t$ is adapted to $\mathcal{F}_{t-1}$. Let $\Theta$ be a compact set of $R^q$ and assume the unknown parameters $\theta \in \Theta$. The least squares estimator (LSE) $\hat{\theta}_n$ of $\theta$, an interior of $\Theta$, is defined by

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \big[y_t - f(x_t, \Delta x_t, ..., \Delta x_{t-d},\ \theta)\big]^2.$$

The asymptotics of $\hat{\theta}_n$ without lags $\Delta x_t, ..., \Delta x_{t-d}$ were initially considered in **?** and then later by **?**, **?**, and **?**. More recently, **?** investigated the asymptotics by allowing for a wide class of commonly used volatility models such as GARCH, time-varying GARCH, and nonlinear GARCH. See also **?** for the related works on weighted nonlinear cointegrating regression. Using Theorem 2.1, this section will explore the asymptotics of $\hat{\theta}_n$ by allowing for certain lags $\Delta x_t, ..., \Delta x_{t-d}$ in the model (3.8).

The remainder of this section is organized as follows. Section 3.1 presents the assumptions on $x_t$ and $u_t$, which are similar to those in **?** and encompass widely used volatility models. Section 3.2 presents the asymptotic results.

## 3.1 Assumptions on $x_t$ and $u_t$

Throughout the section, let $\lambda_i \equiv (\epsilon_{i-1}, \nu_i), i \in \mathbb{Z}$, be a sequence of i.i.d. random vectors with $E\lambda_0 = 0$, $E\epsilon_0^2 = E\nu_0^2 = 1$ and $\rho = E\epsilon_0\nu_1$, and $\{\lambda_i^*\}_{i\in\mathbb{Z}}$ be an independent copy of $\{\lambda_i\}_{i\in\mathbb{Z}}$. We make use of the following assumptions on $x_t$ and $u_t$ in the asymptotic development.

**Assumption 3.1.** *$x_t = x_{t-1} + \xi_t$, where $\xi_t = \sum_{j=0}^{\infty} \phi_j \nu_{t-j}$. The coefficients $\phi_j, j \geq 0$, satisfy one of the following conditions:*

> ***LM.*** *$\phi_j \sim j^{-\mu} l(j), 1/2 < \mu < 1$ and $l(k)$ is a function slowly varying at $\infty$.*
>
> ***SM.*** *$\sum_{j=0}^{\infty} |\phi_j| < \infty$ and $\phi \equiv \sum_{j=0}^{\infty} \phi_j \neq 0$.*

**Assumption 3.2.** *$u_t = \sigma_t \epsilon_t$ with $\sigma_t^2 = \sigma(\lambda_t, \lambda_{t-1}, ...)$, where $\sigma(...)$ is a measurable function satisfying $\mathbb{E}\sigma^2(\lambda_0, \lambda_1, ...) < \infty$ and*

$$E\left|\sigma(\lambda_m, \lambda_{m-1}, ...) - \sigma(\lambda_m, \lambda_{m-1}, ..., \lambda_1, \lambda_0^*, \lambda_1^*, ...)\right|^2 \leq C\, m^{-\alpha}, \tag{3.9}$$

*for any $m \geq 1$ and some* $\alpha > \begin{cases} 4/(2\mu - 1), & \textit{under } \textbf{\textit{LM}}, \\ 4, & \textit{under } \textbf{\textit{SM}}. \end{cases}$

Assumption 3.1 allows for short memory (under **SM**) and long memory (under **LM**) innovations driving the integrated regressor $x_k = \sum_{j=1}^{k} \xi_j$, which is quite general in practice. Define

$$d_n^2 \;=\; \mathbb{E}|\sum_{k=1}^{n} \xi_k|^2 \;\sim\; \begin{cases} c_\mu\, n^{3-2\mu} l^2(n), & \text{under } LM, \\ \phi^2\, n, & \text{under } SM, \end{cases} \tag{3.10}$$

where $c_\mu$ is a constant. Standard functional limit theory (see **?** or Theorem 2.21 in **?** with a minor modification) shows that

$$\begin{aligned} &\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i,\ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \nu_i,\ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \nu_{-i},\ \frac{1}{d_n} x_{[nt]}\Big) \\ &\Rightarrow\ \big(U_{1t},\ U_t,\ U_{-t},\ X_t\big), \end{aligned} \tag{3.11}$$

on $D_{\mathbb{R}^4}[0, \infty)$, where $(U_{1t}, U_t)_{t\geq 0}$ is a bivariate Brownian motion with covariance matrix: $\Omega = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $\{U_{-t}\}_{t\geq 0}$ is an independent copy of $\{U_t\}_{t\geq 0}$ and $X_t$ is defined by

$$X_t \;=\; \begin{cases} U_{3/2-\mu}(t), & \text{under } \textbf{LM}, \\ U_{1/2}(t), & \text{under } \textbf{SM}, \end{cases}$$

where $U_H(t)$ is a fractional Brownian motion that has the following representation: with $a_+ = \max\{a, 0\}$,

$$U_H(t) \;=\; \frac{1}{\Gamma(H+1/2)} \int_{-\infty}^{t} (t-s)_+^{H-1/2} - (-s)_+^{H-1/2} dU_s.$$

In terms of (3.11), there exists a martingale vector difference array $\{(\eta_{1j}, \eta_{2j}), \mathcal{F}_j\}_{j\geq 1}$, where $\mathcal{F}_j = \sigma(\lambda_i, i \leq j)$ so that $\mathbb{E}\big(\epsilon_j \eta_{1j} \mid \mathcal{F}_{j-1}\big) = \mathbb{E}\big(\epsilon_j \eta_{2j} \mid \mathcal{F}_{j-1}\big)$ 0 for all $j \geq 1$ and, on $D_{\mathbb{R}^4}[0, \infty)$,

$$\Big(\frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \epsilon_j,\ \frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \eta_j,\ \frac{1}{d_n} x_{[nt]}\Big) \;\Rightarrow\; \big(B_t, B_{1t}, X_t\big), \tag{3.12}$$

where $(B_t, B_{1t}), t \geq 0$, is a 3−dimensional standard Brownian motion and $X = \{X_t\}_{t\geq 0}$ is a continuous process adapted to the filtration $\mathcal{F} = \{\sigma(B_s, B_{1s}, s \leq t)\}_{t\geq 0}$. Indeed, if we take $\eta_{2j} = \nu_{-j}$ and $\eta_{1j} = \big(\rho\epsilon_j - \nu_j\big)/(1-\rho)$, it is routine to see that

$$\mathbb{E}\big(\epsilon_j \eta_{1j} \mid \mathcal{F}_{j-1}\big) = \mathbb{E}\big(\epsilon_j \eta_{2j} \mid \mathcal{F}_{j-1}\big) = \mathbb{E}\big(\epsilon_j \eta_{1j}\big) = 0$$

and (3.12) is true by (3.11) and the continuous mapping theorem.

Under the settings of Assumption 3.2, we have that $(u_k, \mathcal{F}_k)_{k\geq 1}$, where $\mathcal{F}_k$ is an $\sigma$-field generated by $\lambda_{k+1}, \lambda_k, ...$, forms a martingale difference with $E(u_k^2 | \mathcal{F}_{k-1}) = \sigma_k^2 = \sigma(\lambda_k, \lambda_{k-1}, ...)$. The error process $u_k$ having a martingale difference structure that is similar to that used in **?**, indicating that (3.8) allows for a wide class of models of heteroscedasticity such as GARCH and nonlinear GARCH models. The following exam from **?** provides an illustration. More related results can be found in **?**, **?**, **?**, and **?**.

**Example 3.1.** *(Nonlinear GARCH model) Let $u_t = \sigma_t^{1/2}\epsilon_t$, where $\sigma_t$ is defined by*

$$\sigma_t = \alpha_0 + \alpha\, F(\sigma_{t-1}) + \beta\sigma_{t-1} + \gamma u_{t-1}^2.$$

*This is a nonlinear GARCH model introduced by* **?**. *Suppose that a constant $A_0$ exists such that $|F(x) - F(x')| \leq A_0|x - x'|$ for all $x, x' \in R$. Since we may rewrite $\sigma_t = \alpha_0 + R(\sigma_{t-1}, \epsilon_{t-1})$, where $R(y, \epsilon) = \alpha F(y) + \beta y + \gamma\epsilon^2 y$, it follows from Proposition 3.1 of* **?** *that $u_t$ satisfies Assumption 3.2 if $A_0 > 0, \alpha \geq 0, \beta \geq 0, \gamma \geq 0$ satisfying $\big[\mathbb{E}\big(\alpha A_0 + \beta + \gamma\epsilon_0^2\big)^2\big]^{1/2} < 1$.*

## 3.2 Asymptotics of $\hat{\theta}_n$

This section establishes the limiting distribution of $\hat{\theta}_n$ under certain regular conditions on $f(x, y, \theta)$. To this end, define $\dot{f}(x, y, \theta) = (\dot{f}_1, ..., \dot{f}_q)'$, where $\dot{f}_i = \frac{\partial f(x,y,\theta)}{\partial \theta_i}, i = 1, ..., q$, and write $p(x, y, \theta)$ for either $f(x, y, \theta)$ or one of $\dot{f}_i, i = 1, ..., q$. We need the following restrictions on $p(x, y, \theta)$.

**Assumption 3.3.** *There exist a real continuous function $T_p(x)$ and a constant $\beta \geq 0$ such that*

*(i) for each $\theta_1, \theta_2 \in \Theta$, $(x, y) \in \mathbb{R}^{1+d}$ and for some $0 < \alpha \leq 1$,*

$$|p(x, y, \theta_1) - p(x, y, \theta_2)| \leq ||\theta_1 - \theta_2||^\alpha T_p(x)(1 + ||y||^\beta); \tag{3.13}$$

*(ii) for any bounded $x$,*

$$p(lx, y, \theta) = v_p(l)\, h_p(x, y, \theta) + R_p(lx, y, \theta), \tag{3.14}$$

*where $v_p(l)$ is a positive real function that is bounded away from zero as $l$ is large, $h_p(x, y, \theta)$ for each $\theta \in \Theta$ is a continuous function with respect to $x$ satisfying $|h_p(x, y, \theta)| \leq C_x(1 + ||y||^\beta)$ and*

$$\sup_{\theta \in \Theta, y \in R} \frac{|R_p(lx, y, \theta)|}{T_p(lx)(1 + ||y||^\beta)} = o(1), \quad \text{as } l \to \infty.$$

*(iii) $T_p(lx) \leq v_p(l)\,(1 + |x|^\gamma)$ for some $\gamma > 0$, as $|lx| \to \infty$.*

To introduce the main result in this section, let $\theta_0$ be an interior of $\Theta$ and we make use of the following notations:

$$\begin{aligned}
\widetilde{h}_f(x, \theta) &= \mathbb{E} h_f(x, \xi_0, ..., \xi_{-d}; \theta), \qquad \widetilde{\sigma}_f(x, \theta) = \mathbb{E}\big[h_f(x, \xi_0, ..., \xi_{-d}; \theta)\sigma_0\big], \\
H_{\dot{f}}(x, \theta) &= \big(h_{\dot{f}_1}(x, \xi_0 \ldots \xi_{-d}; \theta), \,...,\, h_{\dot{f}_q}(x, \xi_0, ..., \xi_{-d}; \theta)\big), \\
\widetilde{h_{\dot{f}}}(x, \theta) &= \mathbb{E} H_{\dot{f}}(x, \theta), \qquad \widetilde{\sigma_{\dot{f}}}(x, \theta) = \mathbb{E}\big[H_{\dot{f}}(x, \theta) H_{\dot{f}}(x, \theta)' \sigma_0^2\big].
\end{aligned}$$

It is routine from Assumption 3.3 that these notations are well defined if

$$\mathbb{E}\big(|\xi_{-j}|^{2\beta} \sigma_0^2\big) < \infty, \quad \text{for } j = 0, -1, ..., -d. \tag{3.15}$$

**Theorem 3.1.** *In addition to Assumptions 3.1-3.3, suppose that (3.15) holds and*

$$\int_{|x| \leq \eta} \big[\widetilde{h}_f(x, \theta) - \widetilde{h}_f(x, \theta_0)\big]^2 dx \neq 0, \ \text{for any } \theta \neq \theta_0, \tag{3.16}$$

$$\int_{|x| \leq \eta} \widetilde{h_{\dot{f}}}(x, \theta_0)\, \widetilde{h_{\dot{f}}}'(x, \theta_0)\, dx \ \ \text{is a positive-definite matrix,} \tag{3.17}$$

*for each $\eta > 0$. Then, as $n \to \infty$,*

$$D_n\,(\hat{\theta}_n - \theta_0) \;\; \to_D \;\; \Big(\int_0^1 \Psi(t)\Psi(t)'dt\Big)^{-1}\Big[\int_0^1 \widetilde{G}_t\, dB_t + \Omega^{1/2}\mathcal{N}_q\Big], \tag{3.18}$$

*where $D_n = \sqrt{n}\, diag\big(v_{\dot{f}_1}(d_n), ..., v_{\dot{f}_q}(d_n)\big)$,*

$$\begin{aligned} \Psi(t) &= \widetilde{h_{\dot{f}}}(X_t, \theta_0), \quad \widetilde{G}_t = \widetilde{\sigma}_f(X_t, \theta_0), \quad \widehat{G}_t = \widetilde{\sigma_{\dot{f}}}(X_t, \theta_0) \\ \Omega &= \int_0^1 \big[\widehat{G}_t - \widetilde{G}_t\widetilde{G}_t'\big]dt \end{aligned}$$

*and $\mathcal{N}_q$ is an $q$-dimensional standard normal vector independent of $B$ and $X$.*

**Remark 3.1.** *As indicated in* **?** *and* **?**, *nonlinear cointegrating regressions with structures described in Theorem 3.1 are useful for modeling money demand functions. In such cases, $y_t$ is the logarithm of the real money balance, $x_t$ is the nominal interest rate, and $f$ can either be $f(x, \alpha, \beta) = \alpha + \beta \log|x|$ or $f(x, \alpha, \beta) = \alpha + \beta \log(\frac{1+|x|}{|x|})$. See* **?** *and* **?** *for empirical investigations of the estimation of money demand functions in the United States and Japan, respectively. See also* **?** *for the derivation of these functional forms from the underlying money demand theories studied in macroeconomics.*

# 4 Proofs

Except mentioned explicitly, we use the same notations as in previous sections. We start with the proof of Theorem 2.1. Three lemmas that support the proof of Theorem 2.1 are given in Subsections 4.3 – 4.5, respectively. The proof of Theorem 3.1 is given in Section 4.2. Without loss of generality, in the proof of Theorem 2.1, we assume that $d_1 = 1$ and $H(x, y)$ is continuous with respect to $x$. General extension to $d_1 > 1$ and the piecewise continuity of $H(x, y)$ is obvious from the proof with some notation changes, and hence the details are omitted.

## 4.1 Proof of Theorem 2.1

We first prove (2.4) under an additional restriction:

**R**: $H(x, y)$ is continuous with respect to $x$ and $H(x, y) = 0$ for all $|x| \geq N$, where $N \geq 1$ is a constant.

This restriction will be removed later. Note that, under this additional restriction, the continuous functions $\widetilde{H}(x)$ and $\widehat{H}(x)$ given in Assumption 2.3 can be chosen so that $\widetilde{H}(x)=\widehat{H}(x)=0$ for all $|x|\geq N$. This fact will be used in the proof and the following three lemmas without further explanation. Write

$$S_{1n} \;=\; \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\widetilde{H}(x_{nj})\,\epsilon_{nj} \quad \text{and} \quad S_{2n}=\frac{1}{\sqrt{n}}\sum_{j=1}^{n}K_j(x_{nj},u_{nj})\,\epsilon_{nj},$$

where $K_j(x,y)=H(x,y)-\widetilde{H}(x)$. It suffices to show that

$$\big(S_{1n},\ S_{2n}\big) \;\to_D\; \Big(\int_0^1 \widetilde{H}(X_t)dB_t,\ \sigma(X)\,\mathcal{N}\Big), \tag{4.19}$$

where $\mathcal{N}$ is a standard normal variate independent of $(B_t,X_t)_{t\geq 0}$ or, equivalently, independent of $(B_t,B_{1t})_{t\geq 0}$ since $X=\{X_t\}_{t\geq 0}$ is adapted to the filtration $\mathrm{F}=\{\sigma(B_s,B_{1s},s\leq t)\}_{t\geq 0}$. To this end, for $j,k=1,2,3,...,n$, let $y_{nj}=\frac{1}{\sqrt{n}}K_j(x_{nj},u_{nj})\,\epsilon_{nj}$,

$$V_{nk}^2=\frac{1}{n}\sum_{j=1}^{k}K_j^2(x_{nj},u_{nj}),\quad \widetilde{V}_{nk}^2=\frac{1}{n}\sum_{j=1}^{k}\big(\widehat{H}^2(x_{nj})-\big[\widetilde{H}(x_{nj})\big]^2\big).$$

Suppose that $l_j, j\geq 1$, are iid $N(0,1)$ random variables and $\{l_j\}_{j\geq 1}$ is independent of all other variables $\epsilon_{nj}, u_{nj}$, and $x_{nj}$. We further write

$$\widehat{y}_{nj}=\begin{cases} y_{nj}, & \text{if } j\leq n,\\ \frac{1}{\sqrt{n}}l_i, & \text{if } j>n,\end{cases}\qquad \widehat{V}_{nk}^2 \;=\; \begin{cases} V_{nk}^2, & \text{if } k\leq n,\\ V_{nn}^2+\frac{1}{n}\sum_{j=n+1}^{k}l_j^2, & \text{if } k>n,\end{cases}$$

and $\tau_n(t)=\inf\{k:\widehat{V}_{nk}^2\geq t\}$ for $t>0$. We mention that such a definition of $\widehat{V}_{nk}^2$ ensures that $\widehat{V}_{nk}^2\to\infty, a.s.$ for each fixed $n$. We introduce the following three lemmas. Their proofs will be given in Subsections 4.3 – 4.5, respectively.

**Lemma 4.1.** *If, in addition to Assumptions 2.1 - 2.3, $H(x,y)=0$ for all $|x|\geq N$ where $N\geq 1$ is a constant, then*

$$\max_{1\leq j\leq n}|y_{nj}| \;=\; o_P(1). \tag{4.20}$$

**Lemma 4.2.** *If, in addition to Assumptions 2.1 - 2.3, $H(x,y)=0$ for all $|x|\geq N$ where $N\geq 1$ is a constant, then*

$$\begin{aligned}&\frac{1}{\sqrt{n}}\max_{1\leq k\leq n}\Big|\Big|\sum_{j=1}^{k}\mathbb{E}\big(\xi_{nj}y_{nj}|\mathcal{F}_{n,j-1}\big)\Big|\Big|\\ =\;& \frac{1}{n}\max_{1\leq k\leq n}\Big|\sum_{j=1}^{k}\big[H(x_{nj},u_{nj})-\widetilde{H}(x_{nj})\big]\Big|=o_P(1),\end{aligned}\tag{4.21}$$

*where* $\xi_{nj} = \big(\epsilon_{nj},\ \eta_{nj}\big)$, *and*

$$\max_{1\le k\le n} \big|V_{nk}^2 - \widetilde{V}_{nk}^2\big| \quad = \quad o_P(1). \tag{4.22}$$

**Lemma 4.3.** *If, in addition to Assumptions 2.1 - 2.3, $H(x,y)=0$ for all $|x|\ge N$ where $N\ge 1$ is a constant, then*

$$\sum_{j=1}^{\tau_n(t)} \widehat{y}_{nj} \quad\Rightarrow\quad B_t' \quad \text{on } D_R[0,\infty), \tag{4.23}$$

*where $B'=\{B_t'\}_{t\ge 0}$ is a standard Brownian motion. Furthermore, on $D_{\mathbb{R}^4}[0,\infty)$, we have*

$$\begin{aligned} Z_n(t) \quad &:= \quad \Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\epsilon_{nj},\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\eta_{nj},\ x_{n,[nt]},\ \sum_{j=1}^{\tau_n(t)}\widehat{y}_{nj}\Big) \\ &\Rightarrow \quad \big(B_t,\ B_{1t},\ X_t,\ B_t'\big), \end{aligned} \tag{4.24}$$

*where $B'$ is independent of $X=\{X_t\}_{t\ge 0}$ and $\mathrm{F}=\{\sigma(B_s,B_{1s},s\le t)\}_{t\ge 0}$.*

We now turn back to the proof of (4.19). In fact, in terms of (4.24), the same arguments as in **?** (also, see, **?**) yield that

$$\begin{aligned} &\Big(\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\epsilon_{nj},\ \frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\eta_{nj},\ x_{n,[nt]},\ S_{1n},\ \widetilde{V}_{nn}^2,\ \frac{1}{\sqrt{n}}\sum_{j=1}^{\tau_n(t)}\widehat{y}_{nj}\Big) \\ \Rightarrow\ &\Big(B_t,\ B_{1t},\ X_t,\ \int_0^1 \widetilde{H}(X_t)dB_t,\ \sigma^2(X),\ B_t'\Big). \end{aligned} \tag{4.25}$$

Write $\tau_n^{-1}(k)=\inf\{t\ge 0:\tau_n(t)\ge k\}$. Note that

$$\widehat{H}^2(x_{nn}) + \big[\widetilde{H}(x_{nn})\big]^2 \to_D \widehat{H}^2(X_1) + \big[\widetilde{H}(X_1)\big]^2 = O_P(1),$$

by using $x_{n,[nt]}\Rightarrow X_t$ given Assumption 2.1, It follows from (4.22) that

$$\begin{aligned} |\tau_n^{-1}(n) - \widetilde{V}_{nn}^2| \quad &\le \quad |\tau_n^{-1}(n) - \widehat{V}_{nn}^2| + |V_{nn}^2 - \widetilde{V}_{nn}^2| \\ &\le \quad V_{nn}^2 - V_{n,n-1}^2 + \frac{1}{n}l_{n+1}^2 + |V_{nn}^2 - \widetilde{V}_{nn}^2| \\ &\le \quad 3\max_{1\le k\le n}\big|V_{nk}^2 - \widetilde{V}_{nk}^2\big| + \frac{1}{n}\big(\widehat{H}^2(x_{nn}) - \big[\widetilde{H}(x_{nn})\big]^2\big) + \frac{1}{n}l_{n+1}^2 \\ &\to_{\mathrm{P}} \quad 0. \end{aligned} \tag{4.26}$$

This, together with (4.25), yields that

$$\Big(S_{1n},\ \tau_n^{-1}(n),\ \ W_n(t)\Big) \quad\Rightarrow\quad \Big(\int_0^1 \widetilde{H}(X_t)dB_t,\ \sigma^2(X),\ B_t'\Big), \tag{4.27}$$

on $D_{\mathrm{R}^3}[0,\infty)$, where $W_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{\tau_n(t)} \widehat{y}_{nj}$. Let $D_0$ be the subset of all functions $V = \{V(t)\}_{t\geq 0}$ in $D_{\mathrm{R}}[0,\infty)$ that are nondecreasing with $V(0)\geq 0$. Since $W_n(t)\in D_R[0,\infty)$ and the composition map $\psi(U,V)$ from $D_{\mathrm{R}}[0,\infty)\times D_0$ to $D_{\mathrm{R}}[0,\infty)$ defined by

$$\psi(U,V) := U(V(t)), \quad t\geq 0,$$

is continuous at $(B', \sigma(X))$ (see, **?**, for instance), it follows from (4.27) and the continuous mapping theorem that

$$\begin{aligned}
&\left(S_{1n},\ \tau_n^{-1}(n),\ W_n\big[\tau_n^{-1}(n)\big]\right) \\
\to_D\ &\Big(\int_0^1 \widetilde{H}(B_{1t})dB_t,\ \sigma^2(X),\ B'_{\sigma^2(X)}\Big) =_D \Big(\int_0^1 \widetilde{H}(B_{1t})dB_t,\ \sigma^2(X),\ \sigma(X)\mathcal{N}\Big),
\end{aligned}$$

where $\mathcal{N}$ is a standard normal variate independent of $\{B_t, X_t\}_{t\geq 0}$ and we have used the fact that $X = \{X_t\}_{t\geq 0}$ is adapted to $F = \{\sigma(B_s, B_{1s}), s\leq t\}_{t\geq 0}$ and $B'$ is independent of $F$. Now, by noting

$$\Big|S_{2n} - \frac{1}{\sqrt{n}}\sum_{j=1}^{\tau_n[\tau_n^{-1}(n)]} \widehat{y}_{nj}\Big| \leq \big|y_{nn}\big| \to_{\mathrm{P}} 0$$

owe to (4.20), we have

$$\begin{aligned}
&\Big(S_{1n},\ \tau_n^{-1}(n),\ S_{2n}\Big) = \left(S_{1n},\ \tau_n^{-1}(n),\ W_n\big[\tau_n^{-1}(n)\big]\right) + o_P(1) \\
\to_D\ &\Big(\int_0^1 \widetilde{H}(B_{1t})dB_t,\ \sigma^2(X),\ \sigma(X)\mathcal{N}\Big),
\end{aligned}$$

where $\mathcal{N}$ is a standard normal variate independent of $\{B_t, X_t\}_{t\geq 0}$. This proves (4.19) and hence (2.4) under the additional restriction **R.**

To remove this restriction **R**, for any $K>0$, we assume that $H_K(x,y) = H(x,y)\xi_K(x)$, $\widetilde{H}_K(x) = \widetilde{H}(x)\xi_K(x)$ and $\widehat{H}_K(x) = \widehat{H}(x)\xi_K(x)$ with

$$\xi_K(x) = \begin{cases} 1, & |x|\leq K, \\ 2-|x|/K, & K<|x|<2K, \\ 0, & |x|\geq 2K. \end{cases}$$

It is readily seen that, for each $K>0$, $H_K(x,y)$ is a continuous function with respect to $x$, satisfying Assumption 2.2, $H(x,y)=0$ for all $|x|\geq N=2K$, and Assumption 2.3 with $\widetilde{H}(x)$ and $\widehat{H}(x)$ being replaced by $\widetilde{H}_K(x)$ and $\widehat{H}_K(x)$, respectively. Therefore, it follows from the proof above that, for each $K>0$,

$$S_{nK} \ := \ \frac{1}{\sqrt{n}}\sum_{j=1}^{n} H_K(x_{nj}, u_{nj})\,\epsilon_{nj} \ \to_D \int_0^1 \widetilde{H}_K(X_t)dB_t + \sigma_K(X)\mathcal{N},$$

where $\sigma_K^2(X) = \int_0^1 \big[\widehat{H}_K^2(X_t) - \widetilde{H}_K^2(X_t)\big]dt$ and $\mathcal{N}$ is a standard normal variate independent of $B = \{B_t\}_{t\ge 0}$ and $X = \{X_t\}_{t\ge 0}$. This yields (2.4) since $\mathbb{P}(S_n \neq S_{nK}) \le \mathbb{P}\Big(\max_{1\le k\le n} |x_{nk}| > K\Big) \to 0$ and

$$
\begin{aligned}
&\mathbb{P}\Big(\int_0^1 \widetilde{H}(X_t)dB_t + \sigma(X)\mathcal{N} \neq \int_0^1 \widetilde{H}_K(X_t)dB_t + \sigma_K(X)\mathcal{N}\Big) \\
\le\;& \mathbb{P}\Big(\sup_{0\le t\le 1} |X_t| > K\Big) \to 0,
\end{aligned}
$$

as $n \to \infty$ first and then $K \to \infty$. The proof of Theorem 2.1 is complete. □

## 4.2 Proof of Theorem 3.1

We start with two lemmas on the asymptotics of general nonlinear LS estimators. Consider a nonlinear regression model having the form:

$$
y_k = g_k(\theta) + u_k, \quad k = 1, 2, ..., n, \tag{4.28}
$$

where $\theta = (\theta_1, ..., \theta_q)$ is a vector of unknown parameters, $\{u_k, \mathcal{F}_k\}_{k\ge 1}$ is a martingale difference with $E(u_k^2 \mid \mathcal{F}_{k-1}) < \infty, a.s.$ and $g_k(\theta)$ for each $\theta \in \Theta$, a compact of $\mathbb{R}^q$, is adapted to $\mathcal{F}_{k-1}$. Let $\widehat{\theta}_n$ be the LSE of $\theta$ and $\theta_0$ be an interior of $\Theta$. The following consistency and asymptotic normality results come from **?**. See also **?**.

**Lemma 4.4.** *Suppose that there exist a sequence of constants $0 < k_n \to \infty$ and a sequence of random variables $T_k$ (adapted to $\mathcal{F}_{k-1}$) such that*

*(a) $|g_k(\theta_1) - g_k(\theta_2)| \le ||\theta_1 - \theta_2||^\alpha\, T_k$ for all $\theta_1, \theta_2 \in \Theta$ and for some $\alpha > 0$;*

*(b) $\frac{1}{k_n}\sum_{k=1}^n T_k^2 = O_P(1)$ and $\sum_{k=1}^n T_k^2\big[1 + E(u_k^2 \mid \mathcal{F}_{k-1})\big] = O_P(k_n^2/\log^2 n)$;*

*(c) the finite dimensional distributions of $\frac{1}{k_n}\sum_{k=1}^n \big[g_k(\theta) - g_k(\theta_0)\big]^2$ converge to those of $G(\theta)$, where $G(\theta), \theta \in \Theta$ is a stochastic process of $\theta$ satisfying $P(\inf_{\substack{||\theta-\theta_0||\ge\delta \\ \theta\in\Theta}} G(\theta) > 0) = 1$ for each $\delta > 0$.*

*Then $||\widehat{\theta}_n - \theta_0|| = o_P(1)$, i.e., $\hat{\theta}_n$ is a consistent estimator of $\theta_0$.*

Let $\dot{g}_k(\theta) = \big(\frac{\partial g_k(\theta)}{\partial\theta_1}, ..., \frac{\partial g_k(\theta)}{\partial\theta_q}\big)'$ be the first derivative of $g_k(\theta)$, $Z_n(\theta) = (D_n^{-1})'\sum_{k=1}^n \dot{g}_k(\theta)\, u_k$ and

$$
Y_n = (D_n^{-1})' \sum_{k=1}^n \dot{g}_k(\theta_0)\dot{g}_k(\theta_0)'\, D_n^{-1},
$$

where $D_n = \mathrm{diag}(d_{1n}, ..., d_{qn})$ is a sequence of diagonal matrices satisfying $n^{-\delta}\, \min_{1\le j\le q} d_{jn} \to \infty$ for some $\delta > 0$.

**Lemma 4.5.** *Suppose that*

*(i)* $||D_n^{-1}\big[\dot{g}_k(\theta_1) - \dot{g}_k(\theta_2)\big]|| \le ||\theta_1 - \theta_2||^\alpha T_{nk}$ *for some* $0 < \alpha \le 1$ *and for any* $\theta_1, \theta_2 \in \Theta$*, where* $T_{nk}$ *is adapted to* $\mathcal{F}_{k-1}$ *for each* $n \ge 1$*, satisfying*

$$\sum_{k=1}^n T_{nk}^2\big[1 + E(u_k^2|\mathcal{F}_{k-1})\big] + \sum_{k=1}^n T_{nk}^2 = O_P(1); \tag{4.29}$$

*(ii)* $\widehat{\theta}_n \to_P \theta_0$*,* $Z_n(\theta_0) = O_P(1)$ *and* $Y_n \to_D M$*, where* $M > 0$*, a.s., i.e., the smallest eigenvalue of* $M$ *is almost surely positive.*

*We have*

$$D_n(\hat{\theta}_n - \theta_0) = Y_n^{-1} Z_n(\theta_0) + o_P(1). \tag{4.30}$$

*If in addition* $(Y_n, Z_n) \to_D (M, Z)$ *where* $M > 0$*, a.s., then* $D_n(\hat{\theta}_n - \theta_0) \to_D M^{-1} Z$*.*

We turn back to the proof of Theorem 3.1 by checking the conditions of Lemma 4.5 with

$$g_k(\theta) := f(x_k, \Delta x_k, ..., \Delta x_{k-d},\ \theta) = f(x_k, \xi_k, ..., \xi_{k-d},\ \theta).$$

We first show that $\widehat{\theta}_n \to_P \theta_0$ by using Lemma 4.4 with $k_n = n v_f^2(d_n)$. In this case, it only needs to show the following results: for any $\theta_j \in \Theta$ and $\alpha_j \in R$, $j = 1, 2, ..., q$,

$$\sum_{j=1}^q \alpha_j \frac{1}{n v_f^2(d_n)} \sum_{k=1}^n \big[g_k(\theta_j) - g_k(\theta_0)\big]^2 \ \to_D\ \sum_{j=1}^q \alpha_j\, G(\theta_j), \tag{4.31}$$

where $G(\theta) := \int_0^1 \big[\widetilde{h}_f(X_t, \theta) - \widetilde{h}_f(X_t, \theta_0)\big] dt$ and

$$\frac{1}{n} \sum_{k=1}^n \big[1 + (|x_k|/d_n)^{2\gamma}\big](1 + ||\Gamma_k||^{2\beta})(1 + \sigma_k^2) = O_P(1), \tag{4.32}$$

where $\Gamma_k = (\xi_k, ..., \xi_{k-d})$. Indeed, due to the continuity of $X_t$, it follows from (3.16) that
$P\big[\inf_{\substack{||\theta-\theta_0|| \ge \delta \\ \theta \in \Theta}} G(\theta) > 0\big] = 1$. This, together with (4.31), implies the condition (c) of Lemma 4.4. On the other hand, the conditions (a) and (b) of Lemma 4.4 follow easily from Assumption 3.3[(i) and (iii)] with $T_k = T_f(x_k)(1 + ||\Gamma_k||^\beta)$ and (4.32). We therefore have $\widehat{\theta}_n \to_P \theta_0$.

We only prove (4.31), where the idea is similar to that of **?**. Result (4.32) can be obtained by (2.5) with $H(x, y) = (1 + |x|^{2\gamma})(1 + ||y||^{2\beta})$ and hence the details are omitted. Write $x_t^* = x_t I(|x_t|/d_n \le A)$,

$$
\begin{aligned}
g_k^*(\theta) &= f(x_k^*, \xi_k, ..., \xi_{k-d},\ \theta), \\
R_n(\theta) &= \sum_{k=1}^n \big[g_k(\theta) - v_f(d_n) h_f(x_k/d_n, \xi_k, ..., \xi_{k-d}, \theta)\big]^2, \\
R_n^*(\theta) &= \sum_{t=1}^n \big[g_k^*(\theta) - v_f(d_n) h_f(x_t^*/d_n, \xi_k, ..., \xi_{k-d}, \theta)\big]^2.
\end{aligned}
$$

For any fixed $A > 0$, it follows from Assumption 3.3(ii) and (iii) that, as $n \to \infty$,

$$
\begin{aligned}
\sup_{\theta\in\Theta} |R_n^*(\theta)| &\le o(1)\, v_f^2(d_n) \sum_{t=1}^n T_f^2(x_t^*/d_n)(1 + ||\Gamma_k||^{2\beta}) \\
&= o(1) \sum_{t=1}^n (1 + |x_t^*/d_n|^{2\gamma})(1 + ||\Gamma_k||^{2\beta}) = o_P(1)\, n v_f^2(d_n).
\end{aligned}
$$

This implies that, for any $\epsilon > 0$,

$$
\begin{aligned}
& P\big(\frac{1}{n v_f^2(d_n)} \sup_{\theta\in\Theta} |R_n(\theta)| \ge \epsilon\big) \\
\le\ & P\big(x_k \ne x_k^*, \text{for some k=1,...,n}\big) + P\big(\frac{1}{n v_f^2(d_n)} \sup_{\theta\in\Theta} |R_n^*(\theta)| \ge \epsilon\big) \\
\le\ & P\big(\max_{1\le k\le n} |x_k|/d_n \ge A\big) + P\big(\frac{1}{n v_f^2(d_n)} \sup_{\theta\in\Theta} |R_n^*(\theta)| \ge \epsilon\big) \\
\to\ & 0,
\end{aligned}
$$

as $n \to \infty$ first and then $A \to \infty$, namely, we have

$$
\sup_{\theta\in\Theta} |R_n(\theta)| = o_P\big[n v_f^2(d_n)\big]. \tag{4.33}
$$

Now, by letting

$$
\begin{aligned}
\Delta_n &= \sup_{\theta\in\Theta} |R_n(\theta)| + \sup_{\theta\in\Theta} |R_n(\theta)|^{1/2} \\
&\qquad \Big(\frac{1}{n} \sum_{k=1}^n \big[h_f(x_k/d_n, \xi_k, ..., \xi_{k-d}, \theta) - h_f(x_k/d_n, \xi_k, ..., \xi_{k-d}, \theta_0)\big]^2\Big)^{1/2},
\end{aligned}
$$

and noting that, for any $\theta \in \Theta$,

$$
\begin{aligned}
& \frac{1}{n v_f^2(d_n)} \sum_{k=1}^n \big[g_k(\theta) - g_k(\theta_0)\big]^2 \\
=\ & \frac{1}{n} \sum_{k=1}^n \big[h_f(x_k/d_n, \xi_k, ..., \xi_{k-d}, \theta) - h_f(x_k/d_n, \xi_k, ..., \xi_{k-d}, \theta_0)\big]^2 + 4\,\widetilde{\Delta}_n,
\end{aligned}
$$

where $|\widetilde{\Delta}_n| \le \Delta_n$, result (4.31) follows from (4.33) and (2.5).

We now turn back to the proof of Theorem 3.1 by verifying other conditions of Lemma 4.5. First of all, by letting

$$T_{nk}=\frac{1}{\sqrt{n}}\sum_{j=1}^{q}v_{\dot f_j}^{-1}(d_n)T_{\dot f_j}(x_k)(1+||u_k||^{\beta}),$$

it follows easily from Assumption 3.3 (iii) and (4.32) that

$$\begin{aligned}&\sum_{k=1}^{n}T_{nk}^2\big[1+E(u_k^2|\mathcal{F}_{k-1})\big]+\sum_{k=1}^{n}T_{nk}^2\\ \le\;& \frac{1}{n}\sum_{t=1}^{n}\big[1+(|x_t|/d_n)^{2\gamma}\big](1+||u_t||^{2\beta})(1+\sigma_t^2)=O_P(1).\end{aligned}$$

This yields (4.29). So, together with $\widehat{\theta}_n\to_P\theta_0$ that is proved above, it remains to show that, for any $\alpha_i'=(\alpha_{i1},...,\alpha_{iq})\in\mathbb{R}, i=1,2,3,$

$$\begin{aligned}&(\alpha_1'\,Y_n\alpha_2,\quad\alpha_3'Z_n)\\ &\to_D\;\Big(\alpha_1'\int_0^1\Psi(t)\Psi(t)'\,dt\,\alpha_2,\;\alpha_3'\,Z\Big),\end{aligned}\tag{4.34}$$

where $Z:=\int_0^1\widetilde{G}_t\,dB_t+\Omega^{1/2}\mathcal{N}_q$ and, with $D_n=\sqrt{n}\,\mathrm{diag}\big(v_{\dot f_1}(d_n),...,v_{\dot f_q}(d_n)\big)$,

$$Y_n=(D_n^{-1})'\sum_{k=1}^{n}\dot g_k(\theta_0)\dot g_k(\theta_0)'\,D_n^{-1},\quad Z_n=(D_n^{-1})'\sum_{k=1}^{n}\dot g_k(\theta_0)\,\sigma_k\epsilon_k,.$$

Write $h_{\dot f}(x,y,\theta)=\big(h_{\dot f_1}(x,y;\theta),\;...,\;h_{\dot f_q}(x,y;\theta)\big)$ and

$$\begin{aligned}\widetilde{Y}_n&=\frac{1}{n}\sum_{k=1}^{n}\alpha_1'\,h_{\dot f}(x_k/d_n,\xi_k,...,\xi_{k-d},\theta_0)h_{\dot f}(x_k/d_n,\xi_k,...,\xi_{k-d},\theta_0)'\,\alpha_2\,,\\ \widetilde{Z}_n&=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\alpha_3'\,h_{\dot f}(x_k/d_n,\xi_k,...,\xi_{k-d},\theta_0)\,\sigma_k\epsilon_k.\end{aligned}$$

Note that, by (2.5) in Remark 2.3,

$$(\widetilde{Y}_n,\quad\widetilde{Z}_n)\;\;\to_D\;\;\Big(\alpha_1'\int_0^1\Psi(t)\Psi(t)'\,dt\,\alpha_2,\;\alpha_3'\,Z\Big).$$

and

$$\begin{aligned}\widetilde{Y}_n&=\frac{1}{n}\sum_{k=1}^{n}\alpha_1'\,h_{\dot f}(x_k/d_n,\xi_k,...,\xi_{k-d},\theta_0)h_{\dot f}(x_k/d_n,\xi_k,...,\xi_{k-d},\theta_0)'\,\alpha_2\,,\\ \widetilde{Z}_n&=\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\alpha_3'\,h_{\dot f}(x_k/d_n,\xi_k,...,\xi_{k-d},\theta_0)\,\sigma_k\epsilon_k.\end{aligned}$$

Result (4.34) will follow if we prove: for all $i, j = 1, ..., q$,

$$
\begin{aligned}
R_{1n}(i,j) &:= \frac{1}{n v_{\dot f_i}(d_n) v_{\dot f_j}(d_n)} \sum_{k=1}^n \big[\dot f_i(...)\dot f_j(...) - v_{\dot f_i}(d_n) v_{\dot f_j}(d_n) h_{\dot f_i}(...) h_{\dot f_j}(...)\big] \\
&= o_P(1), \qquad (4.35)\\
R_{2n}(j) &:= \frac{1}{\sqrt{n} v_{\dot f_j}(d_n)} \sum_{k=1}^n \big[\dot f_j(...) - v_{\dot f_j}(d_n) h_{\dot f_j}(...)\big]\sigma_k\,\epsilon_k \;=\; o_P(1). \qquad (4.36)
\end{aligned}
$$

We only prove (4.36). The proof of (4.35) is similar to that of (4.33) and hence the detials are omitted. To this end, for any $A > 0$, let

$$
R_{2n}^*(j) = \frac{1}{\sqrt{n} v_{\dot f_j}(d_n)} \sum_{k=1}^n \big[\dot f_j(...) - v_{\dot f_j}(d_n) h_{\dot f_j}(...)\big] I(|x_k|/d_n \le A)\sigma_k\,\epsilon_k.
$$

Note that, as $n \to \infty$ first and then $A \to \infty$,

$$
P(R_{2n}(j) \ne R_{2n}^*(j)) \;\le\; P\big(\max_{1\le k\le n} |x_k|/d_n \ge A\big) \to 0.
$$

It suffices to show that, for each $A > 0$, $R_{2n}^*(j) = o_P(1)$. This follows easily from the following fact: in term of Assumption 3.3 (ii) and (iii),

$$
\begin{aligned}
\mathbb{E}\big[R_{2n}^*(j)\big]^2 &= \frac{1}{n v_{\dot f_j}^2(d_n)} \sum_{k=1}^n \mathbb{E}\big\{\big[\dot f_j(...) - v_{\dot f_j}(d_n) h_{\dot f_j}(...)\big]^2 I(|x_k|/d_n \le A)\sigma_k^2\big\} \\
&\le \sup_{\theta\in\Theta, y\in R} \frac{|R_{\dot f_j}(d_n\,x, y, \theta)|}{T_{\dot f_j}(d_n x)(1+||y||^\beta)}\, \frac{1+A}{n} \sum_{k=1}^n \mathbb{E}\big[(1+||\Gamma_k||^{2\beta})\sigma_k^2\big] \to 0,
\end{aligned}
$$

as $n \to \infty$, where $\Gamma_k = (\xi_k, ..., \xi_{k-d})$. The proof of Theorem 3.1 is complete. □

## 4.3 Proof of Lemma 4.1

Let $l_{nj} = g_N^2(u_{nj})$. Recall that $H(x, y) = 0$ for all $|x| \ge N$, implying $\widetilde H(x)$ is bounded (by $C_0$, say). It follows from Assumption 2.2(b) that, for any $K > 0$,

$$
\begin{aligned}
\max_{1\le k\le n} |y_{nj}| &\le \frac{1}{\sqrt{n}} \max_{1\le k\le n} \big[\, g_N(u_{nj})\, |\epsilon_{nj}| I_{(|l_{nj}|\ge K)}\big] + \frac{C_0+K}{\sqrt{n}} \max_{1\le j\le n} |\epsilon_{nj}| \\
&:= R_{1n} + (C_0+K)\, R_{2n}. \qquad (4.37)
\end{aligned}
$$

Using the uniformality of $l_{nj}$ and $E(\epsilon_{nj}^2|\mathcal{F}_{n,j-1}) = 1$, we have

$$
\begin{aligned}
\mathbb{E}R_{1n} &\le \Big(\frac{1}{n}\sum_{j=1}^n \mathbb{E}\big[g_N^2(u_{nj})\,\epsilon_{nj}^2 I_{|l_{nj}|\ge K}\big]\Big)^{1/2} \\
&= \Big(\frac{1}{n}\sum_{j=1}^n \mathbb{E} l_{nj}^2 I_{|l_{nj}|\ge K}\Big)^{1/2} \to 0,
\end{aligned}
$$

as $n \to \infty$ first and then $K \to \infty$, i.e, $R_{1n} = o_P(1)$. On the other hand, it follows from $\frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \epsilon_{nj} \Rightarrow B_t$ on $D_R[0, \infty)$ that

$$R_{2n} = \frac{1}{\sqrt{n}} \max_{1 \le j \le n} |\epsilon_{nj}| = o_P(1).$$

See, Proposition VI 3.26 of Jacod and Shiryaev (2001), for instance. Taking these estimates into (4.37), we prove $\max_{1 \le j \le n} |y_{nj}| = o_P(1)$, as required in (4.20). The proof of Lemma 4.1 is now complete. □

## 4.4 Proof of Lemma 4.2

We start with some preliminaries. Since the limit process $X_t$ is path continuous, the convergence $x_{n,[nt]} \Rightarrow X_t$ on $D_{\mathbb{R}}[0, 1]$ can be understood in the sense of the uniform topology. It follows from this fact that

$$\limsup_{N \to \infty} \limsup_{n \to \infty} \Big[ \mathbb{P}\big( \max_{1 \le j \le n} |x_{nj}| \ge N \big) + \mathbb{P}\big( \max_{1 \le j \le n} |x_{nj}| \le 1/N \big) \Big] = 0. \tag{4.38}$$

Furthermore, by the tightness of $\{x_{n,[nt]}\}_{0 \le t \le 1}$, for any $\varepsilon > 0$ and $\delta > 0$, there is some $\tilde{\delta} = \tilde{\delta}(\varepsilon, \delta) > 0$ such that

$$\mathbb{P}\Big( \sup_{|s-t| \le \tilde{\delta}} |x_{n,[nt]} - x_{n,[ns]}| \ge \delta \Big) \le \varepsilon \tag{4.39}$$

for all sufficiently large $n$. In terms of (4.39), for any $\delta > 0$, we have

$$\lim_{n \to \infty} \mathbb{P}\Big( \max_{0 \le k \le n/m} \max_{km \le j \le (k+1)m} |x_{nj} - x_{n,km}| \ge \delta \Big) = 0, \tag{4.40}$$

for any $m := m_n \to \infty$ satisfying $n/m \to \infty$. On the other hand, it follows from (2.2) and the continuities of $H(x, y)$ and $\widetilde{H}(x)$ (with respect to $x$) that, as $m \to \infty$,

$$\max_{m \le k \le n-m} \mathbb{E} \sup_{|x| \le N} \Big| \frac{1}{m} \sum_{j=km+1}^{(k+1)m} H(x, u_{nj}) - \widetilde{H}(x) \Big| \to 0, \tag{4.41}$$

for any fixed $N > 0$.

We are now ready to prove (4.21). Let $H_1(x, y) = H(x, y) - \widetilde{H}(x)$. Note that $H(x, y)$ is uniformly continuous on $|x| \le N$ and $H_1(x, y) = 0$ for all $|x| \ge N$ and $y \in \mathbb{R}$. It follows that, for each $\epsilon > 0$, there exists an $\delta_\epsilon$ depending only on $N, \epsilon$ only such that

$$|H_1(x_1, y) - H_1(x_2, y)| \le |H(x_1, y) - H(x_2, y)| \le \epsilon g_N(y), \tag{4.42}$$

whenever $|x_1 - x_2| \le \delta_\epsilon$, where $g_N(y) = \sup_{|x|\le N} H(x,y)$. Write

$$\Omega_{\delta_\epsilon} = \Big\{\omega : \max_{0\le k\le n/m} \max_{km\le j\le (k+1)m} |x_{nj} - x_{n,km}| \le \delta_\epsilon\Big\}$$

and $T_n = [n/m] - 1$, where $m$ is chosen so that $m \to \infty$ and $n/m \to \infty$. In terms of (4.42), on $\Omega_{\delta_\epsilon}$, we have

$$\begin{aligned}
\widehat{R}_{n,1} &:= \frac{1}{n}\max_{0\le k\le n/m}\Big|\sum_{j=1}^{(k+1)m} H_1(x_{nj},u_{nj})\Big| \;\le\; \frac{1}{n}\sum_{k=0}^{T_n}\Big|\sum_{j=km+1}^{(k+1)m} H_1(x_{nj},u_{nj})\Big| \\
&\le \frac{1}{n}\sum_{k=0}^{T_n}\Big|\sum_{j=km+1}^{(k+1)m} H_1(x_{n,km},u_{nj})\Big| \\
&\qquad + \frac{1}{n}\sum_{k=0}^{T_n}\sum_{j=km+1}^{(k+1)m}\big|H_1(x_{nj},u_{nj}) - H_1(x_{n,km},u_{nj})\big| \\
&\le \frac{1}{n}\sum_{k=0}^{T_n}\sup_{|x|\le N}\Big|\frac{1}{m}\sum_{j=km+1}^{(k+1)m} H_1(x,u_{nj})\Big| + \frac{\epsilon}{n}\sum_{j=1}^{n} g_N(u_{nj}).
\end{aligned}$$

Therefore, for any $\eta, \eta_1 > 0$ and $\epsilon = \eta\,\eta_1$, it follows from Assumption 2.2(b), (4.40) and (4.41) that

$$\begin{aligned}
\mathbb{P}(|\widehat{R}_{n,1}| \ge \eta) &\le \mathbb{P}\big(\bar{\Omega}_{\delta_\epsilon}\big) + \mathbb{P}(|\widehat{R}_{n,1}| \ge \eta, \Omega_{\delta_\epsilon}) \\
&\le \mathbb{P}\big(\bar{\Omega}_{\delta_\epsilon}\big) + \eta^{-1}\max_{m\le k\le n-m}\mathbb{E}\sup_{|x|\le 2N}\Big|\frac{1}{m}\sum_{j=km+1}^{(k+1)m} H_{1N}(x,u_{nj})\Big| \\
&\qquad + \frac{\epsilon\,\eta^{-1}}{n}\sum_{j=1}^{n}\mathbb{E}g_N(u_{nj}) \;\le\; C\,\eta_1 + o(1),
\end{aligned}$$

where $\bar{\Omega}_{\delta_\epsilon}$ denotes the complementary set of $\Omega_{\delta_\epsilon}$. This proves $\widehat{R}_{1n,1} = o_P(1)$.

Similarly, by recalling $|H_1(x,y)| \le g_N(u_{nj})$, it follows from Assumption 2.2(b) that

$$\begin{aligned}
\widehat{R}_{n,2} &:= \frac{1}{n}\max_{0\le k\le n/m}\sum_{j=km+1}^{(k+1)m}\big|H_1(x_{nj},u_{nj})\big| \\
&\le \frac{1}{n}\max_{0\le k\le n/m}\sum_{j=km+1}^{(k+1)m} g_N(u_{nj}) \\
&\le \frac{m}{n}\max_{1\le j\le n} g_N(u_{nj}) = o_P(1).
\end{aligned}$$

In terms of these estimates, we have

$$
\begin{aligned}
& \frac{1}{n}\max_{1\le k\le n}\Big|\sum_{j=1}^{k}\big[H(x_{nj},u_{nj})-\widetilde{H}(x_{nj})\big]\Big| \\
\le\ & \frac{1}{n}\max_{0\le k\le n/m}\ \max_{km\le j\le (k+1)m}\Big|\sum_{k=1}^{j} H_1(x_{nj},u_{nj})\Big| \\
\le\ & \frac{1}{n}\max_{0\le k\le n/m}\Big|\sum_{j=1}^{(k+1)m} H_1(x_{nj},u_{nj})\Big| + \frac{1}{n}\max_{0\le k\le n/m}\sum_{j=km+1}^{(k+1)m}\big|H_1(x_{nj},u_{nj})\big| \\
=\ & \widehat{R}_{n,1}+\widehat{R}_{n,2}=o_P(1),
\end{aligned}
$$

establishing (4.21).

We next prove (4.22). Since $K_j^2(x,y)=H^2(x,y)-2H(x,y)\widetilde{H}(x)+\big[\widetilde{H}(x)\big]^2$, we may write

$$
\begin{aligned}
V_{nk}^2-\widetilde{V}_{nk}^2 &= \frac{1}{n}\sum_{j=1}^{k}\big[H^2(x_{nj},u_{nj})^2-\widehat{H}^2(x_{nj})\big]+\frac{2}{n}\sum_{j=1}^{k}\widetilde{H}(x_{nj})\big[\widetilde{H}(x_{nj})-H(x_{nj},u_{nj})\big] \\
&:= A_{nk}+B_{nk}.
\end{aligned}
$$

The similar argument as in the proof of (4.21), but $H(x,y)$ is replaced by $H^2(x,y)$ and $\widetilde{H}(x)\,H(x,y)$ respectively, yields that

$$
\max_{1\le k\le n}\big|V_{nk}^2-\widetilde{V}_{nk}^2\big| \ \le\ \max_{1\le k\le n}\big|A_{nk}\big|+\max_{1\le k\le n}\big|B_{nk}\big|=o_P(1)
$$

as required. □

## 4.5 Proof of Lemma 4.3

We first prove (4.23). First note that, for each $s\ge 1$,

$$
\frac{1}{n}\sum_{j=n+1}^{[ns]} l_j^2 \to_{a.s.} s-1.
$$

This implies that $\tau_n(t)\le n$ if $0<t\le V_{nn}^2$ and, when $t>V_{nn}^2$,

$$
n\le \tau_n(t)\le n(1+2t),\quad a.s.
$$

In term of these facts, for each $t>0$, we have

$$
\max_{1\le j\le \tau_n(t)}|\widehat{y}_{nj}| \ \le\ \max_{1\le j\le n}|y_{nj}|+\max_{n\le j\le n(1+2t)}|l_j|/\sqrt{n}=o_P(1), \tag{4.43}
$$

wehre we have used (4.20 ) and the fact that, due to $\mathbb{E}|l_j|^3 \le C < \infty$,

$$\max_{n \le j \le n(1+2t)} |l_j|/\sqrt{n} \le \Big(\frac{1}{n^{3/2}} \sum_{j=n}^{n(1+2t)} |l_j|^3\Big)^{1/3} = o_P(1).$$

Similarly, for each $t > 0$, it follow from Assumption 2.2(b) that

$$\begin{aligned} \mathbb{E}\Big|\sum_{j=1}^{\tau_n(t)} \mathbb{E}\big(\widehat{y}_{nj}^2 | \mathcal{F}_{n,j-1}\big) - t\Big| &\le \frac{1}{n}\mathbb{E} \max_{1\le j\le n} |V_{nj}^2 - V_{n,j-1}^2| + \frac{1}{n}\mathbb{E} \max_{n \le j \le n(1+2t)} l_j^2 \\ &\le \frac{C}{n}\mathbb{E} \max_{1\le j\le n} g_N^2(u_{nj}) + \frac{1}{n}\mathbb{E} \max_{n \le j \le n(1+2t)} l_j^2 \\ &\to 0. \end{aligned} \tag{4.44}$$

As a consequence, we have $\sum_{j=1}^{\tau_n(t)} \widehat{y}_{nj} \Rightarrow B_t'$ on $D_R[0, \infty)$ by the classical martingale limit theorem.

We finally prove (4.24). It is readily seen that $\{Z_n(t)\}_{t\ge 0}$ is tight. Therefore, for each $\{n'\} \subseteq \{n\}$, there exists a subsequence $\{n''\} \subseteq \{n'\}$ such that, on $D_{\mathbb{R}^4}[0, \infty)$,

$$Z_{n''}(t) \;\Rightarrow\; Z_t := \big(B_t,\ B_{1t},\ X_t, B_t'\big).$$

As a consequence, (4.24) will follow if we prove that $B'$ is independent of $X$ and $F$. Note that $\big(\epsilon_{nj}, \eta_{nj}, \widehat{y}_{nj}, \mathcal{F}_{nj}\big)_{n\ge 1, j\ge 1}$ forms a martingale array at the same probability space. The independence between $B'$ and F follows the standard martingale limit theorem and the fact: by using (4.21) and definition of $\widehat{y}_{nj}$,

$$\begin{aligned} &\frac{1}{\sqrt{n}} \max_{1\le k\le \tau_n(t)} \Big|\Big|\sum_{j=1}^{k} \mathbb{E}\big(\xi_{nj}\widehat{y}_{nj}|\mathcal{F}_{n,j-1}\big)\Big|\Big| \\ = &\frac{1}{\sqrt{n}} \max_{1\le k\le n} \Big|\Big|\sum_{j=1}^{k} \mathbb{E}\big(\xi_{nj}y_{nj}|\mathcal{F}_{n,j-1}\big)\Big|\Big| \to 0, \end{aligned}$$

where $\xi_{nj} = \big(\epsilon_{nj}, \eta_{nj}\big)$. The fact that $B'$ is independent of $X$ is obvious since $X = \{X_t\}_{t\ge 0}$ is adapted to the filtration $\mathrm{F} = \{\sigma(B_s, B_{1s}, s \le t)\}_{t\ge 0}$. The proof of Lemma 4.3 is complete. $\square$