Least squares estimation for nonlinear regression models with heteroscedasticity

Qiying Wang*

The University of Sydney

July 31, 2020

Abstract

This paper develops an asymptotic theory of nonlinear least squares estimation by establishing a new framework that can be easily applied to various nonlinear regression models with heteroscedasticity. As an illustration, we explore an application of the framework to nonlinear regression models with nonstationarity and heteroscedasticity. In addition to these main results, this paper provides a maximum inequality for a class of martingales, which is of interest in its own right.

JEL Classification: C13, C22.

Keywords and phrases: Nonlinear regression, least squares estimation, nonstationarity, heteroscedasticity, martingale maximum inequality, local time, a mixture of normal distributions.

1 Introduction

Let $\{\mathcal{F}_t\}_{t\geq 0}$ be an increasing sequence of σ-fields on some probability space $(\Omega, \mathcal{F}, P)$ with $\mathcal{F}_0 = \sigma(\phi, \Omega)$ and let Θ be a compact set of R^q. Consider a general nonlinear regression model with the following form:

$$y_t = g_t(\theta) + u_t, \tag{1.1}$$

where $\{u_t, \mathcal{F}_t\}_{t\geq 1}$ forms a martingale difference such that $\sigma_t^2 := E(u_t^2|\mathcal{F}_{t-1}) < \infty, a.s.$, and $g_t(\theta)$ is a $\mathcal{F}_{t-1}$-measurable random function of $\theta \in \Theta$. The unknown parameters θ

*The author thanks Professor Cavaliere (coeditor) and four referees for their very helpful comments on the original version. The author also acknowledges the research support given by the Australian Research Council. Address correspondence to Qiying Wang: School of Mathematics and Statistics, University of Sydney, NSW 2006, Australia. E-mail: qiying.wang@sydney.edu.au

in model (1.1) can be estimated using the least squares (LS) method. Explicitly, the LS estimator $\widehat{\theta}_n$ of a real value $\theta_0 \in \Theta$ is defined by

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \sum_{t=1}^{n} \big[y_t - g_t(\theta)\big]^2. \tag{1.2}$$

The consistency of $\widehat{\theta}_n$ is widely investigated in the literature. Earlier contributions include Jennrich (1969), Wu (1981), Lai (1994), and Skouras (2000). More recently, Jacob (2010) generalized previous works using a nearly necessary condition. Chan and Wang (2015) established a general framework that is useful in nonlinear cointegrating regression models. We also refer to Section 2.1 for further details.

This study is concerned with the asymptotic distribution of $\widehat{\theta}_n$. In this regard, most previous frameworks were established under $\sup_{t\geq 1} \sigma_t^2 \leq C < \infty$ and/or strong smooth conditions on the first- and second-order derivatives of $g_t(\theta)$. For instance, see Wooldridge (1994), Andrew and Sun (2004), Pollard and Rachenko (2006), Jacob (2010), and Chan and Wang (2015). In the fields of econometrics and statistics, the σ_t^2 is commonly referred to as a volatility process. While such a process can be random, the uniform boundedness condition $\sup_{t\geq 1} \sigma_t^2 \leq C < \infty$ is clearly restrictive since it excludes ARCH, GARCH, and many other commonly used models. As a consequence, the applicability of these previous frameworks is limited in financial econometrics due to the fact that many macro-economic variables exhibit evidence of conditional heteroscedasticity. For instance, see Goncalves and Kilian (2004), and Boswijk, Cavaliere, Rahbek, and Taylor (2016).

The first part of this study provides an alternative framework of the asymptotic distribution of $\widehat{\theta}_n$. Our new result has some advantages compared to previous works. First, our framework does not assume $\sup_{t\geq 1} \sigma_t^2 \leq C < \infty$. As argued above, removing the uniform boundedness condition enables our result to be widely applicable in nonlinear regression models with heteroscedasticity. Second, unlike most previous frameworks, the key condition (see Assumption 2.2(i) in Section 2) in the present study is directly related to the Lipschitz difference of the first-order derivative of $g_t(\theta)$ rather than the maximum over parameter space Θ. This simplicity makes our conditions easy to verify, particularly in nonlinear regression models with nonstationarity and heteroscedasticity as seen in Section 3. Furthermore, in the development of our framework, we establish a maximum inequality for a class of martingales. As a technical tool, this maximum inequality is very sharp by taking advantage of the exponential inequality for self-normalized martingales investigated in Bercu and Touati (2008). This maximum inequality has many different applications and is interesting in itself.

The second part of this study considers the asymptotics of least squares estimators in nonlinear regression models with nonstationarity and heteroscedasticity, illustrating the applicability of our main results. Nonlinear regression with integrated time series was initially investigated in Park and Phillips (2001). Since then, many authors have contributed to parametric, nonparametric, and semi-parametric estimation and inference theory in this field. We only refer to Chang, Park, and Phillips (2001), Chang and Park (2003), Bae and De Jong (2007), Wang and Phillips (2009a, 2009b, 2016), Kim and Kim (2012), Gao and Phillips (2013), Chan and Wang (2015), Dong, Gao, and Tjötheim (2016), and Dong and Linton (2018), together with the references cited therein. Although there have been significant developments in the last few decades, these existing studies do not consider the impact of conditional heteroscedasticity on the estimation and inference theory. Using the aforementioned framework, the present study fills the gap. In this work, the volatility process σ_t^2 is set to be $\sigma_t^2 = \sigma(t/n; \lambda_t, \lambda_{t-1}, ...)$, where $\lambda_k, k \in \mathbb{Z}$, is a sequence of i.i.d. random vectors that are possibly unbounded and $\sigma(.; ...)$ is a measurable function satisfying certain regular conditions (we avoid $\sup_{t\geq 1} \sigma_t^2 \leq C < \infty$). These kind of settings allow for time-varying behaviors in volatility processes and includes a wide class of commonly used nonlinear models such as time-varying and nonlinear GARCH, as described in Examples 3.1-3.3 in Section 3. To the best of our knowledge, this work seems to be the first to investigate the impact of conditional heteroscedasticity in nonlinear cointegrating regression. For similar works on linear cointegration and unit root testing, we refer to Cavaliere and Taylor (2007, 2009, 2010), Boswijk, Cavaliere, Rahbek, and Taylor (2016) and the references cited therein.

This paper is organized as follows. Section 2 presents the main framework. We also present a couple of important step results that are required in the proof of the main framework, including the consistency of $\hat{\theta}_n$ and a maximum inequality for a class of martingale. Section 3 considers nonlinear regression models with nonstationarity and heteroscedasticity, extending the existing results established by Park and Phillips (2001) and Chan and Wang (2015). Concluding remarks are provided in Section 4. Technical proofs of the main results in Section 2 are given in Appendix.

In this study, we use the following notation: for $x = (x_{ij})_{1\leq i\leq m, 1\leq j\leq k}$, $||x|| = \sum_{i=1}^m \sum_{j=1}^k |x_{ij}|$. We denote constants by $C, C_1, ...,$, which may be different at each appearance.

2 Main results

This section investigates the asymptotic distribution of the $\widehat{\theta}_n$ defined by (1.2) and provides a couple of important technical results that are required in the proof. Let $\dot{g}_t(\theta) = \big(\frac{\partial g_t(\theta)}{\partial \theta_1}, ..., \frac{\partial g_t(\theta)}{\partial \theta_q}\big)'$ be the first-order derivative of $g_t(\theta)$. We make use of the following assumptions.

Assumption 2.1. $\{u_t, \mathcal{F}_t\}_{t\geq 1}$ forms a martingale difference with $E(u_t^2 \mid \mathcal{F}_{t-1}) < \infty, a.s.$ for each $t \geq 1$.

Assumption 2.2. A matrix $D_n = \text{diag}(d_{1n}, ..., d_{qn})$ satisfying $n^{-\delta} \min_{1\leq j\leq q} d_{jn} \to \infty$ for some $\delta > 0$ exists such that

(i) $||D_n^{-1}\big[\dot{g}_t(\theta_1) - \dot{g}_t(\theta_2)\big]|| \leq ||\theta_1 - \theta_2||^{\alpha}\, T_{nt}$ for some $0 < \alpha \leq 1$ and for any $\theta_1, \theta_2 \in \Theta$, where T_{nt} is adapted to $\mathcal{F}_{t-1}$ for each $n \geq 1$, satisfying

$$\sum_{t=1}^{n} T_{nt}^2\big[1 + E(u_t^2|\mathcal{F}_{t-1})\big] = O_P(1); \tag{2.1}$$

(ii) $Y_n := (D_n^{-1})' \sum_{t=1}^{n} \dot{g}_t(\theta_0)\dot{g}_t(\theta_0)'\, D_n^{-1} \to_D M$, where $M > 0$, a.s., that is, the smallest eigenvalue of M is almost surely positive;

(iii) $Z_n(\theta_0) = O_P(1)$, where $Z_n(\theta) = (D_n^{-1})' \sum_{t=1}^{n} \dot{g}_t(\theta)\, u_t$.

Assumption 2.3. $\widehat{\theta}_n \to_P \theta_0$, where $\widehat{\theta}_n$ is defined by (1.2).

Assumption 2.1 is commonly used in nonlinear regression models, but we do not require the restrictive condition $\sup_t E(u_t^2 \mid \mathcal{F}_{t-1}) \leq C < \infty$, that is, the uniform boundedness of the conditional variance. Instead, we make use of a summability condition of the form (2.1). Since the impact of conditional variance $E(u_t^2 \mid \mathcal{F}_{t-1})$ cannot be eliminated in model (1.1), condition (2.1) is quite natural under the Lipschitz condition for $\dot{g}_t(\theta)$ in Assumption 2.2 (i). Indeed, if $g_t(\theta)$ is linear with respect to θ, (2.1) holds automatically by taking $T_{nt} = 0$ for all $t \geq 1$. If $g_t(\theta)$ ($\dot{g}_t(\theta)$) is stationary, M in Assumption 2.2(ii) is usually a constant and $D_n = diag(\sqrt{n}, ..., \sqrt{n})$ in general. Hence, we may take $T_{nt} = \frac{1}{\sqrt{n}} \sup_{\theta\in\Theta} |\dot{g}_t(\theta)|$, indicating (2.1) is nearly necessary. Condition (2.1) is more involved if $g_t(\theta)$ ($\dot{g}_t(\theta)$) is nonstationary or an $I(1)$ random process. In this situation, M in Assumption 2.2(ii) can be a positive-definite random matrix and D_n usually depends on the shape of $\dot{g}_t(\theta)$. Further discussions on this topic are presented in Section 3 using nonlinear cointegrating regression models with heteroscedasticity. In summary, Assumption 2.2 is applicable to a wide class of nonlinear regression models. We discuss Assumption 2.3 separately in Section 2.1.

The main result is as follows:

Theorem 2.1. *Under Assumptions 2.1-2.3, we obtain*

$$D_n(\hat{\theta}_n - \theta_0) = Y_n^{-1} Z_n(\theta_0) + o_P(1). \tag{2.2}$$

If in addition

$$(Y_n,\ Z_n(\theta_0)) \ \to_D \ (M,\ Z), \tag{2.3}$$

where $M > 0$, a.s., that is, the smallest eigenvalue of the M is almost surely positive, then $D_n(\hat{\theta}_n - \theta_0) \to_D M^{-1} Z$.

Remark 2.1. Given that $g_t(\theta) = E_\theta(y_t|\mathcal{F}_{t-1})$ under Assumption 2.1, the estimation considered herein is essentially the same as the conditional least squares estimation investigated in Wooldridge (1994), Andrew and Sun (2004), Jacob (2010), Chan and Wang (2015), and Wang and Phillips (2016). The conditions in these existing frameworks are usually imposed on the maximum over the parametric space Θ for quantities that are related to the first- and second-order derivatives of $g_t(\theta)$. Using Assumption 2.3, the key step in our proof is only involved in the verifications of the following:

$$\begin{aligned} \sup_{\theta\in\Theta} ||Z_n(\theta) - Z_n(\theta_0)|| &= O_P(\log^{1/2} n), & (2.4)\\ \sup_{||D_n(\theta-\theta_0)||\le \log n} ||Z_n(\theta) - Z_n(\theta_0)|| &= o_P(1), & (2.5) \end{aligned}$$

which are in turn implied by Assumptions 2.1 and 2.2 (i) as seen in Corollary 2.1. Since our primitive condition (2.1) is quite natural, in comparison with these existing results, Theorem 2.1 provides a framework that is simple and easily verified.

Remark 2.2. Theorem 2.1 is still applicable to more general models:

$$y_{nt} = g_{nt}(\theta) + u_{nt}, \quad t = 1, 2, ..., n;\ n \ge 1,$$

where, for each $n \ge 1$, $\{u_{nt}, \mathcal{F}_{nt}\}_{1\le t\le n}$ forms a martingale difference such that $E(u_{nt}^2|\mathcal{F}_{n,t-1}) < \infty, a.s.$, and $g_{nt}(\theta)$ is a $\mathcal{F}_{n,t-1}$-measurable random function of $\theta \in \Theta$. The detailed statement is avoided since it only involves some routine notation changes in Assumptions 2.1-2.3. This remark will be used later without further explanation.

Remark 2.3. The entropy method has been widely used in nonlinear regression models with stationary time series. The chaining argument with stratifications now is standard and seems to be inevitable in investigating stochastic equicontinuity and uniform central limit theory for martingales with stationary conditional variances (i.e., the conditional variance converges to a constant or a deterministic variable). For instance, see Andrews (1994) and Nishiyama (2000a, 2000b, 2007). The $Z_n(\theta)$ defined in Assumption 2.2 (iii) is a martingale, but it allows for a more general structure than that used in existing literature when the $\dot{g}_t(\theta)$ includes an $I(1)$ random process as a part of its components. As shown in Wang (2014), for this class of martingales, the conditional variance may converge in distribution to a random variable rather than a constant and the limit distribution is a mixture of normal distributions or a stochastic integral instead of a standard normal. In terms of the complexity in structure for this new class of martingales, it is currently not clear whether the chaining arguments with stratifications can be utilized to establish (2.4) and (2.5). As a consequence, it seems to be difficult to use the so-called entropy condition rather than the Lipschitz condition in Assumption 2.1 (i).

2.1 Consistency of $\hat{\theta}_n$

The consistency of $\hat{\theta}_n$ is imposed as a preliminary condition for the asymptotic distribution of $\hat{\theta}_n$ in Theorem 2.1, which is usually easy to handle under other conditions. For the earlier contributions in this regard, we refer to Jennrich (1969), Wu (1981), Lai (1994), and Skouras (2000). More recently, Proposition 3.1 in Jacob (2010) established a general result without assuming $\sup_t \sigma_t^2 \le C < \infty$, where $\sigma_t^2 = E(u_t^2 \mid \mathcal{F}_{t-1})$. Let $d_j(\theta) = g_j(\theta) - g_j(\theta_0)$ and $D_{k,\theta} = \sum_{j=1}^k d_j^2(\theta)$. One of the conditions used in Jacob (2010) is the following infinite sum of $D_{k,\theta}$: for any $\delta > 0$,

$$\sup_{||\theta-\theta_0||\ge\delta} \sum_{k=1}^{\infty} \sigma_k^2\, d_k^2(\theta)\, D_{k,\theta}^{-2} < \infty, \quad a.s. \tag{2.6}$$

Although the consistency result given in Jacob (2010) is elegant, the infinite sum of $D_{k,\theta}$ such as (2.6) is usually difficult to verify, particularly in nonlinear cointegrating regression models considered in Section 3. For the purpose of this study, the consistency of $\hat{\theta}_n$ is established under different settings. Our result is similar to Theorem 2.1 in Chan and Wang (2015), but does not assume $\sup_t E(u_t^2 \mid \mathcal{F}_{t-1}) \le C < \infty$.

Theorem 2.2. *In addition to Assumption 2.1, a sequence of constants* $0 < k_n \to \infty$ *exists such that* *(a)* $|g_t(\theta_1) - g_t(\theta_2)| \le ||\theta_1 - \theta_2||^\alpha T_t$, *for some* $0 < \alpha \le 1$ *and for all* $\theta_1, \theta_2 \in \Theta$, *where* T_t *is adapted to* $\mathcal{F}_{t-1}$, *satisfying*

$$\sum_{t=1}^{n} T_t^2 \, E(u_t^2 \mid \mathcal{F}_{t-1}) = O_P(k_n^2/\log^2 n); \tag{2.7}$$

(b) $k_n^{-1}\inf_{||\theta-\theta_0||\ge\delta} D_{n,\theta}$, *for any* $\delta > 0$, *is away from* 0 *with probability one, as* $n \to \infty$.

Then $||\widehat{\theta}_n - \theta_0|| = o_P(1)$.

Condition (2.7) is similar to (2.1), but only depends on $g_t(\theta)$ rather than the first-order derivative of $g_t(\theta)$. Given that Θ is a compact set, we may provide a simple sufficient condition for part (b) using the finite cover theorem. This fact is stated in the following proposition for convenience.

Proposition 2.1. *Part (b) of Theorem 2.2 holds if*

(i) $\frac{1}{k_n}\sum_{t=1}^n \big[g_t(\theta) - g_t(\theta_0)\big]^2 \to_D G(\theta)$ *for any* $\theta \ne \theta_0$, *where* $G(\theta)$ *is a stochastic process of* θ *satisfying that, for each* $\delta > 0$, $P(\inf_{||\theta-\theta_0||\ge\delta} G(\theta) \ge M_\delta) = 1$, *where* $M_\delta > 0$ *is a constant depending only on* δ;

(ii) $|g_t(\theta_1) - g_t(\theta_2)| \le h(||\theta_1 - \theta_2||)\, T_t$ *for all* $\theta_1, \theta_2 \in \Theta$, *where* T_t *is a sequence of random variables satisfying* $\frac{1}{k_n}\sum_{t=1}^n T_t^2 = O_P(1)$ *and* $h(x)$ *is a continuous function satisfying* $\lim_{x\downarrow 0} h(x) = 0$.

2.2 Maximum inequality for a class of martingales

In this section, we establish a new maximum inequality for a class of martingales and hence provide a powerful technical tool to verify (2.4) and (2.5). The result is of interest in its own right, and can be used for different purposes as seen in Remark 2.4. We use the following assumption and the notation is slightly more general than those used in previous sections.

Assumption 2.4. For each $n \ge 1$,

(i) $\{u_{nt}, \mathcal{F}_{nt}\}_{t\ge1}$ forms a martingale difference with $E(u_{nt}^2 \mid \mathcal{F}_{n,t-1}) < \infty, t \ge 1$;

(ii) $y_{nt} = \big\{x_{n1}(t), \cdots, x_{nd}(t)\big\}$ is adapted to $\mathcal{F}_{n,t-1}$, where $d \ge 1$ is an integer;

(iii) Ψ is a set of real measurable functions $f(.)$ on R^d with $\#\Psi \ge 1$, where $\#A$ denotes the number of elements in a set A.

Theorem 2.3. *Suppose Assumption 2.4 holds. If a sequence of positive constants* $\gamma_n \to \infty$ *exists such that*

$$\sup_{f \in \Psi} \sum_{t=1}^{n} f^2(y_{nt}) \big[u_{nt}^2 + E(u_{nt}^2 \mid \mathcal{F}_{n,t-1}) \big] \quad = \quad O_P(\gamma_n), \tag{2.8}$$

then

$$\sup_{f \in \Psi} \Big| \sum_{t=1}^{n} u_{nt}\, f(y_{nt}) \Big| \quad = \quad O_P \big\{ \big[\gamma_n \log^+(\#\Psi) \big]^{1/2} \big\}, \tag{2.9}$$

where $\log^+(\#\Psi) = \max\{1, \log(\#\Psi)\}$.

Theorem 2.3 is a significant extension of Theorem 3.19 in Wang (2015) (see also, Chan and Wang, 2014) and is a very sharp result owing to the following facts:

- $\big[\gamma_n \log^+(\#\Psi)\big]^{1/2}$ is the same rate as in the i.i.d./stationary cases.
- Without assuming the uniform boundedness of $E(u_{nt}^2 \mid \mathcal{F}_{n,t-1}) < \infty$ in t, condition (2.8) is natural (may be necessary) because it only depends on the squared variation $\sum_{t=1}^n f^2(y_{nt})\, u_{nt}^2$ and the quadratic variation (conditional variance) $\sum_{t=1}^n f^2(y_{nt})\, E(u_{nt}^2 \mid \mathcal{F}_{n,t-1})$ of the martingale array $\sum_{t=1}^n u_{nt}\, f(y_{nt})$;
- there are no restrictions on $f(.)$ and γ_n except measurability on R^d and $\gamma_n \to \infty$.

The proof of Theorem 2.3 takes the advantage of an exponential inequality for self-normalized martingales developed in Bercu and Touati (2008). The following corollary is a direct consequence of Theorem 2.3, and verifies (2.4) and (2.5) under the conditions of Theorem 2.1.

Corollary 2.1. *If Assumptions 2.1 and 2.2 (i) hold, we have (2.4) and (2.5).*

Remark 2.4. For a real measurable function $g(x)$ on R^d, a common function of interests $S_n(x)$ of $\{u_{nt}, y_{nt}\}_{t \ge 1, n \ge 1}$ is defined by the following:

$$S_n(x) \quad = \quad \sum_{t=1}^{n} u_{nt}\, g[(y_{nt} + x)/h], \quad x \in \mathrm{R}^d$$

where $h = h_n \to 0$. For instance, in nonparametric estimation problems, g may be a kernel function K or a squared kernel function K^2 and the sequence h is the bandwidth used in nonparametric regression. Using Theorem 2.3, the uniform convergence of $S_n(x)$ can be established under quite general conditions, particularly allowing for both stationary and nonstationary random arrays y_{nt}. As a consequence, we may provide a powerful technical tool to investigate the uniform convergence for nonparametric estimators with nonstationary data, as in Chan and Wang (2014), Gao et al. (2015), and Duffy (2016, 2017). Since this is beyond the scope of this study, we will report the results in a separate work.

3 Nonlinear regression models with nonstationarity and heteroscedasticity

This section considers nonlinear regression models having the following form:

$$y_t = f(x_t, \theta) + u_t, \tag{3.1}$$

where $f(x, ...)$ is a given real function indexed by $\theta = (\theta_1, ..., \theta_q)$, a vector of unknown parameters, x_t is an integrated regressor 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 Θ be a compact set of R^q and assume the unknown parameters $\theta \in \Theta$. As in model (1.1), the least squares estimator $\hat{\theta}_n$ of θ is defined by (1.2) with $g_t(\theta) = f(x_t, \theta)$. The asymptotics of $\hat{\theta}_n$ were initially investigated in Park and Phillips (2001) and then later by Chang, Park, and Phillips (2001), De Jong and Hu (2011), and Chan and Wang (2015). In these existing studies, the uniform boundedness of $E(u_t^2|\mathcal{F}_{t-1})$ in t was usually assumed, excluding a wide class of commonly used volatility models such as GARCH, time-varying GARCH, and nonlinear GARCH.

Since (3.1) is a specified form of model (1.1) with $g_t(\theta) = f(x_t, \theta)$, Theorem 2.1 can be utilized to establish asymptotic distribution of $\hat{\theta}_n$ without assuming $\sup_t E(u_t^2|\mathcal{F}_{t-1}) \leq C < \infty$. Consequently, we may consider nonlinear regression models with nonstationarity and heteroscedasticity, which seems to be new to the literature.

This section is organized as follows. Section 3.1 presents assumptions on x_t and u_t, together with some discussions. In this subsection, we use three examples to illustrate the wide applicability of our settings on u_t. In the following subsections, we establish asymptotic distributions of $\hat{\theta}_n$. Considering that there are essential differences between integrable and nonintegrable regression functions, we present the asymptotics in Sections 3.2 and 3.3 separately.

3.1 Assumptions and Examples

Throughout the section, let $\lambda_i \equiv (\epsilon_{i-1}, \eta_i), i \in \mathbb{Z}$, be a sequence of i.i.d. random vectors with $E\lambda_0 = 0$, $E\epsilon_0^2 = E\eta_0^2 = 1$ and $\rho = E\epsilon_0\eta_1$, 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. $x_t = \gamma x_{t-1} + \xi_t$, where $\gamma = 1 - \tau/n, \tau \geq 0$ and $\xi_t = \sum_{j=0}^{\infty} \phi_j \eta_{t-j}$. 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 ∞.

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 $\sigma_t^2 = \sigma(t/n; \lambda_t, \lambda_{t-1}, ...)$, where $\sigma(.; ...)$ is a measurable function satisfying $\sup_{0\leq u\leq 1} E\sigma^2(u; \lambda_0, \lambda_1, ...) < \infty$, $\int_0^1 E\sigma(u; \lambda_0, \lambda_1, ...)du < \infty$ and

$$\sup_{0\leq u\leq 1} E\big|\sigma(u; \lambda_m, \lambda_{m-1}, ...) - \sigma(u; \lambda_m, \lambda_{m-1}, ..., \lambda_1, \lambda_0^*, \lambda_1^*, ...)\big|^2 \;\leq\; C\, m^{-\alpha}, \tag{3.2}$$

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

Assumption 3.1 allows for short memory (under **SM**) and long memory (under **LM**) innovations driving the nearly integrated regressor x_k, which is quite general in practice. Define

$$d_n^2 \;=\; \mathbb{E}|\sum_{k=1}^{n} \xi_k|^2 \;\sim\; \begin{cases} c_\mu\, n^{3-2\mu} l^2(n), & \text{under } LM, \\ \phi^2\, n, & \text{under } SM, \end{cases} \tag{3.3}$$

where c_μ is a constant. Standard functional limit theory (see Buchmann and Chan (2007) or Theorem 2.21 in Wang (2015) with a minor modification) shows that

$$\begin{aligned} &\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i,\ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \eta_i,\ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \eta_{-i},\ \frac{1}{d_n} x_{[nt]}\Big) \\ &\Rightarrow \big(U_t,\ B_t,\ B_{-t},\ X_t\big), \end{aligned} \tag{3.4}$$

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

$$X_t \;=\; W(t) + \tau \int_0^t e^{-\tau(t-s)} W(s) ds$$

with $W_t = \begin{cases} G_{3/2-\mu}(t), & \text{under } \mathbf{LM}, \\ G_{1/2}(t), & \text{under } \mathbf{SM}, \end{cases}$ where G_β is a fractional Brownian motion that has the following form: with $a_+ = \max\{a, 0\}$,

$$G_\beta(t) \;\; = \;\; \frac{1}{\Gamma(d+1)} \int_{-\infty}^{t} (t-x)_+^d - (-x)_+^\beta dB_t.$$

It should be noted that X_t is an Ornstein-Uhlenbeck process having a continuous local time $L_X(t,s)$ defined by

$$L_X(t,s) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t I(|X_r - s| \le \epsilon) dr.$$

These notations will be used later without further explanations.

Under the settings of Assumption 3.2, we have that $(u_k, \mathcal{F}_k)_{k \ge 1}$, where $\mathcal{F}_k$ is an σ-field generated by $\lambda_{k+1}, \lambda_k, ...$, forms a martingale difference with $E(u_k^2|\mathcal{F}_{k-1}) = \sigma_k^2 = \sigma(k/n; \lambda_k, \lambda_{k-1}, ...)$. The error process u_t having a martingale difference structure has been widely used in previous studies such as Park and Phillips (2001), Chang, Park, and Phillips (2001), De Jong and Hu (2011), and Chan and Wang (2015). Unlike these existing studies, in Assumption 3.2, we do not assume $\sup_t E(u_t^2|\mathcal{F}_{t-1}) \le C < \infty$ since λ_k are possibly unbounded random vectors. This removal of the uniform boundedness of $E(u_t^2|\mathcal{F}_{t-1})$ ensures Assumption 3.2 is applicable for a wide class of models of heteroscedasticity, including the time-varying behaviors in volatility processes. To illustrate, we start with the following proposition, which is a corollary of Theorem 5.1 in Shao and Wu (2007). Let G_k be recursively defined by

$$G_k = R(G_{k-1}, ..., G_{k-p+1}; \lambda_k), \tag{3.5}$$

where $p \ge 1$ and R is a measurable function. Recall that $\lambda_i = (\epsilon_{i-1}, \eta_i), i \in \mathbb{Z}$ are i.i.d. random vectors and $\{\lambda_i^*\}_{i \in \mathbb{Z}}$ is an independent copy of $\{\lambda_i\}_{i \in \mathbb{Z}}$.

Proposition 3.1. *Suppose that $EG_0^2 < \infty$ and functions H_j exist such that*

$$|R(y; \lambda_0) - R(y'; \lambda_0)| \le \sum_{j=1}^{p} H_j(\lambda_0)|x_j - x_j'| \tag{3.6}$$

holds for all $y = (x_1, ..., x_p)$ and $y' = (x_1', ..., x_p')$ and $\sum_{j=1}^p \left[EH_j(\lambda_0)^2\right]^{1/2} < 1$. Then a measurable function $G(...)$ and a constant $0 < \gamma < 1$ exist such that $G_t = G(\lambda_t, \lambda_{t-1}, ...)$ and, for all $t \ge 1$,

$$E|G(\lambda_t, \lambda_{t-1}, ...) - G(\lambda_t, \lambda_{t-1}, ..., \lambda_1; \lambda_0^*, \lambda_1^*, ...)|^2 \le C\gamma^t, \tag{3.7}$$

Due to Proposition 3.1, the following examples satisfy Assumption 3.2, indicating that the class of models of heteroscedasticity within the framework of Assumption 3.2 is quite large.

Example 3.1. (Time-varying $GARCH(p,1)$ model) Let G_t be defined by

$$G_t = \alpha_0 + \alpha_1 \, G_{t-1}\epsilon_{t-1}^2 + \sum_{j=1}^{p} \beta_j G_{t-j}$$

and $u_t = a(t/n)G_t^{1/2}\epsilon_t$, where $a(u)$ is a positive locally Riemann integrable function (i.e., $a(u)$ is Riemann integrable at any finite interval). Assume that $\alpha_0 > 0, \alpha_1 \geq 0, \beta_i \geq 0$ and $\left[E(\alpha_1\epsilon_0^2 + \beta_1)^2\right]^{1/2} + \sum_{j=1}^{p}\beta_j < 1$ and $E\epsilon_0^4 < \infty$. We claim that u_t satisfies Assumption 3.2.

Indeed, it follows from Proposition 3.1 [λ_k is replaced by ϵ_{k-1} in (3.5)] with $H_1(\theta) = \alpha_1\theta^2 + \beta_1$ and $H_j(\theta) = \beta_j, j = 2, ..., p$, that a measurable function $G(...)$ exists such that $G_t = G(\epsilon_{t-1}, \epsilon_{t-2}, ...)$, $EG_0^2 < \infty$ and (3.7) is satisfied. As a consequence, a measurable function $\sigma(.;...)$ so that $\sigma_t^2 := a^2(t/n)G_t = a^2(t/n)G(\epsilon_{t-1}, \epsilon_{t-2}, ...) = \sigma(t/n; \epsilon_{t-1}, \epsilon_{t-2}, ...)$ satisfies (3.2) exists, that is, $u_t = a(t/n)G_t^{1/2}\epsilon_t = \sigma_t\,\epsilon_t$ satisfies Assumption 3.2.

This example allows for time-varying volatility processes. If $a(u) \equiv 1$, then $u_t = G_t^{1/2}\epsilon_t$ is a standard $GARCH(p,1)$ model since $G_t = \alpha_0 + \alpha_1 \, u_{t-1}^2 + \sum_{j=1}^{p}\beta_j G_{t-j}$ in this case.

Example 3.2. (Amplitude-dependent exponential autoregressive (EXPAR) model) Let $u_t = G_t\epsilon_t$, where G_t is defined by

$$G_t = \left[\alpha + \beta \exp(-aG_{t-1}^2)\right]G_{t-1} + \epsilon_{t-1}^2.$$

If $a > 0, \alpha \geq 0, \beta \geq 0$ and $\alpha + \beta < 1$ and $E\epsilon_0^4 < \infty$, it follows from Proposition 3.1 with $H_1(\theta) = \alpha + \beta$ that u_t satisfies Assumption 3.2. Jones (1976) originally considered the EXPAR model. Similar to Example 3.1, if $a(u)$ is a positive locally Riemann integrable function, $u_t = a(t/n)G_t\epsilon_t$ still satisfies Assumption 3.2.

Example 3.3. (Nonlinear GARCH model) Let $u_t = G_t^{1/2}\epsilon_t$, where G_t is defined by

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

This is a nonlinear GARCH model introduced by Lanne and Saikkonen (2005). 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 $G_t = \alpha_0 + R(G_{t-1}, \epsilon_{t-1})$, where $R(y, \epsilon) = \alpha F(y) + \beta y + \gamma\epsilon^2 y$, it follows from Proposition 3.1 that u_t satisfies Assumption 3.2 if $\alpha_0 > 0, \alpha \geq 0, \beta \geq 0, \gamma \geq 0$ satisfying $\left[E\left(\alpha A_0 + \beta + \gamma\epsilon_0^2\right)^2\right]^{1/2} < 1$.

For more examples that satisfy Assumption 3.2, we refer to Wu and Min (2005), Shao and Wu (2007), and Peng and Wang (2018). It should be noted that Assumption 3.2 excludes nonstationary volatility models such as those discussed in Hansen (1995) and Cavaliere and Taylor (2007). The asymptotics for nonlinear cointegrating regression with nonstationary volatility seem to be much more difficult since we have to establish results on weak convergence for a functional of covariance that involves a production of two nonlinear integrated time series. We wish to report the development in the future.

3.2 Integrable regression function

This section considers the limit distribution of $\hat{\theta}_n$ when f is an integrable function, together with some additional smooth conditions on η_t. Write $\dot{f}(x,\theta) = (\dot{f}_1, ..., \dot{f}_q)'$, where $\dot{f}_i = \frac{\partial f(x,\theta)}{\partial \theta_i}, i = 1, ..., q$.

Assumption 3.3. Let $p(x,\theta)$ be one of f and $\dot{f}_i, i = 1, ..., q$.

(i) $p(x,\theta_0)$ is a bounded and integrable real function.

(ii) A bounded and integrable function $T_p : \mathbb{R} \to \mathbb{R}$ exists such that $|p(x,\theta) - p(x,\theta_0)| \le ||\theta - \theta_0||^\alpha T_p(x)$, for each $\theta, \theta_0 \in \Theta$ and some $\alpha > 0$;

(iii) $\Sigma = \int_{-\infty}^{\infty} \dot{f}(s,\theta_0)\dot{f}(s,\theta_0)' ds > 0$ (i.e., Σ is a positive-definite matrix) and $\int_{-\infty}^{\infty} (f(s,\theta) - f(s,\theta_0))^2 ds > 0$ for all $\theta \neq \theta_0$.

Theorem 3.1. *Suppose Assumptions 3.1-3.3 hold and* $\lim_{t\to\infty} |t|^\delta |Ee^{it\eta_1}| < \infty$ *for some* $\delta > 0$. *As* $n \to \infty$,

$$\sqrt{n/d_n}(\hat{\theta}_n - \theta_0) \to_D \Sigma^{-1/2}\, \mathbb{N}\, L_X^{-1}(1,0) \sqrt{\int_0^1 \Lambda^2(s) dL_X(s,0)} \tag{3.8}$$

where $\Lambda^2(u) = E\sigma(u; \lambda_0, \lambda_1, ...)$ *and* $\mathbb{N}$ *is a standard* q*-dimensional normal random vector independent of* X_t.

Remark 3.1. If no impact comes from the time-varying in volatility process σ_t^2 or, in another words, $\Lambda^2 := \Lambda^2(u) \equiv E\sigma(\lambda_0, \lambda_1, ...)$ for all $u \in [0,1]$, result (3.8) can be rewritten as

$$\sqrt{n/d_n}(\hat{\theta}_n - \theta_0) \to_D \Lambda\, \Sigma^{-1/2}\, \mathbb{N}\, L_X^{-1/2}(1,0),$$

which is pivotal upon an estimation of Λ^2. Note that $\Lambda^2 = Eu_1^2$. Λ^2 can be estimated by $\hat{\sigma}_n^2 = \frac{1}{n}\sum_{t=1}^n \left[y_t - g_t(\hat{\theta}_n)\right]^2$ and $\hat{\sigma}_n^2 \to_P \Lambda^2$ under the conditions of Theorem 3.1.

In application, the volatility process σ_t^2 usually depends on some unknown parameters, $\sigma_t^2 = \sigma_t(\eta)$, such as, where $\eta = (\eta_1, ..., \eta_m) \in \Xi$ and Ξ is a compact set of R^m. Note that $\{\epsilon_t^2 - 1, \mathcal{F}_t\}_{t\geq 1}$ still forms a martingale difference and

$$\big[y_t - g_t(\theta)\big]^2 \;=\; \sigma_t(\eta) + \sigma_t(\eta)(\epsilon_t^2 - 1). \tag{3.9}$$

The unknown parameter η can be estimated in a similar manner as that of θ, namely,

$$\hat{\eta}_n \;=\; \arg\min_{\eta\in\Xi} \sum_{t=1}^{n} \Big\{ \big[y_t - g_t(\hat{\theta}_n)\big]^2 - \sigma_t(\eta) \Big\}^2,$$

where $\widehat{\theta}_n$ is the LSE of θ defined by (1.2). Using some standard modifications, we can establish the asymptotics of η, similar to Theorems 2.1 and 3.1. Considering that this is only related to certain repeated arguments, we omit the details.

Remark 3.2. Theorem 3.1 significantly improves Theorem 3.2 of Chan and Wang (2015) by using less restrictive smoothing conditions on $f(x,\theta)$ and allowing for extensive volatility processes as those given in Examples 3.1 - 3.3. It should be mentioned that, to prove Theorem 3.1, we need to establish new results on convergence to local time and a mixture of normal distributions, which will be given in Theorem 3.4 of Section 3.4.

3.3 Nonintegrable regression function

This section establishes the limit distribution of $\hat{\theta}_n$ when f is nonintegrable. As in Section 3.2, write $\dot{f}(x,\theta) = (\dot{f}_1, ..., \dot{f}_q)'$, where $\dot{f}_i = \frac{\partial f(x,\theta)}{\partial \theta_i}$, $i = 1, ..., q$.

Assumption 3.4. Let $p(x,\theta)$ be one of f and $\dot{f}_i, i = 1, ..., q$. A real continuous function $T_p(x)$ exists such that

(i) $|p(x,\theta) - p(x,\theta_0)| \leq ||\theta - \theta_0||^\alpha\, T_p(x)$, for each $\theta, \theta_0 \in \Theta$ and some $\alpha > 0$;

(ii) for any bounded x, $\sup_{\theta\in\Theta} |p(lx,\theta) - v_p(l)\, h_p(x,\theta)|/T_p(lx) = o(1)$, as $l \to \infty$, where $h_p(x,\theta)$ for each $\theta \in \Theta$ is a continuous function and $v_p(l)$ is a positive real function that is bounded away from zero as $l \to \infty$; and

(iii) $T_p(lx) \leq v_p(l)\, T_{1p}(x)$ as $|lx| \to \infty$, where $T_{1p}(x)$ is a continuous function.

Define $\Psi(t) = \dot{h}(X_t, \theta_0)$, where $\dot{h}(a,\theta) = \big(h_{\dot{f}_1}(a,\theta), ..., h_{\dot{f}_q}(a,\theta)\big)$.

Theorem 3.2. *Suppose Assumptions 3.1-3.2 and 3.4 hold. Suppose that, for each $\eta > 0$,*

$$\int_{|x|\le\eta} \big[h_f(x,\theta) - h_f(x,\theta_0)\big]^2 dx \neq 0, \ \textit{for any } \theta \neq \theta_0, \ \textit{ and}$$

$$\int_{|x|\le\eta} \dot{h}(x,\theta_0)\,\dot{h}'(x,\theta_0)dx \quad \textit{is a positive-definite matrix.} \tag{3.10}$$

Then, as $n \to \infty$,

$$D_n\,(\hat{\theta}_n - \theta_0) \to_D \Big(\int_0^1 \Psi(t)\Psi(t)'dt\Big)^{-1} \int_0^1 \Psi(t)\,\Lambda(t)\,dU_t, \tag{3.11}$$

where $\Lambda^2(t) = E\sigma(t;\lambda_0,\lambda_1,...)$ and $D_n = \sqrt{n}\,diag\big(v_{\dot{f}_1}(d_n),...,v_{\dot{f}_q}(d_n)\big)$.

Remark 3.3. As indicated in Chan and Wang (2015) and Park and Phillips (2001), nonlinear cointegrating regressions with structures described in Theorem 3.2 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 Bae and de Jong (2007) and Bae et al. (2006) for empirical studies investigating the estimation of money demand functions in the United States and Japan respectively. See also Bae et al. (2004) for the derivation of these functional forms from the underlying money demand theories studied in macroeconomics.

Remark 3.4. Unlike Theorem 3.1, even in the case that $\Lambda^2 = \Lambda(u)$ is a constant over $u \in [0,1]$ that can be estimated, the limit distribution is not pivotal because it depends on ρ which is hidden in the joint distribution of $\big(\Psi(t), U_t\big)$. It seems to be difficult to estimate this nuisance parameter at the moment and hence we leave it for future work.

Remark 3.5. Theorem 3.4 of Chan and Wang (2015) provided a result that is similar to Theorem 3.2, but imposing strong restrictions on u_k and $f(x,\cdots)$ such that $\sup_{k\ge1} E(u_k^2|\mathcal{F}_{k-1}) \le C < \infty$ and $\frac{\partial^2 f}{\partial\theta_i\partial\theta_j}, i,j = 1,...,q$, satisfy some similar conditions as those of Assumption 3.4 in the present paper. Theorem 3.2 is not only applicable in nonlinear cointegrating regression models with various volatility processes as those discussed in Examples 3.1-3.3, but also the verification of the conditions imposed on $f(x,\cdots)$ is quite straightforward.

3.4 Convergence to stochastic integrals, local time and a mixture of normal distributions

The proofs of Theorems 3.1 and 3.2 depend on certain fundamental results on convergence to stochastic integrals, local time and a mixture of normal distributions, which are summarized in this section. We remark that Theorem 3.4 is new to literature.

Recall that $(\sigma_k\epsilon_k, \mathcal{F}_k)_{k\geq 1}$ forms a martingale difference, where $\sigma_k^2 = \sigma(k/n; \lambda_k, \lambda_{k-1}, ...)$ and $\mathcal{F}_k$ is an σ-field generated by $\lambda_k, \lambda_{k-1}, ...,$. By using a standard argument on functional martingale limit theorem, we have

$$\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\epsilon_i, \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\sigma_i\epsilon_i\Big) \Rightarrow \big(U_t,\ \int_0^t \sqrt{E\sigma(s,\lambda_0,\lambda_1,...)}\,dU_s\big),$$

on $D_{R^2}[0,1]$. This, together with (3.4), implies that

$$\Big(\frac{1}{d_n}x_{[nt]},\ \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor nt\rfloor}\sigma_i\epsilon_i\Big) \Rightarrow \big(X_t,\ \int_0^t \sqrt{E\sigma(s;\lambda_0,\lambda_1,...)}\,dU_s\big),$$

on $D_{R^2}[0,1]$. Let $\Lambda^2(s) = E\sigma(s;\lambda_0,\lambda_1,...)$. An application of Kurtz and Protter (1991) now yields the following result on the convergence to stochastic integrals.

Theorem 3.3. *If $H(x)$ and $H_1(x)$ are continuous functions, then*

$$\begin{aligned}&\Big\{\frac{1}{n}\sum_{k=1}^n H(x_k/d_n),\ \frac{1}{\sqrt{n}}\sum_{k=1}^n H_1(x_k/d_n)\sigma_k\epsilon_k\Big\}\\ &\to_D\ \Big\{\int_0^1 H(X_s)ds,\ \int_0^1 H_1(X_s)\Lambda(s)dU_s\Big\}.\end{aligned} \tag{3.12}$$

If there are more smooth conditions on η_t, we also have the following results on convergence to local time and a mixture of normal distributions.

Theorem 3.4. *Let $g(x)$ and $g_1(x)$ be bounded functions satisfying $\int_{-\infty}^{\infty}(|g(x)|+|g_1(x)|)dx < \infty$. Suppose that $\lim_{t\to\infty}|t|^\epsilon|Ee^{it\eta_1}| < \infty$ for some $\epsilon > 0$.*

(i) We have

$$\begin{aligned}&\Big(\frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\eta_k,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{[nt]}\eta_{-k},\ \frac{d_n}{n}\sum_{k=1}^n g(x_k)\,\sigma_k^2\Big)\\ &\Rightarrow\ \Big(B_t, B_{-t}, \int_{-\infty}^{\infty}g(x)dx\int_0^1\Lambda^2(s)L_X(ds,0)\Big),\end{aligned} \tag{3.13}$$

on $D_{R^3}[0,1]$.

(ii) For any $l(x)$ satisfying $E|l(\epsilon_1)| < \infty$, we have

$$\sum_{k=1}^n g(x_k)\,\sigma_k^{2+\delta}\,l(\epsilon_k) = O_P(n/d_n), \tag{3.14}$$

where $0 \leq \delta < \begin{cases}(2\mu-1)/(3-2\mu), & \text{under LM},\\ 1, & \text{under SM}.\end{cases}$

(iii) We have

$$\Big\{\Big(\frac{d_n}{n}\Big)^{1/2}\sum_{k=1}^n g(x_k)\,\sigma_k\,\epsilon_k,\quad \frac{d_n}{nh}\sum_{k=1}^n g_1(x_k)\Big\}$$
$$\to_D\ \Big\{N\,\widetilde{L}_X^{1/2},\ \int_{-\infty}^{\infty} g_1(s)ds\,L_X(1,0)\Big\}, \tag{3.15}$$

where $\widetilde{L}_X = \int_{-\infty}^{\infty} g^2(x)dx \int_0^1 \Lambda^2(s)L_X(ds,0)$ *and* N *is a standard normal variate independent of* X*.*

Remark 3.6. The past decade has witnessed significant progress in investigating the convergence to local time and the convergence to a mixture of normal distributions. We refer to Wang and Phillips (2009a, 2009b), Jeganathan (2004, 2008), Wang (2014, 2015), Wang, Phillips and Kasparis (2018), Duffy (2019) and references therein. The recent result (3.13) provides an extension of existing results by allowing for σ_k^2 to be a nonlinear process (i.e., $\sigma_k^2 = \sigma(k/n; \lambda_k, \lambda_{k-1}, ...)$. As a consequence, Theorem 3.4 is applicable to time varying GARCH and nonlinear GARCH and many other volatility models, as described in Examples 3.1-3.3.

4 Conclusion

The least squares method is widely used in nonlinear regression analysis and many articles have investigated the asymptotics of least squares estimators. The present paper provides a new framework on asymptotic theory for general nonlinear regression models with heteroscedasticity. Our results apply to various nonlinear models with stationary and nonstationary regressors and the conditions imposed are straightforward and easy to verify. The author hopes the results derived in this paper prove useful in related areas, particularly in nonlinear cointegrating regressions where nonstationarity and nonlinearity play a role.

REFERENCES

Andrew, D. W. K. (1994). Empirical process methods in econometrics. *Chapter 37, Handbook of Econometrics,* **5**, 2248-2294.

Andrew, D. W. K. (1995). Nonparametric kernel estimation for semiparametric models. *Econometric Theory* **11**, 560–596.

Andrews, D. W. K. and Sun, Y. (2004). Adaptive local polynomial Whittle estimation of long-range dependence. *Econometrica*, 72, 569-614.

Bae, Y. and de Jong, R. (2007). Money demand function estimation by nonlinear cointegration. *Journal of Applied Econometrics*, 22(4), 767–793.

Bae, Y. Kakkar, V. and Ogaki, M. (2004). Money demand in Japan and the liquidity trap. Ohio State University Department of Economics Working Paper #04–06.

Bae, Y., Kakkar, V., and Ogaki, M. (2006). Money demand in Japan and nonlinear cointegration. *Journal of Money, Credit, and Banking*, 38(6), 1659–1667.

Bercu, B. and Touati, A. (2008). Exponential inequalities for self-normalized martingales with applications. *Ann. Appl. Probab.* **18**, 1848-1869.

Boswijk, H.P., Cavaliere, G., Rahbek, A. and Taylor, A.M.R. (2016). Inference on cointegration parameters in heteroskedastic vector autoregressions. *Journal of Econometrics*, **192**, 6485.

Cavaliere, G. and A.M.R. Taylor (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics* **140**, 919 - 947.

Cavaliere G. and A.M.R. Taylor (2009). Heteroskedastic time series with a unit root, *Econometric Theory* **25**, 12281276.

Cavaliere, G. and A.M.R. Taylor (2010). Testing for co-integration in vector autoregressions with non-stationary volatility *Journal of Econometrics* **158**, 7-24

Chan, N., Wang, Q. (2014). Uniform convergence for nonparametric estimators with nonstationary data. *Econometric Theory*, **30**, 1110-1133.

Chan, N. and Wang, Q. (2015). Nonlinear regression with nonstationary time series, *Journal of Econometrics*, **185**, 182-195.

Chang, Y., Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear econometric models with cointegrated and deterministically trending regressors. Econometric Journal, 4, 1–36.

Chang, Y. and Park, J. Y. (2003). Index models with integrated time series. *Journal of Econometrics*, **114**, 73-106.

De Jong, R. and Hu, L. (2011). A note on nonlinear models with integrated regressors and convergence order results, *Econometric Letters*, **111**, 23-25.

Duffy, JA (2016). A uniform law for convergence to the local times of linear fractional stable motions. *The Annals of Applied Probability*, **25(1)**, 45-72.

Dong, C. and Linton, O. (2018). Additive nonparametric models with time variables and both stationary and nonstationary regressors. *Journal of Econometrics*, **207**, 212-236.

Dong, C. Gao, J. and Tjstheim, D. (2016). Estimation for single-index and partially linear single-index integrated model. *Ann. Statist*, **44**, 425453.

Duffy, JA (2017). Uniform convergence rates over maximal domains in structural nonparametric cointegrating regression. *Econometric Theory*, **33**, 1387-1417.

Gao, J. Kanaya, S. Li, D. and Tjöstheim, D. (2015). Uniform consistency for nonparametric estimators in null recurrent time series. *Econometric Theory*, **31**, 911-952.

Goncalves, S. and L. Kilian (2004). Bootstrapping autoregressions with conditional heteroscedasticity of unknown form. *Journal of Econometrics* **123**, 89120.

Hansen, B.E. (1995) Regression with nonstationary volatility. *Econometrica*, **63**, 11131132.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.

Kim, S. K. and Kim, I. M. (2012). Partial parametric estimation for nonstationary nonlinear regressions. *Journal of Econometrics*, **167**, 448-457.

Jacob, C. (2010). Conditional least squares estimation in nonstationary nonlinear stochastic regression models. *Ann. Statist.* **38,** 566–597.

Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimation. *Ann. Math. Statist.* **40**, 633-643.

Jeganathan, P. (2004). Convergence of functionals of sums of r.v.s to local times of fractional stable motions, *Ann. Probab.* **32**, 1771-1795.

Jeganathan, P. (2008). Limit theorems for functional sums that converge to fractional Brownian and stable motions. Cowles Foundation Discussion Paper No. 1649, Cowles Foundation for Research in Economics, Yale University.

Kurtz, T. G. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *Annals of Probability*, **19**, 1035–1070.

Lai, T. L. (1994). Asymptotic theory of nonlinear least square estimations. *The Annals of Statistics*, **22**, 1917–1930.

Lanne, M. and Saikkonen, P. (2005). Nonlinear GARCH models for highly persistency volatility, *Econometric Theory*, **8**, 251-276.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis. *Chapter 36, Handbook of Econometrics,* **5**, 2113-2245.

Nishiyama, Y. (2000a). Weak convergence of some classes of martingales with jumps. *Ann. Probab.*, **28**, 685-712.

Nishiyama, Y. (2000b). Entropy methods for martingale. *CWI, Amsterdam.*

Nishiyama, Y. (2007). On the paper "Weak convergence of some classes of martingales with jumps". *Ann. Probab.*, **35**, 1194-1200.

Park, J. Y. and P. C. B. Phillips (2001). Nonlinear regressions with integrated time series, *Econometrica*, **69**, 117-161.

Peng, J. and Wang, Q. (2018). We ak convergence to stochastic integrals under primitive conditions in nonlinear econometric models, *Econometric Theory*, **34**, 1132-1157.

Pollard, D. and Radchenko, P. (2006). Nonlinear least squares estimation. *Journal of Multi. Analysis,* **97**, 548–562.

Shao, X. and W-B. Wu (2007). Asymptotic spectral theory for nonlinear time series. *Ann. Statist,* **35**, 1773-1801.

Skouras, K. (2000). Strong consistency in nonlinear stochastic regression models. *The Annals of Statistics*, **28**, 871–879.

Taniguchi, M and Kakizawa, Y. (2000). *Asymptotic Theory of Statistical Inference for Time Series,* New York: Springer.

Wang, Q. (2014). Martingale limit theorems revisited and non-linear cointegrating regression. *Econometric Theory*, 30, 509-535.

Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression.* World Scientific.

Wang, Q. and Phillips, P.C.B. (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* **25**, 710-738.

Wang, Q. and Phillips, P.C.B. (2009b). Structural Nonparametric Cointegrating Regression, *Econometrica*, **25**, 1901-1948.

Wang, Q. and Phillips, P.C. B.(2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory,* **32,** 359–401.

Wang, Q., Phillips, P.C. B. and Kasparis, I. (2018). Latent Variable Nonparametric Cointegrating Regression, manuscript.

Wang, Q., Lin, Y. X., Gulati, C. M. (2003). Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory*, 19, 143–164.

Wooldridge, J. (1994). Estimation and inference for dependent processes. *Handbook of Econometrics*, R.F. Engle, D.L. McFadden (Eds.), **Vol. IV**, Elsevier, Amsterdam. 2639-2738.

Wu, C. F. (1981). Asymptotic theory of nonlinear least square estimations. *Annals of Statistics*, **9**, 501–513.

Wu, W. B. and Min, W. (2005). On linear processes with dependent innovations. *Stochastic processes and their applications*, **115**, 939-958.

A Proofs of the main results in Section 2

We start with the proof of Theorem 2.3 and then those of Corollary 2.1 because the results of Corollary 2.1 are required to prove Theorem 2.1. The proofs of Theorems 2.1 and 2.2 and Proposition 2.1 will follow. The notations are the same as in the previous sections except that they are mentioned explicitly.

Proof of Theorem 2.3. Let $M_{n,k}(f) = \sum_{t=1}^{k} u_{nt}\, f(y_{nt})$ and $Z_{n,k}(f) = \sum_{t=1}^{k} f^2(y_{nt})\big[u_{nt}^2 + E(u_{nt}^2 \mid \mathcal{F}_{n,t-1})\big]$, where $1 \le k \le n$. Let $M_n(f) = M_{n,n}(f)$ and $Z_n(f) = Z_{n,n}(f)$. For any $\epsilon > 0$, by (2.8), there exist A_0 and n_0 such that, for all $n \ge n_0$,

$$P\big(\sup_{f\in\Psi} Z_n(f) \ge A_0\gamma_n\big) \le \epsilon/2. \tag{A.1}$$

Note that, for any $A > 0$,

$$\begin{aligned} & P\big(\sup_{f\in\Psi} |M_n(f)| \ge A\big[\gamma_n \log^+(\#\Psi)\big]^{1/2}\big) \\ \le\ & P\big(\sup_{f\in\Psi} Z_n(f) \ge A_0\gamma_n\big) + \sum_{f\in\Psi} P\big(|M_n(f)| \ge A\big[\gamma_n \log^+(\#\Psi)\big]^{1/2},\ Z_n(f) \le A_0\gamma_n\big). \end{aligned}$$

Result (2.9) will follow if we prove the following: for each $f \in \Psi$ and $m \ge 2$

$$\begin{aligned} & P\big(|M_n(f)| \ge \sqrt{2m\, A_0}\, \big[\gamma_n \log^+(\#\Psi)\big]^{1/2},\ Z_n(f) \le A_0\gamma_n\big) \\ \le\ & C \min\{(\#\Psi)^{-m}, e^{-m}\}. \end{aligned} \tag{A.2}$$

To prove (A.2), we need the following fact:

F: If X is a random variable such that $E(X \mid \mathcal{F}) = 0$ and $E(X^2 \mid \mathcal{F}) < \infty$, then, for any $t > 0$,

$$E\Big[e^{tX - \frac{t^2}{2}X^2} \mid \mathcal{F}\Big] \ \le\ 1 + \frac{t^2}{2} E\big(X^2 \mid \mathcal{F}\big). \tag{A.3}$$

The fact **F** is a conditional version of Lemma 2.1 in Bercu and Touati (2010). Indeed, by using Jensen's inequality, we obtain

$$\begin{aligned} A(t) \ &:= \ E\Big[e^{tX - \frac{t^2}{2}X^2} \mid \mathcal{F}\Big] \ge e^{-\frac{t^2}{2}E(X^2|\mathcal{F})} \\ &\ge \ 1 - \frac{t^2}{2} E(X^2 \mid \mathcal{F}), \end{aligned}$$

for any $t \in R$. On the other hand, we have

$$A(t) + A(-t) = E\Big[e^{-\frac{t^2}{2}X^2}\big(e^{tX} + e^{-tX}\big) \mid \mathcal{F}\Big] \le 2,$$

indicating that

$$A(t) \le 2 - A(-t) \le 1 + \frac{t^2}{2} E(X^2 \mid \mathcal{F}).$$

We next prove (A.2) by using similar arguments as in the proof of Theorem 2.1 of Bercu and Touati (2010). Let $A_n^+ = \{M_n(f) \ge x,\ Z_n(f) \le y\}$, where $x > 0$ and $y > 0$. By using Cauchy-Schwartz inequality, for all $t > 0$, we have

$$\begin{aligned} P(A_n^+) &\le E\Big[e^{(tM_n(f)-tx)/2}\mathrm{I}(A_n^+)\Big] \\ &= E\Big[e^{\frac{t}{2}M_n(f)-\frac{t^2}{4}Z_n(f)}\, e^{\frac{t^2}{4}Z_n(f)-\frac{tx}{2}}\,\mathrm{I}(A_n^+)\Big] \\ &\le e^{\frac{t^2y}{4}-\frac{tx}{2}}\sqrt{EV_{n,n}(t)\,P(A_n^+)}, \end{aligned} \tag{A.4}$$

where, for $k \ge 1$, $V_{n,k}(t) = e^{t\,M_{n,k}(f)-\frac{t^2}{2}Z_{n,k}(f)}$. Write $x_k = f(y_{nk})u_{nk}$. Following the fact $\mathbf{F}$,

$$\begin{aligned} & E\Big[e^{tx_k-\frac{t^2}{2}x_k^2-\frac{t^2}{2}E(x_k^2|\mathcal{F}_{n,k-1})\big)} \mid \mathcal{F}_{n,k-1}\Big] \\ = \ & e^{-\frac{t^2}{2}E(x_k^2|\mathcal{F}_{n,k-1})}E\Big[e^{tx_k-\frac{t^2}{2}x_k^2} \mid \mathcal{F}_{n,k-1}\Big] \\ \le \ & e^{-\frac{t^2}{2}E(x_k^2|\mathcal{F}_{n,k-1})}\big[1+\frac{t^2}{2}E\big(x_k^2 \mid \mathcal{F}_{n,k-1}\big)\big] \le 1, \end{aligned}$$

we obtain the following: for $k \ge 1$,

$$\begin{aligned} E\big[V_{n,k}(t) \mid \mathcal{F}_{n,k-1}\big] &\le V_{n,k-1}(t)\,E\Big[e^{tx_k-\frac{t^2}{2}x_k^2-\frac{t^2}{2}E(x_k^2|\mathcal{F}_{n,k-1})\big)} \mid \mathcal{F}_{n,k-1}\Big] \\ &\le V_{n,k-1}(t). \end{aligned}$$

It follows that

$$EV_{n,k}(t) \le EV_{n,k-1}(t) \le \ldots \le EV_{n,1}(t) = E\big\{E\big[V_{n,1}(t) \mid \mathcal{F}_{n,0}\big]\big\} \le 1.$$

By taking this estimate into (A.4), we obtain

$$P(A_n^+) \le \inf_{t>0}\exp\big(\frac{t^2y}{2} - tx\big) = e^{-x^2/2y}$$

Let $A_n^- = \{M_n(f) \le -x,\ Z_n(f) \le y\}$, where $x > 0$ and $y > 0$. Similarly, we have $P(A_n^-) \le e^{-x^2/2y}$. As a consequence, we obtain

$$P(|M_n(f)| \ge x,\ Z_n(f) \le y) \le P(A_n^+) + P(A_n^-) \le 2e^{-x^2/2y},$$

yielding (A.2) by taking $x = \sqrt{2m\,A_0}(\gamma_n \log^+(\#\Psi))^{1/2}$ and $y = A_0\gamma_n$. $\square$

Proof of Corollary 2.1. We first show that (2.1) implies that, for $\forall\epsilon > 0$, there exist n_0 and A_0 such that, for all $n \geq n_0$,

$$P\Big(\sum_{t=1}^{n} T_{nt}^2\, u_t^2 \geq A_0^2\Big) \leq \epsilon, \quad \text{i.e.,} \quad \sum_{t=1}^{n} T_{nt}^2\, u_t^2 = O_P(1). \tag{A.5}$$

To prove (A.5), let $\nu_n(A) = \inf\big\{k \geq 1 : \sum_{t=1}^{k} T_{nt}^2\, E(u_t^2|\mathcal{F}_{t-1}) \geq A\big\}$, $k_n(A) = \big[\nu_n(A) - 1\big] \wedge n$ and $R_n(A) = \sum_{t=1}^{k_n(A)} T_{nt}^2\, E(u_t^2|\mathcal{F}_{t-1})$. It is readily seen that

$$R_n(A) \leq \sum_{t=1}^{\nu_n(A)-1} T_{nt}^2\, E(u_t^2|\mathcal{F}_{t-1}) \leq A$$

and, as a consequence,

$$ER_n(A) \;=\; E\Big(\sum_{t=1}^{k_n(A)} T_{nt}^2\, u_t^2\Big) \leq A, \quad \text{for each } A > 0.$$

On the other hand, for $\forall\epsilon > 0$, there exist $A_0 \geq 2/\epsilon$ and n_0 such that, for all $n \geq n_0$,

$$P\Big(\sum_{t=1}^{n} T_{nt}^2\, E(u_t^2|\mathcal{F}_{t-1}) \geq A_0\Big) \leq \epsilon/2,$$

due to (2.1). Now, for $\forall\epsilon > 0$, we have

$$\begin{aligned}
P\Big(\sum_{t=1}^{n} T_{nt}^2\, u_t^2 \geq A_0^2\Big) &\leq P\big[k_n(A_0) \neq n\big] + P\Big(\sum_{t=1}^{k_n(A_0)} T_{nt}^2\, u_t^2 \geq A_0\Big) \\
&\leq P\big(\nu_n(A_0) \leq n\big) + A_0^{-2}\, E\Big(\sum_{t=1}^{k_n(A_0)} T_{nt}^2\, u_t^2\Big) \\
&\leq P\Big(\sum_{t=1}^{n} T_{nt}^2\, E(u_t^2|\mathcal{F}_{t-1}) \geq A_0\Big) + A_0^{-2}\, ER_n(A_0) \\
&\leq \epsilon/2 + A_0^{-1} \leq \epsilon,
\end{aligned}$$

which yields (A.5).

We are now ready to prove Corollary 2.1, starting with (2.4). Let $\theta_j \in \Theta$ be different so that $\Theta_j = \{\theta : ||\theta - \theta_j|| \leq n^{-1/\alpha}\}$, $j = 1, 2, ..., n^k$, for some integer $k \geq 1/\alpha$, covers Θ, where $0 < \alpha \leq 1$ is given as in Assumption 2.2 (i). Note that, by Hölder's inequality,

$$\begin{aligned}
\sup_{\theta\in\Theta_j} ||Z_n(\theta) - Z_n(\theta_j)|| &\leq \sup_{\theta\in\Theta_j} \sum_{t=1}^{n} ||D_n^{-1}\big[\dot{g}_t(\theta) - \dot{g}_t(\theta_j)\big]||\, |u_t| \\
&\leq n^{-1} \sum_{t=1}^{n} |T_{nt}|\, |u_t| \\
&\leq n^{-1/2}\Big(\sum_{t=1}^{n} T_{nt}^2 u_t^2\Big)^{1/2} = O_P(n^{-1/2}),
\end{aligned} \tag{A.6}$$

for any $1 \le j \le n^k$, due to (A.5). Result (2.4) will follow if we prove

$$\max_{1\le j\le n^k} ||Z_n(\theta_j) - Z_n(\theta_0)|| = O_P(\log^{1/2} n). \tag{A.7}$$

In fact (A.7) follows immediately from Theorem 2.3 with $\gamma_n = 1$ owing to the following facts: $Z_n(\theta_j) - Z_n(\theta_0) = \sum_{k=1}^n D_n^{-1}\big[\dot g_t(\theta) - \dot g_t(\theta_j)\big]\, u_k$ and

$$\begin{aligned} I_n &:= \max_{1\le j\le n^k} \sum_{t=1}^n ||D_n^{-1}\big[\dot g_t(\theta) - \dot g_t(\theta_j)\big]||^2 \big[u_t^2 + E(u_t^2 \mid \mathcal{F}_{t-1})\big] \\ &\le \max_{1\le j\le n^k} ||\theta_j - \theta_0||^{2\alpha} \sum_{t=1}^n T_{nt}^2 \big[u_t^2 + E(u_t^2 \mid \mathcal{F}_{t-1})\big] = O_P(1). \end{aligned} \tag{A.8}$$

The proof of (2.5) is similar except that θ_j is chosen so that $||D_n(\theta_j - \theta_0)|| \le \log n$. In this case, instead of (A.8), we have

$$\begin{aligned} I_n &\le \max_{1\le j\le n^k} ||D_n^{-1}\, D_n(\theta_j - \theta_0)||^{2\alpha} \sum_{t=1}^n T_{nt}^2 \big[u_t^2 + E(u_t^2 \mid \mathcal{F}_{t-1})\big] \\ &= O_P(||D_n^{-1}\, \log n||^{2\alpha}) = O_P(\log^{-3} n), \end{aligned}$$

Now it follows from Theorem 2.3 with $\gamma_n = \log^{-3} n$ that

$$\max_{1\le j\le n^k} ||Z_n(\theta_j) - Z_n(\theta_0)|| = O_P(\log^{-1} n) = o_P(1).$$

This, together with (A.6), yields (2.5). □

Proof of Theorem 2.1. Let $\dot Q_n$ be the first derivative of $Q_n(\theta)$ so that $\dot Q_n = \partial Q_n/\partial\theta$. Let $f_k(\theta) = g_k(\theta) - g_k(\theta_0)$ and

$$Q_n(\theta) = \sum_{k=1}^n \big[y_k - g_k(\theta)\big]^2 = \sum_{k=1}^n \big[u_k - f_k(\theta)\big]^2.$$

We have

$$\begin{aligned} \dot Q_n(\theta) &= -\sum_{t=1}^n \dot g_t(\theta)(u_t - f_t(\theta)) \\ &= -\sum_{t=1}^n \dot g_t(\theta_0) u_t + \sum_{t=1}^n \dot g_t(\theta_0)\dot g_t(\theta_0)'(\theta - \theta_0) \\ &\qquad + R_{1n}(\theta) + R_{2n}(\theta) + R_{3n}(\theta), \end{aligned} \tag{A.9}$$

where

$$
\begin{aligned}
R_{1n}(\theta) &= \sum_{t=1}^{n} \big[\dot g_t(\theta_0) - \dot g_t(\theta)\big]\, u_t, \\
R_{2n}(\theta) &= \sum_{t=1}^{n} \big[\dot g_t(\theta) - \dot g_t(\theta_0)\big]\, f_t(\theta), \\
R_{3n}(\theta) &= \sum_{t=1}^{n} \dot g_t(\theta_0)\big[f_t(\theta) - \dot g_t(\theta_0)'(\theta-\theta_0)\big]
\end{aligned}
$$

Given that $f_t(\theta_0) = 0$, it follows from Taylor's expansion and Assumption 2.2(i) that, for some $\widetilde\theta$ between θ_0 and θ,

$$
\begin{aligned}
|f_t(\theta) - \dot g_t(\theta_0)'(\theta-\theta_0)| &\le \big|\big[\dot g_t(\widetilde\theta) - \dot g_t(\theta_0)\big]'(\theta-\theta_0)\big| \\
&\le d(\widetilde\theta, \theta_0)\, T_{nt}\, ||D_n(\theta-\theta_0)||. \qquad \text{(A.10)}
\end{aligned}
$$

As a consequence, by recalling $||\hat\theta_n - \theta_0|| \to_P 0$ and $d(\theta,\theta_0)$ as $\theta \to \theta_0$, we have

$$
\begin{aligned}
& ||D_n^{-1}\, R_{3n}(\hat\theta_n)|| \\
=\ & o_P(1)\, ||D_n(\hat\theta_n - \theta_0)|| \sum_{t=1}^{n} ||D_n^{-1}\dot g_t(\theta_0)||\, |T_{nt}| \\
=\ & o_P(1)\, ||D_n(\hat\theta_n - \theta_0)||\, \Big(\sum_{t=1}^{n} ||D_n^{-1}\dot g_t(\theta_0)||^2\Big)^{1/2} \Big(\sum_{t=1}^{n} T_{nt}^2\Big)^{1/2} \\
=\ & o_P\big(||D_n(\hat\theta_n - \theta_0)||\big),
\end{aligned}
$$

where we have used $\sum_{t=1}^{n} ||D_n^{-1}\dot g_t(\theta_0)||^2 = O_P(1)$ and $\sum_{t=1}^{n} T_{nt}^2 = O_P(1)$ due to Assumption 2.2. Similarly, $||D_n^{-1}\, R_{2n}(\hat\theta_n)|| = o_P\big(||D_n(\hat\theta_n - \theta_0)||\big)$.

Let $\theta = \hat\theta_n$ in (A.9). Given that $\dot Q_n(\hat\theta_n) = 0$, it follows from these estimates and (A.9) that

$$
0 = -D_n^{-1} \sum_{t=1}^{n} \dot g_t(\theta_0) u_t + Y_n\, D_n(\hat\theta_n - \theta_0) + D_n^{-1} R_{1n}(\hat\theta_n) + o_P\big[||D_n(\hat\theta_n - \theta_0)||\big],
$$

i.e. $\big[$by noting $||Y_n^{-1}|| = O_P(1)\big]$,

$$
\begin{aligned}
& D_n(\hat\theta_n - \theta_0) \\
=\ & Y_n^{-1} Z_n(\theta_0) - Y_n^{-1} D_n^{-1} R_{1n}(\hat\theta_n) + o_P\big[||D_n(\hat\theta_n - \theta_0)||\big]. \qquad \text{(A.11)}
\end{aligned}
$$

Given that $||Y_n^{-1} Z_n(\theta_0)|| = O_P(1)$, by using (A.11), (2.2) will follow if we prove

$$
||D_n^{-1} R_{1n}(\hat\theta_n)|| = o_P(1). \qquad \text{(A.12)}
$$

Indeed, if (A.12) holds, then $||D_n(\hat{\theta}_n - \theta_0)|| = O_P(1)$ by (A.11). Invoking this estimate into (A.11) again, we obtain

$$D_n(\hat{\theta}_n - \theta_0) \;=\; Y_n^{-1} Z_n(\theta_0) + o_P(1),$$

as required.

We use Corollary 2.1 to prove (A.12). First note that, by (2.4),

$$||D_n^{-1} R_{1n}(\hat{\theta}_n)|| \le \sup_{\theta\in\Theta} |Z_n(\theta) - Z_n(\theta_0)| = O_P(\log^{1/2} n).$$

By taking this estimate into (A.11), we get $||D_n(\hat{\theta}_n - \theta_0)|| = O_P(\log^{1/2} n)$. Now it follows from (2.5) that, for any $\eta > 0$,

$$\begin{aligned} & P\big(||D_n^{-1} R_{1n}(\hat{\theta}_n)|| \ge \eta\big) \\ \le\; & P\big(||D_n(\hat{\theta}_n - \theta_0)|| \ge \log n\big) + P\Big(\sup_{\theta:||D_n(\theta-\theta_0)||\le \log n} |Z_n(\theta) - Z_n(\theta_0)| \ge \eta\Big) \\ \to\; & 0, \qquad \text{as } n \to \infty, \end{aligned}$$

implying (A.12). The proof of Theorem 2.1 is complete. □

Proof of Theorem 2.2. Write $L_{n,\theta} = \sum_{k=1}^n d_k(\theta)\, u_k$ and $Q_n(\theta) = \sum_{k=1}^n \big[y_k - g_k(\theta)\big]^2$. Given that θ_0 is a real value of the model (1.1), we have

$$\sum u_k^2 \;=\; Q_n(\theta_0) \ge Q_n(\widehat{\theta}_n) = \sum u_k^2 + D_{n,\widehat{\theta}} - 2L_{n,\widehat{\theta}}\,,$$

thus indicating that, for any $\epsilon > 0$,

$$\begin{aligned} P(||\widehat{\theta}_n - \theta_0|| \ge \epsilon) \;&\le\; P\Big(\sup_{||\theta-\theta_0||\ge\epsilon} |L_{n,\theta}| / D_{n,\theta} \ge 1/2\Big) \\ &\le\; P\Big(\sup_{\theta\in\Theta} |L_{n,\theta}| \ge 1/2 \inf_{||\theta-\theta_0||\ge\epsilon} D_{n,\theta}\Big). \end{aligned}$$

Hence, by using condition (b), Theorem 2.2 will follow if we prove

$$\sup_{\theta\in\Theta} |L_{n,\theta}| = o_P(k_n). \tag{A.13}$$

The proof of (A.13) is similar to that of Corollary 2.1 with minor modifications. We omit the details. □

Proof of Proposition 2.1. First note that, for any $\delta > 0$ and $\lambda_0 > 0$, an integer m_0 that depends only on λ_0 and $\theta_j, j = 1, ..., m_0$, exists such that

$$||\theta_j - \theta_0|| \ge \delta \quad \text{and} \quad \Theta \cap \{||\theta - \theta_0|| \ge \delta\} \;\subset \cup_{j=1}^{m_0} \mathcal{N}_{\lambda_0}(\theta_j),$$

where $\mathcal{N}_x(y) = \{\theta : ||\theta - y|| \le x\}$. Given that $|g_t(\theta_1) - g_t(\theta_2)| \le h\big(||\theta_1 - \theta_2||\big)\, T_t$, it is readily seen that, uniformly for $\theta \in \mathcal{N}_{\lambda_0}(\theta_j)$, $j = 1, 2, ..., m_0$,

$$\begin{aligned} |D_{n,\theta} - D_{n,\theta_j}| &\le \Big(\sum_{t=1}^{n} \big[g_t(\theta) - g_t(\theta_j)\big]^2\Big)^{1/2} \Big(\sum_{t=1}^{n} \big[g_t(\theta) + g_t(\theta_j) - 2g_t(\theta_0)\big]^2\Big)^{1/2} \\ &\le 2 \sup_{\theta\in\Theta} h^{1/2}(||\theta||) \max_{1\le j\le m_0} \sup_{\theta\in\mathcal{N}_{\lambda_0}(\theta_j)} h^{1/2}\big(||\theta - \theta_1||\big) \sum_{t=1}^{n} T_t^2. \end{aligned}$$

Hence, by recalling that $k_n^{-1} \sum_{t=1}^{n} T_t^2 = O_P(1)$ and $h(x)$ is continuous with $\lim_{x\downarrow 0} h(x) = 0$, we have

$$k_n^{-1} \max_{1\le j\le m_0} \sup_{\theta\in\mathcal{N}_{\lambda_0}(\theta_j)} |D_{n,\theta} - D_{n,\theta_j}| = o_P(1), \tag{A.14}$$

for any $\delta > 0$, as $\lambda_0 \to 0$ ($m_0 \to \infty$, respectively) and $n \to \infty$. Now, to prove Proposition 2.1, it suffices to show for any $\delta > 0$, $m_0 \ge 1$ and $\eta > 0$ that there exist M_δ (depending only on δ) and n_0 such that for all $n \ge n_0$,

$$P\big(k_n^{-1} \min_{1\le j\le m_0} D_{n,\theta_j} \ge M_\delta\big) \ge 1 - \eta. \tag{A.15}$$

In fact, by using condition (i) and taking $M_\delta > 0$ as in condition (i), there exists an n_0 such that, uniformly for $j = 1, ..., m_0$ and $n \ge n_0$,

$$\begin{aligned} P\big(k_n^{-1} D_{n,\theta_j} \ge M_\delta\big) &\ge P\big[G(\theta_j) \ge M_\delta\big] - \eta/m_0 \\ &\ge P\big[\inf_{\substack{||\theta-\theta_0||\ge\delta \\ \theta\in\Theta}} G(\theta) \ge M_\delta\big] - \eta/m_0 = 1 - \eta/m_0. \end{aligned}$$

for any $m_0 \ge 1$ and $\eta > 0$. We now have

$$P\big(k_n^{-1} \min_{1\le j\le m_0} D_{n,\theta_j} < M_\delta\big) \le \sum_{j=1}^{m_0} P\big(k_n^{-1} D_{n,\theta_j} < M_\delta\big) < \eta,$$

implying (A.15). $\square$

B Proofs of the main results in Section 3

We prove Theorems 3.1 and 3.2 by checking the conditions of Theorem 2.1, where certain fundamental results in Theorems 3.3 and 3.4 are required. Theorem 3.3 is well-known in the literature. The proof of Theorem 3.4 will be given in the end of this section.

Proof of Theorem 3.1. It suffices to show that the consistency of $\hat{\theta}_n$, Assumption 2.2(i) and (2.3) hold. In fact, under the conditions of Theorem 3.1, it follows from Theorem 3.4 that

$$\begin{aligned} &\frac{d_n}{n}\sum_{t=1}^{n}\big[f(x_t,\theta)-f(x_t,\theta_0)\big]^2 \\ &\qquad \to_D G(\theta):=\int_{-\infty}^{\infty}\big[f(x,\theta)-f(x,\theta_0)\big]^2 dx L_X(1,0); \qquad \text{(B.1)} \\ &\frac{d_n}{n}\sum_{t=1}^{n}T_p^2(x_t)(1+\sigma_t^2)=O_P(1), \qquad \text{(B.2)} \end{aligned}$$

where $T_p(x)$ is given in Assumption 3.3(ii) with $p(x,\theta)=f(x,\theta)$ or $\dot{f}_j(x,\theta), j=1,...,q$; and for any $\alpha_i'=(\alpha_{i1},...,\alpha_{iq})\in\mathbb{R}, i=1,2,3,$

$$\begin{aligned} &\Big\{\big(\frac{d_n}{n}\big)^{1/2}\sum_{k=1}^{n}\alpha_3'\dot{f}(x_k,\theta_0)\,\sigma_k\,\epsilon_k,\ \ \frac{d_n}{n}\sum_{k=1}^{n}\alpha_1'\dot{f}(x_k,\theta_0)\dot{f}(x_k,\theta_0)'\alpha_2\Big\} \\ &\quad \to_D\ \Big\{\sqrt{\alpha_3'\Sigma\alpha_3}\,N\sqrt{\int_0^1\Lambda^2(s)dL_X(s,0)},\ \alpha_1'\Sigma\alpha_2\,L_X(1,0)\Big\} \\ &\quad =_D\ \big\{\alpha_3'\Sigma^{1/2}\,\mathbb{N}\sqrt{\int_0^1\Lambda^2(s)dL_X(s,0)},\ \alpha_1'\Sigma\alpha_2\,L_X(1,0)\big\}, \qquad \text{(B.3)} \end{aligned}$$

where $\mathbb{N}$ is a q-dimensional standard normal vector independent of X.

Given that $P(L_X(1,0)>0)=1$, we have $P\big[\inf_{\substack{||\theta-\theta_0||\geq\delta\\ \theta\in\Theta}}G(\theta)\geq M_\delta\big]=1$, where $M_\delta^{-1}=\inf_{\substack{||\theta-\theta_0||\geq\delta\\ \theta\in\Theta}}\int_{-\infty}^{\infty}\big[f(x,\theta)-f(x,\theta_0)\big]^2dx>0$ for each $\delta>0$. The consistency of $\hat{\theta}_n$ follows from (B.1), (B.2) with $p(x,\theta)=f(x,\theta)$ and Proposition 2.1 with $g_t(\theta)=f(x_t,\theta)$ and $k_n=n/d_n$. Assumption 2.2(i) follows from (B.2) with $\dot{g}_t(\theta)=f(x_t,\theta)$ by setting $D_n=\sqrt{n/d_n}\mathrm{diag}(1,...,1)$, and (2.3) follows from (B.3) with $M=\Sigma\,L_X(1,0)$ and $Z=\Sigma^{1/2}\sqrt{\int_0^1\Lambda^2(s)dL_X(s,0)}\,\mathbb{N}$. The proof of Theorem 3.1 is complete. □

Proof of Theorem 3.2. Similar to the proof of Theorem 3.1, it only needs to show the following results under Assumptions 3.1-3.2 and 3.4: (The notations are the same in these assumptions except that they are mentioned explicitly)

$$\begin{aligned} &\frac{1}{nv_f^2(d_n)}\sum_{t=1}^{n}\big[f(x_t,\theta)-f(x_t,\theta_0)\big]^2 \\ &\qquad \to_D G(\theta):=\int_0^1\big[h_f(X_t,\theta)-h_f(X_t,\theta_0)\big]^2dt, \qquad \text{(B.4)} \\ &\frac{1}{n}\sum_{t=1}^{n}T_{1p}^2(x_t/d_n)(1+\sigma_t^2)=O_P(1), \qquad \text{(B.5)} \end{aligned}$$

where $T_{1p}(x)$ is given in Assumption 3.4(iii) with $p(x,\theta) = f(x,\theta)$ or $\dot f_j(x,\theta), j = 1, ..., q$, and for any $\alpha_i' = (\alpha_{i1}, ..., \alpha_{iq}) \in \mathbb{R}, i = 1, 2, 3$,

$$\begin{aligned}
&(\alpha_1' \, Y_n \alpha_2, \quad \alpha_3' W_n) \\
&\to_D \Big(\int_0^1 \alpha_1' \, \Psi(t)\Psi(t)' \, \alpha_2 dt, \; \int_0^1 \alpha_3' \Psi(t) \, \Lambda(t) \, dU_t \Big),
\end{aligned} \tag{B.6}$$

where $Y_n = (D_n^{-1})' \sum_{t=1}^n \dot f(x_t,\theta_0) \dot f(x_t,\theta_0)' \, D_n^{-1}$ and

$$W_n = (D_n^{-1})' \sum_{t=1}^n \dot f(x_t,\theta_0) \, \sigma_t \epsilon_t.$$

Indeed, due to the continuity of X_t, condition (3.10) ensures $P\big[\inf_{\substack{||\theta-\theta_0||\geq\delta \\ \theta\in\Theta}} G(\theta) \geq M_\delta\big] = 1$ for some $M_\delta > 0$, a.s. This, together with (B.5) with $p(x,\theta) = f(x,\theta)$, implies the consistency of $\hat\theta_n$ by Proposition 2.1. On the other hand, (3.10) and (B.6) yields $Y_n \to \int_0^1 \Psi(t)\Psi(t)' \, dt > 0$. Hence Theorem 3.2 follows from an easy application of Theorem 2.1.

We only prove (B.4). Others can be obtained by similar arguments with minor modification and hence the details are omitted. Write $x_t^* = x_t I(|x_t|/d_n \leq A)$,

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

For any fixed $A > 0$, it follows from Assumptions 3.4(ii) and (iii) that

$$\sup_{\theta\in\Theta} |R_n^*(\theta)| \leq o(1) \, v_f^2(d_n) \sum_{t=1}^n T_{1f}^2(x_t^*/d_n) = o(1) n v_f^2(d_n),$$

as $n \to \infty$. This implies that, for any $\epsilon > 0$,

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

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

$$\sup_{\theta\in\Theta} |R_n(\theta)| \; = \; o_P\big[nv_f^2(d_n)\big]. \tag{B.7}$$

Now, by letting

$$\Delta_n = \sup_{\theta\in\Theta} |R_n(\theta)| + \sup_{\theta\in\Theta} |R_n(\theta)|^{1/2} \Big(\frac{1}{n}\sum_{t=1}^{n} \big[h_f(x_t/d_n,\theta) - h_f(x_t/d_n,\theta_0)\big]^2\Big)^{1/2},$$

and noting that

$$\begin{aligned} & \frac{1}{nv_f^2(d_n)}\sum_{t=1}^{n} \big[f(x_t,\theta) - f(x_t,\theta_0)\big]^2 \\ = \; & \frac{1}{n}\sum_{t=1}^{n} \big[h_f(x_t/d_n,\theta) - h_f(x_t/d_n,\theta_0)\big]^2 + 4\,\widetilde{\Delta}_n, \end{aligned}$$

where $|\widetilde{\Delta}_n| \le \Delta_n$, result (B.4) follows from (B.7) and Theorem 3.3. □

Proofs of Theorem 3.4. We start with two preliminary lemmas. Recall that $\{\lambda_k^*\}_{k\in\mathbb{Z}}$ is an independent copy of $\{\lambda_k\}_{k\in\mathbb{Z}}$. Let $\lambda^* = (\lambda_1^*, \lambda_2^*, ...)$.

Lemma B.1. *Let $p(u, x, x_1, ..., x_m; y)$, where $y = (y_1, y_2, ...)$, be a real function of its components and $t_1, ..., t_m \in \mathbb{Z}$, where $m \ge 0$. There exists an $m_0 > 0$ such that the following results hold.*

(i) For any $h > 0$ and $k \ge 2m + m_0$, we have

$$\begin{aligned} & E|\,p(k/n, x_k/h, \lambda_{t_1}, ..., \lambda_{t_m}; \lambda^*)| \\ \le \; & \frac{C\,h}{d_k}\int_{-\infty}^{\infty} E|p(k/n,\, y,\, \lambda_1, ..., \lambda_m; \lambda^*)|dy. \end{aligned} \tag{B.8}$$

(ii) For any $h > 0$, $k - j \ge 2m + m_0$ and $j + 1 \le t_1, ..., t_m \le k$, we have

$$\begin{aligned} & \big|E\big[p(k/n,\, x_k/h, \lambda_{t_1}, ..., \lambda_{t_m}; \lambda^*)\,\big|\, \lambda_j, \lambda_{j-1}, ...; \lambda^*\big]\big| \\ \le \; & \frac{C\,h\sum_{j=0}^{k-\min\{t_1,...,t_m\}}|\phi_j|}{d_{k-j}^2} \\ & \sum_{j=1}^{m}\int_{-\infty}^{\infty} \mathbb{E}\Big\{\big[|p(k/n,\, y,\, \lambda_1, ..., \lambda_m; \lambda^*)\,\eta_j|\big]\,\Big|\,\lambda^*\Big\}\,dy \\ & + \frac{Ch}{d_{k-j}}\int_{-\infty}^{\infty}\big|\mathbb{E}\big[p(k/n,\, y,\, \lambda_1, ..., \lambda_m; \lambda^*)\,\big|\,\lambda^*\big]\big|\,dy. \end{aligned} \tag{B.9}$$

Proof. We only prove (B.9) as the other is similar except simpler. Note that

$$\begin{aligned} x_k - x_j &= \sum_{i=j+1}^{k}\gamma^{k-i}\,\xi_i + \sum_{i=1}^{j}(\gamma^{k-i} - \gamma^{j-i})\,\xi_i \\ &= \sum_{i=j+1}^{k}\gamma^{k-i}\,\Big(\sum_{u=j+1}^{i} + \sum_{u=-\infty}^{j}\Big)\eta_u\phi_{i-u} + \sum_{i=1}^{j}(\gamma^{k-i} - \gamma^{j-i})\,\xi_i. \end{aligned}$$

We may have

$$x_k \;=\; x_{1jk} + x_{2jk}, \tag{B.10}$$

where $x_{1jk} = \sum_{i=j+1}^{k} \eta_i \, a_{k,i}$ with

$$a_{k,i} = \sum_{u=i}^{k} \gamma^{k-u} \phi_{u-i} = a_{k-i},$$

and x_{2jk} depends only on $\eta_j, \eta_{j-1}, ...$

Let $\Lambda_m = \sum_{j=1}^{m} \eta_{t_j} a_{k-t_j}$ and $y^*_{jk} = x_{1jk} - \Lambda_m$. It is readily seen that an $m_0 > 0$ exists such that, whenever $k - j \geq 2m + m_0$, $\mathbb{E}(y^*_{jk})^2 \asymp d^2_{k-j}$. As a consequence, similar arguments to those in the proof of Theorem 2.18 of Wang (2015) yields that

> whenever $k - j \geq 2m + m_0$, y^*_{jk}/d_{k-j} has a density function $\nu_{jk}(x)$, which is uniformly bounded over x by a constant C and

$$\sup_x |\nu_{jk}(x+u) - \nu_{jk}(x)| \;\leq\; C \min\{|u|,\; 1\}. \tag{B.11}$$

This, together with (B.10) and the independence of η_i, implies that

$$\begin{aligned}
& \mathbb{E}\big\{p(k/n, x_k/h, \lambda_{t_1}, ..., \lambda_{t_m}; \lambda^*) \mid \mathcal{F}^*_j\big\} \\
&= \mathbb{E}\big\{p\big[k/n, (x_{2jk} + \Lambda_m + y^*_{jk})/h,\; \lambda_{t_1}, ..., \lambda_{t_m}; \lambda^*\big] \mid \mathcal{F}^*_j\big\} \\
&= \mathbb{E}\Big\{\int_{-\infty}^{\infty} p\big[k/n, (x_{2jk} + \Lambda_m + d_{k-j} y)/h,\; \lambda_{t_1}, ..., \lambda_{t_m}; \lambda^*\big] \, \nu_{jk}(y)\, dy \mid \mathcal{F}^*_j\Big\} \\
&= \frac{h}{d_{k-j}} \int_{-\infty}^{\infty} \mathbb{E}\Big\{p(k/n, y,\; \lambda_{t_1}, ..., \lambda_{t_m}; \lambda^*)\, \nu_{jk}\big(\frac{-x_{2jk} - \Lambda_m + hy}{d_{k-j}}\big) \,\big|\mathcal{F}^*_j\Big\} dy,
\end{aligned} \tag{B.12}$$

where $\lambda^* = (\lambda^*_1, \lambda^*_2, ...)$ and $\mathcal{F}^*_j$ is a σ-field generated by $\lambda_j, \lambda_{j-1}, ...; \lambda^*_1, \lambda^*_2, ...$

As x_{2jk} depends only on $\eta_j, \eta_{j-1}, ...$, and $j + 1 \leq t_1, ..., t_m \leq k$, we have

$$\begin{aligned}
& \Big|\mathbb{E}\Big\{p(k/n, y,\; \lambda_{t_1}, ..., \lambda_{t_m})\, \nu_{jk}\big(\frac{-x_{2jk} + hy}{d_{k-j}}\big) \,\big|\mathcal{F}^*_j\Big\}\Big| \\
\leq\;& C \Big|\mathbb{E}\big\{p(k/n, y,\; \lambda_1, ..., \lambda_m; \lambda^*) \,\big|\lambda^*_1, \lambda^*_2, ...\big\}\Big|.
\end{aligned}$$

By taking this fact into (B.12) and by using (B.11), we have

$$\begin{aligned}
& \big|\mathbb{E}\{p(k/n, x_k/h, \lambda_{t_1}, ..., \lambda_{t_m}; \lambda^*) \mid \mathcal{F}_j^*\}\big| \\
\le\ & \frac{h}{d_{k-j}} \int_{-\infty}^{\infty} \mathbb{E}\Big\{|p(k/n, y,\ \lambda_{t_1}, ..., \lambda_{t_m}; \lambda^*)| \\
& \qquad \Big|\nu_{jk}(\frac{-x_{2jk} - \Lambda_m + hy}{d_{k-j}}) - \nu_{jk}(\frac{-x_{2jk} + hy}{d_{k-j}})\Big| \,\big|\mathcal{F}_j^*\Big\}\, dy \\
& + \frac{Ch}{d_{k-j}} \int_{-\infty}^{\infty} \Big|\mathbb{E}\{p(k/n, y,\ \lambda_1, ..., \lambda_m; \lambda^*) \,\big|\lambda_1^*, \lambda_2^*, ...\}\Big| dy \\
\le\ & \frac{Ch}{d_{k-j}} \int_{-\infty}^{\infty} \mathbb{E}\Big\{|p(k/n, y,\ \lambda_{t_1}, ..., \lambda_{t_m}; \lambda^*)| \,\min\{|\Lambda_m|/d_{k-j},\ 1\} \,\big|\mathcal{F}_j^*\Big\}\, dy \\
& + \frac{Ch}{d_{k-j}} \int_{-\infty}^{\infty} \Big|\mathbb{E}\{p(k/n, y,\ \lambda_1, ..., \lambda_m; \lambda^*) \,\big|\lambda_1^*, \lambda_2^*, ...\}\Big| dy \\
\le\ & \frac{Ch \sum_{j=0}^{k-\min\{t_1,...,t_m\}} |\phi_j|}{d_{k-j}^2} \int_{-\infty}^{\infty} \mathbb{E}\Big\{|p(k/n, y,\ \lambda_1, ..., \lambda_m, \lambda^*)| \sum_{j=1}^{m} |\eta_j| \,\big|\lambda_1^*, \lambda_2^*, ...\Big\}\, dy \\
& + \frac{Ch}{d_{k-j}} \int_{-\infty}^{\infty} \Big|\mathbb{E}\{p(k/n, y,\ \lambda_1, ..., \lambda_m; \lambda^*) \,\big|\lambda_1^*, \lambda_2^*, ...\}\Big| dy,
\end{aligned}$$

as required. □

Lemma B.2. *Let $0 \le m = m_n \le n/2$ be a sequence of integers and, for each $n \ge 1$, $\sigma_n(u, x_1, ..., x_m; y)$, where $y = (y_1, y_2, ...)$, be a real function of its components. For any bounded real function $g(x)$ satisfying $\int_{-\infty}^{\infty} |g(x)|dx < \infty$ and any $h = h_n > 0$, we have*

$$\begin{aligned}
& E \sum_{k=1}^{n} |g(x_k/h)\sigma_n(k/n, \lambda_k, \lambda_{k-1}, ..., \lambda_{k-m}; \lambda^*)| \\
& \quad \le\ C\, h\, (m + n/d_n) \sup_{0 \le u \le 1} E\, |\sigma_n(u, \lambda_1, \lambda_2, ..., \lambda_m; \lambda^*)|. \qquad \text{(B.13)}
\end{aligned}$$

If in addition $E\sigma_n(u, \lambda_1, ..., \lambda_m; \widetilde{\lambda}) = 0$ for each $u \in [0, 1]$, where $\widetilde{\lambda} = (\lambda_{m+1}, \lambda_{m+2}, ...)$, then

$$\begin{aligned}
& E\Big[\sum_{k=1}^{n} g(x_k/h)\sigma_n(k/n, \lambda_k, \lambda_{k-1}, ..., \lambda_{k-m}; \lambda^*)\Big]^2 \\
\le\ & C\, \tau_n^2\, m\, \big[\max\{m^{1/2},\ \log n\}\, h^2 + m d_n/n + h\big]\, (n/d_n) \\
& + C\tau_n\, \tau_{1n}\, h^2 (n/d_n)^2, \qquad \text{(B.14)}
\end{aligned}$$

where $\tau_n^2 = \sup_{0 \le u \le 1} E\sigma_n^2(u, \lambda_1, ..., \lambda_m; \lambda^*)$ and

$$\tau_{1n}^2 = \sup_{0 \le u \le 1} E\Big|\sigma_n(u, \lambda_1, ..., \lambda_m; \lambda^*) - \sigma_n(u, \lambda_1, ..., \lambda_m; \widetilde{\lambda})\Big|^2.$$

Proof . We only prove (B.14). The proof of (B.13) is similar except it is simpler. Write $\sigma_{n,k} = \sigma_n(k/n, \lambda_k, \lambda_{k-1}, ..., \lambda_{k-m}; \lambda^*)$ and $\nu_n = 2m + m_0$, where m_0 is given as in Lemma B.1. We have

$$
\begin{aligned}
& E\Big[\sum_{k=1}^{n} g(x_k/h)\sigma_n(k/n, \lambda_k, \lambda_{k-1}, ..., \lambda_{k-m}, \lambda^*)\Big]^2 \\
& \leq\ 2C\, E\Big[\sum_{k=1}^{\nu_n} |\sigma_{n,k}|\Big]^2 + 2E\Big[\sum_{k=\nu_n+1}^{n} g(x_k/h)\sigma_{n,k}\Big]^2 \\
& \leq\ 2C\, E\Big[\sum_{k=1}^{\nu_n} |\sigma_{n,k}|\Big]^2 + 2\sum_{k=\nu_n+1}^{n} E\big\{g^2(x_k/h)\sigma_{n,k}^2\big\} \\
& +4\sum_{k=\nu_n+1}^{n}\Big[\sum_{j=1}^{\nu_n} + \sum_{j=\nu_n+1}^{n-k}\Big]\big|E\big\{g(x_k/h)\sigma_{n,k} g(x_{k+j}/h)\sigma_{n,k+j}\big\}\big| \\
& :=\ R_{1n} + R_{2n} + R_{3n} + R_{4n}.
\end{aligned} \tag{B.15}
$$

Let $\pi_{n,j}(y) = E\big(|\sigma_n(k/n, \lambda_1, \lambda_2, ..., \lambda_m; y)\, \eta_j|\big), 1 \leq j \leq m$, and $\widehat{\pi}_n(y) = E\sigma_n(k/n, \lambda_1, \lambda_2, ..., \lambda_m; y)$. It is readily seen from (B.8) that, for $k \geq \nu_n$ and $1 \leq j \leq m$

$$
\begin{aligned}
E\big\{|g(x_k/h)\sigma_{n,k}|\pi_{n,j}(\lambda^*)\big\} &\leq Chd_k^{-1}E\big\{|\sigma_{n,k}|\pi_{n,j}(\lambda^*)\big\} \\
&\leq Chd_k^{-1}E\sigma_n^2(k/n, \lambda_1, ..., \lambda_m; \lambda^*) \leq Chd_k^{-1}\tau_n^2.
\end{aligned}
$$

On the other hand, by recalling that $E\sigma_n(k/n, \lambda_1, ..., \lambda_m; \widetilde{\lambda}) = 0$ and $\widetilde{\lambda}$ is independent of λ^*, we have

$$
\begin{aligned}
E|\widehat{\pi}_n(\lambda^*)|^2 &= E\Big|E\Big\{\big[\sigma_n(k/n, \lambda_1, ..., \lambda_m; \lambda^*) - \sigma_n(k/n, \lambda_1, ..., \lambda_m; \widetilde{\lambda})\big] \mid \lambda^*\Big\}\Big|^2 \\
&\leq E\Big|\sigma_n(k/n, \lambda_1, ..., \lambda_m; \lambda^*) - \sigma_n(k/n, \lambda_1, ..., \lambda_m; \widetilde{\lambda})\Big|^2 \leq \tau_{1n}^2,
\end{aligned}
$$

thus indicating

$$
E\big\{|g(x_k/h)\sigma_{n,k}|\,|\widehat{\pi}_n(\lambda^*)|\big\} \ \leq\ Chd_k^{-1}E\big\{|\sigma_{n,k}|\,|\widehat{\pi}_n(\lambda^*)|\big\} \ \leq\ Chd_k^{-1}\tau_n\,\tau_{1n}.
$$

Consequently, by (B.9), we obtain

$$
\begin{aligned}
& \big|E\big\{g(x_k/h)\sigma_{n,k} g(x_{k+j}/h)\sigma_{n,k+j}\big\}\big| \\
\leq\ & E\big\{|g(x_k/h)\sigma_{n,k}|\,\big|E\big[g(x_{k+j}/h)\sigma_{n,k+j}|\mathcal{F}_k^*\big]\big|\big\} \\
\leq\ & \frac{Ch\sum_{s=0}^{m}|\phi_s|}{d_j^2}\sum_{j=1}^{m} E\big\{|g(x_k/h)\sigma_{n,k}|\pi_{n,j}(\lambda^*)\big\} \\
& + \frac{Ch}{d_j}E\big\{|g(x_k/h)\sigma_{n,k}|\,|\widehat{\pi}_n(\lambda^*)|\big\} \\
\leq\ & C\,h^2\,m\sum_{s=0}^{m}|\phi_s|\,d_j^{-2}d_k^{-1}\,\tau_n^2 + Ch^2 d_j^{-1}d_k^{-1}\,\tau_n\,\tau_{1n},
\end{aligned}
$$

for $k, j \geq \nu_n$. Now it is readily seen that

$$
\begin{aligned}
R_{4n} &\leq C\,h^2\,m \sum_{s=0}^{m} |\phi_s| \sum_{k=\nu_n+1}^{n} d_k^{-1} \sum_{j=\nu_n+1}^{n-k} d_j^{-2}\tau_n^2 \\
&\quad + C\,h^2\,\tau_n\,\tau_{1n} \sum_{k=\nu_n+1}^{n} d_k^{-1} \sum_{j=\nu_n+1}^{n-k} d_j^{-1} \\
&\leq C\,h^2\,m\,\max\{m^{1/2}, \log n\}\,(n/d_n)\,\tau_n^2 + Ch^2(n/d_n)^2\tau_n\tau_{1n}.
\end{aligned}
$$

Similarly, by using Lemma B.1 (i) and Hölder's inequality, we have

$$
\begin{aligned}
R_{1n} + R_{2n} + R_{3n} &\leq C\big(\nu_n^2 + \nu_n\,h \sum_{k=1}^{n} d_k^{-1}\big)\,\tau_n^2 \\
&\leq C\,m\,(md_n/n + h)\,(n/d_n)\,\tau_n^2.
\end{aligned}
$$

By taking these estimates into (B.15), we establish (B.14). $\square$

We are now ready to prove (3.13)-(3.15). Let $m = (n/d_n)^{1/2}$ and set

$$
\sigma_{1k}^2 \;=\; \sigma_k^{*2}(m) = \sigma(k/n, \lambda_k, \lambda_{k-1}, ..., \lambda_{k-m}, \lambda_1^*, \lambda_2^*, ...).
$$

First note that $E\sigma_k^2 = E\sigma_{1k}^2 = E\sigma(k/n, \lambda_0, \lambda_1, ...)$. It follows from (3.2) and (B.14) with $h = 1$ that

$$
\begin{aligned}
& E\Big|\frac{d_n}{n} \sum_{k=1}^{n} g(x_k)\;\big[\sigma_{1k}^2 - E\sigma_{1k}^2\big]\Big|^2 \\
\leq\; & C \sup_{0\leq u\leq 1} E\sigma^2(u, \lambda_0, \lambda_1, ..)\,(d_n/n)^{1/4} + C \sup_{0\leq u\leq 1} \big[E\sigma^2(u, \lambda_0, \lambda_1, ..)\big]^{1/2} \\
& \quad \sup_{0\leq u\leq 1} \big(E|\sigma(u, \lambda_1, \lambda_2, ..., \lambda_m, \widetilde{\lambda}) - \sigma(u, \lambda_1, \lambda_2, ..., \lambda_m, \lambda^*)|^2\big)^{1/2} \\
\leq\; & C_1(d_n/n)^{1/4} + C_1(d_n/n)^{\alpha/4} = o(1). \qquad \text{(B.16)}
\end{aligned}
$$

Similarly, for any $0 \leq \delta < \begin{cases} (2\mu-1)/(3-2\mu), & \text{under } LM, \\ 1, & \text{under } SM, \end{cases}$ it follows from (3.2) that

$$
\begin{aligned}
& \frac{d_n}{n} \sum_{k=1}^{n} E\big\{|g(x_k)|\,|\sigma_k^2 - \sigma_{1k}^2|^{1+\delta}\big\} \\
\leq\; & \frac{d_n}{n} \sum_{k=1}^{n} \big(E|g(x_k)|^{2/(1-\delta)}\big)^{(1-\delta)/2} \big(E|\sigma_k^2 - \sigma_{1k}^2|^2\big)^{(1+\delta)/2} \\
\leq\; & \frac{d_n}{n} \sum_{k=1}^{n} d_k^{(\delta-1)/2}\,m^{-\alpha(1+\delta)/2} \leq \big(\frac{d_n}{n}\big)^{1+\alpha/4} \sum_{k=1}^{n} d_k^{(\delta-1)/2} \\
\leq\; & \begin{cases} n^{(1/2-\mu)\alpha/4+1/2}, & \text{under } LM, \\ n^{1/2-\alpha/8}, & \text{under } SM, \end{cases} \\
=\; & o(1), \qquad \text{(B.17)}
\end{aligned}
$$

where we used $E|g(x_k)| \le C d_k^{-1}$ by (B.8) with $h = 1$. By virtue of (B.16) and (B.17) with $\delta = 0$, we have

$$
\begin{aligned}
& \frac{d_n}{n} \Big| \sum_{k=1}^n g(x_k) \ \big[\sigma_k^2 - E\sigma_k^2\big] \Big| \\
\le \ & \frac{d_n}{n} \Big| \sum_{k=1}^n g(x_k) \ \big[\sigma_{1k}^2 - E\sigma_{1k}^2\big] \Big| + \frac{d_n}{n} \sum_{k=1}^n |g(x_k)| \, |\sigma_k^2 - \sigma_{1k}^2(n)| = o_P(1),
\end{aligned}
$$

indicating that

$$
\begin{aligned}
& \frac{d_n}{n} \sum_{k=1}^n g(x_k) \, \sigma_k^2 \\
= \ & \frac{d_n}{n} \sum_{k=1}^n g(x_k) \, E\sigma(k/n, \lambda_0, \lambda_1, ...) + \frac{d_n}{n} \sum_{k=1}^n g(x_k) \ \big[\sigma_k^2 - E\sigma_k^2\big] \\
= \ & \frac{d_n}{n} \sum_{k=1}^n g(x_k) \, E\sigma(k/n, \lambda_0, \lambda_1, ...) + o_P(1).
\end{aligned}
$$

This implies (3.13) by Theorem 2.21 of Wang (2015). Similarly, by using (B.17), we obtain

$$
\begin{aligned}
& \frac{d_n}{n} \sum_{k=1}^n E|g(x_k) \, \sigma_k^{2+\delta} \, \rho(\epsilon_k)| \\
\le \ & E|\rho(\epsilon_1)| \, \frac{d_n}{n} \sum_{k=1}^n E|g(x_k) \, |\sigma_k^2 - \sigma_{1k}^2|^{(2+\delta)/2} + \frac{d_n}{n} \sum_{k=1}^n E\big\{|g(x_k)| \, |\sigma_{1k}^2|^{1+\delta}\big\} \\
= \ & O(1),
\end{aligned}
$$

thus implying (3.14), where we used (B.8) together with some simple calculations.

To prove (3.15), let $x_{nk} = (d_n/n)^{1/2} g(x_k)\sigma_k$. Given that $g(x)$ is bounded by a constant $C > 0$, by virtue of (3.10) with some $\delta > 0$, we obtain the following:

$$
\max_{1\le k\le n} |x_{nk}| \ \le \ \Big[C^{1+\delta} \, \big(\frac{d_n}{n}\big)^{(2+\delta)/2} \sum_{k=1}^n |g(x_k)| \, |\sigma_k|^{2+\delta}\Big]^{1/(2+\delta)} = o_P(1)
$$

and

$$
\frac{1}{\sqrt{n}} \sum_{k=1}^n |x_{nk}| \ \le \ d_n^{-1/2} \, (d_n/n) \sum_{k=1}^n |g(x_k)||\sigma_k| = O_P(d_n^{-1}) = o_P(1).
$$

Now, by recalling (3.13), the result follows easily from an application of the extended martingale limit theorem of Wang (2014, 2015)). □