ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Nonlinear Regression with Harris Recurrent Markov Chains

Degui Li, Dag Tjøstheim and Jiti Gao

July 2012

Working Paper 14/12

Nonlinear Regression with Harris Recurrent Markov Chains[1]

By Degui Li†, Dag Tjøstheim‡ and Jiti Gao†

†*Monash University and* ‡*University of Bergen*

Abstract

In this paper, we study parametric nonlinear regression under the Harris recurrent Markov chain framework. We first consider the nonlinear least squares estimators of the parameters in the homoskedastic case, and establish asymptotic theory for the proposed estimators. Our results show that the convergence rates for the estimators rely not only on the properties of the nonlinear regression function, but also on the number of regenerations for the Harris recurrent Markov chain. We also discuss the estimation of the parameter vector in a conditional volatility function and its asymptotic theory. Furthermore, we apply our results to the nonlinear regression with $I(1)$ processes and establish an asymptotic distribution theory which is comparable to that obtained by Park and Phillips (2001). Some simulation studies are provided to illustrate the proposed approaches and results.

Keywords: Asymptotic distribution, asymptotically homogeneous functions, β-null recurrent Markov chains, Harris recurrence, integrable functions, least squares estimation, nonlinear regression.

JEL Classifications: C13, C22.

[1] The first authors research was supported by an Australian Research Council Discovery Early Career Researcher Award (DE120101130) and the third author's research was supported by an Australian Research Council Discovery Grant (DP1096374). This work was started when the first and third authors visited the second author at Department of Mathematics, University of Bergen in 2011. The authors acknowledge the financial support from the Norwegian Research Council.

1 Introduction

In this paper, we consider a parametric nonlinear regression model defined by

$$\begin{aligned} Y_t &= g(X_t, \theta_{01}, \theta_{02}, \cdots, \theta_{0d}) + e_t \\ &=: g(X_t, \boldsymbol{\theta}_0) + e_t, \quad t = 1, 2, \cdots, n, \end{aligned} \tag{1.1}$$

where $\boldsymbol{\theta}_0$ is the true value of the d-dimensional parameter vector such that

$$\boldsymbol{\theta}_0 = (\theta_{01}, \theta_{02}, \cdots, \theta_{0d})^\tau \in \Theta \subset \mathbb{R}^d$$

and $g(\cdot, \cdot): \mathbb{R}^{d+1} \to \mathbb{R}$ is assumed to be known. Throughout this paper, we assume that Θ is a compact set. How to construct a consistent estimator for the parameter vector $\boldsymbol{\theta}_0$ and derive an asymptotic theory are an important issue in modern statistics and econometrics. When the observations (Y_t, X_t) satisfy stationarity and weak dependence conditions, there is an extensive literature on the asymptotic theory of the parametric nonlinear estimator of $\boldsymbol{\theta}_0$, see, for example, Jennrich (1969), Malinvaud (1970) and Wu (1981) for some early references.

As pointed out in the literature, assuming stationarity is too restrictive and unrealistic in many practical applications. When tackling economic and financial issues from a time perspective, we often deal with nonstationary components. For instance, neither the consumer price index nor the share price index, nor the exchange rates constitute a stationary process. A traditional method to handle such data is to take the first order difference to eliminate possible stochastic or deterministic trends involved in the data, and then do the estimation for a stationary model. However, such differencing may lead to loss of useful information. Thus, the development of a modeling technique that takes both nonstationary and nonlinear phenomena into account in time series analysis is crucial. Without taking differences, Park and Phillips (2001) (hereafter PP) study the nonlinear regression (1.1) with the regressor $\{X_t\}$ satisfying a unit root (or I(1)) structure, and prove that the rates of convergence of the nonlinear least squares (NLS) estimator of $\boldsymbol{\theta}_0$ depend on the properties of $g(\cdot, \cdot)$. For an integrable $g(\cdot,\cdot)$, the rate of convergence is as slow as $n^{1/4}$, and for an asymptotically homogeneous $g(\cdot,\cdot)$, the rate of convergence can achieve the $\sqrt{n}$-rate and even super-n-rate of convergence.

As also pointed out in a recent paper by Myklebust *et al* (2012), the null recurrent Markov process is a nonlinear generalization of the linear unit root process. In fact, under the framework of null recurrent Markov chains, there has been an extensive literature on nonparametric and semiparametric estimation (Karlsen and Tjøstheim 2001; Karlsen *et al* 2007, 2010; Lin *et al* 2009; Gao *et al* 2011a; Schienle 2011; Chen *et al* 2012), by using the technique of the split chain (Nummelin 1984; Meyn and Tweedie 2009), and the generalized ergodic theorem and functional limit theorem developed in Karlsen and Tjøstheim (2001).

As far as we know, however, there is virtually no work on the parametric estimation of the nonlinear regression model (1.1) when the regressor $\{X_t\}$ is generated by a class of Harris recurrent Markov processes that includes both stationary and nonstationary cases. This paper aims to fill this gap. If the function $g(\cdot,\cdot)$ is integrable, we can directly use some existing results for functions of Harris recurrent Markov processes to develop an asymptotic theory for the estimator of $\boldsymbol{\theta}_0$. The case that $g(\cdot,\cdot)$ belongs to a class of asymptotically homogeneous functions is much more challenging, as in this case the function $g(\cdot,\cdot)$ is no longer bounded. In nonparametric or semiparametric estimation theory, we do not have such problems because the kernel function is usually assumed to be bounded and have a compact support. Unfortunately, most of the existing results for the asymptotic theory of the Harris recurrent Markov process focus on the case that $g(\cdot,\cdot)$ is bounded and integrable (see, for example, Chen 1999, 2000). Hence, in this paper, we first modify the conventional NLS estimator for the asymptotically homogeneous $g(\cdot,\cdot)$, and then use a novel method to establish asymptotic distribution as well as rates of convergence for the modified parametric estimator. Our results show that the rates of convergence for the parameter vector in nonlinear cointegrating models rely not only on the properties of the function $g(\cdot,\cdot)$, but also on the magnitude of the regeneration number for the Harris recurrent Markov chain.

In addition, we also study two important issues, which are closely related to nonlinear regression with Harris recurrent Markov chains. The first one is to study the estimation of the parameter vector in a conditional volatility function and its asymptotic theory. As the estimation method is based on the log-transformation, the rates of convergence for the proposed estimator would depend on the property of the derivative of the log-transformed volatility function. Meanwhile, we also discuss the nonlinear regression with $I(1)$ processes when $g(\cdot,\cdot)$ is asymptotically homogeneous. By using Theorem 3.2 in Section 3, we obtain asymptotic normality for the parametric estimator with a stochastic normalized rate, which is comparable to Theorem 5.3 in PP. However, our derivation is done under Markov perspective, which carries with it the potential of extending the theory to nonlinear and nonstationary autoregressive processes.

The rest of this paper is organised as follows. Some preliminary results about Markov theory (especially Harris recurrent Markov chain) and function classes are introduced in Section 2. The main results of this paper are given in Sections 3–5. Some simulation studies are carried out in Section 6. Section 7 concludes this paper. The proofs of the main results and some useful lemmas are provided in Appendices A and B, respectively.

2 Preliminary results

To make the paper self-contained, in this section, we first provide some basic definitions and preliminary results for Harris recurrent Markov process $\{X_t\}$, and then define function classes in a way similar to those introduced in PP.

2.1 Markov theory

Let $\{X_t, t \geq 0\}$ be a ϕ-irreducible Markov chain on the state space $(\mathbb{E}, \mathcal{E})$ with transition probability P. This means that for any set $A \in \mathcal{E}$ with $\phi(A) > 0$, we have $\sum_{t=1}^{\infty} \mathsf{P}^t(x, A) > 0$ for $x \in \mathbb{E}$. We further assume that the ϕ-irreducible Markov chain $\{X_t\}$ is Harris recurrent.

Definition 2.1. *A Markov chain $\{X_t\}$ is Harris recurrent if, given a neighborhood $\mathbb{B}_v$ of v with $\phi(\mathbb{B}_v) > 0$, $\{X_t\}$ returns to $\mathbb{B}_v$ with probability one, $v \in \mathbb{E}$.*

The Harris recurrence allows one to construct a split chain, which decomposes the partial sum of functions of $\{X_t\}$ into blocks of independent and identically distributed (i.i.d.) parts and two asymptotically negligible remaining parts. Let τ_k be the regeneration times, n the number of observations and $N(n)$ the number of regenerations as in Karlsen and Tjøstheim (2001), where they use the notation $T(n)$ instead of $N(n)$. For the process $\{G(X_t) : \ t \geq 0\}$, defining

$$Z_k = \begin{cases} \sum_{t=0}^{\tau_0} G(X_t), & k = 0, \\ \sum_{t=\tau_{k-1}+1}^{\tau_k} G(X_t), & 1 \leq k \leq N(n), \\ \sum_{t=\tau_{N(n)}+1}^{n} G(X_t), & k = N(n) + 1, \end{cases}$$

where $G(\cdot)$ is a real function defined on $\mathbb{R}$, then we have

$$S_n(G) = \sum_{t=0}^{n} G(X_t) = Z_0 + \sum_{k=1}^{N(n)} Z_k + Z_{N(n)+1}. \tag{2.1}$$

From Nummelin (1984), we know that $\{Z_k, k \geq 1\}$ is a sequence of i.i.d. random variables, and Z_0 and $Z_{(n)}$ converge to zero almost surely (a.s.) when they are divided by the number of regenerations $N(n)$ (using Lemma 3.2 in Karlsen and Tjøstheim 2001).

The general Harris recurrence only yields stochastic rates of convergence in asymptotic theory of the parametric and nonparametric estimators (see, for example, Theorems 3.1 and 3.2 below), where distribution and size of the number of regenerations

$N(n)$ have no a priori known structure but fully depend on the underlying process $\{X_t\}$. To obtain a specific rate of $N(n)$ in our asymptotic theory, we next impose some restrictions on the tail behavior of the distribution of the recurrence times of the Markov chain.

DEFINITION 2.2. *A Markov chain $\{X_t\}$ is β-null recurrent if there exist a small nonnegative function f, an initial measure λ, a constant $\beta \in (0, 1)$, and a slowly varying function $L_f(\cdot)$ such that*

$$\mathsf{E}_\lambda(\sum_{i=1}^{n} f(X_t)) \sim \frac{1}{\Gamma(1+\beta)} n^\beta L_f(n), \tag{2.2}$$

where E_λ stands for the expectation with initial distribution λ and $\Gamma(1+\beta)$ is the Gamma function with parameter $1+\beta$.

The definition of a small function f in the above definition can be found in some existing literature (see, for example, pp.15 in Nummelin 1984). Assuming β-null recurrence restricts the tail behavior of the recurrence time of the process to be a regularly varying function. Following a result in Karlsen and Tjøstheim (2001), the regeneration number $N(n)$ of the β-null recurrent Markov chain $\{X_t\}$ has the following asymptotic distribution

$$\frac{N(n)}{n^\beta L_s(n)} \xrightarrow{d} M_\beta(1), \tag{2.3}$$

where $L_s = L_f/(\pi_s f)$, $\pi_s(\cdot)$ is an invariant measure of the Markov chain $\{X_t\}$, and $M_\beta(1)$ is the Mittag-Leffler distribution with parameter β (cf., Kasahara 1984). Since $N(n) < n$ a.s. for the null recurrent case by (2.3), the rates of convergence for the nonparametric kernel estimators are slower than those for the stationary time series cased (see, for example, Karlsen *et al* 2007, Gao *et al* 2011a). However, this is not necessarily the case for the parametric estimator in our model (1.1). In Section 3 below, we will show that our rate of convergence is slower than that for the stationary time series for integrable $g(\cdot, \cdot)$ and may be faster than that for the stationary time series case for asymptotically homogeneous $g(\cdot, \cdot)$. In addition, our rates of convergence also depend on the magnitude of β, which measures the recurrence times of the Markov chain $\{X_t\}$.

2.2 Examples of β-null recurrent Markov chains

For a stationary or positive recurrent process, $\beta = 1$. We next give several examples of β-null recurrent Markov chains.

Example 2.1. (1/2-null recurrent Markov chain).

(i) Let a random walk process be defined as

$$X_t = X_{t-1} + x_t, \quad t = 1, 2, \cdots, \quad X_0 = 0, \tag{2.4}$$

where $\{x_t\}$ is a sequence of i.i.d. random variables. Kallianpur and Robbins (1954) show that $\{X_t\}$ defined by (2.4) is a 1/2-null recurrent Markov chain under weak conditions on the distribution of x_t.

(ii) Consider a parametric threshold autoregressive (TAR) model of the form

$$X_t = \alpha_1 X_{t-1} I\big(X_{t-1} \in \mathbb{S}\big) + \alpha_2 X_{t-1} I\big(X_{t-1} \in \mathbb{S}^c\big) + u_t, \quad X_0 = 0, \tag{2.5}$$

where $\mathbb{S}$ is a compact subset of $\mathbb{R}$, $\mathbb{S}^c$ is the complement of $\mathbb{S}$, $\alpha_2 = 1$, $-\infty < \alpha_1 < \infty$, $\{u_t\}$ is assumed to be i.i.d. with $\mathsf{E}[u_1] = 0$, $0 < \mathsf{E}[u_1^2] < \infty$ and $\mathsf{E}[u_1^4] < \infty$, and the distribution of $\{u_t\}$ is absolutely continuous with respect to Lebesgue measure with $f_u(\cdot)$ being the density function satisfying $\inf_{x\in\mathbb{S}_1} f_u(x) > 0$ for all compact sets $\mathbb{S}_1$. Recently, Gao *et al* (2011b) have shown that such TAR process $\{X_t\}$ is a 1/2-null recurrent Markov chain.

Furthermore, we can generalize the TAR model (2.5) to

$$X_t = H(X_{t-1}, \boldsymbol{\zeta}) I\big(X_{t-1} \in \mathbb{S}\big) + X_{t-1} I\big(X_{t-1} \in \mathbb{S}^c\big) + u_t,$$

where $X_0 = 0$, $\sup_{x\in\mathbb{S}} |H(x, \boldsymbol{\zeta})| < \infty$ and $\boldsymbol{\zeta}$ is a parameter vector. According to Teräsvirta *et al* (2010), the above autoregressive process is also a 1/2-null recurrent Markov chain.

Example 2.2. (β-null recurrent Markov chain with $\beta \neq 1/2$).

Let $\{x_t\}$ be a sequence of i.i.d. random variables taking positive values, and $\{X_t\}$ be defined as

$$X_t = \begin{cases} X_{t-1} - 1, & X_{t-1} > 1, \\ x_t, & X_{t-1} \in [0, 1], \end{cases}$$

for $t \geq 1$, and $X_0 = C_0$ for some positive constant C_0. Myklebust *et al* (2012) prove that $\{X_t, t \geq 0\}$ is β-null recurrent if and only if

$$\mathsf{P}([x_1] > n) \sim n^{-\beta} l^{-1}(n), \quad 0 < \beta < 1,$$

where $[\cdot]$ is the integer function and $l(\cdot)$ is a slowly varying and positive function.

From the above examples, the β-null recurrent Markov chain framework is not restricted to linear processes (see Example 2.1 (ii)). Furthermore, such a null recurrent class has the invariance property that if $\{X_t\}$ is β-null recurrent, then for a one-to-one transformation $\mathcal{T}(\cdot)$, $\{\mathcal{T}(X_t)\}$ is also β-null recurrent (see, for example, Teräsvirta *et al* 2010). Such invariance property does not hold for the $I(1)$ processes. For other examples of the β-null recurrent Markov chain, we refer to Example 1 in Schienle (2011). For some general conditions on diffusion processes to ensure the Harris recurrence is satisfied, we refer to Höpfner and Löcherbach (2003) and Bandi and Phillips (2009).

2.3 *Function classes*

Similarly to Park and Phillips (1999, 2001), we consider two classes of parametric nonlinear functions: integrable functions and asymptotically homogeneous functions, which include many commonly-used functions in nonlinear regression. Let $\|\mathbf{A}\| = \sqrt{\sum_{i=1}^q \sum_{j=1}^q a_{ij}^2}$ for $\mathbf{A} = (a_{ij})_{q \times q}$, and $\|\mathbf{a}\|$ be Euclidean norm of the vector $\mathbf{a}$. A function $h(x) : \mathbb{R} \to \mathbb{R}^d$ is π_s-integrable if

$$\int_{\mathbb{R}} \|h(x)\| \pi_s(dx) < \infty,$$

where $\pi_s(\cdot)$ is the invariant measure of the Harris recurrent Markov chain $\{X_t\}$. When $\pi_s(\cdot)$ is differentiable such that $\pi_s(dx) = p_s(x)dx$, $h(x)$ is π_s-integrable if and only if $h(x)p_s(x)$ is integrable. For the random walk case as in Example 2.1 (i), the π_s-integrability reduces to the conventional integrability as $\pi_s(dx) = dx$.

Definition 2.3. *A d-dimensional vector function $h(x, \boldsymbol{\theta})$ is said to be integrable on Θ if there exists a $\boldsymbol{\theta}_* \in \Theta$ such that $h(x, \boldsymbol{\theta}_*)$ is π_s-integrable, and for each $\boldsymbol{\theta} \in \Theta$, there exist a neighborhood $\mathbb{B}_{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ and $M : \mathbb{R} \to \mathbb{R}$ bounded and π_s-integrable such that $\|h(x, \boldsymbol{\theta}') - h(x, \boldsymbol{\theta})\| \leq \|\boldsymbol{\theta}' - \boldsymbol{\theta}\| M(x)$ for any $\boldsymbol{\theta}' \in \mathbb{B}_{\boldsymbol{\theta}}$.*

The above definition is comparable to Definition 3.3 in PP. However, in our definition, we do not need condition (b) in Definition 3.3 of their paper, which makes the integrable function family in this paper slightly more general. We next introduce a class of asymptotically homogeneous functions.

Definition 2.4. *For a d-dimensional vector function $h(x, \boldsymbol{\theta})$, let $h(\lambda x, \boldsymbol{\theta}) = \kappa(\lambda, \boldsymbol{\theta}) H(x, \boldsymbol{\theta}) + R(x, \lambda, \boldsymbol{\theta})$, where $\kappa(\cdot, \cdot)$ is nonzero. $h(\lambda x, \boldsymbol{\theta})$ is said to be asymptotically homogeneous on Θ if the following two conditions are satisfied: (i) $H(\cdot, \boldsymbol{\theta})$ is locally bounded uniformly for any $\boldsymbol{\theta} \in \Theta$ and continuous with respect to $\boldsymbol{\theta}$; (ii) the remainder term $R(x, \lambda, \boldsymbol{\theta})$ is of order smaller than $\kappa(\lambda, \boldsymbol{\theta})$ as $\lambda \to \infty$ for any $\boldsymbol{\theta} \in \Theta$. As in PP, $\kappa(\cdot, \cdot)$ is the asymptotic order of $h(\cdot, \cdot)$ and $H(\cdot, \cdot)$ is the limit homogeneous function.*

The above definition is quite similar to that of an H-regular function in PP except that the regularity condition (a) in Definition 3.5 of PP is replaced by the local boundness condition (i) in Definition 2.4. Following Definition 3.4 in PP, as $R(x, \lambda, \boldsymbol{\theta})$ is of order smaller than $\kappa(\cdot, \cdot)$, we have either

$$R(x, \lambda, \boldsymbol{\theta}) = a(\lambda, \boldsymbol{\theta}) A_R(x, \boldsymbol{\theta}) \tag{2.6}$$

or

$$R(x, \lambda, \boldsymbol{\theta}) = b(\lambda, \boldsymbol{\theta}) A_R(x, \boldsymbol{\theta}) B_R(\lambda x, \boldsymbol{\theta}), \tag{2.7}$$

where $a(\lambda, \boldsymbol{\theta}) = o(\kappa(\lambda, \boldsymbol{\theta}))$, $b(\lambda, \boldsymbol{\theta}) = O(\kappa(\lambda, \boldsymbol{\theta}))$ as $\lambda \to \infty$, $\sup_{\boldsymbol{\theta} \in \Theta} A_R(\cdot, \boldsymbol{\theta})$ is locally bounded, and $\sup_{\boldsymbol{\theta} \in \Theta} B_R(\cdot, \boldsymbol{\theta})$ is bounded and vanishes at infinity. The conditions (2.6) and (2.7) will be used in the proof of Lemma B.3 in Appendix B.

Note that the above two definitions can be similarly generalized to the case that $h(\cdot,\cdot)$ is a $d\times d$ matrix of functions. Details are omitted here to save space.

3 Main results

In this section, we establish some asymptotic results for the parametric estimators of $\boldsymbol{\theta}_0$ when $g(\cdot,\cdot)$ belongs to the two classes of functions introduced in Section 2.

3.1 *Integrable function on* Θ

We first consider estimating model (1.1) by the nonlinear least squares (NLS) approach, which is also used by PP in the unit root framework. Define the loss function by

$$L_n(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left(Y_t - g(X_t,\boldsymbol{\theta})\right)^2. \tag{3.1}$$

We can obtain the resulting estimator $\widehat{\boldsymbol{\theta}}_n$ by minimizing $L_n(\boldsymbol{\theta})$ over $\boldsymbol{\theta}\in\Theta$, i.e.,

$$\widehat{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta} L_n(\boldsymbol{\theta}). \tag{3.2}$$

For $\boldsymbol{\theta} = (\theta_1,\cdots,\theta_d)^\tau$, let

$$\dot{g}(x,\boldsymbol{\theta}) = \left(\frac{\partial g(x,\boldsymbol{\theta})}{\partial\theta_j}\right)_{d\times 1}, \quad \ddot{g}(x,\boldsymbol{\theta}) = \left(\frac{\partial^2 g(x,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right)_{d\times d}.$$

Before deriving the asymptotic properties of $\widehat{\boldsymbol{\theta}}_n$ when the derivatives of $g(\cdot)$ are integrable on Θ, we give some regularity conditions.

ASSUMPTION 3.1. (i) $\{X_t\}$ is a Harris recurrent Markov chain with invariant measure $\pi_s(\cdot)$.

(ii) $\{e_t\}$ is a sequence of martingale differences satisfying

$$\mathsf{E}(e_t|\mathcal{F}_{n,t-1}) = 0 \text{ and } \mathsf{E}(e_t^2|\mathcal{F}_{n,t-1}) = \sigma^2 < \infty,\ a.s.,$$

where $\mathcal{F}_{n,t} = \sigma\left(X_{s_1}, e_{s_2},\ 1\le s_1\le n,\ 1\le s_2\le t\right)$.

ASSUMPTION 3.2. Both $\dot{g}(x, \boldsymbol{\theta})$ and $\ddot{g}(x, \boldsymbol{\theta})$ are integrable on Θ. Furthermore,

$$\ddot{L}(\boldsymbol{\theta}_0) := \int \dot{g}(x, \boldsymbol{\theta}_0)\dot{g}^{\tau}(x, \boldsymbol{\theta}_0)\pi_s(dx)$$

is a positive definite matrix.

REMARK 3.1. In Assumption 3.1 (i), $\{X_t\}$ is assumed to be Harris recurrent, which includes both positive and null recurrent Markov chains. For the positive recurrent case, the derivative function of the invariant measure reduces to the density function if it exists. For more discussion of Harris recurrence as well as its application, we refer to Karlsen and Tjøstheim (2001), Karlsen *et al* (2007) and Meyn and Tweedie (2009). In Assumption 3.1 (ii), we relax the i.i.d. condition on $\{e_t\}$, and instead assume a martingale difference structure. And $\{e_t\}$ can also be allowed to be stationary and α–mixing dependent at the cost of more tedious proofs to deal with the correlation structure. Furthermore, the homoskedasticity on the error term can also be relaxed. For instance, we can allow the existence of the heteroskedasticity structure defined by

$$\mathsf{E}(e_t^2|\mathcal{F}_{n,t-1}) = \sigma^2(X_t, \boldsymbol{\gamma}_0) \quad a.s., \tag{3.3}$$

where $\sigma^2(\cdot, \boldsymbol{\gamma}_0)$ is a nonlinear conditional variance function indexed by a p–dimensional parameter $\boldsymbol{\gamma}_0$. See Section 4 for more discussion. Assumption 3.2 is quite standard, see, for example, conditions (b) and (c) in Theorem 5.1 in PP.

We next give the asymptotic properties of $\widehat{\boldsymbol{\theta}}_n$. The following theorem is applicable for both stationary (positive recurrent) and nonstationary (null recurrent) time series.

THEOREM 3.1. *Let Assumptions 3.1 and 3.2 hold.*

(a) There exists a solution $\widehat{\boldsymbol{\theta}}_n$ to minimize the loss function $L_n(\boldsymbol{\theta})$ defined in (3.1), and

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_P(1). \tag{3.4}$$

(b) The estimator $\widehat{\boldsymbol{\theta}}_n$ has an asymptotic normal distribution of the form:

$$\sqrt{N(n)}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathsf{N}\left(\mathbf{0}_d, \sigma^2\ddot{L}^{-1}(\boldsymbol{\theta}_0)\right), \tag{3.5}$$

where $\mathbf{0}_d$ *is a d–dimensional null vector.*

Remark 3.2. As we assume that $\{e_t\}$ satisfies the martingale difference structure defined in Assumption 3.1 (ii), $\widehat{\boldsymbol{\theta}}_n$ is asymptotically normal with stochastic convergence rate $\sqrt{N(n)}$ for both the stationary and nonstationary cases. The number of regenerations $N(n)$ plays an important role in constructing the consistent estimators of σ^2 and $\ddot{L}(\boldsymbol{\theta}_0)$. However, as $N(n)$ is not directly observable, we need to link it with a directly observable hitting time. Indeed, if $\mathbb{C} \in \mathcal{E}$ and $I_{\mathbb{C}}$ has a ϕ–positive support, the number of times that the process visits $\mathbb{C}$ up to the time n is defined by $N_{\mathbb{C}}(n) = \sum_{t=1}^{n} I_{\mathbb{C}}(X_t)$. By Lemma 3.2 in Karlsen and Tjøstheim (2001), we have

$$\frac{N_{\mathbb{C}}(n)}{N(n)} \longrightarrow \pi_s I_{\mathbb{C}} \quad a.s. \tag{3.6}$$

From (3.5) in Theorem 3.1 and (2.3) in Section 2.1 above, we have the following corollary.

Corollary 3.1. *Suppose that the conditions in Theorem 3.1 are satisfied. Furthermore,* $\{X_t\}$ *is a* β*-null recurrent Markov chain with* $0 < \beta < 1$*. Then we have*

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_P\Big(\frac{1}{\sqrt{n^{\beta} L_s(n)}}\Big), \tag{3.7}$$

where $L_s(n)$ *is defined below (2.3).*

Remark 3.3. As $\beta < 1$ and $L_s(n)$ is a slowly varying positive function, for the integrable case, the rate of convergence of $\widehat{\boldsymbol{\theta}}_n$ is slower than $\sqrt{n}$, the rate of convergence of the parametric NLS estimator in the stationary time series case. Roughly speaking, the main reason for the reduction in the rate of convergence is because in the β-null recurrent case, the amount of time spent by the Markov chain around any particular point is of order $n^{\beta} L_s(n)$, which is less than n for the stationary (positive recurrent) time series case. Corollary 3.1 complements the existing results on the rates of convergence of nonparametric estimators in β-null recurrent Markov processes (see, for example, Karlsen *et al* 2007; Gao *et al* 2011a). For the random walk case, which corresponds to 1/2-null recurrent Markov chain, the rate of convergence is $n^{1/4}$, which is similar to a result obtained by PP for $\{X_t\}$ processes that are of $I(1)$ type.

3.2 Asymptotically homogeneous function on Θ

We next establish an asymptotic theory for a parametric estimator of $\boldsymbol{\theta}_0$ when the derivatives of $g(\cdot,\cdot)$ belong to a class of asymptotically homogeneous functions. For a unit root process $\{X_t\}$, PP establish the asymptotic consistency and limit distribution of the NLS estimator $\widehat{\boldsymbol{\theta}}_n$ by using the local time technique. Their method relies on the linear framework of the unit root process, the functional limit theorem of the partial sum process, and the continuous mapping theorem. The Harris recurrent Markov chain is a general process and allows for a possibly nonlinear framework, however. In particular, the null recurrent Markov chain can be seen as a nonlinear generalisation of the linear unit root process. Hence, the techniques used by PP for establishing the asymptotic theory is not applicable in such a possibly nonlinear Markov chain framework. Meanwhile, as mentioned in Section 1, the methods used to prove Theorem 3.1 cannot be applied here directly because the asymptotically homogeneous functions usually are not bounded and integrable. This leads to the violation of the conditions in the ergodic theorem of the Harris recurrent Markov process. Hence, it is quite challenging to establish an asymptotic theory for the NLS estimator for the case of asymptotically homogeneous functions.

To address the above concerns, we have to modify the NLS estimator $\widehat{\boldsymbol{\theta}}_n$. Let M_n be a positive and increasing sequence which satisfies $M_n \to \infty$ as $n \to \infty$, but is dominated by certain polynomial rate. We define the modified loss function by

$$Q_{n,g}(\boldsymbol{\theta}) = \sum_{t=1}^{n} \left[Y_t - g(X_t, \boldsymbol{\theta})\right]^2 I\left(|X_t| \leq M_n\right). \tag{3.8}$$

The modified NLS (MNLS) estimator $\overline{\boldsymbol{\theta}}_n$ can be obtained by minimizing $Q_{n,g}(\boldsymbol{\theta})$ over $\boldsymbol{\theta} \in \Theta$,

$$\overline{\boldsymbol{\theta}}_n = \arg\min_{\boldsymbol{\theta}\in\Theta} Q_{n,g}(\boldsymbol{\theta}). \tag{3.9}$$

As $M_n \to \infty$, for the integrable case discussed in Section 3.1, we can easily show that $\overline{\boldsymbol{\theta}}_n$ *has the same asymptotic distribution as* $\widehat{\boldsymbol{\theta}}_n$ *under some regularity conditions.* In Example 6.1 below, we compare the finite sample performance of these two estimators, and find that they are quite similar.

Let

$$B_i(1) = [i-1, i)\,, \quad i = 1, 2, \cdots, [M_n], \quad B_{[M_n]+1}(1) = [[M_n], M_n]\,,$$
$$B_i(2) = [-i, -i+1)\,, \quad i = 1, 2, \cdots, [M_n], \quad B_{[M_n]+1}(2) = [-M_n, -[M_n]]\,,$$

where $[x] \le x$ is the largest integer part of x. It is easy to check that $B_i(k)$, $i = 1, 2, \cdots, [M_n]+1$, $k = 1, 2$, are disjoint, and $[-M_n, M_n] = \cup_{k=1}^{2} \cup_{i=1}^{[M_n]+1} B_i(k)$. Let

$$\begin{aligned}\Lambda_n &= \sum_{i=0}^{[M_n]} \dot{h}_g(\frac{i}{M_n}, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(\frac{i}{M_n}, \boldsymbol{\theta}_0)\pi_s(B_{i+1}(1)) \\ &\quad + \sum_{i=0}^{[M_n]} \dot{h}_g(\frac{-i}{M_n}, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(\frac{-i}{M_n}, \boldsymbol{\theta}_0)\pi_s(B_{i+1}(2)),\end{aligned}$$

where $\dot{h}_g(\cdot, \cdot)$ will be defined in Assumption 3.3 (i) below.

Some additional assumptions are introduced below to establish asymptotic properties of $\overline{\boldsymbol{\theta}}_n$.

Assumption 3.3. (i) $\dot{g}(x, \boldsymbol{\theta})$ is asymptotically homogeneous on Θ with asymptotic order $\dot{\kappa}_g(\cdot, \cdot)$ and limit homogeneous function $\dot{h}_g(\cdot, \cdot)$. $\ddot{g}(x, \boldsymbol{\theta})$ is also asymptotically homogeneous on Θ with asymptotic order $\ddot{\kappa}_g(\cdot, \cdot)$ and limit homogeneous function $\ddot{h}_g(\cdot, \cdot)$.

(ii) Both $\dot{h}_g(\cdot, \boldsymbol{\theta})$ and $\ddot{h}_g(\cdot, \boldsymbol{\theta})$ are continuous on the interval $[-1, 1]$ for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$. There exist a positive definite matrix Λ_g and a sequence of positive numbers $\zeta_g(M_n)$ such that

$$\lim_{n\to\infty} \frac{1}{\zeta_g(M_n)} \Lambda_n = \Lambda_g.$$

(iii) For all $\boldsymbol{\theta} \in \Theta$, $\dot{\kappa}_g(i, \boldsymbol{\theta})$ and $\ddot{\kappa}_g(i, \boldsymbol{\theta})$ are positive and nondecreasing such that $\dot{\kappa}_g(n, \boldsymbol{\theta}) \to \infty$ and $\ddot{\kappa}_g(n, \boldsymbol{\theta}) = O\left(\dot{\kappa}_g(n, \boldsymbol{\theta})\right)$ as $n \to \infty$.

(iv) For each $x \in [-M_n, M_n]$, $\mathcal{N}_x(1) := \{y : x-1 < y < x+1\}$ is a small set.

REMARK 3.4. Assumption 3.3 (i) is quite standard, see, for example, condition (b) in Theorem 5.2 in PP. The explicit forms of Λ_g and $\zeta(M_n)$ in Assumption 3.3 (ii) can be derived for some particular cases. For example, when $\{X_t\}$ is generated by a positive recurrent Markov chain, we have

$$\begin{aligned}\Lambda_n &= (1+o(1))\int_{-M_n}^{M_n} \dot{h}_g(\frac{x}{M_n}, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(\frac{x}{M_n}, \boldsymbol{\theta}_0)\pi_s(dx)\\ &= \frac{(1+o(1))}{\dot{\kappa}_g^2(M_n, \boldsymbol{\theta}_0)}\int_{-\infty}^{\infty} \dot{g}(x, \boldsymbol{\theta}_0)\dot{g}^{\tau}(x, \boldsymbol{\theta}_0)\pi_s(dx).\end{aligned}$$

By letting $\zeta_g(M_n) = \dot{\kappa}_g^{-2}(M_n, \boldsymbol{\theta}_0)$ and assuming that

$$\int_{-\infty}^{\infty} \dot{g}(x, \boldsymbol{\theta}_0)\dot{g}^{\tau}(x, \boldsymbol{\theta}_0)\pi_s(dx) = \int_{-\infty}^{\infty} \dot{g}(x, \boldsymbol{\theta}_0)\dot{g}^{\tau}(x, \boldsymbol{\theta}_0)p_s(x)dx$$

is positive definite, we can verify the second condition in Assumption 3.3 (ii) for the positive recurrent case, where $\zeta_g(M_n) \to 0$ as $n \to \infty$. When $\{X_t\}$ is generated by a null recurrent Markov chain with $\pi_s(M_n dx) = \zeta_g(M_n)\pi_s(dx)$, we then have

$$\begin{aligned}\Lambda_n &= (1+o(1))\int_{-M_n}^{M_n} \dot{h}_g(\frac{x}{M_n}, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(\frac{x}{M_n}, \boldsymbol{\theta}_0)\pi_s(dx)\\ &= (1+o(1))\int_{-1}^{1} \dot{h}_g(x, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(x, \boldsymbol{\theta}_0)\pi_s(M_n dx)\\ &= (1+o(1))\zeta_g(M_n)\int_{-1}^{1} \dot{h}_g(x, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(x, \boldsymbol{\theta}_0)\pi_s(dx).\end{aligned}$$

By assuming that $\int_{-1}^{1} \dot{h}_g(x, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(x, \boldsymbol{\theta}_0)\pi_s(dx)$ is positive definite, we can verify the second condition in Assumption 3.3 (ii) for the null recurrent case. In particular, when $\{X_t\}$ is generated by a random walk process, we have

$$\begin{aligned}\Lambda_n &= (1+o(1))\sum_{i=-[M_n]}^{[M_n]} \dot{h}_g(\frac{i}{M_n}, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(\frac{i}{M_n}, \boldsymbol{\theta}_0)\\ &= (1+o(1))M_n\int_{-1}^{1} \dot{h}_g(x, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(x, \boldsymbol{\theta}_0)\pi_s(dx),\end{aligned}$$

which implies that $\zeta_g(M_n) = M_n$ in the second condition in Assumption 3.3 (ii). In fact, for the random walk case, by modifying the proof of Lemma B.3 in Appendix B, we need only to assume that there exists a $\delta > 0$ such that

$$\int_{-\delta}^{\delta} \dot{h}_g(x, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(x, \boldsymbol{\theta}_0)\pi_s(dx) = \int_{-\delta}^{\delta} \dot{h}_g(x, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(x, \boldsymbol{\theta}_0)dx$$

is positive definite, which is much weaker than condition (d) in Theorem 5.2 in PP where $\int_{-\delta}^{\delta} \dot{h}_g(x, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(x, \boldsymbol{\theta}_0)dx$ is assumed to be positive definite for all $\delta > 0$.

Define

$$\mathbf{J}_g(n, \boldsymbol{\theta}_0) = \dot{\kappa}_g^2(M_n, \boldsymbol{\theta}_0)\zeta_g(M_n)\Lambda_g.$$

We next establish an asymptotic theory for $\overline{\boldsymbol{\theta}}_n$. As for Theorem 3.1, the following theorem is also applicable for both the stationary and nonstationary time series cases.

THEOREM 3.2. *Let Assumptions 3.1 and 3.3 hold.*

(a) There exists a solution $\overline{\boldsymbol{\theta}}_n$ to minimize the loss function $Q_{n,g}(\boldsymbol{\theta})$, and

$$\overline{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_P(1). \tag{3.10}$$

(b) The estimator $\overline{\boldsymbol{\theta}}_n$ has the asymptotic normal distribution,

$$N^{1/2}(n)\mathbf{J}_g^{1/2}(n, \boldsymbol{\theta}_0)\left(\overline{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathsf{N}\left(\mathbf{0}, \sigma^2 I_d\right), \tag{3.11}$$

where I_d is a $d \times d$ identity matrix.

REMARK 3.5. From Theorem 3.2, the asymptotic distribution of $\overline{\boldsymbol{\theta}}_n$ for the asymptotically homogeneous case is quite different from that of $\widehat{\boldsymbol{\theta}}_n$ (and $\overline{\boldsymbol{\theta}}_n$) for integrable case. Such finding is comparable to those in PP. When $\{X_t\}$ is positive recurrent, $\mathbf{J}_g(n, \boldsymbol{\theta}_0)$ is bounded away from zero and infinity, and the role of $N(n)$ is the same as that of the sample size, which implies that the root-n consistency for the parametric estimator in the stationary time series case can be derived from our general asymptotic distribution theory in (3.11). The choice of M_n will be discussed in Corollaries 3.2 and 3.3 below.

From (3.11) in Theorem 3.2 and (2.3) in Section 2 above, we have the following two corollaries. The rate of convergence in (3.12) below is quite general for β-null recurrent Markov processes. When $\beta = 1/2$, it is the same as the convergence rate in Theorem 5.2 of PP.

Corollary 3.2. *Suppose that the conditions in Theorem 3.2 are satisfied. Let $\{X_t\}$ be a β-null recurrent Markov chain with $0 < \beta < 1$ and bounded invariant density function $p_s(\cdot)$. Taking $M_n = M_0 n^{1-\beta} L_s^{-1}(n)$ for some positive constant M_0, we have*

$$\overline{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_P\left((n \dot{\kappa}_g^2(M_n, \boldsymbol{\theta}_0))^{-1/2} \right). \tag{3.12}$$

Corollary 3.3. *Suppose that the conditions in Theorem 3.2 are satisfied. Let $g(x, \boldsymbol{\theta}_0) = x\theta_1$, $\{X_t\}$ be a random walk process and $M_n = M_0 n^{1/2}$ for some positive constant M_0. Then we have*

$$\overline{\theta}_{1n} - \theta_1 = O_P\left(n^{-1}\right), \tag{3.13}$$

where $\overline{\theta}_{1n}$ is the modified NLS estimator of θ_1. Furthermore,

$$\sqrt{M_0^3 N(n) n^{3/2}} \left(\overline{\theta}_{1n} - \theta_1\right) \xrightarrow{d} \mathsf{N}\left(0, 3\sigma^2/2\right). \tag{3.14}$$

Remark 3.6. In the above two corollaries, we give the choice of M_n for some special cases. In fact, for the random walk process $\{X_t\}$ defined as in Example 2.1, where $\{x_t\}$ is a sequence of i.i.d. random variables with zero mean and finite variance, we have

$$\frac{1}{\sqrt{n}} X_{[nr]} = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nr]} x_i \Longrightarrow B(r),$$

where $B(r)$ is a standard Brownian motion and " $\Longrightarrow$ " denotes the weak convergence. Furthermore, by the continuous mapping theorem (see, for example, Billingsley 1968),

$$\sup_{0 \le r \le 1} \frac{1}{\sqrt{n}} X_{[nr]} \Longrightarrow \sup_{0 \le r \le 1} B(r),$$

which implies that it is reasonable to let $M_n = C_\alpha n^{1/2}$, where C_α may be chosen such that

$$\alpha = \mathsf{P}\Big(\sup_{0 \le r \le 1} B(r) > C_\alpha \Big) = \mathsf{P}\Big(|B(1)| > C_\alpha \Big) = 2\left(1 - \Phi\left(C_\alpha\right)\right), \tag{3.15}$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$. This implies that C_α can be obtained when α is given, such as $\alpha = 0.05$.

If $\{X_t\}$ is stationary and positive recurrent and its distribution has a compact support, say $[-M_0, M_0]$, we can let $M_n \equiv M_0$. In addition, for the simple linear regression model with regressors generated by a random walk process, (3.12) and (3.13) imply the existence of super consistency. Corollaries 3.1 and 3.2 imply that the rates of convergence for the parametric estimator in nonlinear cointegrating models rely not only on the properties of the function $g(\cdot, \cdot)$, but also on the magnitude of β.

4 Nonlinear heteroskedastic regression

In this section, we aim to relax the homoskedastic assumption on model (1.1) and extend our methodology to the heteroskedastic case. We will introduce an estimation method for a parameter vector involved in the conditional variance function. As mentioned in Remark 3.1, we consider the following case

$$\mathsf{E}(e_t^2 | \mathcal{F}_{n,t-1}) = \sigma^2(X_t, \boldsymbol{\gamma}_0) \quad a.s. \quad \text{for} \quad \boldsymbol{\gamma}_0 \in \Upsilon \subset \mathbb{R}^p, \tag{4.1}$$

where $\sigma^2(\cdot, \cdot) : \mathbb{R}^{p+1} \to \mathbb{R}$ is positive, and $\boldsymbol{\gamma}_0$ is the true value of the p-dimensional parameter vector involved in the conditional variance function. Estimation of the parametric nonlinear variance function defined in (4.1) is important in empirical applications as many scientific studies depend on understanding the variability of the data. However, so far as we know, there is very little literature on the asymptotic analysis of nonlinear and nonstationary volatility regression (PP only consider the parametric nonlinear regression with unit root regressors for the homoskedastic case). An interesting question is whether there exists a super-consistency result for the nonlinear heteroskedastic regression with null recurrent Markov chains. We will provide a partial answer to this question below.

For simplicity, we focus on the nonlinear and nonstationary heteroskedastic regression defined by

$$Y_t = \sigma(X_t, \boldsymbol{\gamma}_0) e_{t*}, \quad t = 1, \cdots, n, \tag{4.2}$$

where $\{e_{t*}\}$ is a sequence satisfying

$$\mathsf{E}(e_{t*}|\mathcal{F}^*_{n,t-1}) = 0, \quad \mathsf{E}(e^2_{t*}|\mathcal{F}^*_{n,t-1}) = 1 \;\; a.s., \tag{4.3}$$

$\mathcal{F}^*_{nt} = \sigma\left(X_{s_1}, e_{s_2*},\ 1 \leq s_1 \leq n,\ 1 \leq s_2 \leq t\right)$. Letting ϖ_0 be a positive number such that $\mathsf{E}\left[\log(e^2_{t*})|\mathcal{F}^*_{n,t-1}\right] = \log(\varpi_0)$ a.s., we have

$$\begin{aligned} \log(Y_t^2) &= \log\left(\sigma^2(X_t, \boldsymbol{\gamma}_0)\right) + \log(e^2_{t*}) \\ &= \log\left(\sigma^2(X_t, \boldsymbol{\gamma}_0)\right) + \log(\varpi_0) + \log(e^2_{t*}) - \log(\varpi_0) \\ &=: \log\left(\varpi_0 \sigma^2(X_t, \boldsymbol{\gamma}_0)\right) + \zeta_t. \end{aligned} \tag{4.4}$$

It is easy to see that $\mathsf{E}(\zeta_t|\mathcal{F}^*_{n,t-1}) = 0$ a.s. Since our main interest lies in the discussion of the asymptotic theory for the estimator of $\boldsymbol{\gamma}_0$, we first assume that ϖ_0 is known to simplify our discussion.

Model (4.4) can be seen as another nonlinear regression model with parameter vector $\boldsymbol{\gamma}_0$ to be estimated. The log-transformation would make data less skewed, and thus the resulting volatility estimator may be more efficient in terms of dealing with heavy-tailed $\{e_{t*}\}$. Such transformation has been commonly used to estimate the variability of the data in the stationary time series case (see, for example, Peng and Yao 2003; Gao 2007; Chen *et al* 2009). However, any extension to Harris recurrent Markov chains (which may be nonstationary) has not been done in the literature.

Our estimation method will be constructed based on (4.4). Noting that ϖ_0 is assumed to be known, define

$$\sigma^2_*(X_t, \boldsymbol{\gamma}_0) = \varpi_0 \sigma^2(X_t, \boldsymbol{\gamma}_0), \quad g_*(X_t, \boldsymbol{\gamma}_0) = \log(\sigma^2_*(X_t, \boldsymbol{\gamma}_0)).$$

Case (i). If the derivatives of $g_*(X_t, \boldsymbol{\gamma})$ are integrable on Υ, the log-transformed nonlinear least squares (LNLS) estimator $\widehat{\boldsymbol{\gamma}}_n$ can be obtained by minimizing $L_{n,\sigma}(\boldsymbol{\gamma})$

over $\boldsymbol{\gamma} \in \Upsilon$, where

$$L_{n,\sigma}(\boldsymbol{\gamma}) = \sum_{t=1}^{n} \left[\log(Y_t^2) - g_*(X_t, \boldsymbol{\gamma})\right]^2. \tag{4.5}$$

Letting Assumptions 3.1 and 3.2 be satisfied with e_t and $g(\cdot, \cdot)$ replaced by ζ_t and $g_*(\cdot, \cdot)$, respectively, then the asymptotic results developed in Section 3.1 still hold for $\widehat{\boldsymbol{\gamma}}_n$.

Case (ii). If the derivatives of $g_*(X_t, \boldsymbol{\gamma})$ are asymptotically homogeneous on Υ, the log-transformed modified nonlinear least squares (LMNLS) estimator $\overline{\boldsymbol{\gamma}}_n$ can be obtained by minimizing $Q_{n,\sigma}(\boldsymbol{\gamma})$ over $\boldsymbol{\gamma} \in \Upsilon$, where

$$Q_{n,\sigma}(\boldsymbol{\gamma}) = \sum_{t=1}^{n} \left[\log(Y_t^2) - \log\left(\sigma_*^2(X_t, \boldsymbol{\gamma})\right)\right]^2 I\left(|X_t| \leq M_n\right), \tag{4.6}$$

where M_n is defined as in Section 3.2. Then, the asymptotic results developed in Section 3.2 still hold for $\overline{\boldsymbol{\gamma}}_n$ under some regularity conditions such as the slightly modified version of Assumptions 3.1 and 3.3. Hence, *it is possible to achieve the super-consistency result for $\overline{\boldsymbol{\gamma}}_n$ in this case.* In particular, if

$$\frac{\partial g_*(x, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = \frac{\dot{\sigma}^2(x, \boldsymbol{\gamma})}{\sigma^2(x, \boldsymbol{\gamma})}$$

equals a constant vector, where $\dot{\sigma}^2(x, \boldsymbol{\gamma}) = \frac{\partial \sigma^2(x, \boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}$, we can show that the $\sqrt{n}$-rate can be achieved without the process itself being stationary.

We next give some examples to show that different conditions on $\sigma^2(\cdot, \cdot)$ would lead to different properties of $g_*(\cdot, \cdot)$ in the above two cases.

Example 4.1. Let us consider a class of popular volatility functions of the form $\sigma^2(x, \gamma) = (x^2 + \gamma)^q$ with $q, \gamma > 0$. Recall that $\dot{\sigma}^2(x, \gamma) = \frac{\partial \sigma^2(x, \gamma)}{\partial \gamma}$ and note that

$$\frac{\partial g_*(x, \gamma)}{\partial \gamma} = \frac{\dot{\sigma}^2(x, \gamma)}{\sigma^2(x, \gamma)} = \frac{q}{x^2 + \gamma}$$

is integrable. Hence, we have to use the estimation methodology introduced in Case (i), although $\sigma^2(x, \gamma) = (x^2 + \gamma)^q$ belongs to a class of asymptotically homogeneous functions when $q > 0$.

Example 4.2. Let $\sigma^2(x,\gamma) = e^{-\left(x^2\gamma\right)^q}$ with $q, \gamma > 0$, which is another class of popular volatility functions with an exponential rate of decaying to zero when $x \to \infty$. In this case, we have

$$\frac{\partial g_*(x,\gamma)}{\partial \gamma} = -qx^{2q}\gamma^{q-1},$$

which is asymptotically homogeneous. Hence, we have to use the estimation methodology introduced in Case (ii), although $\sigma^2(x,\gamma) = e^{-\left(x^2\gamma\right)^q}$ belongs to a class of integrable functions. Similarly, if $\sigma^2(x,\gamma) = e^{\left(x^2\gamma\right)^q}$ with $q > 1$, which belongs to a class of explosive functions (see Park and Phillips 1999 for definitions), we can also show that $\frac{\partial g_*(x,\gamma)}{\partial \gamma} = qx^{2q}\gamma^{q-1}$ is asymptotically homogeneous, which indicates that we need to apply the estimation methodology discussed in Case (ii).

In practice, however, ϖ_0 is usually unknown and needs to be estimated. We next briefly tackle this issue. By (4.4), define the loss function by

$$Q_n(\boldsymbol{\gamma}, \varpi) = \sum_{t=1}^{n} \left[\log(\overline{e}_t^2) - \log\left(\varpi\sigma^2(X_t, \boldsymbol{\gamma})\right)\right]^2 I\left(|X_t| \le M_n\right).$$

Then, the estimators $\overline{\boldsymbol{\gamma}}_n$ and $\overline{\varpi}_n$ can be obtained by minimizing $Q_n(\boldsymbol{\gamma}, \varpi)$ over $\boldsymbol{\gamma} \in \Upsilon$ and $\varpi \in \mathbb{R}^+$.

5 Nonlinear regression with $I(1)$ processes

As mentioned before, PP consider the nonlinear regression (1.1) with the regressors $\{X_t\}$ generated by

$$X_t = X_{t-1} + x_t, \quad x_t = \sum_{j=0}^{\infty} \phi_j \varepsilon_{t-j}, \tag{5.1}$$

where $\{\varepsilon_j\}$ is a sequence of i.i.d. random variables and $\{\phi_j\}$ satisfies some summability conditions. For simplicity, we assume that $X_0 = 0$ throughout this section. PP establish a suite of asymptotic results for the NLS estimator of the parameter $\boldsymbol{\theta}_0$ involved in (1.1) when $\{X_t\}$ is defined by (5.1). An open problem is how to establish some corresponding results by using the β-null recurrent Markov chain framework.

This is quite challenging as $\{X_t\}$ defined by (5.1) is no longer a Markov process except for some special cases (for example, $\phi_j = 0$ for $j \geq 1$).

The aim of this section is to solve this open problem when $g(\cdot, \cdot)$ is asymptotically homogeneous on Θ and derive an asymptotic theory for $\overline{\boldsymbol{\theta}}_n$ by using Theorem 3.2. The discussion for the integrable case is more complicated, and will be considered in a future study. We first introduce some regularity conditions for the establishment of the asymptotic theory.

ASSUMPTION 5.1. (i) Let $\{\phi_j\}$ satisfy the conditions

$$\sum_{j=0}^{\infty} |\phi_j| < \infty, \quad \phi = \sum_{j=0}^{\infty} \phi_j \neq 0, \quad \Big(\sum_{j=s+1}^{\infty} \phi_j\Big)^2 = O(s^{-(1+\delta)}) \tag{5.2}$$

for some $\delta > 0$.

(ii) $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with zero mean and finite variance.

(iii) The error process $\{e_t\}$ involved in model (1.1) is a sequence satisfying

$$\mathsf{E}(e_t|\mathcal{G}_{n,t-1}) = 0, \quad \mathsf{E}(e_t^2|\mathcal{G}_{n,t-1}) = \sigma^2 < \infty \ a.s.,$$

where $\mathcal{G}_{n,t} = \sigma\left(\varepsilon_{s_1}, e_{s_2}, \ -\infty \leq s_1 \leq n, \ 1 \leq s_2 \leq t\right)$.

(iv) $\mathcal{N}_x(1)$ is a small set for the random walk process $\{X_t^*\}$ defined by $X_t^* = \phi \sum_{s=1}^{t} \varepsilon_s$, $x \in [-M_n, M_n]$.

ASSUMPTION 5.2. (i) Assumptions 3.3 (i) and (iii) are satisfied. Furthermore, both $\dot{h}_g(\cdot, \boldsymbol{\theta})$ and $\ddot{h}_g(\cdot, \boldsymbol{\theta})$ are continuous on the interval $[-1, 1]$ for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$, and the matrix $\int_{-1}^{1} \dot{h}_g(x, \boldsymbol{\theta}_0)\dot{h}_g^{\tau}(x, \boldsymbol{\theta}_0)dx$ is positive definite

(ii) Uniformly for $\boldsymbol{\theta} \in \Theta$, we have

$$\|\dot{g}(x+y, \boldsymbol{\theta}) - \dot{g}(x, \boldsymbol{\theta})\| \leq \|\mathcal{B}_{\dot{g}}(x, \boldsymbol{\theta})y\| \tag{5.3}$$

and

$$\|\ddot{g}(x+y, \boldsymbol{\theta}) - \ddot{g}(x, \boldsymbol{\theta})\| \leq \|\mathcal{B}_{\ddot{g}}(x, \boldsymbol{\theta})y\| \tag{5.4}$$

where $\mathcal{B}_{\dot{g}}(x, \boldsymbol{\theta})$ and $\mathcal{B}_{\ddot{g}}(x, \boldsymbol{\theta})$ are asymptotically homogeneous with asymptotic orders $\kappa_{\mathcal{B}_1}(\cdot, \cdot)$ and $\kappa_{\mathcal{B}_2}(\cdot, \cdot)$, respectively. Let $\kappa_{\mathcal{B}_1}(\cdot, \cdot)$ and $\kappa_{\mathcal{B}_2}(\cdot, \cdot)$ satisfy

$$\frac{\kappa^2_{\mathcal{B}_1}(M_n, \boldsymbol{\theta})}{\dot{\kappa}^2_g(M_n, \boldsymbol{\theta})} = O\left(n^{-\frac{1}{2+\delta}}\right), \quad \frac{\kappa^2_{\mathcal{B}_2}(M_n, \boldsymbol{\theta})}{\dot{\kappa}^2_g(M_n, \boldsymbol{\theta})} = O\left(n^{-\frac{1}{2+\delta}}\right) \tag{5.5}$$

uniformly for $\boldsymbol{\theta} \in \Theta$, where $\dot{\kappa}_g(\cdot, \cdot)$ is defined in Assumption 3.3 (i).

Remark 5.1. Assumption 5.1 contains some mild conditions on the linear process $\{x_t\}$ as well as on the error term $\{e_t\}$. The conditions in (5.2) ensure that $\{x_t\}$ is short-range dependent (see, for example, Wang and Phillips 2009, Wang and Chan 2011), and it can be satisfied by some popular time series models such as an autoregressive model. Assumption 5.2 states some smoothness conditions on $g(\cdot, \cdot)$ as well as its derivatives. In particular, when ϕ_j decays with an exponential rate, δ can be taken arbitrarily large, which implies that condition (5.5) is quite weak.

Define

$$\mathbf{J}_{g*}(n, \boldsymbol{\theta}_0) = \dot{\kappa}^2_g(M_n, \boldsymbol{\theta}_0) M_n \left(\int_{-1}^{1} \dot{h}_g(x, \boldsymbol{\theta}_0) \dot{h}^{\tau}_g(x, \boldsymbol{\theta}_0) dx \right). \tag{5.6}$$

We next give some asymptotic results for $\overline{\boldsymbol{\theta}}_n$ in the case where $\{X_t\}$ is a unit root process.

Theorem 5.1 *Let Assumptions 5.1 and 5.2 hold, and* $n^{-\frac{1}{2(2+\delta)}} M_n \to \infty$.

(a) There exists a solution $\overline{\boldsymbol{\theta}}_n$ *to minimize the loss function* $Q_{n,g}(\boldsymbol{\theta})$ *and*

$$\overline{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = o_P(1). \tag{5.7}$$

(b) The estimator $\overline{\boldsymbol{\theta}}_n$ *has the asymptotic normal distribution,*

$$N^{1/2}_{\varepsilon}(n) \mathbf{J}^{1/2}_{g*}(n, \boldsymbol{\theta}_0) \left(\overline{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right) \xrightarrow{d} \mathsf{N}\left(\mathbf{0}, \sigma^2 I_d\right), \tag{5.8}$$

where $N_{\varepsilon}(n)$ *is the number of regenerations for the random walk process* $\{X^*_t\}$.

Remark 5.2. Theorem 5.1 establishes an asymptotic theory for $\overline{\boldsymbol{\theta}}_n$ for the case where $\{X_t\}$ is a unit root process. Our results are comparable with Theorems 5.2 and 5.3 in PP. However, we establish asymptotic normality in (5.7) with stochastic rate

$N_\varepsilon^{1/2}(n)\mathbf{J}_{g*}^{1/2}(n,\boldsymbol{\theta}_0)$, and PP establish their asymptotic mixed normal distribution theory with a deterministic rate. As $N_\varepsilon^{1/2}(n)\mathbf{J}_{g*}^{1/2}(n,\boldsymbol{\theta}_0) \propto \sqrt{n}\mathbf{J}_{g*}^{1/2}(n,\boldsymbol{\theta}_0)$ in probability, if we take $M_n = M_0\sqrt{n}$ as in Corollary 3.3, we will find that our rate of convergence of $\overline{\boldsymbol{\theta}}_n$ is the same as that derived by PP.

6 Simulated examples

In this section, we provide some simulation studies to compare the finite sample performance of various parametric estimation methods introduced in Sections 3–5 and to illustrate the developed asymptotic theory.

Example 6.1. Consider the generalised linear model defined by

$$Y_t = \exp\{-\theta_0 X_t^2\} + e_t, \quad \theta_0 = 1, \quad t = 1, 2, \cdots, n, \tag{6.1}$$

where $\{X_t\}$ is generated by one of the three Markov processes:

(i) AR(1) process: $X_t = 0.5X_{t-1} + x_t$, where $X_0 = 0$ and $\{x_t\}$ is a sequence of i.i.d. standard normal random variables;

(ii) Random walk process: $X_t = X_{t-1} + x_t$, where $X_0 = 0$ and $\{x_t\}$ is a sequence of i.i.d. standard normal random variables;

(iii) TAR(1) process: $X_t = 0.5X_{t-1}I(|X_{t-1}| \leq 1) + X_{t-1}I(|X_{t-1}| > 1) + x_t$, where $X_0 = 0$ and $\{x_t\}$ is a sequence of i.i.d. standard normal random variables.

The error process $\{e_t\}$ is a sequence of i.i.d. $\mathsf{N}(0, 0.5^2)$ random variables and independent of $\{x_t\}$. In this simulation study, we compare the finite sample behavior of the NLS estimator $\widehat{\theta}_n$ with that of the MNLS estimator $\overline{\theta}_n$, and the sample size n is chosen to be 500, 1000 and 2000. The aim of this example is to illustrate the asymptotic theory developed in Section 3.1 as the regression function in (6.1) is integrable when $\theta_0 \neq 0$. Following the discussion in Section 2.2, the AR(1) process defined in (i) is positive recurrent, and the random process defined in (ii) and the TAR(1) process defined in (iii) are 1/2-null recurrent.

We generate 500 replicated samples for this simulation study, and calculate the means and standard errors for both of the parametric estimators in 500 simulations. In the MNLS estimation procedure, we choose $M_n = C_\alpha n^{1-\beta}$ with $\alpha = 0.01$, where C_α is defined in (3.15), $\beta = 1$ for case (i), and $\beta = 1/2$ for cases (ii) and (iii). It is easy to find that $C_{0.01} = 2.58$.

The simulation results are reported in Table 6.1, where the numbers in the parentheses are the standard errors of the NLS (or MNLS) estimator in 500 replications. From Table 6.1, we have the following interesting findings. (a) The parametric estimators perform better in the stationary case (i) than in the nonstationary cases (ii) and (iii). This is consistent with the asymptotic results obtained in Section 3.1 such as Theorem 3.1 and Corollary 3.1, which indicate that the convergence rates of the parametric estimators can achieve $O_P(\sqrt{n})$ in the stationary case, but only $O_P(n^{1/4})$ in the 1/2-null recurrent case. (b) The finite sample behavior of the MNLS estimator is exactly the same as that of NLS estimator since $\alpha = 0.01$ means little sample information is lost. (c) Both of the two parametric estimators improve as the sample size increases. (d) In addition, for case (i), the ratio of the standard errors between 500 and 2000 is 1.9633 (close to the theoretical ratio $\sqrt{4} = 2$); for case (iii), the ratio of the standard errors between 500 and 2000 is 1.3005 (close to the theoretical ratio $4^{1/4} = 1.4142$). Hence, this again confirms that our asymptotic theory is valid.

Table 6.1. Means and standard errors for the estimators in Example 6.1

sample size	500	1000	2000
The regressor X_t is generated in case (i)			
NLS	1.0036 (0.0481)	1.0002 (0.0339)	1.0000 (0.0245)
MNLS	1.0036 (0.0481)	1.0002 (0.0339)	1.0000 (0.0245)
The regressor X_t is generated in case (ii)			
NLS	0.9881 (0.1783)	0.9987 (0.1495)	0.9926 (0.1393)
MNLS	0.9881 (0.1783)	0.9987 (0.1495)	0.9926 (0.1393)
The regressor X_t is generated in case (iii)			
NLS	0.9975 (0.1692)	1.0028 (0.1463)	0.9940 (0.1301)
MNLS	0.9975 (0.1692)	1.0028 (0.1463)	0.9940 (0.1301)

Example 6.2. Consider the linear model defined by

$$Y_t = (1 + X_t^2)^{\theta_0} + e_t, \quad \theta_0 = 0.5, \quad t = 1, 2, \cdots, n, \tag{6.2}$$

where $\{X_t\}$ is generated either by one of the three Markov processes introduced in Example 6.1, or by (iv) the unit root process:

$$X_t = X_{t-1} + x_t, \quad x_t = 0.2x_{t-1} + v_t,$$

in which $X_0 = x_0 = 0$, $\{v_t\}$ is a sequence of i.i.d. $\mathsf{N}(0, 0.75)$ random variables, and the error process $\{e_t\}$ is defined as in Example 6.1. In this simulation study, we are interested in the finite sample behavior of the MNLS estimator to illustrate the asymptotic theory developed in Sections 3.2 and 5 as the regression function in (6.2) is asymptotically homogeneous. For the comparison purpose, we also investigate the finite sample behavior of the NLS estimation, although we do not establish the related asymptotic theory under the framework of null recurrent Markov chains. The sample size n is chosen to be 500, 1000 and 2000 as in Example 6.1 and the replication number is $R = 500$. In the MNLS estimation procedure, as in the previous example, we choose $M_n = 2.58n^{1-\beta}$, where $\beta = 1$ for case (i), and $\beta = 1/2$ for cases (ii)–(iv).

Table 6.2. Means and standard errors for the estimators in Example 6.2

sample size	500	1000	2000
The regressor X_t is generated in case (i)			
NLS	0.5002 (0.0122)	0.5001 (0.0081)	0.4998 (0.0057)
MNLS	0.5005 (0.0149)	0.5001 (0.0096)	0.4994 (0.0068)
The regressor X_t is generated in case (ii)			
NLS	0.5000 (0.0004)	0.5000 (0.0002)	0.5000 (0.0001)
MNLS	0.5000 (0.0004)	0.5000 (0.0002)	0.5000 (0.0001)
The regressor X_t is generated in case (iii)			
NLS	0.5000 (0.0005)	0.5000 (0.0002)	0.5000 (0.0001)
MNLS	0.5000 (0.0005)	0.5000 (0.0002)	0.5000 (0.0001)
The regressor X_t is generated in case (iv)			
NLS	0.4555 (0.0935)	0.4554 (0.0911)	0.4540 (0.0868)
MNLS	0.4558 (0.0933)	0.4554 (0.0912)	0.4540 (0.0868)

The simulation results are reported in Table 6.2, from which, we have the following conclusions. (a) For the nonlinear cointegration model with asymptotically homogeneous regression function, the parametric estimators perform better in the nonstationary cases (ii) and (iii) than in the stationary case (i). This finding is consistent with the asymptotic results obtained in Section 3.2 such as Theorem 3.2 and Corollaries 3.2 and 3.3, which indicate that the convergence rates of the parametric estimators can achieve $O_P(n)$ in the 1/2-null recurrent case, faster than the corresponding rates in the stationary case. (b) For the case of asymptotically homogeneous regression function, our MNLS estimator performs as well as the NLS estimator. (c) Both the NLS estimation procedure proposed by PP and our MNLS estimation procedure perform poorly for the non-Markov unit root process (iv), which needs to be further studied in future research.

EXAMPLE 6.3. Consider a heteroskedastic model of the form:

$$Y_t = \sigma(X_t, \gamma_0)\ e_{t*}, \quad t = 1, 2, \cdots, n, \tag{6.3}$$

where we consider the following two forms:

Form (i): $\sigma(x, \gamma_0) = \sqrt{x^2 + \gamma_0}$ with $\gamma_0 = 0.2$;

Form (ii): $\sigma(x, \gamma_0) = e^{-\sqrt{x^2\gamma_0}}$ with $\gamma_0 = 0.2$,

$\{X_t\}$ is generated by one of the three Markov processes introduced in Example 6.1, $e_{t*} = \nu_t|\eta_t|$, $\{\eta_t\}$ is a sequence of i.i.d. random variables with $\log(\eta_t^2) \sim U(-1, 1)$, $\{\nu_t\}$ is also i.i.d. and independent of $\{\eta_t\}$ with $\mathsf{P}(\nu_t = 1) = \mathsf{P}(\nu_t = -1) = 1/2$. It is easy to see that $\mathsf{E}[e_{t*}] = 0$ and $\log(\varpi_0) = 0$, where ϖ_0 is defined as in Section 4. In this simulation study, as discussed in Section 4, we are interested in the finite sample behaviors of the LNLS estimator $\widehat{\gamma}_n$ for Form (i) and the LMNLS estimator $\overline{\gamma}_n$ for Form (ii). The sample size n is chosen to be 500, 1000 and 2000 as in the above two examples and the replication number is $R = 500$. Following the analysis in Example 6.2, in the parametric estimation procedure, we choose $M_n = 2.58n^{1-\beta}$.

Table 6.3. Means and standard errors for the estimators in Example 6.3

Estimation	$n = 500$	$n = 1000$	$n = 2000$
The regressor X_t is generated in case (i)			
LNLS $\widehat{\gamma}_n$	0.1997 (0.0109)	0.2006 (0.0077)	0.1998 (0.0054)
LMNLS $\overline{\gamma}_n$	0.2021 (0.0112)	0.1999 (0.0078)	0.2002 (0.0056)
The regressor X_t is generated in case (ii)			
LNLS $\widehat{\gamma}_n$	0.2019 (0.0600)	0.1976 (0.0527)	0.2047 (0.0428)
LMNLS $\overline{\gamma}_n$	0.2000 (0.0011)	0.2000 (0.0006)	0.2000 (0.0003)
The regressor X_t is generated in case (iii)			
LNLS $\widehat{\gamma}_n$	0.1968 (0.0539)	0.1958 (0.0475)	0.2004 (0.0416)
LMNLS $\overline{\gamma}_n$	0.1999 (0.0012)	0.2000 (0.0006)	0.2000 (0.0003)

The simulation results are reported in Table 6.3, from which we can obtain the following conclusions. (a) When the conditional volatility function is defined as in Form (i), the LNLS estimator $\widehat{\gamma}_n$ performs much better in stationary case (i), as there is no super-consistency result in nonstationary cases (ii) and (iii) (see Example 4.1 for detailed discussion). In addition, for the nonstationary cases (ii) and (iii), the standard errors of $\overline{\gamma}_n$ are much smaller than those of $\widehat{\gamma}_n$. (b) When the conditional volatility function is defined as in Form (ii), the LWNLS estimator $\overline{\gamma}_n$ performs much better in the nonstationary case because of the existence of the super-consistency result discussed in Section 4.

7 Conclusion

In this paper, we have systematically studied the nonlinear regression under the general Harris recurrent Markov chain framework, which includes both the stationary and nonstationary cases. Note that the nonstationary null recurrent process considered in this paper is under Markov perspective, which indicates that our methodology has the potential of being extended to the nonlinear autoregressive case. A recent paper by Chan and Wang (2012) also consider a nonlinear cointegrating regression with nonstationary regressors. However, when the regressors are Harris recurrent, they cannot derive the asymptotic distribution theory for the asymptotically homogeneous case. In this paper, we not only develop asymptotic theory for the estimator of $\boldsymbol{\theta}_0$ when $g(\cdot,\cdot)$ is integrable, but also modify the conventional NLS estimator for the asymptotically homogeneous $g(\cdot,\cdot)$ and use a novel method to establish asymptotic theory for the proposed modified parametric estimator. Furthermore, by using the log-transformation, we discuss the estimation of the parameter vector in a conditional volatility function and establish its asymptotic property. Finally, we apply our results to the nonlinear regression with $I(1)$ processes which may be non-Markovian, and establish an asymptotic distribution theory which is comparable to that obtained by

PP. Some simulation studies are provided to illustrate our approaches and results.

A Proofs of the main results

In this appendix, we give the detailed proofs of the main results.

Proof of Theorem 3.1. For (a), it suffices to prove that for $\epsilon > 0$,

$$\lim_{n\to\infty} \mathsf{P}\Big(\inf_{\boldsymbol{\theta}\in S_{\boldsymbol{\theta}_0}(\epsilon)} L_n(\boldsymbol{\theta}) \geq L_n(\boldsymbol{\theta}_0)\Big) = 1, \tag{A.1}$$

where $S_{\boldsymbol{\theta}_0}(\epsilon)$ is a circle centered at $\boldsymbol{\theta}_0$ with radius ϵ. In fact, (A.1) implies that there exists a minimum $\widehat{\boldsymbol{\theta}}_n$ in the interior of $S_{\boldsymbol{\theta}_0}(\epsilon)$, and

$$\mathsf{P}\left(|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0| \leq \epsilon\right) \to 1,$$

which leads to (3.4).

By a second order Taylor expansion of $L_n(\boldsymbol{\theta})$, we have

$$L_n(\boldsymbol{\theta}) = L_n(\boldsymbol{\theta}_0) + \dot{L}_n^{\tau}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^{\tau}\ddot{L}_n(\boldsymbol{\theta}_*)(\boldsymbol{\theta}-\boldsymbol{\theta}_0), \tag{A.2}$$

where $\boldsymbol{\theta}_*$ lies in the line segment connecting $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$,

$$\begin{aligned}\dot{L}_n(\boldsymbol{\theta}) &= -2\sum_{t=1}^{n}\dot{g}(X_t,\boldsymbol{\theta})\big[Y_t - g(X_t,\boldsymbol{\theta})\big],\\ \ddot{L}_n(\boldsymbol{\theta}) &= 2\Big\{\sum_{t=1}^{n}\dot{g}(X_t,\boldsymbol{\theta})\dot{g}^{\tau}(X_t,\boldsymbol{\theta}) - \sum_{t=1}^{n}\ddot{g}(X_t,\boldsymbol{\theta})\big[Y_t - g(X_t,\boldsymbol{\theta})\big]\Big\}.\end{aligned}$$

By (B.4) in Lemma B.2 in Appendix B, using the martingale difference structure on $\{e_t\}$ in Assumption 3.1 (ii) and the integrability condition in Assumption 3.2, we can show that

$$\dot{L}_n(\boldsymbol{\theta}_0) = O_P(\sqrt{N(n)}) = o_P(N(n)). \tag{A.3}$$

By (B.3) in Lemma B.2, we have

$$\frac{1}{N(n)}\sum_{t=1}^{n}\dot{g}(X_t,\boldsymbol{\theta}_*)\dot{g}^{\tau}(X_t,\boldsymbol{\theta}_*) = \int \dot{g}(x,\boldsymbol{\theta}_*)\dot{g}^{\tau}(x,\boldsymbol{\theta}_*)\pi_s(dx) + o_P(1). \tag{A.4}$$

On the other hand, by Definition 2.3 and Lemma B.2, we also have

$$\Big\|\frac{1}{N(n)}\sum_{t=1}^{n}\ddot{g}(X_t,\boldsymbol{\theta}_*)\big[Y_t - g(X_t,\boldsymbol{\theta}_*)\big]\Big\| = O_P(\epsilon), \tag{A.5}$$

where ϵ is the same as that in (A.1).

Let $\lambda_{\min} > 0$ be the smallest eigenvalue of the positive definite matrix $\ddot{L}(\boldsymbol{\theta}_0)$, where $\ddot{L}(\boldsymbol{\theta}_0) = \int \dot{g}(x, \boldsymbol{\theta}_0)\dot{g}^\tau(x, \boldsymbol{\theta}_0)\pi_s(dx)$ is defined as in Assumption 3.2. By (A.4) and (A.5), for $\boldsymbol{\theta}$ located on the circle $S_{\boldsymbol{\theta}_0}(\epsilon)$, we have

$$\mathsf{P}\Big(\inf_{\boldsymbol{\theta}\in S_{\boldsymbol{\theta}_0}(\epsilon)}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^\tau \ddot{L}_n(\boldsymbol{\theta}_*)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) > \frac{1}{2}\lambda_{\min}\epsilon^2\Big) = 1. \tag{A.6}$$

Then (A.2), (A.3) and (A.6) imply that (A.1) holds.

We next turn to the proof of part (b) in Theorem 3.1. By some elementary calculations as above, we have

$$\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = -\ddot{L}_n^+(\boldsymbol{\theta}_0)\dot{L}_n(\boldsymbol{\theta}_0)(1 + o_P(1)), \tag{A.7}$$

where $\ddot{L}_n^+(\boldsymbol{\theta}_0)$ is the Moore-Penrose inverse matrix of $\ddot{L}_n(\boldsymbol{\theta}_0)$.

By (B.3) and (B.4) in Lemma B.2, we have

$$\frac{1}{2N(n)}\ddot{L}_n(\boldsymbol{\theta}_0) = \ddot{L}(\boldsymbol{\theta}_0) + o_P(1). \tag{A.8}$$

By (A.7) and (A.8), to prove (b), we need only to prove that

$$\frac{1}{2\sqrt{N(n)}}\dot{L}_n(\boldsymbol{\theta}_0) \xrightarrow{d} \mathsf{N}\left(\mathbf{0}, \Sigma_1(\boldsymbol{\theta}_0)\right), \tag{A.9}$$

where $\Sigma_1(\boldsymbol{\theta}_0) = \sigma^2\ddot{L}(\boldsymbol{\theta}_0)$. By the so-called Cramér-Wold device, to prove (A.9), it suffices to prove

$$\frac{1}{2\sqrt{N(n)}}\boldsymbol{\alpha}^\tau\dot{L}_n(\boldsymbol{\theta}_0) \xrightarrow{d} \mathsf{N}\left(\mathbf{0}, \boldsymbol{\alpha}^\tau\Sigma_1(\boldsymbol{\theta}_0)\boldsymbol{\alpha}\right), \tag{A.10}$$

where $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_d)^\tau$. Letting

$$Z_{nt} = \frac{1}{\sqrt{N(n)}}\boldsymbol{\alpha}^\tau\dot{g}(X_t, \boldsymbol{\theta}_0)e_t, \quad \mathcal{F}_{nt} = \sigma(X_{s_1}, e_{s_1}, 1 \le s_1 \le n, 1 \le s_2 \le t)$$

in Lemma B.1, it is easy to check that

$$\begin{aligned} & \sum_{t=1}^{n}\mathsf{E}\left(Z_{nt}^2 I(|Z_{nt}| > \epsilon)|\mathcal{F}_{n,t-1}\right) \qquad\qquad (A.11) \\ \le\ & \frac{1}{N(n)}\sum_{t=1}^{n}\big[\boldsymbol{\alpha}^\tau\dot{g}(X_t, \boldsymbol{\theta}_0)\big]^2\mathsf{E}\Big(e_t^2 I\{e_t^2 > \epsilon N(n)\big[\boldsymbol{\alpha}^\tau\dot{g}(X_t, \boldsymbol{\theta}_0)\big]^2\}\Big) \\ =\ & o_P\Big(\frac{1}{N(n)}\sum_{t=1}^{n}\big[\boldsymbol{\alpha}^\tau\dot{g}(X_t, \boldsymbol{\theta}_0)\big]^2\Big) = o_P(1), \end{aligned}$$

as $(\boldsymbol{\alpha}^\tau\dot{g}(x, \boldsymbol{\theta}))^2$ is π_s-integrable.

On the other hand, we have

$$\sum_{t=1}^{n} \mathsf{E}\left(Z_{nt}^2 | \mathcal{F}_{n,t-1}\right) = \frac{\sigma^2}{N(n)} \sum_{t=1}^{n} \left[\boldsymbol{\alpha}^\tau \dot{g}(X_t, \boldsymbol{\theta}_0)\right]^2 = \boldsymbol{\alpha}^\tau \Sigma_1(\boldsymbol{\theta}_0)\boldsymbol{\alpha} + o_P(1). \tag{A.12}$$

By (A.11), (A.12) and Lemma B.1, we can prove (A.10). Hence, the proof of Theorem 3.1 is completed.

Proof of Corollary 3.1. (3.7) can be proved by using (2.3) and (3.5), and following the proof of Lemma A.2 in Gao *et al* (2011a).

By the definition of Mittag-Leffler distribution, there exist two positive constants $0 < C_1 < C_2 < \infty$ such that

$$\mathsf{P}\left(C_1 < M_\beta(1) \le C_2\right) \ge 1 - \frac{\delta}{2} \tag{A.13}$$

for any small $\delta > 0$.

Let $F_n(x) = \mathsf{P}\left\{\frac{N(n)}{n^\beta L_s(n)} \le x\right\}$ and $F(x) = \mathsf{P}\left\{M_\beta(1) \le x\right\}$. Then, equation (2.3) in Section 2.1 implies that, for n large enough, we have

$$F_n(C_2) - F(C_2) \ge -\frac{\delta}{4}, \tag{A.14}$$

$$F_n(C_1) - F(C_1) \le \frac{\delta}{4}. \tag{A.15}$$

Thus, (A.13)–(A.15) imply for large enough n

$$\mathsf{P}\left(\mathbb{J}_n(\beta)\right) \ge 1 - \delta, \tag{A.16}$$

where $\mathbb{J}_n(\beta) = \left\{C_1 n^\beta L_s(n) < N(n) \le C_2 n^\beta L_s(n)\right\}$. Let M_* be a positive constant and $\mathbb{J}_n^c(\beta)$ be the complement of $\mathbb{J}_n(\beta)$. Observe that

$$\begin{aligned} & \mathsf{P}\Big(\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \ge M_*\sqrt{n^\beta L_s(n)}\Big) \\ = & \mathsf{P}\Big(\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \ge M_*\sqrt{n^\beta L_s(n)}, \mathbb{J}_n(\beta)\Big) \\ & + \mathsf{P}\Big(\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \ge M_*\sqrt{n^\beta L_s(n)}, \mathbb{J}_n^c(\beta)\Big) \\ \le & \mathsf{P}\Big(\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \ge M_*\sqrt{n^\beta L_s(n)}, \mathbb{J}_n(\beta)\Big) + \mathsf{P}\big(\mathbb{J}_n^c(\beta)\big). \end{aligned} \tag{A.17}$$

By letting M_* sufficiently large and using (3.5) in Theorem 3.1, we have

$$\mathsf{P}\Big(\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \ge M_*\sqrt{n^\beta L_s(n)}, \mathbb{J}_n(\beta)\Big) \to 0 \tag{A.18}$$

as $n \to \infty$. On the other hand, by (A.16) and letting $\delta \to 0$, we also have

$$\mathsf{P}\big(\mathbb{J}_n^c(\beta)\big) \to 0. \tag{A.19}$$

The proof of (3.7) is completed in view of (A.17)–(A.19).

Proof of Theorem 3.2. We first give the proof of (a). Similarly to the proof of Theorem 3.1 (a), it suffices to prove that for $\epsilon > 0$,

$$\lim_{n\to\infty} \mathsf{P}\Big(\inf_{\boldsymbol{\theta}\in S_{\boldsymbol{\theta}_0}(\epsilon)} Q_{n,g}(\boldsymbol{\theta}) \geq Q_{n,g}(\boldsymbol{\theta}_0)\Big) = 1. \tag{A.20}$$

By a second order Taylor expansion of $Q_n(\boldsymbol{\eta})$, we have

$$Q_{n,g}(\boldsymbol{\theta}) = Q_{n,g}(\boldsymbol{\theta}_0) + \dot{Q}_{n,g}^{\tau}(\boldsymbol{\theta}_0)(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^{\tau}\ddot{Q}_{n,g}(\boldsymbol{\theta}_*)(\boldsymbol{\theta}-\boldsymbol{\theta}_0), \tag{A.21}$$

where

$$\begin{aligned}
\dot{Q}_{n,g}(\boldsymbol{\theta}) &= -2\sum_{t=1}^{n} \dot{g}(X_t,\boldsymbol{\theta})(Y_t - g(X_t,\boldsymbol{\theta}))I\left(|X_t| \leq M_n\right), \\
\ddot{Q}_{n,g}(\boldsymbol{\theta}) &= 2\sum_{t=1}^{n} \left[\dot{g}(X_t,\boldsymbol{\theta})\dot{g}^{\tau}(X_t,\boldsymbol{\theta}) - \ddot{g}(X_t,\boldsymbol{\theta})(Y_t - g(X_t,\boldsymbol{\theta}))\right] I\left(|X_t| \leq M_n\right).
\end{aligned}$$

Note that (B.7) in Lemma B.3 still holds for the d-dimensional vector (or $d \times d$ matrix) function case when Assumption 3.3 (ii) in Section 3.2 is satisfied. Hence, we have

$$\dot{Q}_{n,g}(\boldsymbol{\theta}) = O_P(\mathbf{J}_g(n,\boldsymbol{\theta})\sqrt{N(n)}) = o_P(\mathbf{J}_g(n,\boldsymbol{\theta})N(n)). \tag{A.22}$$

Similarly, by (B.6) in Lemma B.3, we have

$$\frac{1}{N(n)}\sum_{t=1}^{n} \dot{g}(X_t,\boldsymbol{\theta}_*)\dot{g}^{\tau}(X_t,\boldsymbol{\theta}_*)I\left(|X_t| \leq M_n\right) = \mathbf{J}_g(n,\boldsymbol{\theta}_*)(1+o_P(1)),$$

which implies that

$$N^{-1}(n)\mathbf{J}_g^{+}(n,\boldsymbol{\theta}_*)\sum_{t=1}^{n} \dot{g}(X_t,\boldsymbol{\theta}_*)\dot{g}^{\tau}(X_t,\boldsymbol{\theta}_*)I\left(|X_t| \leq M_n\right) = I_d + o_P(1). \tag{A.23}$$

On the other hand, by Definition 2.4, Assumption 3.3 (iii) and (B.7) in Lemma B.3, we also have

$$\begin{aligned}
&\Big\| \sum_{t=1}^{n} \ddot{g}(X_t,\boldsymbol{\theta}_*)\big[Y_t - g(X_t,\boldsymbol{\theta}_*)\big] I\left(|X_t| \leq M_n\right) \Big\| \\
= \;& O_P(\mathbf{J}_g(n,\boldsymbol{\theta}_*)\sqrt{N(n)}) + O_P(\epsilon\mathbf{J}_g(n,\boldsymbol{\theta}_*)N(n)).
\end{aligned} \tag{A.24}$$

By (A.21)–(A.24) and following the proof of Theorem 3.1 (a), we can prove (A.20).

By using Lemmas B.1 and B.3, the proof of Theorem 3.2 (b) is similar to that of Theorem 3.1 (b). We omit the details here.

Proof of Corollary 3.2. By using Theorem 3.2 (b) and following the proof of Corollary 3.1, we can directly prove (3.12). Hence, details are omitted here.

Proof of Corollary 3.3. (3.13) follows from (3.12) in Corollary 3.2 and (3.14) can be proved by using (3.11) in Theorem 3.2 (b).

Proof of Theorem 5.1. By the definition of X_t in (5.1) and the standard argument, we have

$$\begin{aligned} X_t &= \sum_{s=1}^{t} x_s = \sum_{s=1}^{t}(\sum_{j=0}^{\infty}\phi_j\varepsilon_{s-j}) = \sum_{s=1}^{t}(\sum_{j=0}^{s-1}\phi_j\varepsilon_{s-j}) + \sum_{s=1}^{t}(\sum_{j=s}^{\infty}\phi_j\varepsilon_{s-j}) \\ &= \sum_{s=1}^{t}(\sum_{j=0}^{t-s}\phi_j)\varepsilon_s + \sum_{s=0}^{\infty}(\sum_{j=s+1}^{t+s+1}\phi_j)\varepsilon_{-s} \\ &= \sum_{s=1}^{t}(\sum_{j=0}^{\infty}\phi_j)\varepsilon_s - \sum_{s=1}^{t}(\sum_{j=t-s+1}^{\infty}\phi_j)\varepsilon_s + \sum_{s=0}^{\infty}(\sum_{j=s+1}^{t+s+1}\phi_j)\varepsilon_{-s}. \end{aligned} \tag{A.25}$$

Letting $m(t) = [t^{\frac{1}{2+\delta}}]$ where δ is defined in (5.2), we have for sufficiently large t,

$$\begin{aligned} \sum_{s=1}^{t}(\sum_{j=t-s+1}^{\infty}\phi_j)\varepsilon_s &= \sum_{s=t-m(t)+1}^{t}(\sum_{j=t-s+1}^{\infty}\phi_j)\varepsilon_s + \sum_{s=1}^{t-m(t)}(\sum_{j=t-s+1}^{\infty}\phi_j)\varepsilon_s \\ &= O_P(\sqrt{m(t)}) + O_P(\sqrt{t}m^{-\frac{1+\delta}{2}}(t)) \\ &= O_P\left(t^{\frac{1}{2(2+\delta)}}\right). \end{aligned} \tag{A.26}$$

by using Assumption 5.1 (i) and (ii).

On the other hand, for sufficiently large t, we also have

$$\begin{aligned} \sum_{s=0}^{\infty}(\sum_{j=s+1}^{t+s+1}\phi_j)\varepsilon_{-s} &= \sum_{s=0}^{m(t)}(\sum_{j=s+1}^{t+s+1}\phi_j)\varepsilon_{-s} + \sum_{s=m(t)}^{\infty}(\sum_{j=s+1}^{t+s+1}\phi_j)\varepsilon_{-s} \\ &= O_P(\sqrt{m(t)}) + O_P(\sqrt{\sum_{s=m(t)}^{\infty}(\sum_{j=s+1}^{t+s+1}\phi_j)^2}) \\ &= O_P(\sqrt{m(t)}) = O_P\left(t^{\frac{1}{2(2+\delta)}}\right) \end{aligned} \tag{A.27}$$

by using Assumption 5.1 (i) and (ii) again.

By (A.25)–(A.27), we have

$$X_t = \phi \sum_{s=1}^{t} \varepsilon_s + O_P\left(t^{\frac{1}{2(2+\delta)}}\right) = \phi \sum_{s=1}^{t} \varepsilon_s + O_P\left(n^{\frac{1}{2(2+\delta)}}\right) \tag{A.28}$$

for $t \leq n$, where ϕ is defined in (5.2).

Let $V_g(x, \boldsymbol{\theta})$ be $\dot{g}(x, \boldsymbol{\theta})$ or $\ddot{g}(x, \boldsymbol{\theta})$. By Assumption 5.2 (ii) and (A.28),

$$V_g(X_t, \boldsymbol{\theta}) = V_g(\phi \sum_{s=1}^{t} \varepsilon_s, \boldsymbol{\theta}) + \Psi_t, \tag{A.29}$$

where

$$\begin{aligned} \|\Psi_t\| &= \left\| V_g(X_t, \boldsymbol{\theta}) - V_g(\phi \sum_{s=1}^{t} \varepsilon_s, \boldsymbol{\theta}) \right\| \\ &\leq \left\| \mathcal{B}_V(\phi \sum_{s=1}^{t} \varepsilon_s, \boldsymbol{\theta}) n^{\frac{1}{2(2+\delta)}} \right\|, \end{aligned} \tag{A.30}$$

$\mathcal{B}_V(x, \boldsymbol{\theta}) = \mathcal{B}_{\dot{g}}(x, \boldsymbol{\theta})$ if $V_g(x, \boldsymbol{\theta}) = \dot{g}(x, \boldsymbol{\theta})$, and $\mathcal{B}_V(x, \boldsymbol{\theta}) = \mathcal{B}_{\ddot{g}}(x, \boldsymbol{\theta})$ if $V_g(x, \boldsymbol{\theta}) = \ddot{g}(x, \boldsymbol{\theta})$.

Recall that $N_\varepsilon(n)$ is the regeneration number of the random walk process $\{X_t^*\}$, where $X_t^* = \phi \sum_{s=1}^{t} \varepsilon_s$, and the invariant measure for $\{X_t^*\}$ is the Lebesgue measure. Let $\rho_{nt} = I(|X_t| \leq M_n)$ and $\rho_{nt}^* = I(|X_t^*| \leq M_n)$. By (B.6) in Lemma B.3 and Assumption 5.2 (i), and noting that $n^{-\frac{1}{2(2+\delta)}} M_n \to \infty$ assumed in Theorem 5.1, we have,

$$\begin{aligned} & \frac{1}{N_\varepsilon(n)} \sum_{t=1}^{n} \dot{g}(X_t^*, \boldsymbol{\theta}_0) \dot{g}^\tau(X_t^*, \boldsymbol{\theta}_0) \rho_{nt} \\ = & \frac{1 + o_P(1)}{N_\varepsilon(n)} \sum_{t=1}^{n} \dot{g}(X_t^*, \boldsymbol{\theta}_0) \dot{g}^\tau(X_t^*, \boldsymbol{\theta}_0) \rho_{nt}^* \\ = & \mathbf{J}_{g*}(n, \boldsymbol{\theta}_0)(1 + o_P(1)), \end{aligned}$$

which leads to

$$\frac{1}{N_\varepsilon(n)} \mathbf{J}_{g*}^{+}(n, \boldsymbol{\theta}_0) \sum_{t=1}^{n} \dot{g}(X_t^*, \boldsymbol{\theta}_0) \dot{g}^\tau(X_t^*, \boldsymbol{\theta}_0) \rho_{nt} = I_d. \tag{A.31}$$

Note that for $|X_t^*| \leq M_n$,

$$\left\| \dot{g}\left(\frac{X_t^*}{M_n}, \boldsymbol{\theta}_0\right) \mathcal{B}_{\dot{g}}\left(\frac{X_t^*}{M_n}, \boldsymbol{\theta}_0\right) \right\| = O_P(1), \tag{A.32}$$

and

$$\left\| \mathbf{J}_{g*}^{+}(n, \boldsymbol{\theta}_0) \xi_{n0} \right\| = O(\dot{\kappa}_g^{-2}(M_n, \boldsymbol{\theta}_0)), \tag{A.33}$$

where $\xi_{n0} = \int_{-M_n}^{M_n} \pi_s(dx)$. Then, by (5.5) in Assumption 5.2, (A.30)–(A.33), we have

$$\begin{aligned}
& \Big\| \frac{1}{N_\varepsilon(n)} \mathbf{J}_{g*}^{+}(n, \boldsymbol{\theta}_0) \sum_{t=1}^{n} \dot{g}(X_t^*, \boldsymbol{\theta}_0) \big[\dot{g}(X_t, \boldsymbol{\theta}_0) - \dot{g}(\phi \sum_{s=1}^{t} \varepsilon_s, \boldsymbol{\theta}_0)\big] \rho_{nt} \Big\| \\
= \;& O_P\Big(n^{\frac{1}{2(2+\delta)}} \kappa_{\mathcal{B}_1}(M_n, \boldsymbol{\theta}_0) \dot{\kappa}_g(M_n, \boldsymbol{\theta}_0) \left\| \mathbf{J}_{g*}^{+}(n, \boldsymbol{\theta}_0) \right\| \\
& \cdot \frac{1}{N_\varepsilon(n)} \sum_{t=1}^{n} \Big\| \dot{g}(\frac{X_t^*}{M_n}, \boldsymbol{\theta}_0) \mathcal{B}_{\dot{g}}(\frac{X_t^*}{M_n}, \boldsymbol{\theta}_0) \Big\| \rho_{nt}^* \Big) \\
= \;& O_P\Big(n^{\frac{1}{2(2+\delta)}} \kappa_{\mathcal{B}_1}(M_n, \boldsymbol{\theta}_0) \dot{\kappa}_g(M_n, \boldsymbol{\theta}_0) \big\| \mathbf{J}_{g*}^{+}(n, \boldsymbol{\theta}_0) \xi_{n0} \big\| \Big) \\
= \;& O_P\Big(\frac{n^{\frac{1}{2(2+\delta)}} \kappa_{\mathcal{B}_1}(M_n, \boldsymbol{\theta}_0)}{\dot{\kappa}_g(M_n, \boldsymbol{\theta}_0)} \Big) = o_P(1),
\end{aligned}$$

and similarly,

$$\begin{aligned}
& \frac{1}{N_\varepsilon(n)} \mathbf{J}_{g*}^{+}(n, \boldsymbol{\theta}_0) \sum_{t=1}^{n} \big[\dot{g}(X_t, \boldsymbol{\theta}_0) - \dot{g}(\phi \sum_{s=1}^{t} \varepsilon_s, \boldsymbol{\theta}_0)\big] \\
& \cdot \big[\dot{g}(X_t, \boldsymbol{\theta}_0) - \dot{g}(\phi \sum_{s=1}^{t} \varepsilon_s, \boldsymbol{\theta}_0)\big]^{\tau} \rho_{nt} = O_P\Big(\frac{n^{\frac{1}{2+\delta}} \kappa_{\mathcal{B}_1}^2(M_n, \boldsymbol{\theta}_0)}{\dot{\kappa}_g^2(M_n, \boldsymbol{\theta}_0)} \Big) = o_P(1). \qquad \text{(A.34)}
\end{aligned}$$

By (A.29)–(A.34), we have

$$\frac{1}{N_\varepsilon(n)} \mathbf{J}_{g*}^{+}(n, \boldsymbol{\theta}_0) \sum_{t=1}^{n} \dot{g}(X_t, \boldsymbol{\theta}_0) \dot{g}^{\tau}(X_t, \boldsymbol{\theta}_0) \rho_{nt} = I_d. \qquad \text{(A.35)}$$

Similarly to the proofs of (A.25) and (A.26), we can also show that the leading term of $\sum_{t=1}^{n} \dot{g}(X_t, \boldsymbol{\theta}_0) e_t \rho_{nt}$ is $\sum_{t=1}^{n} \dot{g}(X_t^*, \boldsymbol{\theta}_0) e_t \rho_{nt}$. Then, following the proof of Theorem 3.2 by using (A.27) and the martingale central limit theorem in Lemma B.1, we have completed the proof of Theorem 5.1.

B Some technical lemmas

In this appendix, we give some useful lemmas, which have been used in the proofs of the main results in Appendix A. For the establishment of the asymptotic normal distribution theory, we first give a central limit theorem for a martingale difference array, which can be found in Corollary 3.1 of Hall and Heyde (1980).

Lemma B.1. *Let* $\{S_{ni},\ \mathcal{F}_{ni}, 1 \le i \le k_n, n \ge 1\}$ *be a zero-mean, square integrable martingale array with difference* $Z_{ni} := S_{ni} - S_{n,i-1}$, *and let* η^2 *be an a.s. finite random* *variable. Suppose that* $\mathcal{F}_{n-1,i} \subseteq \mathcal{F}_{n,i}$ *for all* $1 \le i \le k_n$ *and for all* $\epsilon > 0$

$$\sum_{i=1}^{k_n} \mathsf{E}\left(Z_{ni}^2 I(|Z_{ni}| > \epsilon)|\mathcal{F}_{n,i-1}\right) \xrightarrow{P} 0, \tag{B.1}$$

and

$$\sum_{i=1}^{k_n} \mathsf{E}\left(Z_{ni}^2|\mathcal{F}_{n,i-1}\right) \xrightarrow{P} \eta^2. \tag{B.2}$$

Then, we have $S_{nk_n} = \sum_{i=1}^{k_n} Z_{ni}$ *converges in distribution to a random variable with characteristic function* $\mathsf{E}\left(\exp\{-\frac{1}{2}\eta^2 t^2\}\right)$.

The next lemma gives a rate of convergence for the sum of the integrable function of the Harris recurrent Markov processes, which is crucial for the proof of Theorem 3.1.

Lemma B.2. *Let* $h_I(x, \boldsymbol{\theta})$ *be the integrable function on* Θ *defined as in Definition 2.3. Assume that* $\{X_t\}$ *is a Harris recurrent Markov chain and* $\pi_s(\cdot)$ *is the invariance measure. Then, uniformly for* $\boldsymbol{\theta} \in \Theta$*, we have*

$$\frac{1}{N(n)} \sum_{t=1}^{n} h_I(X_t, \boldsymbol{\theta}) = \int h_I(x, \boldsymbol{\theta}) \pi_s(dx) + o_P(1). \tag{B.3}$$

Let $\{e_t\}$ *be a sequence satisfying* $\mathsf{E}(e_t|\mathcal{F}_{t-1}) = 0$ *and* $\sigma^2 := \mathsf{E}\left(e_t^2|\mathcal{F}_{t-1}\right) < \infty$ *a.s., where* $\mathcal{F}_t = \sigma(X_{s_1}, e_{s_2}, 1 \le s_1 \le t+1, 1 \le s_2 \le t)$*. Then, we have*

$$\sum_{t=1}^{n} h_I(X_t, \boldsymbol{\theta}) e_t = O_P\left(\sqrt{N(n)}\right) \tag{B.4}$$

uniformly for $\boldsymbol{\theta} \in \Theta$.

Proof. We first prove (B.3). By Definition 2.3, there exists a $\boldsymbol{\theta}_* \in \Theta$ such that $\int h_I(x, \boldsymbol{\theta}_*) \pi_s(dx) < \infty$. By the compactness of Θ and Definition 2.3, we can show that $\int h_I(x, \boldsymbol{\theta}) \pi_s(dx) < \infty$ uniformly for $\boldsymbol{\theta} \in \Theta$. Such π_s-integrability condition is indispensable for the application of the ergodic theorem for the Harris recurrent Markov chain (see, for example, Lemma 3.2 in Karlsen and Tjøstheim 2001).

As Θ is assumed to be a compact set, we can cover Θ by a finite number of subsets $\{\Theta_i\}$ which are centered at $\boldsymbol{\theta}_i$ with radius $\epsilon > 0$. The number of these subsets, $T(\epsilon)$, is bounded by a positive constant which only depends on ϵ. Observe that

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta}\Big|\frac{1}{N(n)}\sum_{t=1}^{n}h_I(X_t,\boldsymbol{\theta})-\int h_I(x,\boldsymbol{\theta})\pi_s(dx)\Big| \\
\leq\ & \max_{1\leq i\leq T(\epsilon)}\Big|\frac{1}{N(n)}\sum_{t=1}^{n}h_I(X_t,\boldsymbol{\theta}_i)-\int h_I(x,\boldsymbol{\theta}_i)\pi_s(dx)\Big| \\
&+\max_{1\leq i\leq T(\epsilon)}\sup_{\boldsymbol{\theta}\in\Theta_i}\Big|\frac{1}{N(n)}\sum_{t=1}^{n}\left(h_I(X_t,\boldsymbol{\theta})-h_I(X_t,\boldsymbol{\theta}_i)\right)\Big| \\
&+\max_{1\leq i\leq T(\epsilon)}\sup_{\boldsymbol{\theta}\in\Theta_i}\Big|\int h_I(x,\boldsymbol{\theta})\pi_s(dx)-\int h_I(x,\boldsymbol{\theta}_i)\pi_s(dx)\Big| \\
=:\ & \Pi_{n1}+\Pi_{n2}+\Pi_{n3}. \qquad\qquad (B.5)
\end{aligned}$$

Noting the fact that $\int h_I(x,\boldsymbol{\theta})\pi_s(dx)<\infty$ uniformly for $\boldsymbol{\theta}\in\Theta$ (by Definition 2.3), we have $\Pi_{n1}=o_P(1)$ for any $\epsilon>0$ by using Lemma 3.2 in Karlsen and Tjøstheim (2001). Furthermore, by Definition 2.3 again and noting that $M(x)$ in that definition is π_s-integrable, it is easy to show that $\Pi_{n2}+\Pi_{n3}\to 0$ by letting $\epsilon\to 0$. Thus, we have completed the proof of (B.3) by letting $n\to\infty$ first and then $\epsilon\to 0$.

We next turn to the proof of (B.4). Observe that

$$\mathsf{E}\Big(\Big[\sum_{t=1}^{n}h_I(X_t,\boldsymbol{\theta})e_t\Big]^2|\mathcal{F}_{t-1}\Big)=\sum_{t=1}^{n}\mathsf{E}\left(h_I^2(X_t,\boldsymbol{\theta})e_t^2|\mathcal{F}_{t-1}\right)=O_P\Big(\sum_{t=1}^{n}h_I^2(X_t,\boldsymbol{\theta})\Big).$$

Note that $h_I^2(x,\boldsymbol{\theta})$ is also integrable on Θ by following the proof of Lemma A6 (b) in PP. Then, by using (B.3), we have proved (B.4).

Lemma B.3 below gives the corresponding convergence results for asymptotically homogeneous function $h_{AH}(\cdot,\cdot):\ \mathbb{R}^{d+1}\to\mathbb{R}$.

LEMMA B.3. *Let $h_{AH}(x,\boldsymbol{\theta})$ be the asymptotically homogeneous function on Θ with asymptotic order $\kappa(\cdot,\cdot)$ and limit homogeneous function $\overline{h}_{AH}(\cdot,\cdot)$ as in Definition 2.4. Assume that $\{X_t\}$ is a Harris recurrent Markov chain, Assumption 3.3 (iv) is satisfied, and $\overline{h}_{AH}(\cdot,\boldsymbol{\theta})$ is continuous on the interval $[-1,1]$ for all $\boldsymbol{\theta}\in\Theta$. Then, uniformly for $\boldsymbol{\theta}\in\Theta$, we have*

$$\frac{1}{N(n)J_1(n,\boldsymbol{\theta})}\sum_{t=1}^{n}h_{AH}(X_t,\boldsymbol{\theta})I\left(|X_t|\leq M_n\right)=1+o_P(1), \qquad (B.6)$$

where

$$J_1(n,\boldsymbol{\theta})=\kappa(M_n,\boldsymbol{\theta})\Big[\sum_{i=0}^{[M_n]}\overline{h}_{AH}(\frac{i}{M_n},\boldsymbol{\theta})\pi_s(B_{i+1}(1))+\sum_{i=0}^{[M_n]}\overline{h}_{AH}(\frac{-i}{M_n},\boldsymbol{\theta})\pi_s(B_{i+1}(2))\Big],$$

$B_i(1)$ and $B_i(2)$ are defined in Section 3.2. Furthermore, letting $\{e_t\}$ be defined as in Lemma B.2, then,

$$\frac{1}{\sqrt{J_2(n,\boldsymbol{\theta})}}\sum_{t=1}^{n} h_{AH}(X_t,\boldsymbol{\theta}) I\left(|X_t|\leq M_n\right) e_t = O_P\big(\sqrt{N(n)}\big) \tag{B.7}$$

uniformly for $\boldsymbol{\theta}\in\Theta$, where

$$J_2(n,\boldsymbol{\theta}) = \kappa^2(M_n,\boldsymbol{\theta})\Big[\sum_{i=0}^{[M_n]}\overline{h}_{AH}^2(\frac{i}{M_n},\boldsymbol{\theta})\pi_s(B_{i+1}(1)) + \sum_{i=0}^{[M_n]}\overline{h}_{AH}^2(\frac{-i}{M_n},\boldsymbol{\theta})\pi_s(B_{i+1}(2))\Big].$$

Proof. We first give the proof of (B.6). By Definition 2.4,

$$h_{AH}(\lambda x,\boldsymbol{\theta}) = \kappa(\lambda,\boldsymbol{\theta})\overline{h}_{AH}(x,\boldsymbol{\theta}) + h_R(x,\lambda,\boldsymbol{\theta}), \tag{B.8}$$

where $h_R(x,\lambda,\boldsymbol{\theta})$ is of order smaller than $\kappa(\lambda,\boldsymbol{\theta})$ as $\lambda\to\infty$ for all $\boldsymbol{\theta}\in\Theta$.

Recall that

$$B_i(1) = [i-1, i)\,,\quad i=1,2,\cdots,[M_n],\quad B_{[M_n]+1}(1) = [[M_n], M_n]\,,$$
$$B_i(2) = [-i, -i+1)\,,\quad i=1,2,\cdots,[M_n],\quad B_{[M_n]+1}(2) = [-M_n, -[M_n]]\,.$$

Then, we have

$$\sum_{t=1}^{n} h_{AH}(X_t,\boldsymbol{\theta}) I\left(|X_t|\leq M_n\right) = \sum_{k=1}^{2}\sum_{i=1}^{[M_n]+1}\sum_{t=1}^{n} h_{AH}(X_t,\boldsymbol{\theta}) I\left(X_t\in B_i(k)\right). \tag{B.9}$$

We first consider the case of $k=1$. For $1\leq i\leq [M_n]+1$, by (B.8), we have

$$\begin{aligned}\sum_{t=1}^{n} h_{AH}(X_t,\boldsymbol{\theta}) I\left(X_t\in B_i(1)\right) &= \kappa(M_n,\boldsymbol{\theta})\sum_{t=1}^{n}\overline{h}_{AH}\big(\frac{X_t}{M_n},\boldsymbol{\theta}\big) I\left(X_t\in B_i(1)\right)\\ &\quad + \sum_{t=1}^{n} h_R\big(\frac{X_t}{M_n}, M_n,\boldsymbol{\theta}\big) I\left(X_t\in B_i(1)\right).\end{aligned}$$

As $\mathcal{N}_x(1)$ is a small set for each $x\in[-M_n,\ M_n]$ and $\overline{h}_{AH}(\cdot,\cdot)$ is continuous on $[-1,\ 1]$, we can prove that

$$\frac{1}{N(n)}\sum_{t=1}^{n}\overline{h}_{AH}\left(\frac{X_t}{M_n},\boldsymbol{\theta}\right) I\left(X_t\in B_i(1)\right) = \overline{h}_{AH}(\frac{i-1}{M_n},\boldsymbol{\theta})\pi_s(B_i(1)) + o_P(1), \tag{B.10}$$

uniformly for $\boldsymbol{\theta}\in\Theta$ and $1\leq i\leq [M_n]+1$. The detailed proof of (B.10) will be provided later in this appendix.

By (B.10), we have, uniformly for $\boldsymbol{\theta} \in \Theta$,

$$\begin{aligned} & \frac{\kappa(M_n, \boldsymbol{\theta})}{N(n) J_1(n, \boldsymbol{\theta})} \sum_{i=1}^{[M_n]+1} \sum_{t=1}^{n} \overline{h}_{AH}\big(\frac{X_t}{M_n}, \boldsymbol{\theta}\big) I\left(X_t \in B_i(1)\right) \qquad \text{(B.11)} \\ = & \frac{\kappa(M_n, \boldsymbol{\theta})}{J_1(n, \boldsymbol{\theta})} \sum_{i=1}^{[M_n]+1} \frac{1}{N(n)} \sum_{t=1}^{n} \overline{h}_{AH}\big(\frac{X_t}{M_n}, \boldsymbol{\theta}\big) I\left(X_t \in B_i(1)\right) \\ = & \frac{\kappa(M_n, \boldsymbol{\theta})}{J_1(n, \boldsymbol{\theta})} \sum_{i=1}^{[M_n]+1} \overline{h}_{AH}\big(\frac{i-1}{M_n}, \boldsymbol{\theta}\big) \pi_s(B_i(1)) + o_P(1) \\ = & \frac{\kappa(M_n, \boldsymbol{\theta})}{J_1(n, \boldsymbol{\theta})} \sum_{i=0}^{[M_n]} \overline{h}_{AH}\big(\frac{i}{M_n}, \boldsymbol{\theta}\big) \pi_s(B_{i+1}(1)) + o_P(1). \end{aligned}$$

Analogously, for $k = 2$, we also have

$$\begin{aligned} & \frac{\kappa(M_n, \boldsymbol{\theta})}{N(n) J_1(n, \boldsymbol{\theta})} \sum_{i=1}^{[M_n]+1} \sum_{t=1}^{n} \overline{h}_{AH}\big(\frac{X_t}{M_n}, \boldsymbol{\theta}\big) I\left(X_t \in B_i(2)\right) \qquad \text{(B.12)} \\ = & \frac{\kappa(M_n, \boldsymbol{\theta})}{J_1(n, \boldsymbol{\theta})} \sum_{i=0}^{[M_n]} \overline{h}_{AH}\big(\frac{i}{M_n}, \boldsymbol{\theta}\big) \pi_s(B_{i+1}(2)) + o_P(1), \end{aligned}$$

uniformly for $\boldsymbol{\theta} \in \Theta$.

For the remaining term $h_R(\cdot, \cdot, \cdot)$, by the fact that it is of order smaller than $\kappa(\cdot, \cdot)$, (2.6) and (2.7), we have either

$$h_R\big(\frac{X_t}{M_n}, M_n, \boldsymbol{\theta}\big) = a(M_n, \boldsymbol{\theta}) \mathcal{A}\big(\frac{X_t}{M_n}, \boldsymbol{\theta}\big) \qquad \text{(B.13)}$$

or

$$h_R\big(\frac{X_t}{M_n}, M_n, \boldsymbol{\theta}\big) = b(M_n, \boldsymbol{\theta}) \mathcal{A}\big(\frac{X_t}{M_n}, \boldsymbol{\theta}\big) \mathcal{B}(X_t, \boldsymbol{\theta}), \qquad \text{(B.14)}$$

where $a(n, \boldsymbol{\theta}) = o(\kappa(n, \boldsymbol{\theta}))$, $b(n, \boldsymbol{\theta}) = O(\kappa(n, \boldsymbol{\theta}))$ as $n \to \infty$, $\sup_{\boldsymbol{\theta} \in \Theta} \mathcal{A}(\cdot, \boldsymbol{\theta})$ is locally bounded, and $\sup_{\boldsymbol{\theta} \in \Theta} \mathcal{B}(\cdot, \boldsymbol{\theta})$ is bounded and vanishes at infinity. When $h_R\big(\frac{X_t}{M_n}, M_n, \boldsymbol{\theta}\big)$ satisfies the structure defined in (B.13), it is easy to show that

$$\begin{aligned} & \frac{1}{N(n) J_1(n, \boldsymbol{\theta})} \sum_{k=1}^{2} \sum_{i=1}^{[M_n]+1} \sum_{t=1}^{n} h_R\big(\frac{X_t}{M_n}, M_n, \boldsymbol{\theta}\big) I\left(X_t \in B_i(k)\right) \qquad \text{(B.15)} \\ = & \frac{a(M_n, \boldsymbol{\theta})}{N(n) J_1(n, \boldsymbol{\theta})} \sum_{k=1}^{2} \sum_{i=1}^{[M_n]+1} \sum_{t=1}^{n} \mathcal{A}\big(\frac{X_t}{M_n}, \boldsymbol{\theta}\big) I\left(X_t \in B_i(k)\right) \\ = & O_P\Big(\frac{a(M_n, \boldsymbol{\theta})}{\kappa(M_n, \boldsymbol{\theta})}\Big) = o_P(1), \end{aligned}$$

as $a(n, \boldsymbol{\theta}) = o(\kappa(n, \boldsymbol{\theta}))$ and $\mathcal{A}\big(\frac{X_t}{M_n}, \boldsymbol{\theta}\big)$ is bounded for $X_t \in B_i(k)$.

When $h_R\big(\frac{X_t}{M_n}, M_n, \boldsymbol{\theta}\big)$ satisfies the structure defined in (B.14), as $\sup_{\boldsymbol{\theta}\in\Theta} \mathcal{B}(\cdot, \boldsymbol{\theta})$ vanishes at infinity, we have

$$\mathcal{B}(X_t, \boldsymbol{\theta}) I\left(X_t \in B_i(k)\right) = o_P(1), \quad i \geq l_n,$$

where $l_n \to \infty$ but $l_n = o(M_n)$. This together with (B.14) and $b(n, \boldsymbol{\theta}) = O(\kappa(n, \boldsymbol{\theta}))$, implies that

$$\begin{aligned}
& \frac{1}{N(n) J_1(n, \boldsymbol{\theta})} \sum_{k=1}^{2} \sum_{i=1}^{[M_n]+1} \sum_{t=1}^{n} h_R\big(\frac{X_t}{M_n}, M_n, \boldsymbol{\theta}\big) I\left(X_t \in B_i(k)\right) \qquad \text{(B.16)} \\
= \; & \frac{b(M_n, \boldsymbol{\theta})}{N(n) J_1(n, \boldsymbol{\theta})} \sum_{k=1}^{2} \sum_{i=1}^{l_n-1} \sum_{t=1}^{n} \mathcal{A}\big(\frac{X_t}{M_n}, \boldsymbol{\theta}\big) \mathcal{B}\left(X_t, \boldsymbol{\theta}\right) I\left(X_t \in B_i(k)\right) \\
& + \frac{b(M_n, \boldsymbol{\theta})}{N(n) J_1(n, \boldsymbol{\theta})} \sum_{k=1}^{2} \sum_{i=l_n}^{[M_n]+1} \sum_{t=1}^{n} \mathcal{A}\big(\frac{X_t}{M_n}, \boldsymbol{\theta}\big) \mathcal{B}\left(X_t, \boldsymbol{\theta}\right) I\left(X_t \in B_i(k)\right) \\
= \; & O_P\Big(\frac{b(M_n, \boldsymbol{\theta})}{J_1(n, \boldsymbol{\theta})} \sum_{k=1}^{2} \sum_{i=1}^{l_n-1} \pi_s(B_i(k))\Big) + o_P\Big(\frac{b(M_n, \boldsymbol{\theta})}{J_1(n, \boldsymbol{\theta})} \sum_{k=1}^{2} \sum_{i=l_n}^{[M_n]+1} \pi_s(B_i(k))\Big) \\
= \; & o_P(1).
\end{aligned}$$

By the definition of $J_1(n, \boldsymbol{\theta})$, (B.9), (B.11), (B.12), (B.15) and (B.16), we can prove (B.6).

By Lemma A6 (c) in PP, $h^2_{AH}(x, \boldsymbol{\theta})$ is asymptotically homogeneous on Θ with asymptotic order $\kappa^2(\cdot, \cdot)$ and limit homogeneous function $\overline{h}^2_{AH}(\cdot, \cdot)$. Then, the proof of (B.7) is analogous to the proof of (B.4) by using (B.6).

Proof of (B.10). Let

$$Z_{nt}(i, \boldsymbol{\theta}) = \overline{h}_{AH}\big(\frac{X_t}{M_n}, \boldsymbol{\theta}\big) I\left(X_t \in B_i(1)\right)$$

and

$$U_k(i, \boldsymbol{\theta}) = \begin{cases} \sum_{t=0}^{\tau_0} Z_{nt}(i, \boldsymbol{\theta}), & k = 0, \\ \sum_{t=\tau_{k-1}+1}^{\tau_k} Z_{nt}(i, \boldsymbol{\theta}), & 1 \leq k \leq N(n), \\ \sum_{t=\tau_{N(n)}+1}^{n} Z_{nt}(i, \boldsymbol{\theta}), & k = N(n) + 1. \end{cases}$$

As mentioned in Section 2.1, $\{U_k(i, \boldsymbol{\theta}),\ k \geq 1\}$ is a sequence of i.i.d. random variables. Then, in order to prove (B.10), we need only to show that, uniformly for $\boldsymbol{\theta}$ and i,

$$\frac{1}{N(n)} \sum_{k=0}^{N(n)+1} U_k(i, \boldsymbol{\theta}) - u(i, \boldsymbol{\theta}) = o_P(1), \qquad \text{(B.17)}$$

where $u(i,\boldsymbol{\theta}) = \overline{h}_{AH}(\frac{i-1}{M_n},\boldsymbol{\theta})\pi_s(B_i(1))$.

As in the proof of Theorem 3.1 in Gao *et al* (2011a), the edge terms $U_0(i,\boldsymbol{\theta})$ and $U_{N(n)+1}(i,\boldsymbol{\theta})$ are asymptotically negligible when divided by $N(n)$. Hence, it suffices to show that

$$\frac{1}{N(n)}\sum_{k=1}^{N(n)} U_k(i,\boldsymbol{\theta}) - u(i,\boldsymbol{\theta}) = o_P(1) \tag{B.18}$$

uniformly for $\boldsymbol{\theta}$ and i. Furthermore, following the argument in the proof of Lemma B.2, we need only to prove (B.18) uniformly for i when $\boldsymbol{\theta} \in \Theta$ is fixed.

Note that for any $\epsilon > 0$,

$$\begin{aligned}
& \mathsf{P}\Big(\max_{1\le i\le M_n}\Big|\frac{1}{N(n)}\sum_{k=1}^{N(n)} U_k(i,\boldsymbol{\theta}) - u(i,\boldsymbol{\theta})\Big| > \epsilon\Big) \\
= \;& \mathsf{P}\Big(\max_{1\le i\le M_n}\Big|\frac{1}{N(n)}\sum_{k=1}^{N(n)} U_k(i,\boldsymbol{\theta}) - u(i,\boldsymbol{\theta})\Big| > \epsilon, \mathbb{I}_n(\beta)\Big) \\
& + \mathsf{P}\Big(\max_{1\le i\le M_n}\Big|\frac{1}{N(n)}\sum_{k=1}^{N(n)} U_k(i,\boldsymbol{\theta}) - u(i,\boldsymbol{\theta})\Big| > \epsilon, \mathbb{I}_n^c(\beta)\Big) \\
\le \;& \mathsf{P}\Big(\max_{1\le i\le M_n}\Big|\frac{1}{N(n)}\sum_{k=1}^{N(n)} U_k(i,\boldsymbol{\theta}) - u(i,\boldsymbol{\theta})\Big| > \epsilon, \mathbb{I}_n(\beta)\Big) + \mathsf{P}\Big(\mathbb{I}_n^c(\beta)\Big),
\end{aligned}$$

where $\mathbb{I}_n(\beta) = I\big(C_\beta n^{\beta-\epsilon_1} \le N(n) \le C_\beta n^{\beta+\epsilon_1}\big)$, C_β is a positive constant, ϵ_1 is a positive and sufficiently small number.

By Lemma 3.4 in Karlsen and Tjøstheim (2001), we have

$$\mathsf{P}\Big(\mathbb{I}_n^c(\beta)\Big) = o(1). \tag{B.19}$$

Letting

$$\begin{aligned}
&\overline{U}_k(i,\boldsymbol{\theta}) = U_k(i,\boldsymbol{\theta}) I\big(|U_k(i,\boldsymbol{\theta})| \le n^{\beta-2\epsilon_1}\big), \\
&\widehat{U}_k(i,\boldsymbol{\theta}) = U_k(i,\boldsymbol{\theta}) - \overline{U}_k(i,\boldsymbol{\theta}), \\
&\overline{u}(i,\boldsymbol{\theta}) = \mathsf{E}\big[\overline{U}_k(i,\boldsymbol{\theta})\big], \quad \widehat{u}(i,\boldsymbol{\theta}) = \mathsf{E}\big[\widehat{U}_k(i,\boldsymbol{\theta})\big],
\end{aligned}$$

we have

$$
\begin{aligned}
& \mathsf{P}\Big(\max_{1\le i\le M_n}\Big|\frac{1}{N(n)}\sum_{k=1}^{N(n)}U_k(i,\boldsymbol{\theta})-u(i,\boldsymbol{\theta})\Big|>\epsilon,\mathbb{I}_n(\beta)\Big) \\
\le\ & \mathsf{P}\Big(\max_{1\le i\le M_n}\Big|\frac{1}{N(n)}\sum_{k=1}^{N(n)}\overline{U}_k(i,\boldsymbol{\theta})-\overline{u}(i,\boldsymbol{\theta})\Big|>\epsilon/2,\mathbb{I}_n(\beta)\Big) \\
& +\mathsf{P}\Big(\max_{1\le i\le M_n}\Big|\frac{1}{N(n)}\sum_{k=1}^{N(n)}\widehat{U}_k(i,\boldsymbol{\theta})-\widehat{u}(i,\boldsymbol{\theta})\Big|>\epsilon/2,\mathbb{I}_n(\beta)\Big). \qquad \text{(B.20)}
\end{aligned}
$$

Following the proof of Lemma B.1 in Gao *et al* (2011a), we have $\mathsf{E}\big[|U_k(i,\boldsymbol{\theta})|^p\big]<\infty$ uniformly for i and $\boldsymbol{\theta}\in\Theta$. By taking $p>\frac{\beta+\epsilon_1}{\beta-2\epsilon_1}$ and the Markov inequality, we can show that

$$
\begin{aligned}
& \mathsf{P}\Big(\max_{1\le i\le M_n}\Big|\frac{1}{N(n)}\sum_{k=1}^{N(n)}\widehat{U}_k(i,\boldsymbol{\theta})-\widehat{u}(i,\boldsymbol{\theta})\Big|>\epsilon/2,\mathbb{I}_n(\beta)\Big) \\
\le\ & \mathsf{P}\Big(\max_{1\le k\le N(n)}\big|U_k(i,\boldsymbol{\theta})\big|>n^{\beta-2\epsilon_1},\mathbb{I}_n(\beta)\Big) \\
\le\ & C_\beta n^{\beta+\epsilon_1}\frac{\mathsf{E}\big[|U_k(i,\boldsymbol{\theta})|^p\big]}{n^{p(\beta-2\epsilon_1)}}=o(1). \qquad \text{(B.21)}
\end{aligned}
$$

On the other hand, by using the Bernstein inequality for i.i.d. random variables, we have

$$
\begin{aligned}
& \mathsf{P}\Big(\max_{1\le i\le M_n}\Big|\frac{1}{N(n)}\sum_{k=1}^{N(n)}\overline{U}_k(i,\boldsymbol{\theta})-\overline{u}(i,\boldsymbol{\theta})\Big|>\epsilon/2,\mathbb{I}_n(\beta)\Big) \\
\le\ & \sum_{i=1}^{M_n}\sum_{l=C_\beta n^{\beta-\epsilon_1}}^{C_\beta n^{\beta+\epsilon_1}}\mathsf{P}\Big(\Big|\frac{1}{l}\sum_{k=1}^{l}\overline{U}_k(i,\boldsymbol{\theta})-\overline{u}(i,\boldsymbol{\theta})\Big|>\epsilon/2\Big) \\
\le\ & \sum_{i=1}^{M_n}\sum_{l=C_\beta n^{\beta-\epsilon_1}}^{C_\beta n^{\beta+\epsilon_1}}\exp\Big\{-\frac{c_0 l^2\epsilon^2}{ln^{\beta-2\epsilon_1}\epsilon}\Big\} \\
=\ & O\big(M_n n^{\beta+\epsilon_1}e^{-c_1 n^{\epsilon_1}}\big)=o(1), \qquad \text{(B.22)}
\end{aligned}
$$

where c_0 and c_1 are two positive constants. By using (B.19)–(B.22), we have completed the proof of (B.18).

References

[1] Bandi, F. and Phillips, P. C. B. (2009). Nonstationary continuous–time processes. *Handbook of Financial Econometrics* (Edited by Y. Aït–Sahalia and L. P. Hansen), Vol **1**, 139–201.

[2] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

[3] Chan, N. and Wang, Q. (2012). Nonlinear cointegrating regressions with nonstationary time series. Working paper, School of Mathematics and Statistics, University of Sydney.

[4] Chen, L., Cheng, M. and Peng, L. (2009). Conditional variance estimation in heteroscedastic regression models. *Journal of Statistical Planning and Inference* **139**, 236–245.

[5] Chen, J., Gao, J. and Li, D. (2012). Estimation in semi-parametric regression with non-stationary regressors. *Bernoulli* **18**, 678–702.

[6] Chen, X. (1999). How often does a Harris recurrent Markov chain occur ? *Annals of Probability* **27**, 1324–1346.

[7] Chen, X. (2000). On the limit laws of the second order for additive functionals of Harris recurrent Markov chains. *Probability Theory and Related Fields* **116**, 89–123.

[8] Gao, J. (2007). *Nonlinear Time Series: Semi– and Non–Parametric Methods.* Chapman & Hall/CRC.

[9] Gao, J., Li, D. and Tjøstheim, D. (2011a). Uniform consistency for nonparametric estimators in null recurrent time series. Working paper available at http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011/13-11.php.

[10] Gao, J., Tjøstheim, D. and Yin, J. (2011b). Estimation in threshold autoregression models with a stationary and a unit root regime. Working paper available at http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2011/wp21-11.pdf. Forthcoming in *Journal of Econometrics.*

[11] Hall, P. and Heyde, C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, New York.

[12] Höpfner, R. and Löcherbach, E. (2003). Limit theorems for null recurrent Markov processes. *Memoirs of the American Mathematical Society*, Vol. **161**, Number **768**.

[13] Jennrich, R. I. (1969). Asymptotic properties of non–linear least squares estimation. *Annals of Mathematical Statistics* **40**, 633–643.

[14] Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Annals of Statistics* **29**, 372–416.

[15] Karlsen, H. A., Myklebust, T. and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* **35**, 252–299.

[16] Karlsen, H. A., Myklebust, T. and Tjøstheim D. (2010). Nonparametric regression estimation in a null recurrent time series. *Journal of Statistical Planning and Inference* **140**, 3619–3626.

[17] Kallianpur, G. and Robbins, H. (1954). The sequence of sums of independent random variables. *Duke Mathematical Journal* **21**, 285–307.

[18] Kasahara, Y. (1984). Limit theorems for Lévy processes and Poisson point processes and their applications to Brownian excursions. *Journal of Mathematics of Kyoto University* **24**, 521–538.

[19] Lin, Z., Li, D. and Chen, J. (2009). Local linear M-estimators in null recurrent time series. *Statistica Sinica* **19**, 1683–1703.

[20] Malinvaud, E. (1970). The consistency of nonlinear regressions. *Annals of Mathematical Statistics* **41**, 956–969.

[21] Meyn, S. and Tweedie, R. L. (2009). *Markov Chains and Stochastic Stability* (2nd Edition). Cambridge University Press.

[22] Myklebust, T., Karlsen, H. A. and Tjøstheim D. (2012). Null recurrent unit root processes. *Econometric Theory* **28**, 1–41.

[23] Nummelin, E. (1984). *General Irreducible Markov Chains and Non–negative Operators*. Cambridge University Press.

[24] Park, J. and Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* **15**, 269–298.

[25] Park, J. and Phillips, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69**, 117–162.

[26] Peng, L. and Yao, Q. (2003). Least absolute deviations estimation for ARCH and GARCH models. *Biometrika* **90**, 967–975.

[27] Schienle, M. (2011). Nonparametric nonstationary regression with many covariates. Working paper, Humboldt–Universität zu Berlin.

[28] Teräsvirta, T., Tjøstheim, D. and Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series.* Oxford University Press.

[29] Wang, Q. and Chan, N. (2011). Uniform convergence for a class of martingales with applications in non–linear cointegrating regression. Working paper, School of Mathematics and Statistics, University of Sydney.

[30] Wang, Q. and Phillips, P. C. B. (2009). Structural nonparametric cointegrating regression. *Econometrica*, **77**, 1901–1948.

[31] Wu, C. F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics* **9**, 501–513.