Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers | Cowles Foundation

12-1-2019

Nonlinear Cointegrating Power Function Regression with Endogeneity

Zhishui Hu

Peter C.B. Phillips

Qiying Wang

Follow this and additional works at: https://elischolar.library.yale.edu/cowles-discussion-paper-series

 Part of the Economics Commons

Recommended Citation

Hu, Zhishui; Phillips, Peter C.B.; and Wang, Qiying, "Nonlinear Cointegrating Power Function Regression with Endogeneity" (2019). *Cowles Foundation Discussion Papers*. 37.
https://elischolar.library.yale.edu/cowles-discussion-paper-series/37

This Discussion Paper is brought to you for free and open access by the Cowles Foundation at EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Cowles Foundation Discussion Papers by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.

NONLINEAR COINTEGRATING POWER FUNCTION REGRESSION WITH ENDOGENEITY

By

Zhishui Hu, Peter C. B. Phillips, and Qiying Wang

December 2019

COWLES FOUNDATION DISCUSSION PAPER NO. 2211



COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Nonlinear cointegrating power function regression with endogeneity*

Zhishui Hu †, Peter C. B. Phillips‡and Qiying Wang §

December 4, 2019

Abstract

This paper develops an asymptotic theory for nonlinear cointegrating power function regression. The framework extends earlier work on the deterministic trend case and allows for both endogeneity and heteroskedasticity, which makes the models and inferential methods relevant to many empirical economic and financial applications, including predictive regression. Accompanying the asymptotic theory of nonlinear regression, the paper establishes some new results on weak convergence to stochastic integrals that go beyond the usual semi-martingale structure and considerably extend existing limit theory, complementing other recent findings on stochastic integral asymptotics. The paper also provides a general framework for extremum estimation limit theory that encompasses stochastically nonstationary time series and should be of wide applicability.

JEL Classification: C13, C22.

Key words and phrases: Nonlinear power regression, Least squares estimation, Nonstationarity, Endogeneity, Heteroscedasticity.

1 Introduction

Since the initial work by Park and Phillips (2001) in this area, the past two decades have witnessed significant developments in nonlinear cointegrating regression, including parametric, nonparametric and semi-parametric specifications of such models. These

*Hu acknowledges research support from NSFC(No. 11671373), Wang acknowledges research support from Australian Research Council and Phillips acknowledges research support from the NSF under Grant No. 18-50860 and the Kelly Fund at the University of Auckland.

†International Institute of Finance, School of Management, University of Science and Technology of China, Hefei, Anhui 230026, China, email: huzs@ustc.edu.cn.

‡University of Auckland, Auckland, New Zealand; Yale University, University of Southampton, Singapore Management University, email: peter.phillips@yale.edu.

§School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia, email: qiying@maths.usyd.edu.au.

developments have provided a framework of econometric estimation and inference for a wide class of nonlinear, nonstationary relationships. Among many other contributions to this research, we may refer to Chang, Park and Phillips (2001), Chang and Park (2003), Bae and De Jong (2007), Wang and Phillips (2009a, b, 2016), Kim and Kim (2012) and Gao and Phillips (2013), together with the references cited therein.

In recent work Chan and Wang (2015) established some general results on nonlinear parametric cointegrating regression. In comparison with previous research, Chan and Wang (2015) employed a different approach to investigating asymptotics in models of this kind. Their approach directly established joint distributional convergence of the martingale of interest in conjunction with its conditional variance, rather than relying on the classical approach to the martingale limit theorem which requires convergence in probability for the conditional variance.[1] The methodology used in Chan and Wang (2015) has important advantages since it is usually difficult to prove convergence in probability without expanding the probability space, particularly in the structure of cointegrating regression settings where the conditional variance typically converges weakly to a random variable rather than in probability to a constant. The latter methodology was used in Park and Phillips (2001) and requires more restrictive conditions as well as expansion of the probability space to secure the required results.

The models considered in Chan and Wang (2015) include integrable and non-integrable regression functionals. However, as is apparent from their Assumption 3.4, power regression functions are excluded, such as those that take the form $f(x) = \beta |x|^{\gamma}$, where $\beta \in \mathbb{R}$ and $\gamma \geq 0$. This shortcoming in coverage is restrictive because power function regression is a commonly used model in many empirical applications. An area of application where such regression has been found particularly useful is in testing the validity and order of polynomial regression (Baek, Cho and Phillips, 2015; Cho and Phillips, 2018.)

One goal of the present paper is to address this omission in coverage. A further goal is to contribute to the general development of asymptotic theory in nonlinear nonstationary regression. First, while this paper focuses on power function regression, our results allow for models that include both endogeneity and heteroskedasticity. Power functions fall within the framework of homogeneous functions that were considered in Park and Phillips (2001), but their results applied to $I(1)$ integrated and weakly exogenous regressor processes and martingale difference equation errors with constant conditional

[1] Readers are referred to Wang (2014) for a recent general exposition and development of limit theory relevant to nonstationary time series regression.

variances, thereby excluding a wide class of nonstationary processes and standard error volatility models such as ARCH and GARCH. Second, accompanying the development of our asymptotic theory for power regression, we provide new results on convergence to stochastic integrals that extend beyond the semimartingale structure. Since the 1980s there has been extensive research in both econometrics and probability on weak convergence to stochastic integrals, yielding a large body of useful theory. But results that extend beyond a semimartingale framework and allow for nonlinear functionals have only recently become available, notably by Liang, et al. (2016) and Peng and Wang (2018). However, the nonlinear functionals considered in the latter papers exclude power functions such as $f(x) = \beta |x|^{\gamma}$, since the first order derivative of $f(x)$ does not everywhere exist or even satisfy a Lipschitz condition in cases such as $-1 < \gamma < 0$, where $f(x) = \beta |x|^{\gamma}$ is locally integrable, but not locally Riemann integrable.

The present paper contributes to this literature by building a framework of theory that accommodates these extensions, thereby helping to complete the limit theory for extremum estimation in nonlinear nonstationary regressions. To achieve this purpose the paper provides a weak convergence result for normalized stochastic processes, associated sample covariance functionals, and quadratic variations at a level of generality that assists in delivering asymptotics for power regression. Further, as in Phillips (2007) where deterministic power function regression was analyzed, we show how different convergence rates apply in corresponding least squares power regressions in the presence of stochastic trends.

The paper is organized as follows. Section 2 develops limit theory for least squares estimation (LSE) in a stochastic power regression model. The technical results concerning weak convergence to stochastic integrals that extend beyond semimartingale formulations are provided in Section 3. Section 4 concludes, proofs of all the main results are given in Section 5, and Appendix A provides a framework of extremum estimation for nonlinear least squares that allows for various convergence rates and asymptotic linearization of LSE with general forms of score and hessian functions that allow for many different forms of limit theory.

Throughout the paper, a function $g(x)$ is called locally integrable if, for all compact sets $K \subset \mathbb{R}$, $\int_K |g(x)|dx < \infty$ or locally Riemann integrable if $g(x)I(x \in K)$ is Riemann integrable. Integrals are generally understood to be in the Lebesgue sense, except when explicitly mentioned.

2 Nonlinear cointegrating power regression

Let $(\xi_k, u_k)_{k\geq 1}$ be a sequence of arbitrary random vectors. Consider a nonlinear cointegrating power regression model defined by

$$y_k = \beta|x_k|^\gamma + u_k, \tag{2.1}$$

where $x_k = \sum_{j=1}^k \xi_j$ and $\theta = (\beta, \gamma) \in \Theta := \mathbb{R} \times [-1/2, \infty)$. The least squares estimator (LSE) $\widehat{\theta}_n$ of θ is defined by the extremum problem

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta} \sum_{k=1}^n \left(y_k - \beta|x_k|^\gamma\right)^2.$$

To develop asymptotics for the estimator $\widehat{\theta}_n$, we denote the true parameter $\theta_0 = (\beta_0, \gamma_0)$, where $\beta_0 \neq 0$ and $\gamma_0 > -1/2$. The power parameter γ is clearly unidentified when $\beta_0 = 0$ and only weakly identified when the true regression coefficient β_0 is local to zero in the sense that $\beta_0 = o(1)$ as $n \to \infty$. The latter case fits within the weak instrument econometric literature and has been analyzed by Shi and Phillips (2012) in the nonstationary regressor case and Andrews and Cheng (2012) in the stationary case. In addition, when $\gamma_0 < -1/2$ there are further difficulties in developing asymptotics for the LSE $\hat{\theta}_n$, as discussed in Remark 2.1 below[2].

Write $x_{nk} = x_k/d_n$ where $d_n^2 = var(x_n)$ and $u_{nk} = \frac{1}{\sqrt{n}}\sum_{j=1}^k u_j$. Throughout the paper, we assume $d_n^2 \simeq n^\mu$ for some $0 < \mu < 2$. This is a minor requirement and holds for usual $I(1)$ processes and a partial sum of a long memory process. We further make use of the following conditions:

A1 (i) $\{u_k, \mathcal{F}_k\}_{k\geq 1}$ forms a martingale difference with $\sup_{k\geq 1} Eu_k^2 < \infty$, where $\{\mathcal{F}_k\}_{k\geq 1}$ is a filtration such that x_k is adapted to $\mathcal{F}_{k-1}$ for all $k \geq 1$ ($\mathcal{F}_0$ is defined to be a trivial σ-field).

(ii) There exists a 2-dimensional continuous Gaussian process (X_t, B_t) with covariance matrix $\Omega_t > 0$ so that

$$(x_{n,[nt]}, u_{n,[nt]}) \;\Rightarrow\; (X_t, B_t), \quad \text{on } D_{\mathbb{R}^2}[0,1], \tag{2.2}$$

Condition **A1** imposes a martingale structure in the model (2.1), which is extensively used in the literature of nonlinear cointegrating regression. See, for instance, Park and

[2]Similar, but less complex non-convergence, issues arise in the deterministic (evaporating) trend case with $x_k = k$ for which $\sum_{k=1}^\infty k^{2\gamma} < \infty$ when $\gamma < -1/2$ and the usual excitation condition for consistency fails.

Phillips (2001) and Chan and Wang (2015). However, unlike these existing results, only $\sup_{k\geq 1} Eu_k^2 < \infty$ rather than $\sup_{k\geq 1} E\big(u_k^2 \mid \mathcal{F}_{k-1}\big) \leq C < \infty$ is used here, which allows for heteroskedasticity in the model (2.1), thereby enhancing wider use of our results in financial econometrics. The martingale structure in model (2.1) will be extended to include endogeneity in Section 2.1 and more general models are considered in Section 2.2.

Let $F_n = diag\big[\sqrt{n}d_n^{\gamma_0}/\log d_n, \sqrt{n}d_n^{\gamma_0}\big]$ be a diagonal rate matrix. Our first result concerning the asymptotic behavior of the extremum estimator $\hat{\theta}_n$ is as follows.

Theorem 2.1. *Suppose **A1** holds. For any $\gamma_0 > 0$, we have*

$$F_n(\hat{\theta}_n - \theta_0) \;\; \to_D \;\; \begin{pmatrix} 1 \\ -1/\beta_0 \end{pmatrix} \frac{U_0 \int_0^1 |X_t|^{\gamma_0} \log|X_t| \, dB_t - U_1 \int_0^1 |X_t|^{\gamma_0} dB_t}{U_1^2 - U_0 \, U_2}, \tag{2.3}$$

where $U_i = \int_0^1 |X_t|^{2\gamma_0} \log^i |X_t| \, dt$ for $i = 0, 1$ and 2.

Since X_t has a continuous path, for $\gamma_0 \geq 0$, the existence of U_i and the limit distribution on the right hand of (2.3) follow immediately. Notably, the limit distribution in (2.3) is degenerate, reflecting the intimate linkage between the roles of the two parameters in θ and the associated singularity of the limiting distribution arising from the asymptotic collinearity of the induced regressors in the linearized system corresponding to the model (2.1). This phenonomenon mirrors the singular limit distribution behavior that was observed in Phillips (2007) in the context of power function deterministic trend regressors.

Theorem 2.1 can be extended to include the extra domain $-1/2 < \gamma_0 \leq 0$ of the power parameter under the following additional condition:

A2 (i) x_k/d_k has a density $p_k(x)$ that is uniformly bounded by a constant K for all $1 \leq k \leq n$ and $x \in \mathbb{R}$,

(ii) X_t has a density $\widetilde{p}_t(x)$ satisfying $\sup_x \int_0^1 \widetilde{p}_t(x) dt < \infty$,

(iii) $\sup_{k\geq 1} E\big(u_k^2 \mid \mathcal{F}_{k-1}\big) \leq C < \infty$.

We mention that **A2** (ii) ensures the existence of U_i for $-1/2 < \gamma_0 \leq 0$ and the smoothness condition on the density of x_k/d_k (i.e., **A1**(i)) is essentially necessary for the convergence of the sample quantities to U_i. See, for instance, Berkes and Horváth (2006). Further discussion is given in Section 3. We also require the more restrictive **A2** (iii) instead of $\sup_{k\geq 1} Eu_k^2 < \infty$ for technical reasons in the proof. This condition can be modified under higher moment conditions on u_k and a narrower interval of validity for

the power parameter γ_0. These extensions involve some complex further calculations and are therefore not pursued in the present work.

Theorem 2.2. *Suppose **A1** and **A2** hold. The limit theory (2.3) continues to hold for* $-1/2 < \gamma_0 \leq 0$.

Remark 2.1. If $\gamma_0 < -1/2$, the random variable U_2 does not exist even in the case where the process X_t is standard Brownian motion B_t because the integral fails to converge. Note, in particular, that $\int_0^1 |B_t|^{2\gamma_0} dt = \infty$ *a.s.* when $\gamma_0 < -1/2$. See, e.g., Ethier and Kurtz (1986, p. 332). In consequence, it is doubtful whether there is any limit distribution of the standardized and centred estimator $F_n(\hat{\theta}_n - \theta_0)$ when $\gamma_0 < -1/2$ under the present settings.

Remark 2.2. The proof of (2.3) heavily depends on the existence of the integral $\int_0^1 |X_t|^{2\gamma_0 - \epsilon} dt$ for some small $\epsilon > 0$ rather than just only on the existence of U_i itself. Since **A2** (ii) is essentially required to ensure the existence of $\int_0^1 |X_t|^{-\epsilon} dt$ for some $\epsilon > 0$, this helps to explain why Theorem 2.1 excludes the case $\gamma_0 = 0$, but this can be established under the additional condition **A2**, as seen in Theorem 2.2.

Remark 2.3. If $\gamma > -1/2$ is fixed and known, the LSE $\hat{\beta}(\gamma)$ of β in model (2.1) is given by

$$\hat{\beta}(\gamma) \;=\; \frac{\sum_{k=1}^n y_k |x_k|^\gamma}{\sum_{k=1}^n |x_k|^{2\gamma}} = \beta_0 + \frac{\sum_{k=1}^n |x_k|^\gamma u_k}{\sum_{k=1}^n |x_k|^{2\gamma}}.$$

Under the given conditions (**A2** is required if $\gamma < 0$), it is readily seen that

$$\sqrt{n}\, d_n^\gamma \big[\hat{\beta}(\gamma) - \beta_0\big] \to_D \frac{\int_0^1 |X_t|^\gamma dB_t}{\int_0^1 |X_t|^{2\gamma} dt} \tag{2.4}$$

In comparison with (2.3), there is now a different convergence rate for the asymptotic distribution of $\hat{\beta}(\gamma)$. This phenomenon was noted by Phillips (2007) in the context of nonlinear power trend regression is investigated.

Remark 2.4. Using (2.3), we have

$$\sqrt{n} d_n^{\gamma_0} (\widehat{\gamma}_n - \gamma_0) \;\; \to_D \;\; \frac{1}{\beta_0} \frac{U_1 \int_0^1 |X_t|^{\gamma_0} dB_t - U_0 \int_0^1 |X_t|^{\gamma_0} \log |X_t| \, dB_t}{U_1^2 - U_0 \, U_2}, \tag{2.5}$$

where $U_i = \int_0^1 |X_t|^{2\gamma_0} \log^i |X_t| \, dt$ for $i = 0, 1$ and 2. Since β_0 is consistently estimable, this limit theory enables model specification of linear cointegration, which involves testing the null hypothesis

$$H_0 : \gamma = 1 \quad \text{vs} \quad H_1 : \gamma \neq 1. \tag{2.6}$$

Indeed a simple test statistic that may be used to test (2.6) can be defined by

$$T_n = \sqrt{n} d_n \, \widehat{\beta}(1) \, (\widehat{\gamma}_n - 1), \tag{2.7}$$

where $\widehat{\beta}(1)$ is given as in Remark 2.3. From (2.5) and (2.4) it follows that under the null H_0,

$$T_n \to_D \frac{\widetilde{U}_1 \int_0^1 |X_t| dB_t - \widetilde{U}_0 \int_0^1 |X_t| \log |X_t| \, dB_t}{\widetilde{U}_1^2 - \widetilde{U}_0 \, \widetilde{U}_2},$$

where $\widetilde{U}_i = \int_0^1 |X_t| \log^i |X_t| \, dt$ for $i = 0, 1$ and 2. This test is therefore asymptotically pivotal and consistent with $P_{H_1}(|T_n| \geq t_0) \to 1$ for any $t_0 > 0$, where $P_{H_1}(\cdots)$ denotes the probability under the alternative H_1. But some aspects of inference, such as confidence interval construction, are more difficult. The limit distribution of $\hat{\theta}_n$, given in (2.3) depends jointly on the parameter vector $\theta_0 = (\beta_0, \gamma_0)$, making direct inference about θ_0 in power regression more complex. It is not clear at present whether or not an asymptotic theory might be developed for $\hat{\theta}_n$ using a different approach such as a self-normalized quantity in place of the use of rate matrix scaling like F_n so that the unknown parameter θ_0 on the right hand of (2.3) can be eliminated and an asymptotically pivotal approach developed.

Remark 2.5. In a natural setting amenable to a linear cointegrated structure, it may be desirable to consider the following nonlinear power function cointegrating regression model

$$y_k = \alpha' z_k + \beta |x_k|^\gamma + \, u_k, \tag{2.8}$$

where $\alpha = (\alpha_1, ..., \alpha_d)'$, $\beta \in \mathbb{R}$, and $-1/2 < \gamma < 1$ are unknown parameters, z_k is a $d-$dimensional regressor whose differences $\Delta z_k = z_k - z_{k-1}$ are stationary, and x_k and u_k are defined as in Theorem 2.1. In applications related to cointegration analysis and forecasting based on usual linear regression formulations, the power term $\beta |x_k|^\gamma$ in model (2.8) may provide an extra precision correction term that admits nonlinear effects that are relevant in certain empirical examples. It is also interesting to consider the impact of the presence of a power regressor term in cointegrating regression as a simple mechanism for testing linearity, as was done in Baek et al. (2015) and Cho and Phillips (2018) in stationary and deterministic trend model settings. The limit theory in the present paper provides a foundation for a general study of such formulations and tests, in addition to the approach based on the test T_n given in (2.7). Full investigation of this topic in the present context requires challenging new limit theorems, which are deferred to later work.

2.1 Extension to endogeneity

The data generating process in model (2.1) is assumed to have a martingale structure. This is used in many articles in parametric cointegrating regression. See, for instance, Chang, Park and Phillips (2001), Park and Phillips (2001) and Chan and Wang (2015). From the viewpoint of empirical applications, however, this martingale structure is restrictive. The aim of this section is to remove the restriction so that endogeneity is allowed in the model. Explicitly, we are concerned with the model:

$$y_k = \beta|x_k|^\gamma + w_k, \tag{2.9}$$

where x_k is the partial sum process $x_k = \sum_{j=1}^k \xi_j$,

$$w_k = u_k + z_{k-1} - z_k, \tag{2.10}$$

the u_k are assumed to satisfy **A1**, and the z_k satisfy certain regularity conditions that are specified as follows:

A3 (i) $\sup_{k\geq 1} E\big[(1+|z_{k-1}|)(1+|\xi_k|)\big]^\alpha < \infty$ for some $\alpha > 1$;

(ii) $Ez_{k-1}\xi_k \to A_0$, as $k \to \infty$;

(iii) $\sup_{k\geq 2m} |E(\lambda_k|\mathcal{F}_{k-m})| = o_P(1)$, as $m \to \infty$, where $\lambda_k = z_{k-1}\xi_k - Ez_{k-1}\xi_k$.

The process $\{w_k\}_{k\geq 1}$ in (2.10) was used by Peng and Wang (2018) in an investigation of weak convergence to stochastic integrals beyond the usual semimartingale structure. The extension arises because $\{w_k\}_{k\geq 1}$ is usually not a martingale difference, but the partial sum process $\sum_{k=1}^n w_k = \sum_{k=1}^n v_k + z_0 - z_n$ provides an approximation to a martingale, just as in the decomposition of Phillips and Solo (1992). Such martingale approximations have been widely studied in the literature. As shown in Peng and Wang (2018), **A3** allows for both w_k and ξ_k to be a causal process, admits near-epoch dependence in the model, and introduces endogeneity by virtue of **A3**(ii).

Within this framework and asymptotic theory for the estimator $\hat{\theta}_n$ can be developed under model (2.9). We set $F_n = diag\big[\sqrt{n}d_n^{\gamma_0}/\log d_n, \sqrt{n}d_n^{\gamma_0}\big]$ as earlier, as have the following result.

Theorem 2.3. *Suppose that $d_n^2/n \to \infty$, **A1** and **A3** hold. For $\gamma_0 > 1$, we have*

$$F_n(\hat{\theta}_n - \theta_0) \;\to_D\; \begin{pmatrix} 1 \\ -1/\beta_0 \end{pmatrix} \frac{U_0 \int_0^1 |X_t|^{\gamma_0} \log|X_t|\, dB_t - U_1 \int_0^1 |X_t|^{\gamma_0} dB_t}{U_1^2 - U_0\, U_2}, \tag{2.11}$$

*where, for $i = 0, 1, 2$, U_i are defined as in Theorem 2.1. If **A2** holds in addition, we still have (2.11) for any $1/\alpha < \gamma_0 \leq 1$.*

The condition that $d_n^2/n \to \infty$ is satisfied if ξ_k is a long memory process, in which case we usually have $d_n^2 = var(\sum_{k=1}^n \xi_k) \sim C\, n^\mu$ for some $1 < \mu < 2$. See, Wang, et al. (2003), for instance. Due to the fast convergence rate involving d_n in F_n, the additional term involving z_k in the equation error (2.10) does not produce a bias term in the limit distribution. But elimination of the bias term requires a more restrictive condition on the interval in which the real parameter γ_0 is located. More explanation can be found in Remark 2.6 given discussed below.

The situation is different if $d_n^2/n \to \sigma^2 < \infty$, which generally holds if x_k is a partial sum of a short memory process ξ_k. In this case, as the following theorem shows, the additional term z_k has an essential impact on the limit distribution. Explicitly, when $0 < \sigma < \infty$, the additional term z_k contributes a bias term in comparison with (2.11). It is interesting to notice that, when $\sigma = 0$ (i.e., $d_n/\sqrt{n} \to 0$), the additional term z_k dominates and the convergence rate of $\hat{\theta}_n - \theta_0$ becomes slow. It seems that this phenomenon was unnoticed in previous research even in the case of linear cointegrating regression.

Theorem 2.4. *Suppose that $d_n/\sqrt{n} \to \sigma$ with $0 \le \sigma < \infty$, **A1** and **A3** hold. For $\gamma_0 > 1$, we have*

$$\frac{d_n}{\sqrt{n}} F_n(\hat{\theta}_n - \theta_0) \;\; \to_D \;\; \begin{pmatrix} 1 \\ -1/\beta_0 \end{pmatrix} \frac{U_0\, W_1 \; - U_1\, W_0}{U_1^2 - U_0\, U_2}, \tag{2.12}$$

where, for $i = 0, 1, 2$, U_i are defined as in Theorem 2.1 and

$$W_i \;\; = \;\; \sigma \int_0^1 |X_t|^{\gamma_0} \log^k |X_t| dB_t + \; A_0 \int_0^1 |X_t|^{\gamma_0 - 1} \big(\gamma_0 \log^k |X_t| + k\big)\, sign(X_t)\, dt$$

*for $i = 1, 2$. If in addition that **A2** holds, we still have (2.12) for any $1/\alpha < \gamma_0 \le 1$.*

Remark 2.6. As explained in Remark 2.1, to ensure the existence of W_k, some restriction on the range of the real parameter γ_0 is essentially necessary, in the present case amounting to the condition $\gamma_0 > 0$ because of the presence of the factor $|X_t|^{\gamma_0 - 1}$ in the integrand of the second component of W_k. Moreover, there is a trade off between the condition $\gamma_0 > 1/\alpha$ used in Theorem 2.4 and the moment condition on z_k as is apparent from the condition assumed in **A3** (i). It is not clear at the moment whether or not the moment condition on z_k can be improved without sacrificing the interval where γ_0 is satisfied.

2.2 Further extension and remarks

Phillips (2007) considered a regression model in the following form:

$$y_k = \alpha + \beta\, l(k) + u_k, \tag{2.13}$$

where $l(x)$ is a function slowly varying at ∞. When $l(x) = \log x$, (2.13) becomes the semilogarithmic growth model, which raises naturally in the study of growth convergence problems and economic transition. Since the sample moment matrix of the regressors is asymptotically singular, model (2.13) fails to fit within the usual framework. Phillips (2007) investigated asymptotics of LSE $(\hat{\alpha}, \hat{\beta})$ of (α, β) by using a second order approximation of $l(xn)$ by $l(n)$ for any $x \in \mathbb{R}$.

Using similar arguments as in Phillips (2007) and the results developed in Section 3, model (2.13) can be extended to a stochastic slowly evolving trend model defined as follows

$$y_k = \alpha + \beta\, l(|x_k|) + u_k, \tag{2.14}$$

where $x_k = \sum_{j=1}^{k} \xi_j$ and the u_k are assumed to satisfy **A1**. Let (α_0, β_0) be the true parameter of (α, β). We have the following theorem.

Theorem 2.5. *Suppose **A1** and **A2** hold and $l(x)$ satisfies the following condition: there exists $\epsilon(\lambda) \to 0$ as $\lambda \to \infty$ such that*

$$\sup_{0<|x|\le A} \Big| \frac{l(|x|\lambda)}{l(\lambda)} - 1 - \epsilon(\lambda) \log |x| \Big| = o\big[\epsilon(\lambda)\big], \tag{2.15}$$

for any fixed $A > 0$. For the LSE $(\hat{\alpha}, \hat{\beta})$ of (α, β), we have

$$F_{1n}\Big[\begin{pmatrix} \hat{\alpha}_n \\ \hat{\beta}_n \end{pmatrix} - \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} \Big] \;\to_D\; \begin{pmatrix} 1 \\ -1 \end{pmatrix} \frac{\int_0^1 \log(|X_t|) dB_t - B_1 \int_0^1 \log(|X_t|) dt}{\int_0^1 \log^2(|X_t|) dt - \big[\int_0^1 \log(|X_t|) dt\big]^2}, \tag{2.16}$$

where $F_{1n} = diag\{\sqrt{n} l(d_n)\epsilon(d_n), \sqrt{n}\epsilon(d_n)\}$.

We remark that condition (2.15) is weak and is satisfied by the majority of slowly varying functions. For details, see Phillips (2007). The martingale structure given in **A1** is essential for the establishment of Theorem 2.5. Indeed, in the proof of Theorem 2.5, we need to handle sample covariances of the type $S_n = \sum_{k=1}^{n} \log |x_{nk}|\, u_k$. Since $d \log x/dx = 1/x$ is not locally integrable, as seen in Section 3, we cannot provide asymptotics for S_n in the case where u_k is replaced by w_k. This was also noticed in de Jong (2002). If $l(x)$ satisfies certain continuity conditions rather than being slowly varying at ∞, it is possible to modify model (2.14) so that endogeneity is allowed. For details, we refer to Peng and Wang (2018).

Several papers have studied the general nonlinear parametric cointegrating regression model

$$y_k = f(x_k, \theta) + u_k, \tag{2.17}$$

where $f : \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$ is a known nonlinear function, x_k is a non-stationary regressor, u_k is a regression error, and θ is an m-dimensional parameter vector that lies in the parameter set Θ. See, for instance, Park and Phillips (2001), Chang, et al. (2001), de Jong (2002), Chan and Wang (2015), Wang and Phillips (2016) and Wang (2018). For extensions to semiparametric models, we refer to Kim and Kim (2012) and Gao and Phillips (2013).

Various settings have been specified for the known nonlinear function $f(x, \theta)$ in the literature. Park and Phillips (2001) and many subsequent papers considered the situation where $f(x, \theta)$ belongs to a class of asymptotically homogeneous functions, i.e.,

$$f(\lambda x, \theta) = v(\lambda, \theta)h(x, \theta) + R(x, \lambda, \theta), \tag{2.18}$$

where $v(\lambda, \theta)$ is not singular, $h(x, \theta)$ is regular[3] on Θ and $R(x, \lambda, \theta)$ is of order smaller than $v(\lambda, \theta)$ for all $\theta \in \Theta$. Chan and Wang (2015) [also see Wang and Phillips (2016) and Wang (2018)] made use of Lipschitz type conditions on $f(x, \theta)$ and $\partial f(x, \theta)/\partial\theta$ with respect to θ. As mentioned in the introduction, Chan and Wang (2015) excluded the power function $f(x, \beta, \gamma) = \beta\,|x|^\gamma$ in their treatment and one aim of the present paper is to fill in the gap in the literature. On the other hand, while the power function $f(x, \beta, \gamma) = \beta\,|x|^\gamma$ is included in (2.18), Park and Phillips (2001) essentially required the following condition after changing the probability space and using Skorohod embedding:

$$\frac{1}{n}\sum_{k=1}^{n} f(x_k/\sqrt{n}, \theta_0) \to_{a.s} \int_0^1 f(B_t, \theta_0)dt, \tag{2.19}$$

in their main results (i.e., Theorems 5.2 and 5.3) following a detailed check of their proofs. It seems difficult to prove (2.19) even for the case where x_k is a partial sum of i.i.d. normal variables without changing the probability space. In consequence, using (2.19) is restrictive and it is desirable to have a direct proof of joint weak convergence of the relevant functionals like $\frac{1}{n}\sum_{k=1}^{n} f(x_k/\sqrt{n}, \theta_0)$ together with corresponding sample covariance functions and normalized processes, such as those given in Theorem 3.1 in the following section.

It would be interesting to consider model (2.17) within nonlinear function settings like (2.18) under alternative conditions that can be more easily verified than (2.19). To do so, new limit theorems need to be developed for regular functions like $h(x, \theta)$ along the

[3]A function H is called regular on Θ if (a) for all $x \in \mathbb{R}$, $H(x, \cdot)$ is equicontinuous in a neighborhood of x; (b) for each $\theta \in \Theta$, $H(., \theta)$ is regular, i.e., for any compact subset K of $\mathbb{R}$, there exist for each $\epsilon > 0$ continuous functions $\underline{H}_\epsilon$, $\overline{H}_\epsilon$, and a constant $\delta_\epsilon > 0$ such that $\underline{H}_\epsilon(x, \theta) \le H(y, \theta) \le \overline{H}_\epsilon(x, \theta)$ for all $|x - y| < \delta_\epsilon$ on K, and such that $\int_K (\overline{H}_\epsilon - \underline{H}_\epsilon)(x, \theta)dx \to 0$ as $\epsilon \to 0$.

lines of Pötscher (2004), Berkes and Horvath (2006), Wang (2014), Liang, et al. (2016) and Peng and Wang (2018). This more general development is beyond the scope of the current paper and is left for future work.

3 Convergence to stochastic integrals: beyond the semimartingale structure

We maintain the same notation as in Section 2, except when explicitly mentioned. Our first result provides a framework of joint weak convergence to stochastic integrals that accommodates the normalized process, sample moments of nonlinear functions and sample covariances. This result, which follows in a long tradition of similar results, delivers the technical tools needed to establish the main limit Theorems 2.1 - 2.4 given in Section 2 because of its allowance for locally integrable functions and hence power functions in regression models.

Theorem 3.1. *Suppose **A1** holds. For any locally Riemann integrable functions $g(s)$ and $f(s)$, we have*

$$\begin{aligned} & \Big(x_{n,[nt]},\ u_{n,[nt]},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})\, u_k \Big) \\ \Rightarrow\ & \Big(X_t,\ B_t,\ \int_0^1 g(X_t)dt,\ \int_0^1 f(X_t)\, dB_t \Big), \end{aligned} \tag{3.1}$$

*on $D_{\mathbb{R}^4}[0,1]$. If **A2** holds in addition, we still have (3.1) whenever $g(x)$ and $f^2(x)$ are locally integrable.*

Aspects of the first part of Theorem 3.1 are known in the existing literature, particularly for situations where the functions $g(x)$ and $f(x)$ are continuous. See, for instance, Kurtz and Protter (1991). Extension to locally integrable functions seems to be new and in such cases the condition **A2** is essentially necessary to ensure the existence of the stochastic integrals in the limit. Applying Theorem 3.1 to the functions $g(x) = |x|^{2\gamma} \log^{m_1} |x|$ and $f(x) = |x|^{\gamma} \log^{m_2} |x|$, where $m_1, m_2 \geq 0$ are integers, we have the following corollary, which plays a key role in the proofs of the main results in the paper.

Corollary 3.1. *Suppose **A1** holds. For all $\gamma > 0$, we have*

$$\begin{aligned} W_n(\gamma) \quad := \quad & \Big(\frac{1}{n}\sum_{k=1}^{n} |x_{nk}|^{2\gamma} \log^{m_1} |x_{nk}|,\ \frac{1}{\sqrt{n}}\sum_{k=1}^{n} |x_{nk}|^{\gamma} \log^{m_2} |x_{nk}|\, u_k \Big) \\ \to_D \quad & \Big(\int_0^1 |X_t|^{2\gamma} \log^{m_1} |X_t| dt,\ \int_0^1 |X_t|^{\gamma} \log^{m_2} |X_t| dB_t \Big), \end{aligned} \tag{3.2}$$

jointly for all integers $m_1, m_2 \geq 0$. *If* ***A2*** *holds in addition, we still have (3.2) for* $-1/2 < \gamma \leq 0$.

Let $C[a, \infty)$ denote the set of all continuous real-valued functions defined on the interval $[a, \infty)$ endowed with the uniform norm topology. With an index γ that satisfies $\gamma \geq a > -1/2$, it is readily seen that $\{W_n(\gamma), n \geq 1\}$ is a sequence of random processes defined on the space $C[a, \infty)$. Consequently, we may extend Corollary 3.1 to the following form of functional convergence for the process $W_n(\gamma)$.

Theorem 3.2. *Suppose* ***A1*** *and* ***A2*** *hold. On* $C[a, \infty)$ *with* $a > -1/2$, *for any integers* $m_1, m_2 \geq 0$ *we have*

$$W_n(\gamma) \;\Rightarrow\; \Big(\int_0^1 |X_t|^{2\gamma} \log^{m_1} |X_t| dt,\; \int_0^1 |X_t|^{\gamma} \log^{m_2} |X_t| dB_t \Big). \tag{3.3}$$

In related work to Theorem 3.1 and Corollary 3.1 on convergence to stochastic integrals that sought generality beyond a semimartingale structure, Liang, et al. (2016) and Wang (2015, Section 4.5) obtained weak convergence results of sample quantities such as $\sum_{k=0}^{n-1} f(x_{nk})\, w_k$, where $w_k = \sum_{j=0}^{\infty} \varphi_j\, u_{k-j}$, with $\varphi = \sum_{j=0}^{\infty} \varphi_j \neq 0$ and $\sum_{j=0}^{\infty} j\, |\varphi_j| < \infty$, and u_k defined as in **A1**(i). More recently, Peng and Wang (2018) provided another result on such sample covariances by using the error process representation $w_k = u_k + z_{k-1} - z_k$ given in (2.10) instead of the martingale difference u_k. While these results are useful, the functions $f(x)$ that are employed satisfy strong smoothness conditions that require $f'(x)$ to be continuous and satisfy a Lipschitz condition. Such conditions are clearly not satisfied for functions that arise in power regression of the form $f(x) = |x|^{\gamma} \log^{k} |x|$, where $k \geq 0$ are integers, particularly, in the case where $f(x)$ is locally integrable (i.e., $\gamma < 0$) rather than locally Riemann integrable.

The aim of the following theorems is to fill this gap, providing new results on convergence to stochastic integrals for the purpose of this paper. We mention that these extensions are non-trivial. To resolve the limit theory, we need to introduce new techniques involving truncation and functional approximation. We mention here that the ideas developed in the proofs seem promising for use in even more general situations such as convex functions, although those extensions are not pursued in the present work.

For use in the following, recall that $w_k = u_k + z_{k-1} - z_k$, as defined in (2.10).

Theorem 3.3. *Suppose that* $d_n^2/n \to \infty$ *and* ***A1*** *and* ***A3*** *hold. Then, for any* $\gamma > 1$, *any*

integer $m \geq 0$, and any locally Riemann integrable function $g(s)$, we have

$$\begin{aligned} &\Big(x_{n,[nt]},\ u_{n,[nt]},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} |x_{nk}|^{\gamma}\ \log^{m}|x_{nk}|\, w_k\Big) \\ \Rightarrow\ &\Big(X_t,\ B_t,\ \int_0^1 g(X_t)dt,\ \int_0^1 |X_t|^{\gamma}\ \log^{m}|X_t|\, dB_t\Big), \end{aligned} \tag{3.4}$$

*on $D_{\mathbb{R}^4}[0,1]$. If **A2** holds in addition, then (3.4) remains valid for any $1/\alpha < \gamma \leq 1$, any integer $m \geq 0$, and any locally integrable function $g(x)$, where α is given in **A3**(i).*

As noted in Section 2, the rate condition $d_n^2/n \to \infty$ usually holds if ξ_k is a long memory process. The result is different if ξ_k is a short memory process or equivalently $d_n^2/n \to \sigma^2 < \infty$ ($\sigma = 0$ is allowed), as seen in the following theorem.

Theorem 3.4. *Suppose that $d_n/\sqrt{n} \to \sigma$ with $0 \leq \sigma < \infty$, and **A1** and **A3** hold. For any $\gamma > 1$, any integer $m \geq 0$ and any locally Riemann integrable function $g(s)$, we have*

$$\begin{aligned} &\Big(x_{n,[nt]},\ u_{n,[nt]},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{d_n}{n}\sum_{k=1}^{n-1} |x_{nk}|^{\gamma}\ \log^{m}|x_{nk}|\, w_k\Big) \\ \Rightarrow\ &\Big(X_t,\ B_t,\ \int_0^1 g(X_t)dt,\ \sigma\int_0^1 |X_t|^{\gamma}\ \log^{m}|X_t|\, dB_t + A_0\int_0^1 f'(X_t)dt\Big), \end{aligned} \tag{3.5}$$

on $D_{\mathbb{R}^4}[0,1]$, where $f'(x) = |x|^{\gamma-1}\log^{m-1}|x|(\gamma\log|x| + m)sign(x)$.

*If **A2** holds in addition, (3.5) remains valid for any $1/\alpha < \gamma \leq 1$, any integer $m \geq 0$ and any locally integrable function $g(x)$, where α is given in **A3**(i).*

Remark 3.1. As mentioned by Peng and Wang (2018), results such as Theorems 3.3 and 3.4 have application to the following processes that are relevant in much time series econometric work:

(i) ξ_k is a long memory process and w_k is a stationary causal process such as time series generated by TAR and bilinear models;

(ii) both ξ_k and w_k are stationary causal processes; and

(iii) $(\xi_k, w_k)_{k\geq 1}$ is near-epoch dependent, particularly a sequence of α-mixing random variables.

Just as in Theorem 3.2, Theorems 3.3 and 3.4 may be extended as follows to functional weak convergence results involving the index γ.

Theorem 3.5. *Suppose that* ***A1****,* ***A2*** *and* ***A3*** *hold. Let* $A > 1/\alpha$ *be a real number, where* α *is given in* ***A3****(i), and* $m \geq 0$ *is an integer.*

(a). If $d_n^2/n \to \infty$*, on* $C[A, \infty)$ *we have*

$$Z_n(\gamma) := \frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} |x_{nk}|^{\gamma} \log^m |x_{nk}| \, w_k \Rightarrow \int_0^1 |X_t|^{\gamma} \log^m |X_t| \, dB_t.$$

(b). If $d_n/\sqrt{n} \to \sigma$ *with* $0 \leq \sigma < \infty$*, on* $C[A, \infty)$ *we have*

$$Z_n(\gamma) := \frac{d_n}{n} \sum_{k=1}^{n-1} |x_{nk}|^{\gamma} \log^m |x_{nk}| \, w_k \Rightarrow \sigma \int_0^1 |X_t|^{\gamma} \log^m |X_t| \, dB_t + A_0 \int_0^1 f'(X_t) dt,$$

where $f'(t)$ *is defined in Theorem 3.4.*

Remark 3.2. The functional limit theorems for the process $Z_n(\gamma)$ appearing in Theorems 3.2 and 3.5 are useful in testing linearity or polynomial regression using power transformations of regressors. See, for example, Baek et al (2015) and Cho and Phillips (2018). Further, in models like (2.8) we may be interested in testing $\beta = 0$. In such inference problems, the unknown power parameter γ is identified (or semi-identified) only under the alternative (local alternative) hypothesis and is unidentified under the null. Weak identification occurs in such cases because the loading coefficient parameter β of the nonlinear function may be close to zero and limit theory under the alternative typically fails to provide a good approximation to finite sample behavior close to the null. Development of a local limit theory that improves the approximation uniformly well irrespective of the strength of the identification relies on uniform weak convergence of sample covariance functionals to stochastic integral limits. In such cases, the related test statistic is required to satisfy certain functional limit theorems with respect to γ. See Shi and Phillips (2012) for a development of such a theory that involves nonstationary data and Andrews and Cheng (2012) for a general discussion.

4 Conclusion

Power function regressions provide a simple way of generalizing simpler polynomial representations and offer potential for constructing general omnibus tests for specification, as shown in Cho and Phillips (2018). These characteristics extend to nonlinear cointegrating regression models with power function regressors. The present paper provides

new limit theory that enables the development of an asymptotic theory of estimation in such models, allowing for both endogeneity in the regressors and for heterogeneity in the errors. As in earlier research on nonlinear nonstationary regression models, a key element in the asymptotics is the establishment of stochastic integral limit theory that goes beyond standard martingale and semimartingale structures. The findings in the present work add to that literature and provide a broader foundation for estimation and inference in models with these characteristics.

5 Proofs of the main results

This section provides proofs of the main theorems. We first prove Theorems 3.1 - 3.4, since these theorems provide technical support for the proofs of Theorems 2.1 - 2.4.

We start with some basic preliminaries. Recalling $x_{n,[nt]} \Rightarrow X_t$ on $D_{\mathbb{R}}[0,1]$ and the limit process $X(t)$ is path continuous, we have $x_{n,[nt]} \Rightarrow X_t$ on $D_{\mathbb{R}}[0,1]$ in the sense of uniform topology. See, for instance, Section 18 of Billingsley (1968). This fact implies that

$$\limsup_{N\to\infty} \limsup_{n\to\infty} P\Big(\max_{1\le k\le n} |x_{nk}| \ge N \Big) = 0, \tag{5.1}$$

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

$$P(\sup_{|s-t|\le\tilde{\delta}} |x_{n,[nt]} - x_{n,[ns]}| \ge \delta) \le \varepsilon \tag{5.2}$$

holds for all sufficiently large n. In terms of (5.2), for any $\delta > 0$, we have

$$\lim_{n\to\infty} P(\max_{0\le j\le n/m} \max_{jm\le k\le (j+1)m} |x_{nk} - x_{n,jm}| \ge \delta) = 0, \tag{5.3}$$

for any $m := m_n \to \infty$ satisfying $n/m \to \infty$.

We next introduce a lemma, which play a key role in the proof of the main theorems. Let v_k be a sequence of arbitrary stochastic processes satisfying the following condition.

A4. $\sup_{k\ge 1} E|v_k| < \infty$ and there exist $A_0 \in \mathbb{R}$ and $0 < m := m_n \to \infty$ satisfying $n/m \to \infty$ so that $\max_{m\le j\le n-m} E\left|\frac{1}{m}\sum_{k=j+1}^{j+m} v_k - A_0\right| = o(1)$.

Lemma 5.1. *Suppose **A4** holds. For any locally Riemann integrable function $H(x)$ we have*

$$\frac{1}{n}\sum_{k=1}^{n} H(x_{nk})v_k \;\; = \;\; \frac{A_0}{n}\sum_{k=1}^{n} H(x_{nk}) + o_P(1). \tag{5.4}$$

If in addition $\sup_{k\ge 1} E|v_k|^{\alpha} < \infty$ for some $\alpha > 1$ and x_k/d_k has a density $p_k(x)$ that is uniformly bounded by a constant K for all $1 \le k \le n$ and $x \in \mathbb{R}$, then (5.4) remains valid if $H^{\alpha/(\alpha-1)}(x)$ is a locally integrable function.

Remark 5.1. If we are only interested in the boundedness of $\sum_{k=1}^{n} H\big(x_{nk}\big)v_k$, condition **A4** is not necessary. Indeed, following the proof of Lemma 5.1, it is easy to see that

$$\sum_{k=1}^{n} H\big(x_{nk}\big)v_k = O_p(n), \tag{5.5}$$

if one of the following conditions holds:

(a) $\sup_{k\geq 1} E|v_k| < \infty$ and $H(x)$ is a locally Riemann integrable function;

(b) $\sup_{k\geq 1} E|v_k|^\alpha < \infty$ for some $\alpha > 1$, x_k/d_k has a density $p_k(x)$ that is uniformly bounded by a constant K for all $1 \leq k \leq n$ and $x \in \mathbb{R}$, and $H^{\alpha/(\alpha-1)}(x)$ is locally integrable.

Proof. Let $\widetilde{\lambda}_k = v_k - A_0$, $H_N(x) = H(x)I(|x| \leq N)$, $R_n = \frac{1}{n}\sum_{k=1}^n H(x_{nk})\,\widetilde{\lambda}_k$ and $R_{1n} = \frac{1}{n}\sum_{k=1}^n H_N(x_{nk})\,\widetilde{\lambda}_k$. Due to (5.1), we have

$$P(R_n \neq R_{1n}) \leq P\Big(\max_{1\leq k\leq n} |x_{nk}| > N\Big) \to 0,$$

as $n \to \infty$ first and then $N \to \infty$, Lemma 5.1 will follow if we prove

$$R_{1n} = \frac{1}{n}\sum_{k=1}^n H_N(x_{nk})\,\widetilde{\lambda}_k = o_P(1), \tag{5.6}$$

for each fixed $N \geq 1$.

We first assume that $H(x)$ is locally Riemann integrable. In this situation, $H_N(x)$ is uniformly bounded and, for any $\epsilon > 0$, there exist continuous functions $H_{N,\epsilon}(x)$ with bounded support such that $|H_N(x) - H_{N,\epsilon}(x)| \leq \epsilon$. Furthermore, since $H_{N,\epsilon}(x)$ is uniformly continuous, for any $\epsilon > 0$ there exists a $\delta_\epsilon > 0$ so that whenever $|x - y| \leq \delta_\epsilon$ we have

$$|H_N(x) - H_N(y)| \leq |H_N(x) - H_{N,\epsilon}(x)| + |H_{N,\epsilon}(x) - H_{N,\epsilon}(y)| + |H_N(y) - H_{N,\epsilon}(y)| \leq 3\epsilon.$$

Write $\Omega_{\delta_\epsilon} = \{\omega : \max_{0\leq j\leq n/m} \max_{jm\leq k\leq (j+1)m} |x_{nk} - x_{n,jm}| \leq \delta_\epsilon\}$ and $T_n = [n/m] - 1$, where m is chosen so that $m \to \infty$ and $n/m \to \infty$. By virtue of the above facts, it is readily seen that, on Ω_{δ_ϵ},

$$\begin{aligned}
|R_{1n}| &\leq \frac{1}{n}\sum_{j=0}^{T_n} \Big|\sum_{k=jm+1}^{(j+1)m} H_N(x_{nk})\,\widetilde{\lambda}_k\Big| + \frac{1}{n}\sum_{k=mT_n+1}^{n} |H_N(x_{nk})\,\widetilde{\lambda}_k| \\
&\leq \frac{1}{n}\sum_{j=0}^{T_n} |H_N(x_{n,jm})| \Big|\sum_{k=jm+1}^{(j+1)m} \widetilde{\lambda}_k\Big| + \frac{C_N}{n}\sum_{k=mT_n+1}^{n} |\widetilde{\lambda}_k| \\
&\quad + \max_{0\leq j\leq T_n} \max_{jm+1\leq k\leq (j+1)m} |H_N(x_{nk}) - H_N(x_{n,jm})|\,\frac{1}{n}\sum_{k=1}^n |\widetilde{\lambda}_k| \\
&\leq \frac{C_N}{n}\sum_{j=0}^{T_n} \Big|\sum_{k=jm+1}^{(j+1)m} \widetilde{\lambda}_k\Big| + \frac{C_N}{n}\sum_{k=mT_n+1}^{n} |\widetilde{\lambda}_k| + \frac{3\epsilon}{n}\sum_{k=1}^n |\widetilde{\lambda}_k|,
\end{aligned} \tag{5.7}$$

where C_N is a constant depending only on N. Now, for any $\eta_1 > 0$ and $\eta_2 > 0$, it follows from (5.3) and **A4** that, for all sufficiently large n,

$$\begin{aligned} P(|R_{1n}| \geq \eta_1) &\leq P\big(\bar{\Omega}_{\delta_\epsilon}\big) + \eta_1^{-1} C_N \max_{m \leq j \leq n-m} E\big|\frac{1}{m}\sum_{k=j+1}^{j+m} v_k - A_0\big| \\ &\quad + \frac{\eta_1^{-1} C_N}{n} \sum_{k=mT_n+1}^{n} E|\widetilde{\lambda}_k\big| + \frac{3\eta_2}{n}\sum_{k=1}^{n} E|\widetilde{\lambda}_k| \;\leq\; C_{1N}\,\eta_2, \end{aligned}$$

by taking $\epsilon = \eta_1\eta_2$, where $\bar{\Omega}_{\delta_\epsilon}$ denotes the complementary set of Ω_{δ_ϵ} and C_{1N} is a constant depending only on N. This proves (5.6) for a locally Riemann integrable function $H(x)$.

We next assume that $H^{\alpha/(\alpha-1)}(x)$ is locally integrable. In this situation, for any $\epsilon > 0$, since both $\int_x H_N(x)dx < \infty$ and $\int_x H_N^{\alpha/(\alpha-1)}(x)dx < \infty$, there exists a continuous function $H_{N,\epsilon}(x)$ such that

$$\int_x |H_N(x) - H_{N,\epsilon}(x)|dx \leq \epsilon \quad \text{and} \quad \int_x |H_N(x) - H_{N,\epsilon}(x)|^{\alpha/(\alpha-1)}dx \leq \epsilon. \tag{5.8}$$

See, for instance, Lemma 5.1.2 of Stein and Shakarchi (2005). Write

$$\begin{aligned} R_{1n} &= \frac{1}{n}\sum_{k=1}^{n} H_{N,\epsilon}\big(x_{nk}\big)\widetilde{\lambda}_k + \frac{1}{n}\sum_{k=1}^{n}\big[H_N(x_{nk}) - H_{N,\epsilon}\big(x_{nk}\big)\big]\widetilde{\lambda}_k \\ &:= S_{n1} + S_{n2}. \end{aligned}$$

For any $\epsilon > 0$, using the fact shown in the first part of the proof, we have $S_{n1} = o_P(1)$.

It suffices to show that $E|S_{n2}| \to 0$ as $n \to \infty$ first and then $\epsilon \to 0$. Note that, by using (5.8) and the fact that x_k/d_k has a density $p_k(x)$ that is uniformly bounded by a constant K, we have

$$\begin{aligned} E\big|H_N(x_{nk}) - H_{N,\epsilon}\big(x_{nk}\big)\big|^{\alpha/(\alpha-1)} &= \int \big|H_N(y/d_{nk}) - H_{N,\epsilon}\big(y/d_{nk}\big)\big|^{\alpha/(\alpha-1)} p_k(y)dy \\ &\leq K \int \big|H_N(y/d_{nk}) - H_{N,\epsilon}\big(y/d_{nk}\big)\big|^{\alpha/(\alpha-1)}dy \\ &\leq K\, d_{nk} \int \big|H_N(y) - H_{N,\epsilon}\big(y\big)\big|^{\alpha/(\alpha-1)}dy \\ &\leq K\,\epsilon\, d_{nk}, \end{aligned}$$

where $d_{nk} = d_n/d_k$. It follows from Hölder's inequality and $d_n^2 \simeq n^\mu$ for some $0 < \mu < 2$ that

$$\begin{aligned} (E|S_{n2}|)^\alpha &\leq \sup_{k\geq 1} E|\tilde{\lambda}_k|^\alpha \Big\{\frac{1}{n}\sum_{k=1}^{n} E\big|H_N(x_{nk}) - H_{N,\epsilon}\big(x_{nk}\big)\big|^{\alpha/(\alpha-1)}\Big\}^{\alpha-1} \\ &\leq C\,\big(\frac{\epsilon}{n}\sum_{k=1}^{n} d_n/d_k\big)^{\alpha-1} \to 0. \end{aligned}$$

This proves $E|S_{n2}| \to 0$, as $n \to \infty$ first and then $\epsilon \to 0$, and hence completes the proof of Lemma 5.1. □

5.1 Proof of Theorem 3.1

We only provide a outline for (3.1) when $g(x)$ and $f^2(x)$ are locally integrable. The other proofs are similar and details are omitted.

The idea is similar to that of Lemma 5.1. Let $f_N(x) = f(x)I(|x| \le N)$ and $g_N(x) = g(x)I(|x| \le N)$. Due to the local integrability of $g(x)$ and $f^2(x)$, for any $\epsilon > 0$, there exist continuous functions $g_{N,\epsilon}(x)$ and$f_{N,\epsilon}(x)$ such that $\int_x |g_N(x) - g_{N,\epsilon}(x)|dx \le \epsilon$,

$$\int_x |f_N(x) - f_{N,\epsilon}(x)|dx \le \epsilon \quad \text{and} \quad \int_x |f_N(x) - f_{N,\epsilon}(x)|^2 dx \le \epsilon. \tag{5.9}$$

It follows from Kurtz and Protter (1991) that, for each $N \ge 1$ and $\epsilon > 0$,

$$\begin{aligned} & \Big(x_{n,[nt]},\ u_{n,[nt]},\ \frac{1}{n}\sum_{k=1}^{n} g_{N,\epsilon}(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f_{N,\epsilon}(x_{nk})\, u_k\Big) \\ \Rightarrow\ & \Big(X_t,\ B_t,\ \int_0^1 g_{N,\epsilon}(X_s)ds,\ \int_0^1 f_{N,\epsilon}(X_s)\, dB_s\Big). \end{aligned}$$

Note that, uniformly for $0 \le t \le 1$,

$$\begin{aligned} & P\Big[\Big(X_t, B_t, \int_0^1 g(X_s)ds, \int_0^1 f(X_s)\, dB_s\Big) \ne \Big(X_t, B_t, \int_0^1 g_N(X_s)ds, \int_0^1 f_N(X_s)\, dB_s\Big)\Big] \\ \le\ & P\Big(\sup_{0\le s\le 1} |X_s| > N\Big) \to 0, \\ & P\Big[\Big(x_{n,[nt]},\ u_{n,[nt]},\ \frac{1}{n}\sum_{k=1}^{n} g(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f(x_{nk})\, u_k\Big) \\ & \qquad \ne \Big(x_{n,[nt]},\ u_{n,[nt]},\ \frac{1}{n}\sum_{k=1}^{n} g_N(x_{nk}),\ \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f_N(x_{nk})\, u_k\Big)\Big] \\ \le\ & P\Big(\max_{1\le k\le n} |x_{nk}| > N\Big) \to 0, \end{aligned}$$

as $n \to \infty$ first and then $N \to \infty$. It is readily seen from these facts that (3.1) will hold

if we prove the following: for each $N \geq 1$,

$$\frac{1}{n}\sum_{k=1}^{n} g_{N,\epsilon}(x_{nk}) - \frac{1}{n}\sum_{k=1}^{n} g_N(x_{nk}) = o_P(1), \tag{5.10}$$

$$\int_0^1 g_{N,\epsilon}(X_t)dt - \int_0^1 g_N(X_t)dt = o_P(1); \tag{5.11}$$

$$I_n := \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f_{N,\epsilon}(x_{nk})\, u_k - \frac{1}{\sqrt{n}}\sum_{k=1}^{n-1} f_N(x_{nk})\, u_k = o_P(1), \tag{5.12}$$

$$I_{1n} := \int_0^1 f_{N,\epsilon}(X_t)\, dB_t - \int_0^1 f_N(X_t)\, dB_t = o_P(1). \tag{5.13}$$

as $n \to \infty$ first and then $\epsilon \to 0$.

As in the proof of Lemma 5.1, by using **A2** (i) and (iii), we have

$$\begin{aligned} EI_n^2 &= \frac{1}{n}\sum_{k=1}^{n-1} E\big\{\big[f_{N,\epsilon}(x_{nk}) - f_N(x_{nk})\big]\, u_k\big\}^2 \\ &\leq \frac{1}{n}\sum_{k=1}^{n-1} E\Big(\sup_{k\geq 1} E\big(u_k^2 \mid \mathcal{F}_{k-1}\big)\, \big[f_{N,\epsilon}(x_{nk}) - f_N(x_{nk})\big]^2\Big) \\ &\leq C\,\frac{1}{n}\sum_{k=1}^{n} d_n/d_k \int_x |f_N(x) - f_{N,\epsilon}(x)|^2 dx \leq C_1\epsilon. \end{aligned}$$

Hence (5.12) holds. Similarly, it follows from **A2** (ii) that

$$\begin{aligned} EI_{1n}^2 &\leq \int_0^1 E\big[f_{N,\epsilon}(X_t) - f_N(X_t)\big]^2 dt \\ &\leq \int_0^1 \int_y \big[f_{N,\epsilon}(y) - f_N(y)\big]^2 p_t(y) dy dt \\ &\leq \sup_y \int_0^1 p_t(y)dt \int_y \big[f_{N,\epsilon}(y) - f_N(y)\big]^2 dy \leq C\,\epsilon, \end{aligned}$$

yielding (5.13). The proofs of (5.10) and (5.11) are similar and this completes the proof of Theorem 3.1. □

5.2 Proof of Theorem 3.2

To prove (3.3), let

$$Y_{1n}(\gamma) = \frac{1}{n}\sum_{k=1}^{n} |x_{nk}|^{2\gamma}\, \log^{m_1} |x_{nk}|, \quad Y_{2n}(\gamma) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} |x_{nk}|^{\gamma}\, \log^{m_2} |x_{nk}|\, u_k.$$

By virtue of Theorem 3.1 and the Cramér-Wold device, it suffices to show the tightness of $Y_{1n}(\gamma)$ and $Y_{2n}(\gamma)$ on $C[a, K]$ for each $K > 0$.

Without loss of generality, assume that $a = 0$ and $K = 1$. It is sufficient to show only tightness for $Y_{2n}(\gamma)$, i.e., that, for every ϵ and $\eta > 0$, there exists an k_0 such that

$$\limsup_{n\to\infty} P\Big[\sup_{\gamma_1,\gamma_2\in[0,1],|\gamma_1-\gamma_1|\le 2^{-k_0}} |Y_{2n}(\gamma_1) - Y_{2n}(\gamma_2)| \ge \epsilon\Big] \le \eta. \tag{5.14}$$

To prove (5.14), for $s = 1, 2, \cdots$, we define $\gamma^{(s)} = j2^{-s}$ if $\gamma \in [j2^{-s}, (j+1)2^{-s})$ for some $j \in \mathbb{N}$. Take a sequence k_n so that

$$n2^{-2k_n} \to 0. \tag{5.15}$$

Then for any $\gamma_1, \gamma_2 \in [0, 1]$ with $|\gamma_1 - \gamma_2| \le 2^{-k_0}$, we have

$$\begin{aligned}
& |Y_{2n}(\gamma_1) - Y_{2n}(\gamma_2)| \\
\le\ & 2\sup_{0\le\gamma\le1} |Y_{2n}(\gamma) - Y_{2n}(\gamma^{(k_0)})| + \sup_{1\le j\le 2^{k_0}} |Y_{2n}(j2^{-k_0}) - Y_{2n}((j-1)2^{-k_0})| \\
\le\ & 2\sup_{0\le\gamma\le1} |Y_{2n}(\gamma) - Y_{2n}(\gamma^{(k_n)})| + 2\sum_{s=k_0}^{k_n} \sup_{1\le j\le 2^s} |Y_{2n}(j2^{-s}) - Y_{2n}((j-1)2^{-s})|.
\end{aligned}$$

Hence, by noting that $P\Big(\max_{1\le k\le n} |x_{nk}| > N\Big) \to 0$ as $N \to \infty$, the result (5.14) will follow if we prove that for every $\delta > 0$ and $N > 0$,

$$\lim_{n\to\infty} E\Big(I_{1n} I\Big(\max_{1\le k\le n} |x_{nk}| \le N\Big)\Big) = 0, \tag{5.16}$$

and, there exists an integral $k_0 > 0$ such that

$$\limsup_{n\to\infty} E\Big(I_{2n} I\Big(\max_{1\le k\le n} |x_{nk}| \le N\Big)\Big) \le \delta, \tag{5.17}$$

where

$$I_{1n}(n) = \sup_{0\le\gamma\le1} |Y_{2n}(\gamma) - Y_{2n}(\gamma^{(k_n)})|,\ I_{2n} = \sum_{s=k_0}^{k_n} \sup_{1\le j\le 2^s} |Y_{2n}(j2^{-s}) - Y_{2n}((j-1)2^{-s})|.$$

We first prove (5.16). Note that, for any $c \le \gamma_1 < \gamma_2 \le d$ and $x \ne 0$,

$$\begin{aligned}
||x|^{\gamma_1} - |x|^{\gamma_2}| &\le |\gamma_1 - \gamma_2|\ \max\{|x|^{\gamma_1}, |x|^{\gamma_2}\} |\log|x|| \\
&\le |\gamma_1 - \gamma_2| \left(|x|^c + |x|^d\right) |\log(|x|)|.
\end{aligned} \tag{5.18}$$

It follows from the definition of $\gamma^{(s)}$ that, for any $\gamma \in [0, 1]$,

$$\begin{aligned}
|Y_{2n}(\gamma) - Y_{2n}(\gamma^{(k_n)})| &\le \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |u_k| |x_{nk}|^{\gamma} - |x_{nk}|^{\gamma^{(k_n)}}| \left|\log^{m_2} |x_{nk}|\right| \\
&\le \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |u_k| |\gamma - \gamma^{(k_n)}|\ \max\{|x_{nk}|^{\gamma}, |x_{nk}|^{\gamma^{(k_n)}}\} |\log^{1+m_2} |x_{nk}|| \\
&\le \frac{2^{-k_n}}{\sqrt{n}} \sum_{k=1}^{n} |u_k| (|x_{nk}| + 1) \left|\log^{1+m_1} |x_{nk}|\right|.
\end{aligned} \tag{5.19}$$

Let $f_N(x) = (|x|+1)|\log^{1+m_2}|x||I(|x| \le N)$. By virtue of (5.19) and using the condition **A2**, we have

$$\begin{aligned} E\Big(I_{1n}I\Big(\max_{1\le k\le n}|x_{nk}| \le N\Big)\Big) &\le \frac{2^{-k_n}}{\sqrt{n}}\sum_{k=1}^{n} E\big(f_N(x_{nk})|u_k|\big) \\ &\le \frac{2^{-k_n}}{\sqrt{n}}\sum_{k=1}^{n} E\big(f_N(x_{nk})E(|u_k||\mathcal{F}_{k-1})\big) \\ &\le \frac{\sqrt{C}\,2^{-k_n}}{\sqrt{n}}\sum_{k=1}^{n} d_n/d_k \int f_N(x)dx \le C_N\sqrt{n}2^{-k_n} \to 0, \end{aligned}$$

as $n \to \infty$, where C_N is a constant depending only on N and we have used the fact that $\sup_k E(u_k^2|\mathcal{F}_{k-1}) \le C < \infty$ and $d_n^2 \backsimeq n^\mu$ for some $0 < \mu < 2$. This yields (5.16).

We next prove (5.17). Let $f_{N,j,s}(x) = (|x|^{j2^{-s}} - |x|^{(j-1)2^{-s}})|\,|\log^{m_2}|x|\,|I(|x| \le N)$, $1 \le j \le 2^s$. It follows from (5.18) that

$$|f_{N,j,s}(x)| \le 2^{-s}\max\{|x|^{j2^{-s}}, |x|^{(j-1)2^{-s}}\}\,\big|\log^{m_2+1}|x|\big|I(|x| \le N) \le 2^{-s}f_N(x).$$

Thus, by recalling that $\{u_k, \mathcal{F}_k\}_{k\ge1}$ is a martingale difference, we have

$$\begin{aligned} & E\max_{1\le j\le 2^s}|Y_{2n}(j2^s) - Y_{2n}((j-1)2^s)|I\Big(\max_{1\le k\le n}|x_{nk}| \le N\Big) \\ \le\ & n^{-1/2}E\max_{1\le j\le 2^s}\Big|\sum_{k=1}^{n} u_k f_{N,j,s}(x_{nk})\Big| \\ \le\ & 2^{s/2}n^{-1/2}\max_{1\le j\le 2^s}\Big(E\Big(\sum_{t=1}^{n} u_k f_{N,j,s}(x_{nk})\Big)^2\Big)^{1/2} \\ =\ & 2^{s/2}n^{-1/2}\max_{1\le j\le 2^s}\Big(\sum_{k=1}^{n} Eu_k^2 f_{N,j,s}^2(x_{nk})\Big)^{1/2} \\ \le\ & 2^{-s/2}n^{-1/2}\Big(\sum_{k=1}^{n} E\big[f_N^2(x_{nk})E(|u_k^2||\mathcal{F}_{k-1})\big]\Big)^{1/2} \\ \le\ & C2^{-s/2}n^{-1/2}\Big(\sum_{k=1}^{n} d_n/d_k\int f_N^2(x)dx\Big)^{1/2} \le C_N 2^{-s/2}, \end{aligned}$$

indicating that, for every $\delta > 0$ and $N > 0$, there exists an integral $k_0 > 0$ such that

$$\limsup_{n\to\infty} E\Big(I_{2n}I\Big(\max_{1\le k\le n}|x_{nk}| \le N\Big)\Big) \le C_N\sum_{s=k_0}^{k_n} 2^{-s/2} \le \frac{C_N 2^{-k_0/2}}{1-2^{-1/2}} \le \delta.$$

This proves (5.17) and thereby completes the proof of the tightness for $Y_{2n}(\gamma)$. □

5.3 Proofs of Theorems 3.3 and 3.4

Let $f(x) = |x|^\gamma \log^m |x|$ where $\gamma > 0$ and integer $m \geq 0$, and

$$f'(x) = |x|^{\gamma-1} \log^{m-1} |x| \, (\gamma \log |x| + m) \, \mathrm{sign}(x).$$

Obviously, $f'(x)$ is locally Riemann integrable for $\gamma > 1$ and $[f'(x)]^{\alpha/(\alpha-1)}$ is locally integrable for $\gamma > 1/\alpha$. These facts will be used in the proof without further indication. To prove Theorems 3.3 and 3.4, we write

$$\begin{aligned} & \frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} |x_{nk}|^\gamma \, \log^m |x_{nk}| \, w_k \\ = \; & \frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} f(x_{nk}) \, u_k + \frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} [f(x_{nk}) - f(x_{n,k-1})] \, z_{k-1} + o_P(1) \\ = \; & \frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} f(x_{nk}) \, u_k + \frac{\sqrt{n}}{d_n} \frac{1}{n} \sum_{k=1}^{n-1} z_{k-1} \xi_k f'(x_{n,k-1}) + R_n + o_P(1), \end{aligned} \tag{5.20}$$

where

$$R_n \; := \; \frac{1}{\sqrt{n}} \sum_{k=1}^{n-1} \left[f(x_{nk}) - f(x_{n,k-1}) - \frac{\xi_k}{d_n} f'(x_{n,k-1}) \right] z_{k-1}.$$

Recalling **A3**, it is easy to show that $v_k = z_{k-1}\xi_k$ satisfies **A4** and $\sup_{k \geq 1} E|v_k|^\alpha < \infty$. It follows from Lemma 5.1 that

$$\frac{1}{n} \sum_{k=1}^{n-1} z_{k-1} \xi_k f'(x_{n,k-1}) \; = \; \frac{A_0}{n} \sum_{k=1}^{n-1} f'(x_{n,k-1}) + o_P(1), \tag{5.21}$$

for all $\gamma > 1/\alpha$. We mention that, if $1/\alpha < \gamma \leq 1$, to prove (5.21) we need the fact that x_k/d_k has a density $p_k(x)$ that is uniformly bounded by a constant K for all $1 \leq k \leq n$ and $x \in \mathbb{R}$, which is imposed in **A2** (i).

Due to (5.20) and (5.21), together with Theorem 3.1 and the continuous mapping theorem, simple algebra shows that Theorem 3.3 will follow if we have

$$R_n \; = \; o_P(1), \quad \text{for all } \gamma > 1/\alpha. \tag{5.22}$$

where there is no bias term due to the fact that $\sqrt{n}/d_n \to 0$. Similarly, Theorem 3.4 will follow if we prove

$$\frac{d_n}{\sqrt{n}} R_n \; = \; o_P(1), \quad \text{for all } \gamma > 1/\alpha. \tag{5.23}$$

Indeed, for Theorem 3.4, the result comes from

$$\begin{aligned}
&\frac{d_n}{n}\sum_{k=1}^{n-1}|x_{nk}|^{\gamma}\,\log^m|x_{nk}|\,w_k\\
=\;&\frac{d_n}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}f(x_{nk})\,u_k+\frac{1}{n}\sum_{k=1}^{n-1}z_{k-1}\xi_k f'(x_{n,k-1})+\frac{d_n}{\sqrt{n}}R_n+o_P(1),
\end{aligned}$$

which is a minor modification of (5.20).

We next prove (5.23) under the conditions of Theorem 3.4, where we assume $d_n^2/n\to\sigma^2<\infty$. The proof of (5.22) is similar but simpler. To do this, we need the following lemma, which will be established in Appendix B.

Lemma 5.2. *Suppose that $\delta>0$ is a sufficiently small constant. For any x,y with $0<|x|,|y|<N$ and $|x-y|<\delta$, we have that, for any $\alpha>1$,*

$$\begin{aligned}
&\big|f(x)-f(y)-(x-y)f'(y)\big|\\
\le\;&3|x-y|\Big[|f'(|x|)|I(|x|<2\delta)+|f'(|y|)|I(|y|<2\delta)\Big]+C_{\delta,N}|x-y|^{\min\{\alpha,2\}},
\end{aligned}$$

where $C_{\delta,N}$ is a constant only depending on γ,m,α,δ and N.

We turn back to the proof of (5.23). By Lemma 5.2, on $\Omega_{N,\delta}:=\Big\{\max\limits_{1\le k\le n}|x_{nk}|\le N\Big\}\cap\Big\{\max\limits_{1\le k\le n}|x_{nk}-x_{n,k-1}|<\delta\Big\}$, we have

$$\frac{d_n}{\sqrt{n}}|R_n|\;\le\;R_{1n}+R_{2n}+R_{3n},$$

where

$$\begin{aligned}
|R_{1n}|&\le\frac{3}{n}\sum_{k=1}^{n-1}|z_{k-1}\xi_k||f'(x_{nk})|I(|x_{nk}|\le 2\delta),\\
|R_{2n}|&\le\frac{3}{n}\sum_{k=1}^{n-1}|z_{k-1}\xi_k||f'(x_{n,k-1})|I(|x_{n,k-1}|\le 2\delta),\\
|R_{3n}|&\le\frac{C_{\delta,N}}{n{d_n}^{\min\{\alpha-1,1\}}}\sum_{k=1}^{n-1}|z_{k-1}||\xi_k|^{\min\{\alpha,2\}}.
\end{aligned}$$

Since $\lim\limits_{N\to\infty}\lim\limits_{n\to\infty}P(\bar\Omega_{N,\delta})=0$ by (5.1) and (5.2), where $\bar\Omega_{N,\delta}$ denotes the complementary set of $\Omega_{N,\delta}$, (5.23) will follow if we prove that, for any fixed $N\in\mathbb{N}$ and $\zeta>0$,

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}P(|R_{in}|>\zeta,\;\Omega_{N,\delta})=0,\quad i=1,2,3. \tag{5.24}$$

Let $\eta > 0$ be small enough so that $\gamma - 2\eta > 1$ or $\gamma - 2\eta > 1/\alpha$ whenever $\gamma > 1$ or $1 \geq \gamma > 1/\alpha$, respectively. For this $\eta > 0$, there exists a constant c_0, which only depends on r and m, such that $|f'(x)| \leq c_0|x|^{\gamma-\eta-1}$ for all $0 < x < 2\delta$. Now, recalling that $\sup_{k\geq 1} E|z_{k-1}\xi_k|^\alpha < \infty$, it follows from (5.5) with $v_k = z_{k-1}\xi_k$ that, for all $\gamma > 1/\alpha$,

$$\begin{aligned} |R_{1n}| &\leq \frac{3c_0}{n}\sum_{k=1}^{n-1} |z_{k-1}\xi_k||x_{nk}|^{\gamma-\eta-1} I(|x_{nk}| \leq 2\delta) \\ &\leq \frac{3c_0(2\delta)^\eta}{n}\sum_{k=1}^{n-1} |z_{k-1}\xi_k||x_{nk}|^{\gamma-2\eta-1} = O_p(\delta^\eta), \end{aligned}$$

where we have used the fact that, when $1/\alpha < \gamma \leq 1$, **A2** (i) holds and $x^{(\gamma-2\eta-1)\alpha/(\alpha-1)}$ is locally integrable due to $\gamma - 2\eta > 1/\alpha$. This implies (5.24) for $i = 1$.

Similarly, (5.24) holds for $i = 2$. Since $\alpha > 1$ and

$$\sup_k E\big(|z_{k-1}|\,|\xi_k|^{\min\{\alpha,2\}}\big) \leq \sup_k \Big(E\big[(1+|z_{k-1}|)|\xi_k|\big]^\alpha\Big)^{\frac{\min\{\alpha,2\}}{\alpha}} < \infty,$$

it is readily seen that

$$\frac{1}{n}\sum_{k=1}^{n-1} |z_{k-1}||\xi_k|^{\min\{\alpha,2\}} = O_P(1),$$

and then (5.24) holds for $i = 3$ since $d_n^2 \simeq n^\mu$ for some $0 < \mu < 2$. Combining all these results gives (5.23). The proof of Theorem 3.3 is then complete. □

5.4 Proof of Theorem 3.5

Similar to the proof of Theorem 3.2, by virtue of Theorems 3.3 and 3.4, we only need to prove tightness, i.e., to show that, for any $\varepsilon > 0$, $\eta > 0$ and $M > A$, there exists an k_0 such that

$$\limsup_{n\to\infty} P\Big[\sup_{\gamma_1,\gamma_2\in[A,M],|\gamma_1-\gamma_2|\leq 2^{-k_0}} |Z_n(\gamma_1) - Z_n(\gamma_2)| \geq \varepsilon\Big] \leq \eta, \tag{5.25}$$

where

$$Z_n(\gamma) := \frac{a_n}{\sqrt{n}}\sum_{k=1}^{n-1} |x_{nk}|^\gamma \log^m |x_{nk}|\, w_k,$$

with $a_n = 1$ in (a) and $a_n = d_n/\sqrt{n}$ in (b).

As in (5.20), we may write

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}|x_{nk}|^{\gamma}\,\log^{m}|x_{nk}|\,w_k\\
=\;&\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}f_\gamma(x_{nk})\,u_k+\frac{\sqrt{n}}{d_n}\frac{1}{n}\sum_{k=1}^{n-1}z_{k-1}\xi_k f'_\gamma(x_{n,k-1})+S_{1n}(\gamma)+S_{2n}(\gamma)\\
=\;&\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}f_\gamma(x_{nk})\,u_k+\frac{\sqrt{n}}{d_n}\frac{A_0}{n}\sum_{k=1}^{n-1}f'_\gamma(x_{n,k-1})+S_{1n}(\gamma)+S_{2n}(\gamma)+S_{3n}(\gamma),
\end{aligned}$$

where $f_\gamma(x)=|x|^\gamma\log^m|x|$, $f'^{\gamma-1}_\gamma\log^{m-1}|x|\,(\gamma\log|x|+m)\,\mathrm{sign}(x)$,

$$\begin{aligned}
S_{1n}(\gamma)&:=\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}\left[f_\gamma(x_{nk})-f_\gamma(x_{n,k-1})-\frac{\xi_k}{d_n}f'_\gamma(x_{n,k-1})\right]z_{k-1},\\
S_{2n}(\gamma)&:=\frac{1}{\sqrt{n}}f_\gamma(x_{n,n-2})\,z_{n-2},
\end{aligned}$$

and

$$S_{3n}(\gamma):=\frac{\sqrt{n}}{d_n}\frac{1}{n}\sum_{k=1}^{n-1}(z_{k-1}\xi_k-A_0)f'_\gamma(x_{n,k-1}).$$

Following the same arguments as in the proof of Theorem 3.2, we deduce tightness of $\frac{1}{\sqrt{n}}\sum_{k=1}^{n-1}f_\gamma(x_{nk})\,u_k$ and $\frac{A_0}{n}\sum_{k=1}^{n-1}f'_\gamma(x_{n,k-1})$ on $C[A,\infty)$. Thus, to prove (5.25), it suffices to show that

$$a_n\sup_{\gamma\in[A,M]}|S_{in}(\gamma)|=o_p(1),\qquad i=1,2,3. \tag{5.26}$$

(5.26) for $i=2$ is obvious, following from

$$\sup_{\gamma\in[A,M]}|S_{2n}(\gamma)|\le\frac{1}{\sqrt{n}}(|x_{n,n-2}|^M+|x_{n,n-2}|^A)\log^m|x_{n,n-2}|\,|z_{n-2}|=o_P(1).$$

We next prove (5.26) for $i=1$. Note that

$$\sup_{\gamma\in[A,M]}|f'_\gamma(x)|\le(|x|^{M-1}+|x|^{A-1})|\log^{m-1}|x||\,(M|\log|x||+m):=g_M(x). \tag{5.27}$$

By Lemma 5.2, on $\Omega_{N,\delta}:=\Big\{\max_{1\le k\le n}|x_{nk}|\le N\Big\}\cap\Big\{\max_{1\le k\le n}|x_{nk}-x_{n,k-1}|<\delta\Big\}$, we have

$$\frac{d_n}{\sqrt{n}}|S_{1n}(\gamma)|\;\le\;R_{1n}(\gamma)+R_{2n}(\gamma)+R_{3n}(\gamma),$$

where

$$\begin{aligned}
\sup_{\gamma\in[A,M]}|R_{1n}(\gamma)| &\le \frac{3}{n}\sum_{k=1}^{n-1}|z_{k-1}\xi_k|g_M(x_{nk})I(|x_{nk}|\le 2\delta),\\
\sup_{\gamma\in[A,M]}|R_{2n}(\gamma)| &\le \frac{3}{n}\sum_{k=1}^{n-1}|z_{k-1}\xi_k|g_M(x_{n,k-1})I(|x_{n,k-1}|\le 2\delta),\\
\sup_{\gamma\in[A,M]}|R_{3n}(\gamma)| &\le \frac{C_{\delta,N}}{nd_n^{\min\{\alpha-1,1\}}}\sum_{k=1}^{n-1}|z_{k-1}||\xi_k|^{\min\{\alpha,2\}},
\end{aligned}$$

where $C_{\delta,N}$ is a constant only depending on A, M, m, α, δ and N. Now, as in the proof of (5.24), we obtain the following: for any fixed $N\in\mathbb{N}$ and $\zeta>0$,

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}P(\sup_{\gamma\in[A,M]}|R_{in}(\gamma)|>\zeta,\ \Omega_{N,\delta})=0,\quad i=1,2,3. \tag{5.28}$$

This, together with the fact that $\lim_{N\to\infty}\lim_{n\to\infty}P(\bar\Omega_{N,\delta})=0$, implies (5.26).

We finally prove (5.26) for $i=3$. First note that, for any $\varepsilon>0$, there exists a $\delta_\varepsilon\in(0,\varepsilon)$ such that $\sup_{\gamma\in[A,M]}|f'_\gamma(x)-f'_\gamma(y)|<\varepsilon$ holds for any $\varepsilon\le|x|,|y|\le N$ with $|x-y|<\delta_\varepsilon$. Thus, if $|x|,|y|\le N$ and $|x-y|<\delta_\varepsilon<\varepsilon$, then

$$\begin{aligned}
&|f'_\gamma(x)I(\varepsilon\le|x|\le N)-f'_\gamma(y)I(\varepsilon\le|y|\le N)|\\
\le\ &|f'_\gamma(x)-f'_\gamma(y)|I(\varepsilon\le|y|\le N)+|f'_\gamma(x)||I(\varepsilon\le|x|\le N)-I(\varepsilon\le|y|\le N)|\\
\le\ &\varepsilon+2|f'_\gamma(x)|I(|x|<2\varepsilon).
\end{aligned}$$

Write $\tilde\lambda_k=z_{k-1}\xi_k-A_0$, $T_n=[n/l]-1$ and

$$\tilde\Omega_{N,\varepsilon}=\{\max_{1\le k\le n}|x_{nk}|\le N\}\cap\{\max_{0\le j\le n/l}\max_{jl\le k\le(j+1)l}|x_{nk}-x_{n,jl}|\le\delta_\epsilon\},$$

where l is chosen so that $l\to\infty$ and $n/l\to\infty$. On $\tilde\Omega_{N,\varepsilon}$, it is readily seen that

$$\begin{aligned}
&\sup_{\gamma\in[A,M]}\frac{1}{n}\Big|\sum_{k=1}^{n-1}(z_{k-1}\xi_k-A_0)f'_\gamma(x_{n,k-1})\Big|\\
\le\ &\sup_{\gamma\in[A,M]}\frac{1}{n}\sum_{k=1}^{n-1}|\tilde\lambda_k||f'_\gamma(x_{n,k-1})|I(|x_{n,k-1}|<\varepsilon)\\
&+\frac{1}{n}\sup_{\gamma\in[A,M]}\Big|\sum_{k=1}^{n-1}\tilde\lambda_k f'_\gamma(x_{n,k-1})I(\varepsilon\le|x_{n,k-1}|\le N)\Big|\\
=:\ &T_{1n}(\varepsilon)+T_{2n}(\varepsilon),
\end{aligned}\tag{5.29}$$

and similar arguments to those in the proof of Lemma 5.1 [see (5.7) there] yields

$$\begin{aligned} T_{2n}(\varepsilon) &\leq \frac{C_{N,\varepsilon}}{n}\sum_{j=0}^{T_n}\Big|\sum_{k=jl+1}^{(j+1)l}\widetilde{\lambda}_k\Big| + \frac{C_{N,\varepsilon}}{n}\sum_{k=lT_n+1}^{n}|\widetilde{\lambda}_k| + \frac{\epsilon}{n}\sum_{k=1}^{n}|\widetilde{\lambda}_k| \\ &\quad + \frac{2}{n}\sup_{\gamma\in[A,M]}\sum_{k=1}^{n-1}|\tilde{\lambda}_k||f'_\gamma(x_{n,k-1})|I(|x_{n,k-1}|<2\varepsilon) \\ &= T_{3n}(\epsilon) + 2T_{1n}(2\epsilon), \quad \text{say,} \end{aligned} \tag{5.30}$$

where $C_{N,\varepsilon}$ is a constant only depending on M, m, ε and N.

Recalling (5.27), we have

$$T_{1n}(2\epsilon) \leq \frac{1}{n}\sum_{k=1}^{n-1}|\tilde{\lambda}_k|g_M(x_{n,k-1})I(|x_{n,k-1}|<2\varepsilon)$$

and, as in (5.28),

$$\limsup_{\delta\to 0}\limsup_{n\to\infty}P(T_{1n}(2\epsilon)>\zeta,\ \tilde{\Omega}_{N,\varepsilon}) = 0, \tag{5.31}$$

for any fixed $N \in \mathbb{N}$ and $\zeta > 0$. On the other hand, by recalling **A3**, we find that $v_k = z_{k-1}\xi_k$ satisfies **A4** and $\sup_{k\geq 1}E|v_k|^\alpha < \infty$, implying that, for any fixed $N \in \mathbb{N}$ and $\zeta > 0$,

$$\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}P(|T_{3n}(\varepsilon)|>\zeta,\ \tilde{\Omega}_{N,\varepsilon}) = 0. \tag{5.32}$$

Combining (5.29) to (5.32), for any fixed $N \in \mathbb{N}$ and $\zeta > 0$, we have

$$\limsup_{\varepsilon\to 0}\limsup_{n\to\infty}P(a_n\sup_{\gamma\in[A,M]}|S_{3n}(\gamma)|\geq\zeta,\ \tilde{\Omega}_{N,\varepsilon}) = 0.$$

This implies (5.26) for $i = 3$, since, for any $\epsilon > 0$,

$$P\big(\bar{\tilde{\Omega}}_{N,\varepsilon}\big) \leq P(\max_{1\leq k\leq n}|x_{nk}|>N\}) + P\{\max_{0\leq j\leq n/l}\max_{jl\leq k\leq(j+1)l}|x_{nk}-x_{n,jl}|>\delta_\epsilon\} \to 0,$$

as $N \to \infty$ and $n \to \infty$ by (5.3), where $\bar{\tilde{\Omega}}_{N,\varepsilon}$ denotes the complementary set of $\tilde{\Omega}_{N,\varepsilon}$. The proof of Theorem 3.5 is now completed. □

5.5 Proofs of Theorems 2.1 - 2.4

We prove Theorems 2.1 - 2.4 by verifying the conditions of Theorem A.1 in Appendix A with $g_k(\theta) = \beta|x_k|^\gamma, \theta = (\beta,\gamma)$, and $F_n = diag\big[\sqrt{n}d_n^{\gamma_0}/\log d_n, \sqrt{n}d_n^{\gamma_0}\big]$. Theorems 3.1 - 3.4 are highly involved in providing necessary technical support in the derivation.

We start with some preliminaries. Let $\dot{g}_k(\theta) = \frac{\partial g_k(\theta)}{\partial \theta}$ and $\ddot{g}_k(\theta) = \frac{\partial^2 g_k(\theta)}{\partial \theta \partial \theta'}$, and set $H_n(\theta) = \sum_{k=1}^n \dot{g}_k(\theta)\dot{g}_k(\theta)$ and

$$U_{nm} = \frac{1}{n}\sum_{k=1}^{n} |x_{nk}|^{2\gamma_0} \log^m |x_{nk}|, \quad m = 0, 1, 2,$$

where $x_{nk} = x_k/d_n$. Since $F_n^{-1} = diag\big[\big(\sqrt{n}d_n^{\gamma_0}\big)^{-1} \log d_n, \big(\sqrt{n}d_n^{\gamma_0}\big)^{-1}\big]$ and

$$\begin{aligned}
\dot{g}_k(\theta)\dot{g}_k(\theta)' &= \begin{pmatrix} |x_k|^{2\gamma} & \beta|x_k|^{2\gamma}\log|x_k| \\ \beta|x_k|^{2\gamma}\log|x_k| & \beta^2|x_k|^{2\gamma}\log^2|x_k| \end{pmatrix} \\
&= \begin{pmatrix} 1 & \beta\log d_n \\ \log d_n & \beta^2\log^2 d_n \end{pmatrix} |x_k|^{2\gamma} \\
&\quad + \begin{pmatrix} 0 & \beta \\ \beta & 2\beta^2\log d_n \end{pmatrix} |x_k|^{2\gamma}\log\frac{|x_k|}{d_n} + \begin{pmatrix} 0 & 0 \\ 0 & \beta^2 \end{pmatrix} |x_k|^{2\gamma}\log^2\frac{|x_k|}{d_n},
\end{aligned}$$

we may write

$$\begin{aligned}
F_n^{-1}H_n(\theta_0)F_n^{-1} &= F_n^{-1}\sum_{k=1}^{n}\dot{g}_k(\theta_0)\dot{g}_k(\theta_0)'F_n^{-1} \\
&= \begin{pmatrix} 1 & \beta_0 \\ \beta_0 & \beta_0^2 \end{pmatrix} U_{n0}\log^2 d_n + \begin{pmatrix} 0 & \beta_0 \\ \beta_0 & 2\beta_0^2 \end{pmatrix} U_{n1}\log d_n + \begin{pmatrix} 0 & 0 \\ 0 & \beta_0^2 \end{pmatrix} U_{n2} \\
&=: \begin{pmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{pmatrix},
\end{aligned}$$

where

$$\begin{aligned}
&H_{11} = U_{n0}\log^2 d_n, \quad H_{12} = \beta_0\big(U_{n0}\log^2 d_n + U_{n1}\log d_n\big), \\
&H_{22} = \beta_0^2(U_{n0}\log^2 d_n + U_{n1}\log d_n + U_{n2}).
\end{aligned}$$

It is easy to show that

$$\det|F_n^{-1}H_n(\theta_0)F_n^{-1}| = H_{11}H_{22} - H_{12}^2 = \beta_0^2\big(U_{n0}U_{n2} - U_{n1}^2\big)\log^2 d_n > 0 \ \ a.s.,$$

and

$$\begin{aligned}
(F_n^{-1}H_n(\theta_0)F_n^{-1})^{-1} &= \frac{1}{H_{11}H_{22} - H_{12}}\begin{pmatrix} H_{22} & -H_{12} \\ -H_{12} & H_{11} \end{pmatrix} \\
&= \frac{U_{n0}}{\beta_0^2\big(U_{n0}U_{n2} - U_{n1}^2\big)}\begin{pmatrix} \beta_0^2 & -\beta_0 \\ -\beta_0 & 1 \end{pmatrix} + \frac{U_{n1}}{\beta_0^2\big(U_{n0}U_{n2} - U_{n1}^2\big)\log d_n}\begin{pmatrix} 2\beta_0^2 & -\beta_0 \\ -\beta_0 & 0 \end{pmatrix} \\
&\quad + \frac{U_{n2}}{\beta_0^2\big(U_{n0}U_{n2} - U_{n1}^2\big)\log^2 d_n}\begin{pmatrix} \beta_0^2 & 0 \\ 0 & 0 \end{pmatrix}. \qquad (5.33)
\end{aligned}$$

Furthermore, under one of the conditions imposed in Theorems 2.1 - 2.4, we have

$$\begin{aligned}
\lambda_{min}(F_n^{-1}H_n(\theta_0)F_n^{-1}) &= \frac{H_{11}+H_{22}-\sqrt{(H_{11}+H_{22})^2-4(H_{11}H_{22}-H_{12}^2)}}{2} \\
&= \frac{4(H_{11}H_{22}-H_{12}^2)}{2(H_{11}+H_{22})+2\sqrt{(H_{11}+H_{22})^2-4(H_{11}H_{22}-H_{12}^2)}} \\
&= \frac{H_{11}H_{22}-H_{12}^2}{H_{11}+H_{22}}+o_p(1) \\
&= \frac{\beta_0^2(U_{n0}U_{n2}-U_{n1}^2)}{(\beta_0^2+1)U_{n0}}+o_p(1) \\
&\to_D \frac{\beta_0^2(U_0U_2-U_1^2)}{(\beta_0^2+1)U_0}>0, \quad \text{for } \gamma_0>-1/2, \qquad (5.34)
\end{aligned}$$

by using Corollary 3.1 and the continuous mapping theorem, where $U_i=\int_0^1|X_t|^{2\gamma_0}\log^i|X_t|dt$ defined as in Theorem 2.1. We recall that **A2** (i) is required to establish (5.34) only for $-1/2<\gamma_0\le 0$.

After these preliminaries, we are now ready to prove Theorems 2.1 - 2.4. For convenience of reading, we adopt the same notation used in Theorem A.1, namely, we let $Q_n(\theta)=\sum_{k=1}^n(y_k-g_k(\theta))^2$, $S_n(\theta)=\frac{\partial Q_n(\theta)}{\partial\theta}$, $W_n(\theta)=\frac{\partial^2 Q_n(\theta)}{\partial\theta\partial\theta'}$, $Z_n=F_n^{-1}S_n(\theta_0)$ and $Y_n=F_n^{-1}H_n(\theta_0)F_n^{-1}$.

Proofs of Theorems 2.1 and 2.2.

Since $\lambda_{\min}^{-1}(Y_n)=O_p(1)$ by (5.34), by using Theorem A.1, Theorems 2.1 and 2.1 will follow if we prove that, for $\gamma_0>-1/2$ (**A2** is required only for $0\ge\gamma_0>-1/2$),

$$Y_n^{-1}Z_n\to_D\begin{pmatrix}1\\-1/\beta_0\end{pmatrix}\frac{U_0\int_0^1|X_t|^{\gamma_0}\log|X_t|\,dB_t-U_1\int_0^1|X_t|^{\gamma_0}dB_t}{U_0U_2-U_1^2}, \qquad (5.35)$$

where $U_i=\int_0^1|X_t|^{2\gamma_0}\log^i|X_t|dt$, $i=0,1,2$, and

$$\sup_{\theta:||F_n(\theta-\theta_0)||\le\log d_n}||F_n^{-1}\big[W_n(\theta)-H_n(\theta_0)\big]F_n^{-1}||=o_P(\log^{-2}d_n). \qquad (5.36)$$

The proof of (5.35) follows from an application of Corollary 3.1 and the continuous mapping theorem. Indeed, under model (2.1), we have

$$S_n(\theta) = -\sum_{k=1}^n\dot g_k(\theta)\,u_k+\sum_{k=1}^n\dot g_k(\theta)\,d_k(\theta),$$

where $d_k(\theta)=g_k(\theta)-g_k(\theta_0)$. Hence, some simple algebra shows that

$$\begin{aligned}
F_n^{-1}S_n(\theta_0) &= -\log d_n\begin{pmatrix}\frac{1}{\sqrt n}\sum_{k=1}^n|x_{nk}|^{\gamma_0}u_k\\ \frac{\beta_0}{\sqrt n}\sum_{k=1}^n|x_{nk}|^{\gamma_0}u_k+\frac{\beta_0}{\sqrt n\log d_n}\sum_{k=1}^n|x_{nk}|^{\gamma_0}\log|x_{nk}|\,u_k\end{pmatrix} \\
&= -\log d_n\begin{pmatrix}1\\ \beta_0\end{pmatrix}\frac{1}{\sqrt n}\sum_{k=1}^n|x_{nk}|^{\gamma_0}u_k-\begin{pmatrix}0\\ \beta_0\end{pmatrix}\frac{1}{\sqrt n}\sum_{k=1}^n|x_{nk}|^{\gamma_0}\log|x_{nk}|\,u_k.
\end{aligned}$$

This, together with (5.33), yields that

$$\begin{aligned}
Y_n^{-1} Z_n &= (F_n^{-1} H_n(\theta_0) F_n^{-1})^{-1} F_n^{-1} S_n(\theta_0) \\
&= -\frac{U_{n0}}{\beta_0^2 \big(U_{n0}\, U_{n2} - U_{n1}^2\big)} \begin{pmatrix} \beta_0^2 & -\beta_0 \\ -\beta_0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \beta_0 \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |x_{nk}|^{\gamma_0} \log |x_{nk}| u_k \\
&\quad -\frac{U_{n1}}{\beta_0^2 \big(U_{n0}\, U_{n2} - U_{n1}^2\big)} \begin{pmatrix} 2\beta_0^2 & -\beta_0 \\ -\beta_0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ \beta_0 \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |x_{nk}|^{\gamma_0} u_k + O_P(\log^{-1} d_n) \\
&= \frac{U_{n0}}{U_{n0}\, U_{n2} - U_{n1}^2} \begin{pmatrix} 1 \\ -1/\beta_0 \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |x_{nk}|^{\gamma_0} \log |x_{nk}| u_k \\
&\quad -\frac{U_{n1}}{U_{n0}\, U_{n2} - U_{n1}^2} \begin{pmatrix} 1 \\ -1/\beta_0 \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |x_{nk}|^{\gamma_0} u_k + O_P(\log^{-1} d_n), \qquad (5.37)
\end{aligned}$$

implying (5.35) by Corollary 3.1 and the continuous mapping theorem.

Note that $W_n(\theta) = \sum_{k=1}^n \dot{g}_k(\theta)\dot{g}_k(\theta)' + \sum_{k=1}^n \ddot{g}_k(\theta)\big[d_k(\theta) - u_k\big]$, where $d_k(\theta) = g_k(\theta) - g_k(\theta_0)$. The proof of (5.36) follows by verification of the following facts: for $\gamma_0 > -1/2$ (**A2** is required for $0 \geq \gamma_0 > -1/2$),

$$\sup_{\theta: \|F_n(\theta-\theta_0)\| \leq \log d_n} \| F_n^{-1} \sum_{k=1}^{n} \big[\dot{g}_k(\theta)\dot{g}_k(\theta)' - \dot{g}_k(\theta_0)\dot{g}_k(\theta_0)'\big] F_n^{-1} \| = o_P(\log^{-2} d_n), \quad (5.38)$$

$$\sup_{\theta: \|F_n(\theta-\theta_0)\| \leq \log d_n} \| F_n^{-1} \sum_{k=1}^{n} \ddot{g}_k(\theta) \big[g_k(\theta) - g_k(\theta_0)\big] F_n^{-1} \| = o_P(\log^{-2} d_n), \quad (5.39)$$

$$\sup_{\theta: \|F_n(\theta-\theta_0)\| \leq \log d_n} \| F_n^{-1} \sum_{k=1}^{n} \ddot{g}_k(\theta)\, u_k\, F_n^{-1} \| = o_P(\log^{-2} d_n). \quad (5.40)$$

Let $(A)_{ij}$ be the (i,j) entry of the matrix A. To prove (5.38), it is sufficient to prove that

$$\sup_{\theta: \|F_n(\theta-\theta_0)\| \leq \log d_n} \Big(F_n^{-1} \sum_{k=1}^{n} \big[\dot{g}_k(\theta)\dot{g}_k(\theta)' - \dot{g}_k(\theta_0)\dot{g}_k(\theta_0)'\big] F_n^{-1}\Big)_{ij} = o_P(\log^{-2} d_n) \quad (5.41)$$

for all $i, j = 1, 2$. Here we only prove the case $i = j = 2$ since the other cases are similar.

Let $\varepsilon > 0$ be a constant satisfying that $\gamma_0 - \varepsilon > 0$ if $\gamma_0 > 0$ or $\gamma_0 - \varepsilon > -1/2$ if $-1/2 < \gamma_0 \leq 0$. For any $|\gamma - \gamma_0| < \varepsilon/3$, we have

$$(\log |x_k|)^2 ||x_k|^{2\gamma} - |x_k|^{2\gamma_0}| \leq C_\varepsilon |\gamma - \gamma_0| (|x_k|^{2\gamma_0+\varepsilon} + |x_k|^{2\gamma_0-\varepsilon}),$$

where C_ε is a constant only depending on ε. By using Corollary 3.1,

$$\frac{1}{n d_n^{2\gamma_0+\varepsilon}} \sum_{k=1}^{n} |x_k|^{2\gamma_0+\varepsilon} = O_p(1), \qquad \frac{1}{n d_n^{2\gamma_0-\varepsilon}} \sum_{k=1}^{n} |x_k|^{2\gamma_0-\varepsilon} = O_p(1).$$

Hence, for sufficiently large n, we have

$$\begin{aligned}
&\sup_{\theta:\|F_n(\theta-\theta_0)\|\le\log d_n}\Big|\Big(F_n^{-1}\sum_{k=1}^n\big[\dot g_k(\theta)\dot g_k(\theta)'-\dot g_k(\theta_0)\dot g_k(\theta_0)'\big]\,F_n^{-1}\Big)_{22}\Big|\\
=\;& n^{-1}d_n^{-2\gamma_0}\sup_{\theta:\|F_n(\theta-\theta_0)\|\le\log d_n}\Big|\sum_{k=1}^n(\log|x_k|)^2(\beta^2|x_k|^{2\gamma}-\beta_0^2|x_k|^{2\gamma_0})\Big|\\
\le\;& n^{-1}d_n^{-2\gamma_0}\sup_{\theta:\|F_n(\theta-\theta_0)\|\le\log d_n}\sum_{k=1}^n(\log|x_k|)^2\big(\beta^2||x_k|^{2\gamma}-|x_k|^{2\gamma_0}|+|\beta^2-\beta_0^2||x_k|^{2\gamma_0}\big)\\
=\;& O_P(1)\times\, n^{-3/2}d_n^{-3\gamma_0}\log^2 d_n\Big(\sum_{k=1}^n|x_k|^{2\gamma_0-\varepsilon}+\sum_{k=1}^n|x_k|^{2\gamma_0+\varepsilon}\Big)\\
=\;& O_p(n^{-1/2}d_n^{-\gamma_0+\varepsilon}\log^2 d_n).
\end{aligned}$$

This proves (5.41) for $i=j=2$ by noting that $d_n^2\simeq n^\mu$ for some $0<\mu<2$. The proof of (5.39) is similar and details are omitted.

As for (5.40), by recalling

$$\ddot g_k(\theta)\;:=\;\frac{\partial^2 g_k(\theta)}{\partial\theta\partial\theta'}=\begin{pmatrix}0 & |x_k|^\gamma\log|x_k|\\ |x_k|^\gamma\log|x_k| & \beta^2|x_k|^\gamma\log^2|x_k|\end{pmatrix},$$

we only need to prove that, for $m=1,2$,

$$\frac{\log^2 d_n}{nd_n^{2\gamma_0}}\sup_{|\gamma-\gamma_0|\le n^{-1}d_n^{-\gamma_0}\log^2 d_n}\Big|\sum_{k=1}^n|x_k|^\gamma\log^m|x_k|\,u_k\Big|=o_P(\log^{-2}d_n).$$

By using Corollary 3.1, we have

$$\frac{\log^2 d_n}{nd_n^{2\gamma_0}}\Big|\sum_{k=1}^n|x_k|^{\gamma_0}\log^m|x_k|\,u_k\Big|=o_P(\log^{-2}d_n),\qquad m=1,2.$$

Hence, it is sufficient to show that

$$\frac{\log^2 d_n}{nd_n^{2\gamma_0}}\sup_{|\gamma-\gamma_0|\le n^{-1}d_n^{-\gamma_0}\log^2 d_n}\Big|\sum_{k=1}^n(|x_k|^\gamma-|x_k|^{\gamma_0})\log^m|x_k|\,u_k\Big|=o_P(\log^{-2}d_n).$$

Similar to the proof of (5.38), for sufficiently large n, it follows from (5.5) that

$$\begin{aligned}
&\frac{\log^2 d_n}{nd_n^{2\gamma_0}}\sup_{|\gamma-\gamma_0|\le n^{-1}d_n^{-\gamma_0}\log^2 d_n}\Big|\sum_{k=1}^n(|x_k|^\gamma-|x_k|^{\gamma_0})\log^m|x_k|\,u_k\Big|\\
\le\;&\frac{C'_\varepsilon\log^4 d_n}{n^2d_n^{3\gamma_0}}\sum_{k=1}^n(|x_k|^{\gamma_0+\varepsilon}+|x_k|^{\gamma_0-\varepsilon})\,|u_k|\\
=\;& o_P(\log^{-2}d_n),
\end{aligned}$$

where C'_ε is a constant only depending on ε. The proofs of Theorems 2.1 and 2.2 are complete. □

Proofs of Theorems 2.3 and 2.4.

The argument is the same as that of Theorem 2.1. To illustrate, we consider an outline of the proof of Theorem 2.4. The proof of Theorem 2.3 is the same except that we replace u_k by $w_k = u_k + z_{k-1} - z_k$ and Corollary 3.1 by Theorem 3.3, rather than Theorem 3.4.

First note that, following the proof of (5.40) but replacing Corollary 3.1 by Theorem 3.4, we have

$$\sup_{\theta:\|F_n(\theta-\theta_0)\|\le \log d_n} \| F_n^{-1} \sum_{k=1}^{n} \ddot{g}_k(\theta)\, w_k\, F_n^{-1} \| \;=\; o_P(\log^{-2} d_n).$$

Hence, using the same argument and notation as those of Theorem 2.1, it follows that

$$\begin{aligned} F_n(\hat{\theta}_n - \theta_0) &= -\, Y_n^{-1}\, Z_n + o_P(1) \\ &= -\frac{U_{n0}}{U_{n0}\, U_{n2} - U_{n1}^2} \begin{pmatrix} 1 \\ -1/\beta_0 \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |x_{nk}|^{\gamma_0}\, \log|x_{nk}| w_k \\ &\quad + \frac{U_{n1}}{U_{n0}\, U_{n2} - U_{n1}^2} \begin{pmatrix} 1 \\ -1/\beta_0 \end{pmatrix} \frac{1}{\sqrt{n}} \sum_{k=1}^{n} |x_{nk}|^{\gamma_0}\, w_k + o_P(1). \end{aligned}$$

Now, by using Theorem 3.4 again and the continuous mapping theorem, we have

$$\begin{aligned} \frac{d_n}{\sqrt{n}} F_n(\hat{\theta}_n - \theta_0) &= -\frac{U_{n0}}{U_{n0}\, U_{n2} - U_{n1}^2} \begin{pmatrix} 1 \\ -1/\beta_0 \end{pmatrix} \frac{d_n}{n} \sum_{k=1}^{n} |x_{nk}|^{\gamma_0}\, \log|x_{nk}| w_k \\ &\quad + \frac{U_{n1}}{U_{n0}\, U_{n2} - U_{n1}^2} \begin{pmatrix} 1 \\ -1/\beta_0 \end{pmatrix} \frac{d_n}{n} \sum_{k=1}^{n} |x_{nk}|^{\gamma_0}\, w_k + o_P(1) \\ &\to_D \begin{pmatrix} 1 \\ -1/\beta_0 \end{pmatrix} \frac{U_0\, W_1 \; - U_1\, W_0}{U_1^2 - U_0\, U_2}, \end{aligned}$$

as required. □

5.6 Proof of Theorem 2.5

It is readily seen that

$$\begin{aligned} \hat{\beta} &= \frac{\sum_{k=1}^{n} y_k \big[l(|x_k|) - n^{-1} \sum_{j=1}^{n} l(|x_j|) \big]}{\sum_{k=1}^{n} \big[l(|x_k|) - n^{-1} \sum_{j=1}^{n} l(|x_j|) \big]^2} \\ &= \beta_0 + \frac{\sum_{k=1}^{n} u_k \big[l(|x_k|) - n^{-1} \sum_{j=1}^{n} l(|x_j|) \big]}{\sum_{k=1}^{n} \big[l(|x_k|) - n^{-1} \sum_{j=1}^{n} l(|x_j|) \big]^2}, \qquad (5.42) \\ \hat{\alpha} &= \frac{1}{n} \sum_{k=1}^{n} y_k - \frac{\hat{\beta}}{n} \sum_{k=1}^{n} l(|x_k|) \\ &= \alpha_0 + \frac{1}{n} \sum_{k=1}^{n} u_k - \frac{\hat{\beta} - \beta_0}{n} \sum_{k=1}^{n} l(|x_k|). \qquad (5.43) \end{aligned}$$

Let $l^*_\lambda(x) = \frac{l(|x|\lambda)}{l(\lambda)} - 1 - \epsilon(\lambda)\log|x|$, $a_{nk} = \log|x_{nk}| - n^{-1}\sum_{k=1}^n \log(|x_{nk}|)$, and

$$b_{nk} = l^*_\lambda(|x_{nk}|) - n^{-1}\sum_{k=1}^n l^*_\lambda(|x_{nk}|),$$

where $x_{nk} = x_k/d_n$. We may write

$$\frac{1}{l^2(d_n)}\sum_{k=1}^n \Big[l(|x_k|) - n^{-1}\sum_{j=1}^n l(|x_j|)\Big]^2 \;=\; \epsilon^2(d_n)\sum_{k=1}^n a_{nk}^2 + \sum_{k=1}^n b_{nk}^2 + 2\epsilon(d_n)\sum_{k=1}^n a_{nk}b_{nk},$$

and

$$\frac{1}{l(d_n)}\sum_{k=1}^n u_k\Big[l(|x_k|) - n^{-1}\sum_{j=1}^n l(|x_j|)\Big] \;=\; \epsilon(d_n)\sum_{k=1}^n u_k a_{nk} + \sum_{k=1}^n u_k b_{nk}.$$

Taking these facts into (5.42) and (5.43), we obtain

$$\begin{aligned}
\sqrt{n}l(d_n)\epsilon(d_n)\big(\hat\beta_n - \beta_0\big) &= \frac{\frac{1}{\sqrt{n}}\sum_{k=1}^n u_k a_{nk} + \frac{1}{\sqrt{n}\,\epsilon(d_n)}\sum_{k=1}^n u_k b_{nk}}{\frac{1}{n}\sum_{k=1}^n a_{nk}^2 + \frac{1}{n\,\epsilon^2(d_n)}\sum_{k=1}^n b_{nk}^2 + \frac{2}{n\,\epsilon(d_n)}\sum_{k=1}^n a_{nk}b_{nk}},\\
\sqrt{n}\epsilon(d_n)\big(\hat\alpha_n - \alpha_0\big) &= \frac{\epsilon(d_n)}{\sqrt{n}}\sum_{k=1}^n u_k - \sqrt{n}l(d_n)\epsilon(d_n)\big(\hat\beta - \beta_0\big)\frac{1}{nl(d_n)}\sum_{k=1}^n l(|x_k|).
\end{aligned}$$

Since $\epsilon(d_n)\to 0$ and $\log^2|x|$ is locally integrable, by using Theorem 3.1 and the continuous mapping theorem, Theorem 2.5 will follow if we prove

$$\frac{1}{n\,\epsilon^2(d_n)}\sum_{k=1}^n b_{nk}^2 \;=\; o_P(1), \tag{5.44}$$

$$\frac{1}{\sqrt{n}\,\epsilon(d_n)}\sum_{k=1}^n u_k b_{nk} \;=\; o_P(1), \tag{5.45}$$

$$\frac{1}{nl(d_n)}\sum_{k=1}^n l(|x_k|) \;=\; 1 + o_P(1). \tag{5.46}$$

The idea to prove (5.44)-(5.46) is quite similar to that of Theorem 3.1. We only provide an outline for (5.45). In fact, by letting $b^*_{nk} = b_{nk}I(|x_{nk}|\le N)$, we have

$$P\Big(\frac{1}{\sqrt{n}\,\epsilon(d_n)}\sum_{k=1}^n u_k b_{nk} \ne \frac{1}{\sqrt{n}\,\epsilon(d_n)}\sum_{k=1}^n u_k b^*_{nk}\Big) \le P\big(\max_{1\le k\le n}|x_{nk}|\ge N\big)\to 0,$$

as $n\to\infty$ first and then $N\to\infty$. This result, together with the fact that due to (2.15) and **A2** (iii)

$$\frac{1}{n\,\epsilon^2(d_n)}E\Big(\sum_{k=1}^n u_k b^*_{nk}\Big)^2 \le C\,\frac{1}{n\,\epsilon^2(d_n)}\sum_{k=1}^n Eb^{*2}_{nk} = o(1), \quad \text{as } n\to\infty,$$

for any fixed $N\ge 1$, implies (5.45). This completes the proof of Theorem 2.5. □

REFERENCES

Andrews, D. W. K. and Cheng X. (2012). Estimation and inference with weak, semistrong, and strong identification. *Econometrica*, **80**, 2153-2211.

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

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

Baek, Y. In., Cho, H. S. and Phillips P. C. B. (2015). Testing linearity using power transforms of regressors. *Journal of econometrics*, 187, 376-384.

Berkes, I. and Horváth, L. (2006). Convergence of integral functionals of stochastic processes. *Econometric Theory*, **22**, 304-322.

Billingsley, P. (1968) *Convergence of probability measures*. Wiley.

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

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

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

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

Cho, J. S., and Phillips, P. C. B. (2018) Sequentially testing polynomial model hypotheses using power transforms of regressors. *Journal of Applied Econometrics* 33, 141-159.

Ethier, S. N. and Kurtz, T. G. (1986). *Markov Processes: Characterization and Convergence.* Wiley, New York.

Gao, J. and Phillips, P. C. B. (2013). Semiparametric estimation in triangular system equations with nonstationarity. *Journal of Econometrics,* 59-79.

de Jong, R. (2002). Nonlinear regression with integrated regressors but without Exogeneity,. Mimeograph, Department of Economics, Michigan State University.

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

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

Liang, H. Phillips, P. C. B., Wang, H. and Wang, Q. (2016). Weak convergence to stochastic integrals for econometric applications.*Econometric Theory*, **32**, 1349-1375.

Park, J. Y. and Phillips, P. C. B. (2000). Nonstationary binary choice. *Econometrica*, **68**, 1249-1280.

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

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

Phillips, P. C. B. (2007). Regression with slowly varying regressors and nonlinear trends. *Econometric Theory*, **23**, 557-614.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *Ann. Statist.* **20**, 971-1001.

Pötscher, B. M. (2004). Nonlinear functions and convergence to Brownian motion: beyond the continuous mapping theorem. *Econometric Theory*, **20**, 1–22.

Shi, X. and Phillips, P. C. B. (2012). Nonlinear Cointegrating regression under weak identification. *Econometric Theory*, **28**, 509-547.

Stein, E. M. and Shakarchi, R. (2005). *Real Analysis, Measure Theory, Integration and Hilbert Spaces.* Princeton University Press.

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

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

Wang, Q. (2018). Least squares estimation for nonlinear regression models with heteroscedasticity, *Econometric Theory*, under revision.

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

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

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

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

143–164.

Wooldridge, J. M. (1994). Estimation and Inference for Dependent Processes. *Handbook of Econometrics*, Vol. IV, ed. by Engle, R. F. and McFadden, D. L. Amsterdam: North Holland.

A A general framework for nonlinear least squares estimation

We consider the general nonlinear parametric regression model

$$y_k = g_k(\theta) + u_k, \tag{A.1}$$

where $\theta \in \Theta$, Θ is a subset of $\mathbb{R}^m$, $g_k(\theta)$ is a sequence of measurable random functions on Θ and u_k is a sequence of error variables. This section considers extremum estimation of the unknown parameters θ in model (A.1) by nonlinear least squares (NLS). The approach taken here is similar to that used in Park and Phillips (2000, 2001) in the development of nonlinear nonstationary regression, which in turn utilizes the framework of Wooldridge (1994).

Let $Q_n(\theta) = \sum_{k=1}^n \left[y_k - g_k(\theta)\right]^2$. The NLS estimator $\hat{\theta}_n$ of θ is defined as the extremum estimator that minimizes $Q_n(\theta)$ over $\theta \in \Theta$, viz.,

$$\hat{\theta}_n = \arg\min_{\theta\in\Theta} Q_n(\theta).$$

Let $S_n(\theta) = (1/2)\partial Q_n(\theta)/\partial\theta$, $W_n(\theta) = (1/2)\partial^2 Q_n(\theta)/\partial\theta\partial\theta'$ and

$$H_n(\theta) \;=\; \sum_{k=1}^n \dot{g}_k(\theta)\dot{g}_k(\theta)', \quad \text{where } \dot{g}_k(\theta) = \partial g_k(\theta)/\partial\theta.$$

For later use, we define $\ddot{g}_k(\theta) = \partial^2 g_k(\theta)/\partial\theta\partial\theta'$ and assume that these quantities exist whenever they are introduced.

To develop asymptotics for $\hat{\theta}_n$, we employ the following framework, which is a generalization of Theorem 8.1 of Wooldridge (1994). Wooldridge dealt with an abstract extremum estimation problem for possibly deterministically trending and weakly dependent time series. The approach involved a smooth objective function and regularity conditions that enabled consistency and asymptotic normality for extremum estimators to be obtained within the same framework. That framework was extended to time trend power regression in Phillips (2007) and to stochastically nonstationary time series in Park and Phillips

(2000, 2001), with related subsequent work in Andrews and Sun (2004), Chan and Wang (2014) and Wang and Phillips (2016).

For a sequence of matrices $F_n = diag\big[a_1(n), ..., a_m(n)\big]$, we define

$$Z_n = F_n^{-1} S_n(\theta_0), \quad Y_n = F_n^{-1} H_n(\theta_0) F_n^{-1}.$$

With these components we are able to state the main result.

Theorem A.1. *Suppose that θ_0 is a finite interior point of Θ, and $\lambda_{\min}^{-1}(Y_n) = O_p(1)$, where $\lambda_{\min}(A)$ denotes the smallest eigenvalue of A, and there exists a sequence of constants $\{k_n, n \geq 1\}$ satisfying $k_n \to \infty$ and $k_n \max_{1\leq j\leq m} a_j(n)^{-1} \to 0$ such that $Y_n^{-1} Z_n = o_P(k_n)$ and*

$$\sup_{\theta: ||F_n(\theta-\theta_0)||\leq k_n} ||F_n^{-1}\big[W_n(\theta) - H_n(\theta_0)\big]F_n^{-1}|| = o_P(k_n^{-2}). \tag{A.2}$$

Then there exists a sequence of estimators $\hat{\theta}_n$ such that $S_n(\hat{\theta}_n) = 0$ with probability that goes to one and

$$F_n(\hat{\theta}_n - \theta_0) \;\; = \;\; -Y_n^{-1} Z_n + o_P(1). \tag{A.3}$$

Proof. The proof follows the same arguments as that of Theorem 4.1 in Wang and Phillips (2016), see also Andrews and Sun (2004). We provide an outline here for completeness and convenience for future reference. Let $\Theta_0 = \{\theta : \; ||F_n(\theta - \theta_0)|| \leq k_n\}$. As $k_n||F_n^{-1}|| = o(1)$, we may take n sufficiently large so that $\Theta_0 \subset \{\theta : ||\theta - \theta_0|| \leq \delta\} \subset \Theta$, for some $\delta > 0$. Recall that $Q_n(\theta)$ is twice differentiable whenever $\| \theta - \theta_0 \| \leq \delta$. It follows by Taylor expansion that

$$\begin{aligned} Q_n(\theta) - Q_n(\theta_0) &= 2(\theta - \theta_0)' S_n(\theta_0) + (\theta - \theta_0)' W_n(\theta_1)(\theta - \theta_0) \quad (\text{for some } \theta_1 \in \Theta_0) \\ &= 2(\theta - \theta_0)' S_n(\theta_0) + (\theta - \theta_0)' H_n(\theta_0)(\theta - \theta_0) + R_n(\theta, \theta_0) \\ &= \big[F_n(\theta - \theta_0) + Y_n^{-1} Z_n\big]' Y_n \big[F_n(\theta - \theta_0) + Y_n^{-1} Z_n\big] \\ &\quad - Z_n' Y_n^{-1} Z_n + R_n(\theta, \theta_0), \end{aligned} \tag{A.4}$$

uniformly for $\theta \in \Theta_0$, where, due to (A.2),

$$\begin{aligned} \sup_{\theta\in\Theta_0} |R_n(\theta, \theta_0)| &\leq \sup_{\theta\in\Theta_0} \sup_{\theta_1\in\Theta_0} |(\theta - \theta_0)'\big[W_n(\theta_1) - H_n(\theta_0)\big](\theta - \theta_0)| \\ &\leq k_n^2 \sup_{\theta_1\in\Theta_0} \| F_n^{-1}\big[W_n(\theta_1) - H_n(\theta_0)\big] F_n^{-1} \| = o_P(1). \end{aligned} \tag{A.5}$$

Let $\widetilde{\theta}_n = \theta_0 - F_n^{-1}Y_n^{-1}\,Z_n$. Using $Y_n^{-1}Z_n = o_P(k_n)$, we have

$$P(\widetilde{\theta}_n \notin \Theta_0) \leq P(||Y_n^{-1}\,Z_n|| \geq k_n) \to 0. \tag{A.6}$$

This, together with (A.4), yields

$$Q_n(\widetilde{\theta}_n) - Q_n(\theta_0) = -\,Z_n'Y_n^{-1}Z_n + R_n(\widetilde{\theta}_n, \theta_0), \tag{A.7}$$

where $R_n(\widetilde{\theta}_n, \theta_0) = o_P(1)$. For any $\epsilon > 0$ and $n \geq 1$, now let

$$\Theta_n(\epsilon) = \{\theta \in \Theta : ||F_n(\theta - \theta_0) + Y_n^{-1}Z_n|| \leq \epsilon\}.$$

Using $Y_n^{-1}Z_n = o_P(k_n)$ again, we get $P[\Theta_n(\epsilon) \subset \Theta_0] \to 1$, as $n \to \infty$. Hence, for any $\theta \in \partial\Theta_n(\epsilon)$, where $\partial\Theta_n(\epsilon)$ denotes the boundary of $\Theta_n(\epsilon)$, it follows from (A.4) and (A.7) that

$$Q_n(\theta) - Q_n(\widetilde{\theta}_n) = \nu_n'Y_n\nu_n + o_P(1), \tag{A.8}$$

where ν_n is a vector with $||\nu_n|| = \epsilon > 0$. Since $\nu_n'Y_n\nu_n \geq \lambda_{\min}(Y_n)||\nu_n||^2 = \epsilon^2\lambda_{\min}(Y_n)$, and $\widetilde{\theta}_n \in \Theta_n(\epsilon)$, equation (A.8) implies that, for each $\epsilon > 0$, the event that the minimum of $Q_n(\theta)$ over $\Theta_n(\epsilon)$ is in the interior of $\Theta_n(\epsilon)$ has probability that goes to one as $n \to \infty$. In particular, for each $\epsilon > 0$, there exists a point $\hat{\theta}_n(\epsilon) \in \Theta_n(\epsilon)$ (not necessary unique) such that $P(\dot{Q}_n[\hat{\theta}_n(\epsilon)] = 0) \to 1$, as $n \to \infty$. In consequence, there exists a sequence of $\hat{\theta}_n = \hat{\theta}_n(1/J_n) \in \Theta_n(1/J_n)$ where $J_n \to \infty$ so that $P(\dot{Q}_n(\hat{\theta}_n) = 0) \to 1$, as $n \to \infty$, and (A.3) holds. □

B Proof of Lemma 5.2

For $f(x) = |x|^\gamma \log^m|x|$ and any $x, y \neq 0$, we have

$$\big|f(x) - f(y) - (|x| - |y|)f'(y)\big| \;=\; \frac{1}{2}(|x| - |y|)^2|f''(z_0)| \leq \frac{1}{2}(x-y)^2|f''(z_0)|,$$

where z_0 lies between $|x|$ and $|y|$, and

$$f''(z) = z^{\gamma-2}\log^{m-2} z[\gamma(\gamma-1)\log^2 z + m(2\gamma-1)\log z + m(m-1)], \qquad z > 0.$$

Hence, if $0 < |x|, |y| < N$ and $|x - y| \leq \delta$, then

$$\begin{aligned} &\big|f(x) - f(y) - (|x| - |y|)f'(y)\big|I(|x| > \delta, |y| > \delta) \\ \leq\; &\frac{1}{2}\sup_{\delta < z < N}|f''^2 \leq C_{\delta,N}|x - y|^{\min\{\alpha,2\}}, \end{aligned} \tag{B.1}$$

where $C_{\delta,N} = \frac{1}{2}\,\delta^{\max\{2-\alpha,0\}}\,\sup_{\delta<z<N}|f''(z)| < \infty$.

For sufficiently small $\delta > 0$, we have either $f''(z) > 0$ for all $0 < z < 2\delta$ or $f''(z) < 0$ for all $0 < z < 2\delta$. Thus, if $|x-y| \le \delta$, then

$$\begin{aligned}
& \big|f(x) - f(y) - (|x| - |y|)f'(y)\big| I(|x| < \delta \text{ or } |y| < \delta) \\
\le\ & \big|(|x| - |y|)(f'(z_1) - f'(y))\big| I(|x| < 2\delta, |y| < 2\delta) \\
\le\ & |x-y||f'(x) - f'(|y|)| I(|x| < 2\delta, |y| < 2\delta) \\
\le\ & |x-y|\Big[|f'(x)| I(|x| < 2\delta) + |f'(y)| I(|y| < 2\delta)\Big], \qquad \text{(B.2)}
\end{aligned}$$

where z_1 lies between $|x|$ and $|y|$. Note that for any $x, y \neq 0$, we may write

$$\begin{aligned}
(|x| - |y|)f'(y) &= (|x| - |y|)f'(y)I(xy > 0) + (|x| - |y|)f'(y)I(xy < 0) \\
&= (x-y)\text{sign}(x)f'(y)I(xy > 0) + (|x| - |y|)f'(y)I(xy < 0) \\
&= (x-y)\text{sign}(x)f'(y) - (x-y)\text{sign}(x)f'(y)I(xy < 0) \\
&\quad +(|x| - |y|)f'(y)I(xy < 0).
\end{aligned}$$

Hence, for any $|x-y| < \delta$,

$$\begin{aligned}
& \big|(|x| - |y|)f'(y) - (x-y)\text{sign}(x)f'(y)\big| \\
\le\ & 2|x-y||f'(y)|[I(x > 0, y < 0) + I(x \le 0, y > 0)] \\
\le\ & 2|x-y||f'(y)| I(|x| \le \delta). \qquad \text{(B.3)}
\end{aligned}$$

Lemma 5.2 now follows from (B.1)-(B.3). □