Statistical Analysis of Nonlinear Time Series

by

YOSHIMASA UEMATSU

Submitted in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in

Economics

Department of Economics

Hitotsubashi University

2013

Preface

This dissertation contains the results of my four and a half years of research undertaken at the Graduate School of Economics of Hitotsubashi University. The thesis is composed of three articles on nonlinear time series analysis, and is organized as follows. The first presents a model that has an asymptotically efficient ordinary least squares estimator even though it does not satisfy the well-known Grenander conditions. The second considers estimation and inference for the unit root model with asymptotically collinear regressors. The third studies estimation and inference for nonlinear regression models with integrated time series by quantile regression.

Although the contributions might be small in view of the immense body of literature on nonlinear time series analysis, I hope that the findings become a basis for further study, including for my future works.

In the course of writing this dissertation, I have received a tremendous amount of advice and support from many individuals. Most of all, I am deeply grateful to my supervisor, Prof. Eiji Kurozumi, and my sub-supervisor, Prof. Toshio Honda, who always provided carefully considered feedback and valuable comments. Without their continuing support and encouragement, I would never have completed this dissertation. I am also obliged to Prof. Katsuto Tanaka, Prof. Naoyuki Ishimura, Prof. Yoichi Kuwana, Prof. Yohei Yamamoto, Prof. Katsumi Shimotsu, Prof. Peter C. B. Phillips, Prof. Mototsugu Shintani, Prof. Jana Jurečková, Dr. Shinya Tanaka, Dr. Ryota Yabe, and all the participants of my presentations for their constructive comments. Their suggestions helped me improve the dissertation. I also thank my parents, Ryomei and Ayako, for their enduring financial support.

Finally, this research has been partially supported by a Japan Society for the Promotion of Science Research Fellowship for Young Scientists.

Yoshimasa Uematsu

September, 2013

Contents

Chapter 1

Overview

1.1 Introduction

The statistical properties of both stationary and nonstationary linear time series have been intensively analyzed. From an applicational point of view, empirical econometricians frequently use linear models because of their manageability. Although such models are known to provide good approximations of economic phenomena in many cases, there are some exceptions where a linear model does not adequately capture economic behavior; this is typically observed in financial time series. It may also be natural to employ nonlinear models to determine the model best suited for capturing nonlinearities in economic activities. Nonlinear models in time series appear in Teräsvirta et al. (2010), for instance. In terms of the estimation and inference of nonlinear models, intrinsic difficulties arise in each model that a researcher wants to employ. This is because complexity increases as a model becomes more precise. One of the objectives for statisticians is to supply empirical researchers with valid methods to identify the statistical properties of models and make greater inferences.

There are many ways to extend simple linear models to nonlinear ones even if we restrict our attention to parametric models. In particular, this dissertation focuses on the following two formulations that may capture some of the nonlinearities that arise in the analysis of time series.

The first is characterized by a model with deterministic regressors as follows.

$$y_t = \alpha + \beta f(t) + u_t,$$

where the model is formulated to be linear-in-parameter, but the trend term $f(t)$ is specified as a nonlinear function of time t. A number of economic time series, such as economic growth data, are distributed around deterministic functions of time. Therefore, this modeling seems to be valid for determining the features of observations in a flexible manner. Many articles, such as Vogelsang (1998), utilized a general l-dimensional vector of time-trending regressors $f(t) = [f_1(t), \ldots, f_l(t)]'$ for $t = 1, \ldots, n$ to construct a deterministically time-trending model. In this case, it is assumed that there exist a diagonal normalization matrix D_n and a vector of functions $F(r)$ for $r \in [0,1]$ so that

$$D_n f(t) = F(t/n) + o(1), \quad \int_0^1 F(r)dr < \infty, \quad \det\left[\int_0^1 F(r)F(r)'dr\right] > 0.$$

These conditions are similar to a part of Grenander's conditions and are general enough to include many trend formulations, such as polynomials of t, possibly with structural changes. However, they rule out regressors, which are asymptotically collinear in spite of their possible applications, meaning that they limit the scope of empirical analysis. Accordingly, time series models with such regressors are investigated in Chapter 2 and 3. Their roles in econometric analysis are summarized in Section 2.1.

The second formulation is marked by the parametric nonlinear model of the form

$$y_t = \alpha + g(x_t, \beta) + u_t.$$

This type of formulation includes many econometric models, such as discrete choice models, logistic models, models with Box-Cox transformation, and smooth transition models. In a time series context, numerous works have been produced with a focus on stationary covariates. Park and Phillips (2001) considered models of this type with the covariate x_t

as an integrated time series. That paper sought to develop the asymptotic properties of the estimators obtained by nonlinear least squares estimation. Their asymptotic distributions depend heavily on the formulation of the regression function g. Following this seminal work of Park and Phillips, a large number of related papers were published from theoretical and applied perspectives. The reader may find a portion of these studies in Section 4.1. Given this background, this thesis deals with this type of model and investigates the asymptotic properties of the quantile regression estimators in Chapter 4.

1.2 Overview: Chapter 2

Chapter 2 presents a model that has an asymptotically efficient ordinary least squares (OLS) estimator, irrespective of the singularity of its limiting sample moment matrix. In the literature on time series, Grenander and Rosenblatt's result is necessary to judge the asymptotic efficiency of the OLS estimator with requiring that the regressors satisfy Grenander's conditions. Without the conditions, however, it is not obvious whether the estimator is efficient. In Chapter 2, we introduce such a model by analyzing the regression model with a *slowly varying* (SV) regressor under a quite general assumption on errors. These regressors are known to display asymptotic singularity in the sample moment matrices; that is, Grenander's condition fails.

A positive-valued function L on $\mathbb{R}_+$ is called SV if it satisfies, for any $r > 0$, $L(rn)/L(n) \to 1$ as $n \to \infty$. To deal with an SV function L, we suppose that L has the following *Karamata's representation*

$$L(n) = c_L \exp\left(\int_B^n \frac{\varepsilon(s)}{s} ds\right) \quad \text{for} \quad n \geq B$$

for some $B > 0$. Here, $c_L > 0$, ε is continuous and $\varepsilon(n) \to 0$ as $n \to \infty$. Note that any SV function is of order $o(n^{\alpha})$ for all $\alpha > 0$. With some additional conditions on L, we consider the following regression model

$$y_t = \beta_0 + \beta_1 L(t) + u_t \quad \text{for} \quad t = 1, \ldots, n, \quad \text{or} \quad y = X\beta + u,$$

where $y = [y_1, \ldots, y_n]'$, $\beta = [\beta_0, \beta_1]'$ and $X = [\iota, L]$ with $\iota = [1, \ldots, 1]'$ and $L = [L(1), \ldots, L(n)]'$. The term u_t represents the regression error and is modeled to include a very wide class of stationary processes with a positive and continuous spectrum. If we write $u = [u_1, \ldots, u_n]'$, the variance is given by

$$\mathrm{Var}(u) = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{bmatrix} = [\Gamma_0, \Gamma_1, \ldots, \Gamma_{n-1}] = \Gamma,$$

where Γ_t is the tth column vector of $\mathrm{Var}(u)$. We further denote the long-run variance of $\{u_t\}$ as σ^2. Using these notations, we may define the OLS and GLS estimators as $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$ and $\hat{\beta}_{GLS} = (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}y$, respectively.

The OLS estimator is said to be asymptotically efficient if

$$F_n \left[\mathrm{Var}(\hat{\beta}_{OLS}) - \mathrm{Var}(\hat{\beta}_{GLS}) \right] F_n \to 0$$

is true for some common standardizing matrix F_n. In Chapter 2, the convergence is proved for the SV model. That is, if regularity conditions hold, then $\mathrm{Var}(\hat{\beta}_{OLS})$ and $\mathrm{Var}(\hat{\beta}_{GLS})$ of the model have the same asymptotic variance of the form

$$\sigma^2 \begin{bmatrix} \dfrac{1}{n\varepsilon(n)^2} & -\dfrac{1}{nL(n)\varepsilon(n)^2} \\ -\dfrac{1}{nL(n)\varepsilon(n)^2} & \dfrac{1}{nL(n)^2\varepsilon(n)^2} \end{bmatrix} (1 + o(1)),$$

implying that the OLS estimator is asymptotically efficient with the standardizing matrix $F_n = diag[n^{1/2}\varepsilon(n), n^{1/2}L(n)\varepsilon(n)]$.

1.3 Overview: Chapter 3

Chapter 3 considers a unit root test in the presence of a SV regressor L. The definition of the SV function L follows from the preceding section, but the assumptions to be imposed here are somewhat different. Consider the model

$$y_t = \alpha + \beta L(t) + u_t \quad \text{and} \quad u_t = \rho u_{t-1} + v_t \quad \text{for} \quad t = 1, \ldots, n, \tag{1.1}$$

where $\rho = 1$ and $\{v_t\}$ is assumed to be a one-summable linear process with $E|v_t|^p < \infty$ for some $p > 2$. The regressor $L(t)$ is given by an SV function that satisfies some regularity conditions. The first result is to derive the limiting distribution of the OLS estimator:

$$\frac{1}{n^{1/2}} \begin{bmatrix} \varepsilon(n)(\hat{\alpha}_n - \alpha) \\ L(n)\varepsilon(n)(\hat{\beta}_n - \beta) \end{bmatrix} \xrightarrow{d} N\left(0, \frac{2\sigma_L^2}{27} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\right),$$

where σ_L^2 is the long-run variance of v_t. This weak convergence is obtained by Mynbaev's CLT (see Mynbaev 2009, 2011a), which is applicable for a weighted sum of linear processes. Because any SV function possesses an asymptotic order of $o(n^{1/2})$, the OLS estimators cannot be consistent. This fact contrasts with the case where the simple trend t is employed. Considering models with an SV regressor, we therefore remark that the existence of a unit root leads to a meaningless regression and that testing for a unit root is indispensable.

Let $W(\cdot)$ denote the standard Brownian motion obtained in the limit of the partial sum process $\sigma_L^{-1} n^{-1/2} \sum_{t=1}^{[\cdot n]} v_t$. The next result we investigate is the asymptotic behaviors of the unit root test statistics, that is, the estimated regression coefficient $\hat{\rho}_n$ and corresponding t-statistic $t_{\hat{\rho}_n}$, based on the regression residuals. Under the regularity conditions, we obtain

$$n\varepsilon(n)^2 \left(\hat{\rho}_n - 1\right) \xrightarrow{d} -\frac{U_1}{2V_1} \quad \text{and} \quad \varepsilon(n)^2 t_{\hat{\rho}_n} \xrightarrow{d} -\frac{\sigma_L}{\sigma_S} \frac{U_1}{2\sqrt{V_1}},$$

where σ_S^2 is the short-run variance of v_t and U_1 and V_1 are given by

$$U_1 = \left\{\int_0^1 (1 + \log r) W(r) dr\right\}^2 \quad \text{and}$$

$$V_1 = \int_0^1 W(r)^2 dr - \left(\int_0^1 W(r) dr\right)^2 - \left\{\int_0^1 (1 + \log r) W(r) dr\right\}^2.$$

These results are derived by an application of both Mynbaev's CLT and FCLT for a linear process. Testing a unit root by these statistics, however, will be useless because the finite sample approximation is poor.

To overcome this difficulty, we first consider the following no-constant model

$$y_t = \beta L(t) + u_t \quad \text{and} \quad u_t = \rho u_{t-1} + v_t \quad \text{for} \quad t = 1, \ldots, n, \tag{1.2}$$

where the same assumptions on $L(t)$ and v_t continue to hold. Then, we have the similar weak convergence results

$$\frac{L(n)}{n^{1/2}} \left(\hat{\beta}_n - \beta\right) \xrightarrow{d} N\left(0, \frac{\sigma_L^2}{3}\right),$$

and

$$n\left(\hat{\rho}_n - 1\right) \xrightarrow{d} \frac{U_2 - \sigma_S^2/\sigma_L^2}{2V_2} \quad \text{and} \quad t_{\hat{\rho}_n} \xrightarrow{d} \frac{\sigma_L}{\sigma_S} \frac{U_2 - \sigma_S^2/\sigma_L^2}{2\sqrt{V_2}},$$

where

$$U_2 = \left\{ W(1) - \int_0^1 W(r)dr \right\}^2 \quad \text{and} \quad V_2 = \int_0^1 W(r)^2 dr - \left(\int_0^1 W(r)dr \right)^2.$$

In practice, it may not be appropriate to suppose that the true model has no constant term. It is worth analyzing the situation where the true model is given by (1.1), which possesses a constant term, but the no-constant model (1.2) is employed for regression. Then, we still have the same asymptotic result based on the no-constant model with the effect of a constant term declining at the rate $O(n^{-1/2})$. This manipulation brings about a significant improvement in terms of size and power in finite sample situations, and a unit root test based on this procedure is recommended. Applying the result, we finally give general Phillips and Perron type test statistics.

1.4 Overview: Chapter 4

Chapter 4 studies estimation and inference for nonlinear regression models with integrated time series by quantile regression. Suppose that the scalar-valued random variable y_t is generated from the following nonlinear model

$$y_t = \alpha_0 + g(x_t, \beta_0) + u_t \quad \text{for} \quad t = 1, \ldots, n, \tag{1.3}$$

where $g : \mathbb{R} \times \mathbb{R}^\ell \to \mathbb{R}$ is a known regression function, and x_t and u_t are the covariate and error, respectively. In particular, x_t is specified as a simple AR(1) unit root model $x_t = x_{t-1} + v_t$, where v_t is stationary. The ℓ-dimensional true parameter vector $\theta_0 = (\alpha_0, \beta_0')'$ is assumed to lie in the parameter set $\Theta = A \times B \subset \mathbb{R} \times \mathbb{R}^\ell$. Moreover, let F and f denote the cumulative distribution function (CDF) and pobability density function (PDF) of u_t, respectively. Note that the τth quantile of u_t for a fixed $\tau \in (0,1)$ is simply denoted by $F^{-1}(\tau)$ under some regularity conditions on u_t. Let $u_{t\tau} = u_t - F^{-1}(\tau)$ for $t = 1, \ldots, n$. We may also rewrite the parameter so that $\alpha_0(\tau) = \alpha_0 + F^{-1}(\tau)$ in response to the error term $u_{t\tau}$ and define the new parameter vector $\theta_0(\tau) = \big(\alpha_0(\tau), \beta_0'\big)'$.

The regression function $(x, \beta) \mapsto g(x, \beta)$ is classified into two functional classes as in park and Phillips (2001). The first is the class of *H-regular* functions, which are defined by

$$g(\lambda x, \beta) = \kappa(\lambda) h(x, \beta) + R(x, \lambda, \beta),$$

where the functions κ and h are said to be the *asymptotic order* and *limit homogeneous function* of g, respectively. The last term, $R(x, \lambda, \beta)$, is a remainder. Polynomial functions, distribution-like functions and logarithmic function are included in this *H-regular* class. The second is the class of *I-regular* functions, which are characterized as bounded and integrable functions with respect to x with sufficient smoothness in β. An example is an exponential function of the form $\beta \exp(-x^2)$.

With this setting, we may obtain the nonlinear quantile regression (NQR) estimator $\hat{\theta}_n(\tau) = \Big(\hat{\alpha}_n(\tau), \hat{\beta}_n(\tau)'\Big)'$ of $\theta_0(\tau)$ by solving the minimization problem

$$\hat{\theta}_n(\tau) = \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} \rho_\tau(y_t - \alpha - g(x_t, \beta)),$$

where $\rho_\tau(u) = u\psi_\tau(u)$ with $\psi_\tau(u) = \tau - 1(u < 0)$.

To develop the analysis, we are required to make assumptions on errors u_t and v_t. We construct two partial sum processes

$$U_n^{\psi}(\tau,r) = \frac{1}{n^{1/2}}\sum_{t=1}^{[nr]}\psi_\tau(u_{t\tau}) \quad \text{and} \quad V_n(r) = \frac{1}{n^{1/2}}\sum_{t=0}^{[nr]} v_{t+1}.$$

The process $\{\psi_\tau(u_{t\tau})\}$ is assumed to be a martingale difference sequence. Further, we suppose that the vector $(U_n^{\psi}, V_n)(r)$ converges weakly to a vector Brownian motion $(U^{\psi}, V)(r)$ whose covariance matrix is $\Omega(r)$. In addition, we assume that v_t is a linear process for an I-regular g.

One of the main contributions of Chapter 4 is that the asymptotic distribution of the NQR estimator $\hat{\theta}_n(\tau)$ is derived with restricting our attention to the class of H-regular functions. Note that the class of I-regular functions cannot be treated in the same framework due to the irregular convergence rate $n^{1/4}$; this class is investigated later with a restriction to model (1.3). The regression function $g(x,\cdot)$ is always supposed to be twice continuously differentiable. Define notation of the first and second order derivatives as

$$\dot{g}(x,\beta) = \frac{\partial g}{\partial \beta}(x,\beta) \quad \text{and} \quad \ddot{G}(x,\beta) = \frac{\partial^2 g}{\partial \beta \partial \beta'}(x,\beta),$$

and we further write $\ddot{g} = \mathrm{vec}(\ddot{G})$. Corresponding to the ℓ-dimensional vector $\dot{g}$ and ℓ^2-dimensional vector $\ddot{g}$, the asymptotic order matrices $\dot{\kappa}_n$ $(\ell \times \ell)$ and $\ddot{\kappa}_n$ $(\ell^2 \times \ell^2)$ and the vector of the limit homogeneous functions $\dot{h}$ and $\ddot{h}$ are introduced when $\dot{g}$ and $\ddot{g}$ are H-regular. We further let $\tilde{g} = (1, \dot{g}')'$, $\tilde{h} = (1, \dot{h}')'$ and $\tilde{\kappa}_n = \mathrm{diag}(1, \dot{\kappa}_n)$. To obtain the limiting distribution, we need to suppose an additional assumption on the parameter vector θ so that $\theta = \theta_0(\tau) + n^{-1/2}\tilde{\kappa}_n'^{-1}\pi$, where π lies in a compact set $\Pi \subset \mathbb{R} \times \mathbb{R}^{\ell}$.

Define the derivative from the right of the objective function as

$$z_t(\theta) = \tilde{g}_t(\beta)\psi_\tau(y_t - \alpha - g_t(\beta)).$$

Utilizing this function, we may derive the limiting distribution by considering the "first order condition"

$$n^{-1/2}\tilde{\kappa}_n^{-1}\sum_{t=1}^{n} z_t(\hat{\theta}_n(\tau)) = o_p(1).$$

This estimating equation leads to the Bahadur representation of the NQR estimator $\hat{\theta}_n(\tau)$. In consequence, we obtain the result. Let $\dot{g}$ and $\ddot{g}$ be H-regular on B. Then, under some regularity conditions, we have

$$\begin{aligned} & n^{1/2}\tilde{\kappa}_n'(\hat{\theta}_n(\tau)-\theta_0(\tau)) \\ & \xrightarrow{d} \frac{1}{f(F^{-1}(\tau))}\left[\int_0^1 \tilde{h}(V(r),\beta_0)\tilde{h}(V(r),\beta_0)'dr\right]^{-1}\int_0^1 \tilde{h}(V(r),\beta_0)dU^{\psi}(r). \end{aligned}$$

Note that the limiting distribution is not usually (mixed) normal because of the possible nonzero correlation between U^{ψ} and V. Therefore standard inferences are not applicable in this case.

To overcome the difficulty, we suggest fully-modified NQR (FM-NQR) estimator based on the results of Phillips and Hansen (1990) and de Jong (2002). The FM-NQR estimator is constructed by

$$\hat{\theta}_n^+(\tau) = \hat{\theta}_n(\tau) - \frac{n^{-1/2}\tilde{\kappa}_n^{-1}}{\widehat{f(F^{-1}(\tau))}}\frac{\hat{\omega}_{\psi v}}{\hat{\omega}_v^2}S_n^{-1}T_n,$$

where $1/\widehat{f(F^{-1}(\tau))}$, $\hat{\omega}_{\psi v}$ and $\hat{\omega}_v^2$ are consistent estimators. The statistics S_n and T_n satisfy

$$S_n \xrightarrow{p} \int_0^1 \tilde{h}(V(r),\beta_0)\tilde{h}(V(r),\beta_0)'dr \quad \text{and} \quad T_n \xrightarrow{d} \int_0^1 \tilde{h}(V(r),\beta_0)dV(r).$$

If some additional conditions of de Jong (2002) are satisfied, we then have

$$\begin{aligned} & n^{1/2}\tilde{\kappa}_n'\left(\hat{\theta}_n^+(\tau)-\theta_0(\tau)\right) \\ & \xrightarrow{d} \frac{1}{f(F^{-1}(\tau))}\left[\int_0^1 \tilde{h}(V(r),\beta_0)\tilde{h}(V(r),\beta_0)'dr\right]^{-1}\int_0^1 \tilde{h}(V(r),\beta_0)dU^{\psi+}(r), \end{aligned}$$

where $U^{\psi+}(r) = U^{\psi}(r) - \omega_{\psi v}\omega_v^{-2}V(r)$. Therefore, the mixed normality of the limiting random variable is brought to light immediately because $U^{\psi+}$ is easily found to be uncorrelated with V and, hence, independent of V. Because of the asymptotic normality of the FM-NQR estimator, we can consider testing linear restrictions on the parameter vector.

We have considered the NQR estimator of the nonlinear model only in the case of H-regular $\dot{g}$ and $\ddot{g}$. We then investigate the so-called linear-in-parameter model obtained by confining model (1.3) to

$$g(x_t, \beta_0) = \beta_0 g(x_t). \tag{1.4}$$

The regression function g is either I-regular or H-regular and write $g_t = g(x_t)$. The parameter β_0 is allowed ℓ-dimensional, but is assumed $\ell = 1$ for the sake of brevity. Because the model is linear in parameter, the asymptotics can be derived even if g is an I-regular function as well as an H-regular one.

First, we consider model (1.3) under restriction (1.4) with I-regular regression function derivative $\dot{g}$. The limiting distribution of the NQR estimator $\hat{\theta}_n(\tau)$ is summarized as follows. Let g be I-regular on B, and suppose some regularity conditions. Then we have

$$D_n^I(\hat{\theta}_n(\tau) - \theta_0(\tau)) \xrightarrow{d} \frac{1}{f(F^{-1}(\tau))} \begin{bmatrix} U^{\Psi}(1) \\ \left(L(1,0)\int_{-\infty}^{\infty} g(x)^2 dx\right)^{-1/2} W(1) \end{bmatrix},$$

where $D_n^I = \mathrm{diag}(n^{1/2}, n^{1/4})$ and the Brownian motion $W(r)$ has variance $r\tau(1-\tau)$, which is the same as the variance of $U^{\Psi}(r)$. This implies that $\hat{\alpha}_n(\tau)$ and $\hat{\beta}_n(\tau)$ are asymptotically independent; consequently, the limiting joint distribution is mixed normal of the form

$$MN\left(0,\ \frac{\omega_{\psi}^2}{f(F^{-1}(\tau))^2} \begin{bmatrix} 1 & 0 \\ 0 & L(1,0)\int_{-\infty}^{\infty} g(x)^2 dx \end{bmatrix}^{-1}\right).$$

Hence, standard inferences are applicable in an asymptotic sense. For the case of H-regular functions, we certainly have the same result obtained above.

Finally, we investigate finite sample performances of the NQR estimators with $\tau = 0.5$ via comparison to nonlinear least squares (NLS) estimators by simulations. We observe from simulations that our suggested NQR estimators are preferable to the NLS estimators in terms of estimation accuracy and powers of tests when distributions of regression errors possess fat tails.

Chapter 2

Asymptotic Efficiency of the OLS Estimator with a Singular Limiting Sample Moment Matrix

This chapter presents a model that has an asymptotically efficient ordinary least squares estimator irrespective of the singularity of its limiting sample moment matrices. In the literature on time series, Grenander and Rosenblatt's result is necessary to judge the efficiency with requiring Grenander's conditions. Without the conditions, however, it is not obvious whether the estimator is efficient. In this chapter, we introduce such a model by analyzing a model with a slowly varying regressor under quite general assumptions on errors. These regressors are known to display asymptotic singularity in the sample moment matrices, or namely Grenander's condition fails.

2.1 Introduction

Discussion on the asymptotic relative efficiency of ordinary least squares (OLS) estimators of a time series regression dates from the middle of the twentieth century. When the regressors are deterministic functions of time and the disturbances may be serially autocorrelated, Grenander (1954), Rosenblatt (1956), and Grenander and Rosenblatt (1957) (GR) have shown the necessary and sufficient condition for OLS estimators to be asymptotically efficient. Following the seminal work of GR, much attention has been given to stochastic regressors. Krämer (1986) proved the asymptotic equivalence of the OLS and GLS estimators when the regressor is a univariate integrated process. Phillips and Park (1988) extended these results to multiple regressions. Krämer and Hassler (1998) studied the case where the regressors are fractionally integrated. Shin and Oh (2002) generalized the class of regressors to the class of unstable regressors containing a seasonally integrated process.

The present chapter reconsiders a model with deterministic regressors. In order to judge its efficiency, the GR theorem requires the regressors to satisfy the well-known Grenander's conditions in the first place. In general, it is unclear whether the OLS estimator is efficient or not unless Grenander's conditions are satisfied and the GR theorem is applied.

Given this background, this chapter shows the existence of a model that has asymptotically efficient OLS estimators even though it does not satisfy the conditions. The result is derived through investigating the regression model with a slowly varying (SV) regressor. Grenander's conditions are known to be valid for many types of regressors, including a polynomial of time, and many studies have restricted the regressors to this class. However, this is not true for SV regressors (see Phillips (2007)), although there are many applications of such SV regressors. These include the log-periodogram regression of long memory (Robinson (1995), Hurvich et al. (1998), Phillips (1999), and references therein), nonlinear least squares estimation (Wu (1981), Phillips (2007), Mynbaev (2011)), and the study of growth convergence (Barro and Sala-i-Martin (2004)). Concerning economic convergence and transition modeling, Phillips and Sul (2007, 2009) designed a new model that represents the behavior of economies in transition and proposed an associated test for convergence, utilizing SV functions explicitly. Therefore, it is worth analyzing such SV models from an applicational point of view. Furthermore, in view of a theoretical contribution, this yields a simple but significant example of models, as the results complement the GR theory because the assumption to be made on the error term is identical to that of GR.

The remainder of the chapter is organized as follows: Section 2.2 includes some assumptions and provides some preliminary theory. We especially consider models with SV regressors as investigated by Phillips (2007) and Mynbaev (2009). Section 2.3 states the main theorem for asymptotic efficiency and Section 2.4 concludes. Section 2.5 contains the proofs for the results.

2.2 Preliminaries

2.2.1 Slowly varying regressor

We start with a definition of an SV function. A positive-valued function L on $\mathbb{R}_+$ is called SV if it satisfies, for any $r > 0$, $L(rn)/L(n) \to 1$ as $n \to \infty$. In order to deal with an SV function L, the following Karamata representation theorem is essential. That is, the function L is SV if and only if it may be written in the form

$$L(n) = c(n) \exp\left(\int_a^n \frac{\varepsilon(s)}{s} ds\right) \quad \text{for} \quad n > a$$

for some $a > 0$, where $c(n) \to c \in (0, \infty)$ and $\varepsilon(n) \to 0$ as $n \to \infty$. Considering regression theory, however, we require a stronger assumption.

Assumption 1 Suppose all the conditions below:

(a) The function L is SV and has Karamata's representation

$$L(x) = c_L \exp\left(\int_B^x \frac{\varepsilon(s)}{s} ds\right) \quad \text{for} \quad x \geq B$$

for some $B > 0$. Here $c_L > 0$, ε is continuous and $\varepsilon(x) \to 0$ as $x \to \infty$. Hereafter, this part of the assumption is shortened to $L = K(\varepsilon)$.

(b) $\varepsilon = K(\eta)$, $\eta = K(\mu)$, where $\mu = (\varepsilon + \eta)/2$, and $\eta(n)^2 = o(\varepsilon(n))$.

(c) L is monotonically increasing.

Remark 1 Condition (a) imposes more restrictive assumptions than the Karamata representation. This condition also appeared in Phillips (2007) and Mynbaev (2009). Condition (b) is essential for Lemma 1 below (see Mynbaev (2011a, 4.2.4)). Condition (c) seems rather stringent, but this reduces the burden of calculation in the proof. Note that any SV function is of order $o(n^{\alpha})$ for all $\alpha > 0$.

An application of Assumption 1 leads to the following lemma.

Lemma 1 (Phillips) *If $L(t)$ satisfies Assumption 1, then, for $k = 0, 1, 2, \ldots$ we have*

$$\frac{1}{n}\sum_{t=1}^{n} L(t)^k = L(n)^k - kL(n)^k\varepsilon(n) + k^2L(n)^k\varepsilon(n)^2 + kL(n)^k\varepsilon(n)\eta(n) + o\left(L(n)^k\varepsilon(n)\left[\varepsilon(n) + \eta(n)\right]\right).$$

2.2.2 Regression model

Consider the following regression model:

$$y_t = \beta_0 + \beta_1 L(t) + u_t \quad \text{for} \quad t = 1, \ldots, n, \quad \text{or} \quad y = X\beta + u, \tag{2.1}$$

where $y = [y_1, \ldots, y_n]'$, $\beta = [\beta_0, \beta_1]'$, and $X = [\iota, L]$ with $\iota = [1, \ldots, 1]'$ and $L = [L(1), \ldots, L(n)]'$. Using these notations, we may define the OLS and GLS estimators as $\hat{\beta}_{OLS} = (X'X)^{-1}X'y$ and $\hat{\beta}_{GLS} = (X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}y$, respectively. The OLS estimator is said to be asymptotically efficient if

$$F_n\left[\mathrm{Var}(\hat{\beta}_{OLS}) - \mathrm{Var}(\hat{\beta}_{GLS})\right]F_n \to 0 \tag{2.2}$$

is true for some common standardizing matrix F_n.

An assumption on the regression errors u_t is required. As the objective here is not to derive the asymptotic distribution of the OLS estimator, but to prove its asymptotic efficiency, we need an assumption that is sufficiently general to include a very wide class of stationary processes.

Assumption 2 The error process $\{u_t\}$ is real and stationary with $Eu_t = 0$ and $Eu_tu_{t+h} = \gamma_h$. Furthermore, $\{u_t\}$ has a spectral density $f(\lambda)$ that is positive and continuous in $\lambda \in [-\pi, \pi]$.

Let $u = [u_1, \ldots, u_n]'$ and

$$\mathrm{Var}(u) = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n-1} & \gamma_{n-2} & \cdots & \gamma_0 \end{bmatrix} = [\Gamma_0, \Gamma_1, \ldots, \Gamma_{n-1}] = \Gamma,$$

where Γ_{t-1} is the tth column vector of $\mathrm{Var}(u)$ for $t = 1, \ldots, n$.

Remark 2 The error term $\{u_t\}$ with Assumption 2 is quite general in that the GR result requires the same condition. In particular, no summability condition is provided on γ_h. In contrast, Phillips (2007) and Mynbaev (2009) assume that $\{u_t\}$ is a linear process with more restrictive conditions because they attempt to derive the asymptotic distributions of the OLS estimator.

Remark 3 Throughout the chapter, we let

$$\sigma_n^2 = n\mathrm{Var}(\bar{u}) = \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\gamma_{s-t} \quad \text{and} \quad \sigma^2 = 2\pi f(0),$$

where σ^2 is the long-run variance. If Assumption 2 holds, then we have $\sigma_n^2 \to \sigma^2$ as $n \to \infty$ (see Fuller (1996, p. 310)). The boundedness of the limit σ^2 is ensured by the continuity of f. In addition, both the GR result and our subsequent analysis require the positiveness of f because its reciprocal is needed for the expression of the variance of the GLS estimator.

Using Lemma 1, Phillips (2007) proved that

$$D_n^{-1}X'XD_n^{-1} \to \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad F_n\left(X'X\right)^{-1}F_n \to \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

where $D_n = diag(n^{1/2}, n^{1/2}L(n))$ and $F_n = diag(n^{1/2}\varepsilon(n), n^{1/2}L(n)\varepsilon(n))$. This implies that the covariance and its inverse have singular limits after standardization, thereby failing

Grenander's conditions; the GR result may thus not be applied directly. For further discussion on the GR result, see Anderson (1971, Sect. 10.2.3), for example.

2.3 Result

We now derive the asymptotic expression for the variance of the OLS and GLS estimators of model (2.1). The first result yields the asymptotic variance of the OLS estimator.

Lemma 2 *If Assumptions 1 and 2 are satisfied, then the OLS estimator of model (2.1) has asymptotic variance*

$$\operatorname{Var}(\hat{\beta}_{OLS}) = \sigma^2 \begin{bmatrix} \dfrac{1}{n\varepsilon(n)^2} & -\dfrac{1}{nL(n)\varepsilon(n)^2} \\ -\dfrac{1}{nL(n)\varepsilon(n)^2} & \dfrac{1}{nL(n)^2\varepsilon(n)^2} \end{bmatrix} (1+o(1)).$$

The asymptotic expression of $\operatorname{Var}(\hat{\beta}_{OLS})$ is the same as that derived by Phillips (2007, Theorem 3.1) as a by-product of proving asymptotic normality, although the result of Lemma 2 is valid for a wider class of short-memory stationary errors.

We next derive the asymptotic variance of the GLS estimator.

Lemma 3 *If Assumptions 1 and 2 are satisfied, then the GLS estimator of model (2.1) has the same asymptotic variance as that given in Lemma 2.*

To summarize, we state the theorem below.

Theorem 1 *If Assumptions 1 and 2 are satisfied, then the OLS estimator of model (2.1) is asymptotically efficient. In fact, (2.2) is true with the normalizing matrix* $F_n = diag(n^{1/2}\varepsilon(n), n^{1/2}L(n)\varepsilon(n))$.

Remark 4 This result is given for the simple regression (2.1), although Phillips (2007) dealt with polynomial and multiple regressions of SV regressors as well. In the same manner, our result can be extended to the polynomial models since a power of L is also SV. However, the computation required would be burdensome. For a multiple regression with different SV regressors, it may be difficult to verify the same result; see Mynbaev (2011b).

2.4 Conclusion

In the literature of stationary time series analysis and classical GR results, regressors are required to satisfy the Grenander conditions when researchers want to know whether the OLS estimator is asymptotically efficient. In the present chapter, however, we have revealed the existence of a model that has an asymptotically efficient OLS estimator even though it does not satisfy Grenander's conditions. The model and result have been given by proving the asymptotic equivalence of the OLS and GLS estimators of the model with an SV regressor. The assumption needed for the error term is quite general, and is identical to that for GR.

2.5 Appendix

2.5.1 Useful lemmas

Lemma 4 *If Assumption 2 is satisfied, then, for all $m,n \geq 1$, it follows that*

$$\sum_{u=1}^{m}\sum_{v=1}^{n}\gamma_{u-v} = \int_{-\pi}^{\pi}\frac{\sin(m\lambda/2)\sin(n\lambda/2)\cos((m-n)\lambda/2)}{\sin^2(\lambda/2)}f(\lambda)d\lambda.$$

Proof From Assumption 2, we first note that the autocovariance γ_{u-v} is written as

$$\gamma_{u-v} = \int_{-\pi}^{\pi} e^{i(u-v)\lambda}f(\lambda)d\lambda. \tag{2.3}$$

Because

$$\begin{aligned}\sum_{u=1}^{m} e^{iu\lambda} &= \frac{e^{i\lambda}\left(1-e^{im\lambda}\right)}{1-e^{i\lambda}}\\ &= \frac{e^{i\lambda}\left(e^{-im\lambda/2}-e^{im\lambda/2}\right)e^{im\lambda/2}}{\left(e^{-i\lambda/2}-e^{i\lambda/2}\right)e^{i\lambda/2}} = \frac{\sin(m\lambda/2)}{\sin(\lambda/2)}e^{i(m+1)\lambda/2},\end{aligned}$$

it follows from (2.3) that

$$\begin{aligned}
\sum_{u=1}^{m}\sum_{v=1}^{n}\gamma_{u-v} &= \int_{-\pi}^{\pi}\sum_{u=1}^{m}e^{iu\lambda}\sum_{v=1}^{n}e^{-iv\lambda}f(\lambda)d\lambda \\
&= \int_{-\pi}^{\pi}\frac{\sin(m\lambda/2)\sin(n\lambda/2)}{\sin^2(\lambda/2)}e^{i(m-n)\lambda/2}f(\lambda)d\lambda \\
&= \int_{-\pi}^{\pi}\frac{\sin(m\lambda/2)\sin(n\lambda/2)}{\sin^2(\lambda/2)}\{\cos((m-n)\lambda/2)+i\sin((m-n)\lambda/2)\}f(\lambda)d\lambda.
\end{aligned}$$

Since γ_{u-v} is real, the result follows. ■

Lemma 5 *Define* $g_n(u)=\sum_{s=1}^{n}\gamma_{s-u}-\sigma_n^2$, $\tilde{g}_n(t)=\sum_{u=1}^{t}g_n(u)$, $h_n(u)=\sum_{s=1}^{n}L(s)\gamma_{s-u}-\sigma_n^2L(u)$ *and* $\tilde{h}_n(t)=\sum_{u=1}^{t}h_n(u)$. *If Assumptions 1 and 2 are satisfied, then, for a sufficiently large n, the following statements are true:*

(a) $\max\limits_{1\le t\le n}|\tilde{g}_n(t)|\le C$ *for some constant* $C>0$,

(b) $\max\limits_{1\le t\le n}\left|\tilde{h}_n(t)\right|=o(nL(n)\varepsilon(n)^2)$.

Proof (a) Recalling $\sigma_n^2=n^{-1}\sum\sum_{s,v=1}^{n}\gamma_{s-v}$ and applying Lemma 4, we have

$$\begin{aligned}
&|\tilde{g}_n(t)| \\
&= \left|\sum_{u=1}^{t}\sum_{s=1}^{n}\gamma_{s-u}-\frac{t}{n}\sum_{s=1}^{n}\sum_{v=1}^{n}\gamma_{s-v}\right| \\
&= \left|\int_{-\pi}^{\pi}\left\{\frac{\sin(t\lambda/2)\sin(n\lambda/2)\cos((t-n)\lambda/2)}{\sin^2(\lambda/2)}-\frac{t}{n}\frac{\sin^2(n\lambda/2)}{\sin^2(\lambda/2)}\right\}f(\lambda)d\lambda\right| \\
&\le \max_{-\pi\le\lambda\le\pi}f(\lambda)\int_{-\pi}^{\pi}\left|\sin\frac{t\lambda}{2}\sin\frac{n\lambda}{2}\cos\frac{(t-n)\lambda}{2}-\frac{t}{n}\sin^2\frac{n\lambda}{2}\right|\frac{d\lambda}{\sin^2(\lambda/2)}.
\end{aligned} \tag{2.4}$$

Let the last integrand in (2.4) denote $S_{t,n}(\lambda)$. Then, $S_{t,n}(\lambda)$ is clearly bounded on $[-\pi,\pi]$ uniformly in $t=1,\dots,n$, and $n\ge 1$ except at $\lambda=0$. At this point only, the ratio is of indeterminate form. Even if the point $\lambda=0$ is included, however, boundedness can be proved uniformly on $[-\pi,\pi]$ as follows.

Transforming the product of the trigonometric functions to their summations and applying l'Hôpital's rule twice, we obtain

$$\begin{aligned}
&\lim_{\lambda\to 0} S_{t,n}(\lambda) \\
&= \lim_{\lambda\to 0}\left|-\cos n\lambda + \cos(t-n)\lambda + 1 - \cos t\lambda - \frac{2t}{n}(1-\cos n\lambda)\right| \frac{1}{2(1-\cos\lambda)} \\
&= \lim_{\lambda\to 0}\left|n\sin n\lambda - (t-n)\sin(t-n)\lambda + t\sin t\lambda - 2t\sin n\lambda\right| \frac{1}{2\sin\lambda} \\
&= \lim_{\lambda\to 0}\left|n^2\cos n\lambda - (t-n)^2\cos(t-n)\lambda + t^2\cos t\lambda - 2nt\cos n\lambda\right| \frac{1}{2\cos\lambda} \\
&= \left|n^2 - (t-n)^2 + t^2 - 2nt\right| \frac{1}{2} = 0
\end{aligned} \tag{2.5}$$

identically for all $t = 1,\dots,n$ and $n \geq 1$. Thus, the integral in (2.4) is uniformly bounded in $t = 1,\dots,n$ and $n \geq 1$, and this gives the proof of (a).

(b) Without loss of generality, we can set $L(1) = 0$. Applying summation by parts and collecting terms gives

$$\begin{aligned}
\left|\tilde{h}_n(t)\right| &= \left|\sum_{v=1}^{n} L(v)\sum_{u=1}^{t}\gamma_{v-u} - \frac{\sum_{u=1}^{t}L(u)}{n}\sum_{u=1}^{n}\sum_{v=1}^{n}\gamma_{v-u}\right| \\
&= \left|L(n)\sum_{v=1}^{n}\sum_{u=1}^{t}\gamma_{v-u} - \sum_{s=2}^{n}\sum_{v=1}^{s-1}\sum_{u=1}^{t}\gamma_{v-u}\Delta L(s) - \frac{\sum_{u=1}^{t}L(u)}{n}\sum_{u=1}^{n}\sum_{v=1}^{n}\gamma_{v-u}\right| \\
&\leq \sum_{s=2}^{n}\left|\sum_{v=1}^{n}\sum_{u=1}^{t}\gamma_{v-u} - \sum_{v=1}^{s-1}\sum_{u=1}^{t}\gamma_{v-u} - \frac{\sum_{u=1}^{t}L(u)}{nL(n)}\sum_{u=1}^{n}\sum_{v=1}^{n}\gamma_{v-u}\right|\Delta L(s).
\end{aligned}$$

If we use Lemma 4 as in (a), $\left|\tilde{h}_n(t)\right|$ may be bounded by

$$\begin{aligned}
\left|\tilde{h}_n(t)\right| \leq \max_{-\pi\leq\lambda\leq\pi} f(\lambda)\sum_{s=2}^{n}\int_{-\pi}^{\pi}&\left|\sin\frac{t\lambda}{2}\sin\frac{n\lambda}{2}\cos\frac{(n-t)\lambda}{2}\right. \\
-\sin\frac{t\lambda}{2}\sin\frac{(s-1)\lambda}{2}\cos\frac{(s-t-1)\lambda}{2} &- \left.\frac{\sum_{u=1}^{t}L(u)}{nL(n)}\sin^2\frac{n\lambda}{2}\right|\frac{d\lambda}{\sin^2(\lambda/2)}\Delta L(s).
\end{aligned} \tag{2.6}$$

Let the integrand of (2.6) denote $T_{s,t,n}(\lambda)$. Since $L(t)$ is positive and monotonically increasing, we have the bound

$$0 \leq \frac{\sum_{u=1}^{t}L(u)}{nL(n)} \leq \frac{\sum_{u=1}^{n}L(u)}{nL(n)} \leq 1 \tag{2.7}$$

uniformly in $t = 1,\dots,n$ and $n \geq 1$. Therefore, as in the proof of (a), $T_{s,t,n}(\lambda)$ is bounded on $[-\pi,\pi]$ uniformly in $s,t = 1,\dots,n$ and $n \geq 1$ except at the indeterminate point $\lambda = 0$. By the same computational method as in (2.5), using bound (2.7) again, we can observe for a large n that

$$\lim_{\lambda\to 0} T_{s,t,n}(\lambda) = \frac{1}{2}\left|2nt - t(s-1) - 2n\frac{\sum_{u=1}^{t} L(u)}{L(n)}\right| = O(n^2)$$

uniformly in $s,t = 1,\ldots,n$. Returning to the integral in (2.6), we split the area of the integral $I_n = (-\delta\varepsilon(n)^2/n, \delta\varepsilon(n)^2/n)$ and $J_n = [-\pi,\pi]\backslash A_n$ for any fixed $\delta > 0$. Then we have

$$\begin{aligned}\int_{-\pi}^{\pi} T_{s,t,n}(\lambda)d\lambda &= \int_{I_n} T_{n,s,t}(\lambda)d\lambda + \int_{J_n} T_{n,s,t}(\lambda)d\lambda \\ &= \frac{\delta\varepsilon(n)^2}{n}O(n^2) + O(1) = o(n\varepsilon(n)^2)\end{aligned} \tag{2.8}$$

uniformly in $s,t = 1,\ldots,n$. From (2.6) and (2.8), we obtain the result that $\left|\tilde{h}_n(t)\right| = o(nL(n)\varepsilon(n)^2)$ uniformly in $t = 1,\ldots,n$. ■

2.5.2 Proofs of the results

Proof of Lemma 2 We can write

$$X'\Gamma X = \begin{bmatrix} \iota'\Gamma\iota & \iota'\Gamma L \\ \iota'\Gamma L & L'\Gamma L \end{bmatrix} \quad \text{and} \quad \sigma_n^2 = \frac{1}{n}\sum_{s=1}^{n}\sum_{t=1}^{n}\gamma_{s-t}$$

and first show that

(i) $\iota'\Gamma\iota = \sigma_n^2 n,$

(ii) $\iota'\Gamma L = \sigma_n^2\iota' L + o\left(nL(n)\varepsilon(n)^2\right),$

(iii) $L'\Gamma L = \sigma_n^2 L' L + o\left(nL(n)^2\varepsilon(n)^2\right).$

If these equations are true, the variance leads to

$$
\begin{aligned}
&\mathrm{Var}(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\Gamma X(X'X)^{-1} \\
&= \begin{bmatrix} n & \iota' L \\ \iota' L & L'L \end{bmatrix}^{-1} \begin{bmatrix} \iota'\Gamma\iota & \iota'\Gamma L \\ L'\Gamma\iota & L'\Gamma L \end{bmatrix} \begin{bmatrix} n & \iota' L \\ L'\iota & L'L \end{bmatrix}^{-1} \\
&= \frac{1}{(nL'L-(\iota'L)^2)^2} \begin{bmatrix} L'L & -\iota'L \\ -L'\iota & n \end{bmatrix} \\
&\quad \times \sigma_n^2 \begin{bmatrix} n & \iota'L + o(nL(n)\varepsilon(n)^2) \\ L'\iota + o(nL(n)\varepsilon(n)^2) & L'L + o(nL(n)^2\varepsilon(n)^2) \end{bmatrix} \begin{bmatrix} L'L & -\iota'L \\ -L'\iota & n \end{bmatrix} \\
&= \frac{\sigma_n^2}{n^4L(n)^4\varepsilon(n)^4(1+o(1))} \\
&\quad \times \begin{bmatrix} n^3L(n)^4\varepsilon(n)^2(1+o(1)) & -n^3L(n)^3\varepsilon(n)^2(1+o(1)) \\ -n^3L(n)^3\varepsilon(n)^2(1+o(1)) & n^3L(n)^2\varepsilon(n)^2(1+o(1)) \end{bmatrix} \\
&= \frac{\sigma_n^2}{1+o(1)} \begin{bmatrix} \dfrac{1}{n\varepsilon(n)^2} & -\dfrac{1}{nL(n)\varepsilon(n)^2} \\ -\dfrac{1}{nL(n)\varepsilon(n)^2} & \dfrac{1}{nL(n)^2\varepsilon(n)^2} \end{bmatrix} (1+o(1)),
\end{aligned}
$$

where we have used the fact that $nL'L-(\iota'L)^2 = n^2L(n)^2\varepsilon(n)^2(1+o(1))$ by Lemma 1, and the result follows. Thus, it suffices to prove (i), (ii) and (iii).

(i) It is trivial by the definition of σ_n^2.

(ii) Define $g_n(t) = \sum_{s=1}^n \gamma_{s-t} - \sigma_n^2$ and $\tilde{g}_n(t) = \sum_{u=1}^t g_n(u)$. Then, we see that

$$
\begin{aligned}
\left|\iota'\Gamma L - \sigma_n^2\iota'L\right| &= \left|\sum_{t=1}^n L(t)\left(\sum_{s=1}^n \gamma_{s-t} - \sigma_n^2\right)\right| = \left|\sum_{t=1}^n L(t)g_n(t)\right| \\
&= \left|L(n)\tilde{g}_n(n) - \sum_{t=2}^n \tilde{g}_n(t-1)\Delta L(t)\right| \\
&\le \sum_{t=2}^n |\tilde{g}_n(t-1)|\,\Delta L(t),
\end{aligned}
\tag{2.9}
$$

where the third equality holds by summation by parts, and the last inequality follows from $L(n)\tilde{g}_n(n) = 0$ and the monotonicity of $L(t)$ with the triangle inequality. By Lemma 5

(a), $\max_{1\le t\le n}|\tilde{g}_n(t)| \le C$ for some C. Thus, the last term of (2.9) is $O(L(n))$, which is $o(nL(n)\varepsilon(n)^2)$.

(iii) It can be proved in a similar way. Define $h_n(t) = \sum_{s=1}^n L(s)\gamma_{s-t} - \sigma_n^2 L(t)$ and $\tilde{h}_n(t) = \sum_{u=1}^t h_n(u)$. Then, we see that

$$\begin{aligned}
\left|L'\Gamma L - \sigma_n^2 L'L\right| &= \left|\sum_{t=1}^n L(t)\left(\sum_{s=1}^n L(s)\gamma_{s-t} - \sigma_n^2 L(t)\right)\right| = \left|\sum_{t=1}^n L(t)h_n(t)\right| \\
&= \left|L(n)\tilde{h}_n(n) - \sum_{t=2}^n \tilde{h}_n(t-1)\Delta L(t)\right| \\
&\le L(n)\left|\tilde{h}_n(n)\right| + \sum_{t=2}^n \left|\tilde{h}_n(t-1)\right|\Delta L(t) \\
&= A(n) + B(n), \quad \text{say,}
\end{aligned}$$

where the third equality holds by summation by parts, and the inequality follows from the monotonicity of $L(t)$ and the triangle inequality. Because of the symmetry of γ_{s-t} and the result of (ii), the first term $A(n)$ is evaluated by

$$\begin{aligned}
A(n) = L(n)\left|\tilde{h}_n(n)\right| &= L(n)\left|\sum_{t=1}^n\left(\sum_{s=1}^n L(s)\gamma_{s-t} - \sigma_n^2 L(t)\right)\right| \\
&= L(n)\left|\sum_{t=1}^n L(t)\left(\sum_{s=1}^n \gamma_{s-t} - \sigma_n^2\right)\right| = L(n)\left|\sum_{t=1}^n L(t)g_n(t)\right| = O(L(n)^2).
\end{aligned} \tag{2.10}$$

Owing to Lemma 5 (b), the second term $B(n)$ reduces to

$$B(n) = \sum_{t=2}^n \left|\tilde{h}_n(t-1)\right|\Delta L(t) = o(nL(n)\varepsilon(n)^2)O(L(n)) = o(nL(n)^2\varepsilon(n)^2). \tag{2.11}$$

From (2.10) and (2.11), $\left|L'\Gamma L - \sigma_n^2 L'L\right|$ is found to be $o(nL(n)^2\varepsilon(n)^2)$. ■

Proof of Lemma 3 Let γ^{s-t} denote the (s,t)th element of Γ^{-1} and define

$$\omega_n^2 = \frac{1}{n}\sum_{s=1}^n\sum_{t=1}^n \gamma^{s-t}.$$

First, it should be proved that

(i) $\iota'\Gamma^{-1}\iota = \omega_n^2 n$,

(ii) $\iota'\Gamma^{-1}L = \omega_n^2\iota'L + o\left(nL(n)\varepsilon(n)^2\right)$,

(iii) $L'\Gamma^{-1}L = \omega_n^2 L'L + o\left(nL(n)^2\varepsilon(n)^2\right)$.

These can be shown from the proof of Lemma 2 as long as the spectral density of γ^{s-t} is continuous on $[-\pi,\pi]$. However, this is true for a sufficiently large n because of the fact in Shaman (1975) that Γ^{-1} is asymptotically replaced by the matrix Γi, whose (s,t)th element is defined by

$$\gamma i_{s-t} = \frac{1}{(2\pi)^2}\int_{-\pi}^{\pi} e^{i(s-t)\lambda}\frac{1}{f(\lambda)}d\lambda.$$

Here, $f(\lambda)^{-1}$ is well-defined since f is positive on $[-\pi,\pi]$ by Assumption 2, implying that the spectral density of γ^{s-t} is asymptotically given by $(2\pi)^{-2}f(\lambda)^{-1}$. Because Assumption 2 ensures its continuity, (i)–(iii) hold.

Utilizing (i), (ii) and (iii), we have

$$\begin{aligned}
\mathrm{Var}(\hat{\beta}_{GLS}) = (X'\Gamma^{-1}X)^{-1} &= \begin{bmatrix} \iota'\Gamma^{-1}\iota & \iota'\Gamma^{-1}L \\ L'\Gamma^{-1}\iota & L'\Gamma^{-1}L \end{bmatrix}^{-1} \\
&= \frac{\omega_n^{-2}}{nL'L-(\iota'L)^2+o(n^2L(n)^2\varepsilon(n)^2)} \\
&\qquad \times \begin{bmatrix} L'L+o(nL(n)^2\varepsilon(n)^2) & -\iota'L+o(nL(n)\varepsilon(n)^2) \\ -L'\iota+o(nL(n)\varepsilon(n)^2) & n \end{bmatrix} \\
&= \frac{\omega_n^{-2}}{n^2L(n)^2\varepsilon(n)^2(1+o(1))}\begin{bmatrix} nL(n)^2(1+o(1)) & -nL(n)(1+o(1)) \\ -nL(n)(1+o(1)) & n \end{bmatrix} \\
&= \frac{\omega_n^{-2}}{1+o(1)}\begin{bmatrix} \dfrac{1}{n\varepsilon(n)^2} & -\dfrac{1}{nL(n)\varepsilon(n)^2} \\ -\dfrac{1}{nL(n)\varepsilon(n)^2} & \dfrac{1}{nL(n)^2\varepsilon(n)^2} \end{bmatrix}(1+o(1)).
\end{aligned}$$

Finally, we specify the asymptotic form of ω_n^2, but we know that the value converges to the long-run variance, or 2π times the spectrum evaluated at the origin. Therefore, it follows that

$$\omega_n^2 \to 2\pi\frac{1}{4\pi^2 f(0)} = \frac{1}{2\pi f(0)} = \frac{1}{\sigma^2},$$

which yields the desired result. ■

Chapter 3

Testing for a Unit Root in the Presence of a Slowly Varying Regressor

This chapter considers the unit root model with a slowly varying (SV) regressor. This regressor is known to be asymptotically collinear with the constant term, so a standard asymptotic theory is not directly applied. In this chapter, the estimated regression coefficients of the constant term and SV regressor are proved to be asymptotically normal, but neither is consistent if the error term has a unit root. Further, we derive the limiting distribution of the unit root test statistic. Because of the influence of the collinear regressor, however, the finite sample approximation turns out to be poor and asymptotic tests seem to be impracticable. To overcome this difficulty, we recommend dropping the constant term intentionally from the regression and constructing the statistics. This procedure provides a consistent estimator even if the true model has the constant term. The powers and sizes of these statistics are found to be significantly improved.

3.1 Introduction

So far, considerable research on deterministic time-trending models has been produced. Focusing on the formulations of trends, most of these studies (e.g., Vogelsang (1998)) employed a general vector of time-trending regressors like polynomials, but ruled out regressors that are asymptotically collinear with a constant term like slowly varying (SV) regressors. Nevertheless, there are many application examples of models with such regressors, including the log-periodogram regression of long memory (Robinson (1995), Hurvich et al. (1998), Phillips (1999) and references therein), nonlinear least squares estimation (Wu (1981), Phillips (2007), Mynbaev (2011)), and the study of growth convergence (Barro and Sala-i-Martin (2004)). Focusing on economic convergence and transition modeling, Phillips and Sul (2007, 2009) designed a new model that represents the behavior of economies in transition and proposed an associated test for convergence, utilizing SV functions explicitly.

Given this background, Phillips (2007) established the theory on stationary models possessing slowly varying (SV) regressors, which is now reviewed briefly. The typical model is given by the logarithmic trend model

$$y_t = \alpha + \beta \log t + v_t \quad \text{for } t = 1, \ldots, n, \tag{3.1}$$

where $\{v_t\}$ is stationary and supposed to satisfy some regularity conditions. If the scaling matrix F_n^{-1} is given by $\text{diag}[n^{1/2}\log^{-1} n, n^{1/2}]$, then the scaled sample covariance matrix $(X'X)^{-1}$, where $X = [X_1', \ldots, X_n']'$ with $X_t = [1, \log t]$, behaves like

$$F_n^{-1}(X'X)^{-1}F_n^{-1} \to \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \tag{3.2}$$

indicating the asymptotic collinearity of the regressors and singularity of the limiting matrix. Phillips (2007) also demonstrated that the OLS estimator $(\hat{\alpha}_n, \hat{\beta}_n)$ is consistent and asymptotically normally distributed, but the convergence rate is affected by the presence of the logarithmic trend. In view of asymptotic theories, Phillips (2007) relied on uniform strong approximation of partial sums by Brownian motions, but the condition was rather restrictive and the proof is partially insufficient in dealing with SV regressors. Mynbaev (2009) then applied the central limit theorem (CLT) based on the L_p-*approximation* technique to make the proof rigorous under less stringent conditions. This sophisticated idea is utilized in this chapter, and is reviewed in Section 3.2 and 3.6.1.

The present chapter extends Phillips' stationary model to the integrated one and investigates the properties of unit root tests, but a few problems arise. First, the OLS estimator of the trend coefficient $\hat{\beta}_n$ is inconsistent, as is that of the constant term $\hat{\alpha}_n$. This means that the regression is almost meaningless if the model has a unit root (Section 3.3.1). This phenomenon may be understood to mean that an SV regressor is classified into a constant in the asymptotic sense, and this feature is amplified under integrated errors. Of course, an analysis like that of Canjels and Watson (1997) no longer makes sense. We must then emphasize the necessity of unit root tests when it comes to employing an SV regressor. Second, when we test a null of a unit root, if we construct a Phillips and Perron (PP)-type test statistic, the finite sample distribution hardly approaches the limiting one. This makes it difficult to test a unit root on the basis of the limiting critical values (Section 3.4.1). We present a solution to this problem by using the misspecified regression model, in which we intentionally drop a constant term. Such an intentionally misspecified procedure is asymptotically justified even if the true model has a constant term. This manipulation brings about a significant improvement in terms of size and power in finite sample situations (Section 3.4.2).

The rest of this chapter is as follows. Section 3.2 includes some assumptions and preliminary theories for SV functions based on the results obtained by Phillips (2007) and Mynbaev (2009). Section 3.3 and 3.4 state the main analytical results; the limiting distributions of the OLS estimator and the unit root test statistics are derived in Section 3.3, and 3.4 studies the finite sample properties of the unit root test statistics derived in Section 3.3 through Monte Carlo simulations involving a procedure to improve the performance of the tests. A general PP-type test statistic is also presented. Section 3.5 concludes. Section 3.6 is the appendix; Section 3.6.1 provides a summary of L_p-approximability, Section 3.6.2 gives lemmas used in the proofs, and the proofs of the analytical results derived in Sections 3.3 and 3.4 are given in Section 3.6.3.

3.2 Assumptions and Preliminary Results

Our main objective in the chapter is to analyze the regression model with an SV regressor under a unit root assumption. For this purpose, we start with a discussion about SV functions and then give assumptions on the error term, including a review on the corresponding asymptotic theory based on Phillips (2007) and Mynbaev (2009, 2011).

3.2.1 Slowly varying regressor

A positive function L on $[A,\infty)$, $A>0$ is called *slowly varying* (SV) if it satisfies, for any $r>0$, $L(rx)/L(x)\to 1$ as $x\to\infty$. To deal with such an SV function L, *Karamata's representation theorem* is well-known and essential. That is, the function L varies slowly if and only if it is written in the form

$$L(x)=c(x)\exp\left(\int_B^x \frac{\varepsilon(s)}{s}ds\right) \quad \text{for} \quad x\geq B\geq A \tag{3.3}$$

for some $B>0$, where $c(x)\to c\in(0,\infty)$ and $\varepsilon(x)\to 0$ as $x\to\infty$. Considering regression theory, however, we require a stronger assumption on L in terms of its smoothness and behavior for a large x.

Definition 1 We say $L=K(\varepsilon,\phi_\varepsilon)$ if the function L satisfies all the conditions below:

(a) The function L is SV and has Karamata's representation

$$L(x)=c\exp\left(\int_B^x \frac{\varepsilon(s)}{s}ds\right) \quad \text{for} \quad x\geq B$$

for some $B>0$. Here $c>0$, ε is continuous and $\varepsilon(x)\to 0$ as $x\to\infty$. Hereafter, this part of the assumption is shortened to $L=K(\varepsilon)$.

(b) The function $|\varepsilon|$ is SV.

(c) There exists a function ϕ_ε on $[0,\infty)$ called a *remainder* that satisfies the following properties,

- The function ϕ_ε is positive, nondecreasing on $[0,\infty)$, $\phi_\varepsilon(x) \to \infty$, and there exist positive numbers θ and X such that $x^{-\theta}\phi_\varepsilon(x)$ is nonincreasing on $[X,\infty)$.
- There exists a positive constant c satisfying

$$\frac{1}{c\phi_\varepsilon(x)} \leq |\varepsilon(x)| \leq \frac{c}{\phi_\varepsilon(x)} \quad \text{for} \quad x \geq c.$$

Assumption 1 $L = K(\varepsilon, \phi_\varepsilon)$ and $\varepsilon = K(\eta, \phi_\eta)$.

Remark 1 Conditions (a) and (b) are more restrictive assumptions than in Karamata's representation (3.3), but they also appeared in Phillips (2007) and Mynbaev (2009) to cope with asymptotics. Mynbaev (2009) introduced condition (c) to ensure that the asymptotic analysis of the regressions was more rigorous. Many SV functions, including all the $L(x)$ tabulated in Table 3.1, possess the remainder $\phi_\varepsilon(x) = 1/|\varepsilon(x)|$. For further discussion of SV with a remainder, see Mynbaev (2009, 2011) and Bingham, Goldie and Teugels (1987) in Sections 2.3 and 3.12. Note that any SV function is of order $o(n^\alpha)$ for all $\alpha > 0$.

Under Assumption 1, we have an important result that is useful for deriving asymptotic results as follows:

$$\varepsilon(n) = \frac{nL'(n)}{L(n)} \to 0 \quad \text{and} \quad \eta(n) = \frac{n\varepsilon'(n)}{\varepsilon(n)} \to 0 \quad \text{as} \quad n \to \infty. \tag{3.4}$$

This is easily obtained by the representation theorem. Consequently, (3.4) produces some examples of $L(t)$ in Table 3.1. Conversely, typical SV functions L in Table 3.1 satisfy Assumption 1. Another application of Assumption 1 leads to the following lemma due to Phillips (2007).

Lemma 1 *If Assumption 1 is satisfied, we have*

$$\frac{1}{n}\sum_{t=1}^{n} L(t)^k = L(n)^k \left(1 - k\varepsilon(n)[1+o(1)]\right).$$

Lemma 1 is used for asymptotic expansion of the sum of SV functions to evaluate the limiting behavior of estimators and is frequently employed in the proofs of our results.

We next make a more restrictive assumption on L.

Assumption 2 L is monotonically increasing.

When unit root test statistics are considered in Section 3.3.2 and 3.4, Assumption 2 is supposed in addition to Assumption 1 in order to avoid a cumbersome argument on SV regressors and improve the outlook for the proofs.

3.2.2 Disturbances and preliminary results on asymptotics

Next, we impose an assumption on the error term and review the CLT of a weighted sum of a linear process achieved by Mynbaev, which is stated in Theorem 1 later. The CLT works well under the assumption below, but is slightly stronger than necessary.

Assumption 3 The error sequence $\{v_t\}$ is modeled by the linear process

$$v_t = C(L)e_t = \sum_{j=0}^{\infty} c_j e_{t-j}, \quad \sum_{j=0}^{\infty} j|c_j| < \infty, \quad C(1) \neq 0,$$

where $\{e_t\}$ is a sequence of martingale difference random variables with respect to the natural filtration $\mathscr{F}_{t-1}$. Moreover, the sequence $\{e_t^2\}$ is uniformly integrable and $E[e_t^2|\mathscr{F}_{t-1}] = \sigma_e^2 < \infty$ for all t.

Remark 2 (a) In order to make a functional CLT (FCLT) that holds for the process $\{v_t\}$ as well as Mynbaev's CLT, Assumption 3 requires the one-summability of the coefficients $\{c_j\}$. In fact, the CLT holds with the absolutely-summable condition, which is less restrictive than the assumed one-summable condition. Meanwhile, the uniform integrability of $\{e_t^2\}$ is necessary for the CLT although it is redundant for the FCLT. Assumption 3 thus makes unit root asymptotics with an SV regressor tractable in that it enables us to deal with the FCLT and CLT simultaneously in subsequent sections. For further information on an FCLT and a CLT of (a weighted sum of) a linear process, see Phillips and Solo (1992) and Mynbaev (2009, 2011), respectively.

(b) A sufficient condition for the uniform integrability of $\{e_t^2\}$ is $E|e_t|^p < \infty$ for some $p > 2$.

Let $\bar{x}$ denote the sample mean of some variables $x_1, \ldots, x_n$. Under Assumption 3, we let $\sigma_L^2 = \lim_{n\to\infty} \mathrm{Var}(n^{1/2}\bar{v}) = \sigma_e^2 C(1)^2$ and $\sigma_S^2 = \mathrm{Var}(v_t) = \sigma_e^2 \sum_{j=0}^{\infty} c_j^2$; these are long-run and short-run variance, respectively. If we suppose $\{v_t\}$ to be a sequence of serially uncorrelated random variables, we know that $\sigma_L^2 = \sigma_S^2$. This simplification is used in simulation studies to exclude the effect of the long-run variance estimation error and focus on the influence of an SV regressor.

As is mentioned in Remark 2, Assumption 3 is sufficient for Mynbaev's CLT for a weighted sum of linear processes. The key concept for the CLT is called *L_p-approximability* (or *L_p-closeness*) of the weight to the continuous argument, and is reviewed in Appendix A. The following theorem yields the CLT as in Mynbaev (2009, 2011).

Theorem 1 *Let $\{v_t\}$ satisfy Assumption 3 and a sequence of weights $w_n = (w_{n1}, \ldots, w_{nn})$ be L_2-close to $f \in L_2$. Then, as $n \to \infty$, we have*

$$\sum_{t=1}^{n} w_{nt} v_t \xrightarrow{d} N\left(0, \sigma_L^2 \int_0^1 f(r)^2 dr\right).$$

We then prepare for the asymptotic analysis of regression with an SV regressor. Utilizing Theorem 1, Mynbaev (2009) gave a rigorous proof for the following lemma.

Lemma 2 *Under Assumptions 1 and 3, we have, as $n \to \infty$,*

$$(i) \quad \frac{1}{n^{1/2}L(n)} \sum_{t=1}^{n} L(t) v_t \xrightarrow{d} N(0, \sigma_L^2),$$

$$(ii) \quad \frac{1}{n^{1/2}L(n)\varepsilon(n)} \sum_{t=1}^{n} (L(t) - \bar{L}) v_t \xrightarrow{d} N(0, \sigma_L^2).$$

Lemma 2 was first proved by Phillips (2007) by way of the strong approximation of the partial sum process to the Brownian motion, although Mynbaev (2009) pointed out that it was partially incomplete. If we consider the following regression model with stationary errors,

$$y_t = \alpha + \beta L(t) + v_t \quad \text{for} \quad t = 1, \ldots, n, \tag{3.5}$$

then the asymptotic distribution of the OLS estimator $(\hat{\alpha}_n, \hat{\beta}_n)$ is obtained under Assumptions 1 and 3; see Phillips (2007) and Mynbaev (2009, 2011) for the proof and discussion.

Theorem 2 *If all the assumptions in Lemma 2 are satisfied, then, as $n \to \infty$, the OLS estimator of model (3.5) has the following limiting distribution:*

$$n^{1/2} \begin{bmatrix} \varepsilon(n)(\hat{\alpha}_n - \alpha) \\ L(n)\varepsilon(n)(\hat{\beta}_n - \beta) \end{bmatrix} \xrightarrow{d} N\left(0, \sigma_L^2 \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\right).$$

We may observe the singularity of the asymptotic covariance matrix in Theorem 2. This is caused by the asymptotic collinearity of the regressors demonstrated in (3.2).

Instead of Assumption 3, we adopt the following assumption in the rest of this chapter in order to consider a unit root case.

Assumption 4 The process $\{u_t\}$ possesses a unit root under the null hypothesis $\rho = 1$ in $u_t = \rho u_{t-1} + v_t$, where $\{v_t\}$ is the same linear process as in Assumption 3.

3.3 Limit Distributions

The main purpose of this section is to reveal the asymptotic behavior of the OLS estimator of the model with an SV regressor in the presence of a unit root. Specifically, the following regression model is considered:

$$y_t = \alpha + \beta L(t) + u_t, \quad t = 1, \ldots, n, \tag{3.6}$$

where the SV regressor $L(t)$ and disturbance u_t are supposed to satisfy Assumptions 1 and 4, respectively. The limiting distribution of the OLS estimator $(\hat{\alpha}_n, \hat{\beta}_n)$ is obtained first in Section 3.3.1, and the limit behavior of unit root test statistics is analyzed in Section 3.3.2.

3.3.1 Limit distribution of the OLS estimator

The derivation of the limiting distribution of the OLS estimator $(\hat{\alpha}_n, \hat{\beta}_n)$ from model (3.6) requires a lemma that is an $I(1)$ analogue of Lemma 2. In fact, the next lemma provides the foundation for the rest of the chapter. Let $W(\cdot)$ denote the standard Brownian motion obtained in the limit of the partial sum process $\sigma_L^{-1} n^{-1/2} \sum_{t=1}^{[\cdot n]} v_t$.

Lemma 3 *Under Assumptions 1 and 4, we have, as $n \to \infty$,*

$$(i) \quad \frac{1}{n^{3/2} L(n)} \sum_{t=1}^{n} L(t) u_t \xrightarrow{d} \sigma_L \int_0^1 W(r) dr,$$

$$(ii) \quad \frac{1}{n^{3/2} L(n) \varepsilon(n)} \sum_{t=1}^{n} (L(t) - \bar{L}) u_t \xrightarrow{d} \sigma_L \int_0^1 (1 + \log r) W(r) dr.$$

Remark 3 (a) The limiting distributions of (i) and (ii) in Lemma 3 turn out to be $N(0, \sigma_L^2/3)$ and $N(0, 2\sigma_L^2/27)$, respectively, by simple calculation. Furthermore, the integral in (ii) is equivalent to $\int_0^1 (W(r) - F_1(r)/r) dr$, where $F_1(r) = \int_0^r W(s) ds$ is the one-folded integrated Brownian motion.

(b) In order to handle unit root test statistics later, Lemma 3 clarifies not only the limit laws but the form of the limit random variables by an application of both Mynbaev's CLT and the FCLT for linear processes. Result (i) is indeed obtained by proving the asymptotic equivalence of the left-hand side of (i) and $n^{-3/2} \sum_{t=1}^{n} u_t$, which is achieved by an application of Theorem 1 that shows that the weight of the difference between them is L_2-close to zero. Then, (i) follows because $n^{-3/2} \sum_{t=1}^{n} u_t$ converges in distribution to the right-hand side of (i) by the FCLT. Result (ii) also holds by a similar manipulation.

According to Lemma 1, the sum $\sum_{t=1}^{n} (L(t) - \bar{L})^2$ is approximated by $nL(n)^2 \varepsilon(n)^2 [1 + o(1)]$ for a large n. Thus, a direct application of Lemma 3 with the fact in Remark 3 (a) gives the limiting distribution of the OLS estimator.

Theorem 3 *If all the assumptions in Lemma 3 are satisfied, then, as $n \to \infty$, the OLS estimator of model (3.6) has the following limiting distribution:*

$$n^{-1/2}\begin{bmatrix} \varepsilon(n)\left(\hat{\alpha}_n-\alpha\right) \\ L(n)\varepsilon(n)\left(\hat{\beta}_n-\beta\right) \end{bmatrix} \xrightarrow{d} N\left(0, \frac{2\sigma_L^2}{27}\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\right).$$

Remark 4 It should be emphasized that, because any SV function possesses the asymptotic order of $o(n^{1/2})$, the OLS estimators, $\hat{\alpha}_n$ and $\hat{\beta}_n$, cannot be consistent. This result contrasts with the case where the simple trend t is employed. Considering models with an SV regressor, we therefore remark that the existence of a unit root leads to a meaningless regression and that testing for a unit root is thus essential.

3.3.2 Limit distribution of the unit root test statistic

Using the results in Theorem 3, we then derive the limiting distribution of the OLS estimator $\hat{\rho}_n$ in the residual-based regression of $\hat{u}_t$ on $\hat{u}_{t-1}$ with $\rho = 1$ under Assumption 4. These residuals $\hat{u}_t$ are obtained from the regression model (3.6), so that we have

$$\hat{u}_t = u_t - \bar{u} - (L(t)-\bar{L})\left(\hat{\beta}_n-\beta\right) \quad \text{for } t = 1,\ldots,n. \tag{3.7}$$

From (3.7), the scaled OLS estimator $\hat{\rho}_n$ is obtained by

$$\begin{aligned} n(\hat{\rho}_n-1) &= \frac{1}{n}\sum_{t=2}^{n}\hat{u}_{t-1}\left(\hat{u}_t-\hat{u}_{t-1}\right) \Bigg/ \left[\frac{1}{n^2}\sum_{t=2}^{n}\hat{u}_{t-1}^2\right] \\ &= \left[\frac{1}{2n}(\hat{u}_n^2-\hat{u}_1^2) - \frac{1}{2n}\sum_{t=2}^{n}(\hat{u}_t-\hat{u}_{t-1})^2\right] \Bigg/ \left[\frac{1}{n^2}\sum_{t=1}^{n}\hat{u}_t^2 - \frac{1}{n^2}\hat{u}_n^2\right], \end{aligned}$$

where the scale coefficient n is tentative. The asymptotic behavior depends on the four terms $\hat{u}_n^2$, $\hat{u}_1^2$, $\sum_{t=2}^{n}(\hat{u}_t-\hat{u}_{t-1})^2$ and $\sum_{t=1}^{n}\hat{u}_t^2$. Their asymptotic behaviors are presented in the following lemma.

Lemma 4 *If Assumptions 1, 2 and 4 are satisfied in model (3.6), we have, as $n \to \infty$,*

$$(i) \quad \frac{\varepsilon(n)^2}{n}\hat{u}_n^2 = o_p(1),$$

$$(ii) \quad \frac{\varepsilon(n)^2}{n}\hat{u}_1^2 \xrightarrow{d} \sigma_L^2 U_1,$$

$$(iii) \quad \frac{\varepsilon(n)^2}{n}\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2 = o_p(1),$$

$$(iv) \quad \frac{1}{n^2}\sum_{t=1}^{n}\hat{u}_t^2 \xrightarrow{d} \sigma_L^2 V_1,$$

where

$$U_1 = \left\{\int_0^1 (1+\log r)W(r)dr\right\}^2 \quad \text{and}$$

$$V_1 = \int_0^1 W(r)^2 dr - \left(\int_0^1 W(r)dr\right)^2 - \left\{\int_0^1 (1+\log r)W(r)dr\right\}^2.$$

Remark 5 From Lemma 4, the dominated term is found not to be $\hat{u}_n^2$, but $\hat{u}_1^2$. This can be accounted for by (3.7). That is, $\hat{u}_1$ includes the difference $L(1) - \bar{L} = O(L(n))$, whereas $\hat{u}_n$ has the difference $L(n) - \bar{L} = O(L(n)\varepsilon(n))$, which is indeed smaller than $O(L(n))$.

Since the dominated term is found to be $\hat{u}_1^2$ by Lemma 4, we conclude that the limiting distribution of the unit root coefficient test statistic, $n\varepsilon(n)^2(\hat{\rho}_n - 1)$, is given by the next theorem.

Theorem 4 *If all the assumptions in Lemma 4 are satisfied, then, as $n \to \infty$, it follows that*

$$n\varepsilon(n)^2(\hat{\rho}_n - 1) \xrightarrow{d} -\frac{U_1}{2V_1}.$$

Since the limiting distribution in Theorem 4 is free from the nuisance parameters σ_S^2 and σ_L^2, it appears to be manageable. However, it will turn out to be useless because the finite sample approximation is poor; see Section 3.4.1. The corresponding scaled t-statistic $\varepsilon(n)^2 t_{\hat{\rho}_n} = \varepsilon(n)^2(\hat{\rho}_n - 1)/s.e.(\hat{\rho}_n)$ for testing the null hypothesis, $H_0 : \rho = 1$, is also obtained.

Corollary 1 *If all the assumptions in Lemma 4 are satisfied, then, as $n \to \infty$, it follows that*

$$\varepsilon(n)^2 t_{\hat{\rho}_n} \xrightarrow{d} -\frac{\sigma_L}{\sigma_S}\frac{U_1}{2\sqrt{V_1}}.$$

Remark 6 The test statistic $t_{\hat{\rho}_n}$ in Corollary 1 requires the computation of $s.e.(\hat{\rho}_n)$ or $\hat{\sigma}_S^2$, which is a consistent estimator of $\sigma_S^2 = \mathrm{Var}(v_t)$. However, the natural candidate estimator, $n^{-1}\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2$, is not consistent for σ_S^2, but has a nondegenerate limit distribution, as in Lemma 4 (iii). We will suggest a method to construct a consistent estimator, $\hat{\sigma}_S^2$, in Remark 7.

3.4 Properties of the Unit Root Test Statistics

Theorem 4 and Corollary 1 have yielded the limiting distributions of the unit root regression coefficient and corresponding t-statistic. In Section 3.4.1, we examine the effect of the SV regressor on these statistics by simulation studies, but find that there is a serious problem in testing. In Section 3.4.2, we remedy the problem. Finally, the general test statistics are proposed in Section 3.4.3.

3.4.1 Finite sample behaviors

We first observe the finite sample behaviors of the simulated cumulative distribution functions (CDFs) of the statistics investigated in Theorem 4 and Corollary 1. To do this, we employ Assumption 4, but restrict $\{v_t\}$ to a sequence of i.i.d. Gaussian random variables with mean zero and variance unity to exclude the influence caused by the long-run variance estimation. If we need to embody the form of the regressor $L(t)$ for simulation studies, we always use $\log t$. The number of replications is 10000 unless otherwise noted.

Figures 3.1 and 3.2 show the finite sample and limiting CDFs of the two statistics. For each figure, the finite sample CDFs are expressed in dotted, dashed and solid lines for each sample size 100, 500 and 50000, and the limiting CDF is expressed in bold lines, respectively.

These figures indicate that the approaching manner of the finite sample CDFs to the limiting ones is not monotonic. That is, finite sample CDFs approach the limiting ones from the left to the right in their upper tails first. Then, in their lower tails, the CDFs are attracted from below to the above limits more slowly. As a consequence, fatal size distortions are provoked in unit root testing based on the limiting critical values in their lower tails. Table 3.2 shows the percentage points of the limiting distributions. From these graphs and the limiting percentage points, we are convinced that 5% empirical sizes are very close to zero even if the sample size is 50000. This means that tests based on them are almost impossible.

So far, we have assumed that the regression model contains a constant term. In this case, of course, the OLS estimator includes the sum of the squared deviations from the sample mean, or $n^{-1}\sum_{t=1}^{n}(L(t)-\bar{L})^2$. In view of Lemma 1, the first and second leading terms asymptotically offset each other and only the third term survives. Therefore, this asymptotic order is $O(L(n)^2\varepsilon(n)^2)$, which fluctuates greatly. One solution may be obtained by avoiding such computation.

3.4.2 Improvement of the finite sample performance

In this subsection, we first consider the regression without the constant term. The model is defined as

$$y_t = \beta L(t) + u_t \quad \text{for } t = 1, \ldots, n, \tag{3.8}$$

with the error term u_t satisfying Assumption 4. The following results are parallel to those in the preceding section.

Theorem 5 *If all the assumptions in Lemma 3 are satisfied, then, as $n \to \infty$, the OLS estimator of model (3.8) has the following limiting distribution:*

$$\frac{L(n)}{n^{1/2}}(\hat{\beta}_n - \beta) \xrightarrow{d} N\left(0, \frac{\sigma_L^2}{3}\right).$$

Lemma 5 *If all the assumptions in Lemma 4 are satisfied in model (3.8), we have, as $n \to \infty$,*

$$(i) \quad \frac{1}{n}\hat{u}_n^2 \xrightarrow{d} \sigma_L^2 U_2,$$

$$(ii) \quad \frac{1}{n}\hat{u}_1^2 = o_p(1),$$

$$(iii) \quad \frac{1}{n}\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2 \xrightarrow{p} \sigma_S^2,$$

$$(iv) \quad \frac{1}{n^2}\sum_{t=1}^{n}\hat{u}_t^2 \xrightarrow{d} \sigma_L^2 V_2,$$

where

$$U_2 = \left\{W(1) - \int_0^1 W(r)dr\right\}^2 \quad \text{and} \quad V_2 = \int_0^1 W(r)^2 dr - \left(\int_0^1 W(r)dr\right)^2.$$

Theorem 6 *If all the assumptions in Lemma 4 are satisfied in model (3.8), we have, as $n \to \infty$,*

$$n(\hat{\rho}_n - 1) \xrightarrow{d} \frac{U_2 - \sigma_S^2/\sigma_L^2}{2V_2}.$$

Corollary 2 *If all the assumptions in Lemma 4 are satisfied in model (3.8), we have, as $n \to \infty$,*

$$t_{\hat{\rho}_n} \xrightarrow{d} \frac{\sigma_L}{\sigma_S}\frac{U_2 - \sigma_S^2/\sigma_L^2}{2\sqrt{V_2}}.$$

In practice, it may not be appropriate to suppose that the true model has no constant term. However, it is worth analyzing the situation where the true model is given by (3.6), which possesses a constant term, but the no-constant model (3.8) is employed for regression. We find that it works well asymptotically from the following theorem.

Theorem 7 *Assume that all the Assumptions in Lemma 4 are satisfied. Furthermore, we suppose that the true data-generating process (DGP) is given by (3.6), but (3.8) is employed for regression. Then we still have the same asymptotic result given in Theorem 5 with the effect of a constant term declining at the rate $O(n^{-1/2})$.*

By Theorem 7, the test statistics constructed in the same way also have the same asymptotic behavior as in Theorem 6 and Corollary 2. This suggests that even if the DGP includes a nonzero constant term, regression without it may still be beneficial under integrated errors, and may provide a good test statistic. We study the finite sample properties of such test statistics in the rest of this subsection.

Remark 7 (a) If we let $\hat{\sigma}_S^2 = n^{-1}\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2$ in Lemma 5, which is constructed on the basis of the philosophy of Theorem 7, $\hat{\sigma}_S^2$ is used as a consistent estimator of σ_S^2.
(b) If we use the regression model $y_t = \alpha + u_t$ against the true model (3.6), a similar result is obtained. In this case, however, the declining rate of the irrelevant constant term becomes $O(L(n)n^{-1/2})$, which is slightly slower than that obtained in Theorem 7, $O_p(n^{-1/2})$.

Using the no-constant model (3.8) for regression with (3.6) being the DGP (α is set to 0, 1, and 5 for the simulation), we have Table 3.3, Tables 3.4 to 3.6, Figures 3.3 to 3.8, and Figures 3.9 to 3.14 under Assumption 4 with v_t i.i.d., just as in the preceding subsection. Figures 3.3 to 3.8 display the finite sample behaviors of the CDFs. Tables 3.4 to 3.6 indicate the empirical sizes and Figures 3.9 to 3.14 show the size-adjusted powers. From Figures 3.3 to 3.8 and Tables 3.4 to 3.6, we may observe the significant improvement of their finite sample approximation although the size distortion tends to increase when α is large. We further notice that the power tends to be low when α is large and n is small. However, it is asymptotically justified that the data are shifted to intersect around the origin when $|\alpha|$ is expected to be large. In consequence, the influence of an SV regressor on the test statistics can be removed and the tests are suitable for practical use even though some slight size distortion remains.

3.4.3 Generalization of the unit root test statistics

The test statistics in the preceding subsection have nuisance parameters in their limit despite their good performances. Finally, we introduce Phillips and Perron (1988) (PP)-type statistics on the basis of the discussion in the preceding subsection.

Theorem 8 *Construct the following two statistics from the no-constant regression model (3.8) against the natural DGP (3.6):*

$$Z_{\hat{\rho}_n} = n(\hat{\rho}_n - 1) - \frac{n^2(\hat{\sigma}_L^2 - \hat{\sigma}_S^2)}{2\sum_{t=1}^n \hat{u}_{t-1}^2} \quad \text{and} \quad Z_t = \frac{\hat{\sigma}_S}{\hat{\sigma}_L} t_{\hat{\rho}_n} - \frac{n(\hat{\sigma}_L^2 - \hat{\sigma}_S^2)}{2\hat{\sigma}_L \sqrt{\sum_{t=1}^n \hat{u}_{t-1}^2}},$$

where

$$\hat{\sigma}_S^2 = \frac{1}{n}\sum_{t=2}^n (\Delta \hat{u}_t)^2 \quad \text{and} \quad \hat{\sigma}_L^2 = \hat{\sigma}_S^2 + \frac{2}{n}\sum_{j=1}^k \left(1 - \frac{j}{k+1}\right) \sum_{t=j+1}^n \Delta \hat{u}_t \Delta \hat{u}_{t-j}$$

with $k = o(n^{1/4})$. If all the Assumptions in Lemma 4 are satisfied, then, as $n \to \infty$, it follows that

$$Z_{\hat{\rho}_n} \xrightarrow{d} \frac{U_2 - 1}{2V_2} \quad \text{and} \quad Z_t \xrightarrow{d} \frac{U_2 - 1}{2\sqrt{V_2}}.$$

These limiting distributions have the same percentage points of $n(\hat{\rho}_n - 1)$ and $t_{\hat{\rho}_n}$, as shown in Table 3.3.

Remark 8 It may be possible to reduce the size distortion caused by the estimation of the long-run variance by using a method such as that of Perron and Ng (1996). However, this is beyond the scope of this chapter, and is left to future studies.

3.5 Conclusion

We have studied the model with an SV regressor in the presence of integrated errors and found three main results. First, the estimated regression coefficients are asymptotically normally distributed, but they are not consistent. We thus observe that there is a contrast between a simple time trend and an SV regressor. Second, ordinary unit root test statistics based on the residuals behave badly because of the coexistence of a constant term and an SV regressor, and it is not recommended to conduct tests based on them. Third, in spite of this difficulty, correction of sizes is possible by intensionally eliminating the constant term from the regression model. It is shown that the statistics constructed by this approach are still consistent even when the DGP has a constant term. Applying this result, we give PP-type test statistics.

3.6 Appendix

3.6.1 L_p-approximability

The key concept of the CLT in Theorem 1 is the L_p-*approximability* of the weight, and is reviewed here. Let $\|f\|_p = \left(\int_0^1 |f(x)|^p dx\right)^{1/p}$ for $p \in [1, \infty)$ and let L_p denote the space of measurable functions on $[0,1)$ satisfying $\|f\|_p < \infty$. Further, we partition the interval $[0,1)$ into subintervals $i_t = [(t-1)/n, t/n)$ for $t = 1, \ldots, n$. Then we obtain a *discretization operator* $\delta_{np} : L_p \to \mathbb{R}^n$ defined by

$$(\delta_{np} f)_t = n^{1-1/p} \int_{i_t} f(x) dx \quad \text{for} \quad t = 1, \ldots, n.$$

Let l_p denote a discrete analogue of L_p with the norm $\|w\|_p = (\sum_t |w_t|^p)^{1/p}$. Then, we say that a sequence $\{w_n\}$, $w_n \in \mathbb{R}^n$ is L_p-*approximable* (or L_p-*close*) to $f \in L_p$ if it satisfies, as $n \to \infty$,

$$\|w_n - \delta_{np} f\|_p \to 0. \tag{3.9}$$

Instead of the discretization operator δ_{np}, it is sometimes convenient to use the *interpolation operator* $\Delta_{np} : \mathbb{R}^n \to L_p$ defined by

$$\Delta_{np} w_n = n^{1/p} \sum_{t=1}^{n} w_{nt} 1_{i_t}, \tag{3.10}$$

where 1_{i_t} is the indicator of i_t. It is known that $\|\Delta_{np} w_n - f\|_p \to 0$ is equivalent to (3.9).

3.6.2 Useful lemmas

Lemma 6 *Under all the assumptions of Lemma 3 (ii) with the same notations as in that proof, we have, as $n \to \infty$,*

$$\left\| \frac{1}{n} \sum_{s=[nu]+1}^{n} \left(1+\log\frac{s}{n}\right) + u\log u \right\|_{2,(0,1)} \to 0. \tag{3.11}$$

Proof of Lemma 6 We use the same B_b as in Lemma 3 (i) and let $\delta \in (0,1/2]$ be fixed. The left-hand side of (3.11) is bounded above by

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{s=[nu]+1}^{n} \left(1+\log\frac{s}{n}\right) + u\log u \right\|_{2,(\delta,1)} \\ &\qquad + \|u\log u\|_{2,(0,\delta)} + \left\| \frac{1}{n} \sum_{s=[nu]+1}^{n} \left(1+\log\frac{s}{n}\right) \right\|_{2,(0,\delta)}. \end{aligned} \tag{3.12}$$

We then prove that the three terms are asymptotically negligible.

Consider the case of $u \in [\delta,1)$. For a large n and $r \in [\delta, 1+1/(2B_b)]$ satisfying $nr \in \mathbb{N}$, we have

$$\frac{1}{n}\sum_{s=nr}^{n}\left(1+\log\frac{s}{n}\right) = -r\log r + o(1)$$

uniformly in $r \in [\delta, 1+1/(2B_b)]$ because of the expansion of $\sum \log s$ based on Lemma 1. If we define $r = [nu+1]/n$, the same argument in Case 1 in the proof of Lemma 3 (i) leads to the convergence that

$$\frac{1}{n}\sum_{s=[nu]+1}^{n}\left(1+\log\frac{s}{n}\right) + u\log u \to 0$$

uniformly in $u \in (\delta,1)$, and hence the first term in (3.12) tends to zero.

Consider the case of $u \in [0,\delta)$. The second term in (3.12) can be made as small as desired if we choose a small $\delta > 0$. For the third term in (3.12), note that

$$\begin{aligned} \left| \frac{1}{n}\sum_{s=[nu]+1}^{n}\left(1+\log\frac{s}{n}\right) \right| &\le \left|1-\frac{[nu]}{n}\right| + \left|\frac{1}{n}\sum_{s=[nu]+1}^{n}\log\frac{[nu]+1}{n}\right| \\ &= \left|1-\frac{[nu]}{n}\right| \left\{1+\left|\log\frac{[nu]+1}{n}\right|\right\} \le 2\left\{1 \vee \left|\log\frac{[nu]+1}{n}\right|\right\}. \end{aligned}$$

In this upper bound, $\log(([nu]+1)/n)$ is identical to the G-function $G([nu]+1,n)$ in the case that $L(x)$ is given by $\log x$. Thus, Cases 2 and 3 in the proof of Theorem 4.4.1 in Mynbaev (2011) directly provide the result. ■

Lemma 7 *Under all the assumptions of Lemma 3 (ii) with the same notations as in that proof, we have, as $n \to \infty$,*

$$\left\| \frac{1}{n} \sum_{s=[nu]+1}^{n} \frac{L(s)-\bar{L}}{L(n)\varepsilon(n)} + \frac{[nu]}{n} G([nu],n) \right\|_{2,(0,1)} \to 0. \tag{3.13}$$

Proof of Lemma 7 We see that

$$\begin{aligned} \frac{1}{n} \sum_{s=[nu]+1}^{n} \frac{L(s)-\bar{L}}{L(n)\varepsilon(n)} &= \frac{1}{nL(n)\varepsilon(n)} \left(\sum_{s=1}^{n} L(s) - \sum_{s=1}^{[nu]} L(s) - n\bar{L} + [nu]\bar{L} \right) \\ &= -\frac{1}{nL(n)\varepsilon(n)} \left(\sum_{s=1}^{[nu]} L(s) - \frac{[nu]}{n} \sum_{s=1}^{n} L(s) \right) \\ &= -\frac{[nu]}{n} \left\{ G([nu],n) - \left(\frac{L([nu])}{L(n)} \frac{\varepsilon([nu])}{\varepsilon(n)} - 1 \right) + \frac{L([nu])}{L(n)\varepsilon(n)} o\left(\varepsilon([nu])\right) - o(1) \right\}, \end{aligned} \tag{3.14}$$

where the last equality follows from the expansions of the two sums based on Lemma 1 and the definition of a G-function. From (3.14) with $[nu]/n \le 1$ and the triangle inequality, the left-hand side of (3.13) reduces to and is bounded by

$$\begin{aligned} &\left\| \frac{[nu]}{n} \left\{ \left(\frac{L([nu])}{L(n)} \frac{\varepsilon([nu])}{\varepsilon(n)} - 1 \right) - \frac{L([nu])}{L(n)\varepsilon(n)} o\left(\varepsilon([nu])\right) + o(1) \right\} \right\|_{2,(0,1)} \\ &\le \left\| \left(\frac{L([nu])}{L(n)} \frac{\varepsilon([nu])}{\varepsilon(n)} - 1 \right) - \frac{L([nu])}{L(n)\varepsilon(n)} o\left(\varepsilon([nu])\right) \right\|_{2,(0,1)} + o(1) \\ &\le \left\| \frac{L([nu])}{L(n)} - 1 \right\|_{2,(0,1)} \left(\left\| \frac{\varepsilon([nu])}{\varepsilon(n)} \right\|_{2,(0,1)} (1+o(1)) \right) \\ &\qquad + \left\| \frac{\varepsilon([nu])}{\varepsilon(n)} - 1 \right\|_{2,(0,1)} + o\left(\left\| \frac{\varepsilon([nu])}{\varepsilon(n)} \right\|_{2,(0,1)} \right) + o(1). \end{aligned} \tag{3.15}$$

From the proof of Lemma 4.4.6 (i) in Mynbaev (2011), $\|L([nu])/L(n)-1\|_{2,(0,1)}$ converges to zero. Similarly, the convergence of $\|\varepsilon([nu])/\varepsilon(n)-1\|_{2,(0,1)}$ to zero is ensured by $\varepsilon = K(\eta,\phi_\eta)$ in Assumption 1. Therefore, the right-hand side of (3.15) converges to zero and

(3.13) follows. ■

Lemma 8 *Let Assumptions 1 and 2 hold. For a sufficiently large n and $j = 0, 1, \ldots, k < n$ for some fixed k, it follows that*

$$\sum_{t=j+1}^{n} \Delta L(t)\Delta L(t-j) = o(L(n)^2).$$

Proof of Lemma 8 For all j and $t = j+1, \ldots, n$, Assumptions 1 and 2 and (3.4) imply that $\Delta L(t-j) \geq 0$ and $\Delta L(t-j) \leq K$ for some $K \geq 0$, so we have

$$\begin{aligned} \left| \sum_{t=j+1}^{n} \Delta L(t)\Delta L(t-j) \right| &\leq K \sum_{t=j+1}^{n} \Delta L(t-j) \\ &\leq K(L(n) - L(1)) = O(L(n)) = o(L(n)^2), \end{aligned}$$

which gives the result. ■

3.6.3 Proofs of the main results

Proof of Lemma 3 (i) Because the partial sum process $n^{-3/2}\sum_{t=1}^{n} u_t$ converges weakly to the random variable $\sigma_L \int_0^1 W(r)dr$, it suffices to show the asymptotic equivalence

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} u_t - \frac{1}{n^{3/2}L(n)} \sum_{t=1}^{n} L(t)u_t \xrightarrow{p} 0. \tag{3.16}$$

The left-hand side of (3.16) can be rewritten as the weighted sum of the linear process $\{v_t\}$:

$$\sum_{t=1}^{n} \left\{ \sum_{s=t}^{n} \frac{L(n) - L(s)}{n^{3/2}L(n)} \right\} v_t. \tag{3.17}$$

In order to get result (3.16), we have only to prove the L_2-closeness of the weight $w_n = (w_{n1}, \ldots, w_{nn})$ to the function $f = 0$, where $w_{nt} = \sum_{s=t}^{n} \{L(n) - L(s)\} / \{n^{3/2}L(n)\}$ appears in (3.17), or to prove $\|\Delta_{n2} w_n - f\|_{2,(0,1)} = \|\Delta_{n2} w_n\|_{2,(0,1)} \to 0$ as $n \to \infty$. Here, $\|\cdot\|_{2,(0,1)}$ is the L_2 norm on $(0,1)$; see the discussion in Section 3.2.2 and 3.6.1. Then, the convergence in

(3.16) follows immediately by the CLT of Theorem 1 because the convergence in probability to zero is equivalent to the convergence in distribution to a random variable that equals zero a.s.

Following the proof of Theorem 4.4.1 in Mynbaev (2011), we can write

$$(\Delta_{n2}w_n)(u) = \frac{1}{nL(n)} \sum_{s=[nu]+1}^{n} \{L(n) - L(s)\} \quad \text{for} \quad 0 \le u < 1.$$

Let $\delta \in (0, 1/2]$ be fixed. With the number B_b from Lemma 4.3.5 of Mynbaev (2011), the interval $(B_b/n, \delta)$ is not empty for $n > n_1 \equiv B_b/\delta$, and the bound is obtained by the triangle inequality:

$$\|\Delta_{n2}w_n\|_{2,(0,1)} \le \|\Delta_{n2}w_n\|_{2,(\delta,1)} + \|\Delta_{n2}w_n\|_{2,(0,B_b/n)} + \|\Delta_{n2}w_n\|_{2,(B_b/n,\delta)}. \tag{3.18}$$

For the three terms of the upper bound in (3.18), we consider three cases.

Case 1: $u \in [\delta, 1)$. For a large n and $r \in [\delta, 1 + 1/(2B_b)]$ satisfying $rn \in \mathbb{N}$, we have

$$\frac{1}{nL(n)} \sum_{s=rn}^{n} \{L(n) - L(s)\} = 1 - r - \frac{1}{nL(n)} \sum_{s=rn}^{n} L(s). \tag{3.19}$$

The last term in (3.19) may be evaluated as

$$\frac{1}{nL(n)} \sum_{s=rn}^{n} L(s) = \frac{1}{nL(n)} \left(\sum_{s=1}^{n} L(s) - \sum_{s=1}^{rn} L(s) \right) = 1 - r + o(1)$$

uniformly in $r \in [\delta, 1 + 1/(2B_b)]$ because of Lemma 1 and the uniform convergence theorem (see Mynbaev (2011, 4.1.2)). Thus, the right-hand side of (3.19) tends to zero uniformly in $r \in [\delta, 1 + 1/(2B_b)]$. If we define $r = [nu + 1]/n$ with the inequality $nu < [nu + 1] \le nu + 1$, we have

$$\delta \le u < r \le u + \frac{1}{n} < 1 + \frac{1}{n_1} \le 1 + \frac{1}{2B_b}. \tag{3.20}$$

This implies that $r = u + o(1)$ and $r \in [\delta, 1 + 1/(2B_b)]$. Consequently, we have $(\Delta_{n2}w_n)(u) = o(1)$ uniformly in $u \in [\delta, 1)$. This proves, as $n \to \infty$,

$$\|\Delta_{n2}w_n\|_{2,(\delta,1)} \to 0. \tag{3.21}$$

Case 2: $u \in [B_b/n, \delta)$. For a large $n > n_2 \equiv (n_1 \vee 2)$, we have

$$\begin{aligned} |(\Delta_{n2}w_n)(u)| &\leq \frac{1}{nL(n)} \sum_{s=[nu]+1}^{n} \{L(n)+L(s)\} \\ &\leq \frac{1}{nL(n)} \left(nL(n) + \sum_{s=1}^{n} L(s) \right) = 2 + o(1) \end{aligned} \tag{3.22}$$

uniformly in $u \in [B_b/n, \delta)$ because of $L(s) > 0$ and Lemma 1. Integrating the terms gives

$$\|(\Delta_{n2}w_n)(u)\|^2_{2,(B_b/n,\delta)} \leq \int_{B_b/n}^{\delta} (4+o(1))du \leq 4\delta(1+o(1)). \tag{3.23}$$

Case 3: $u \in (0, B_b/n)$. In this case, the same bound as in (3.22) holds uniformly in $u \in (0, B_b/n)$. Therefore, integrating the terms yields

$$\|(\Delta_{n2}w_n)(u)\|^2_{2,(0,B_b/n)} \leq \int_{0}^{B_b/n} (4+o(1))du = \frac{B_b}{n}(4+o(1)). \tag{3.24}$$

From (3.21), (3.23) and (3.24), we can choose a small δ and large n to make the left-hand side of (3.18) as small as desired. This, in turn, implies (3.16). ■

(ii) We basically take the same manipulation as in (i). Because the partial sum process $n^{-3/2}\sum_{t=1}^{n}(1+\log(t/n))u_t$ converges weakly to the random variable $\sigma_L \int_0^1 (1+\log r)W(r)dr$ by Phillips (2007, Eq.(9)), it suffices to show

$$\frac{1}{n^{3/2}} \sum_{t=1}^{n} \left(1+\log\frac{t}{n}\right) u_t - \frac{1}{n^{3/2}L(n)\varepsilon(n)} \sum_{t=1}^{n} (L(t)-\bar{L})\, u_t \xrightarrow{p} 0. \tag{3.25}$$

The left-hand side of (3.25) can be rewritten as the weighted sum of the linear process $\{v_t\}$:

$$\sum_{t=1}^{n} \left\{ \sum_{s=t}^{n} \frac{L(n)\varepsilon(n)\,(1+\log(s/n)) - (L(s)-\bar{L})}{n^{3/2}L(n)\varepsilon(n)} \right\} v_t. \tag{3.26}$$

In order to get result (3.25), we have only to prove the L_2-closeness of the weight $w_n = (w_{n1}, \ldots, w_{nn})$ to the function $f = 0$, where w_{nt} is defined as the sum in curly brackets $\{\cdot\}$ in (3.26); that is, we prove $\|\Delta_{n2}w_n - f\|_{2,(0,1)} = \|\Delta_{n2}w_n\|_{2,(0,1)} \to 0$ as $n \to \infty$.

Following the proof of Theorem 4.4.1 in Mynbaev (2011), we can write

$$(\Delta_{n2}w_n)(u) = \frac{1}{n} \sum_{s=[nu]+1}^{n} \left\{ \left(1+\log\frac{s}{n}\right) - \frac{L(s)-\bar{L}}{L(n)\varepsilon(n)} \right\} \quad \text{for} \quad 0 \leq u < 1.$$

Hereafter, we use the notation $G(t,n) = (L(t) - L(n))/(L(n)\varepsilon(n))$, called the *G-function*; see Mynbaev (2011, 4.2.6). We first observe from the triangle inequality that

$$\begin{aligned}\|\Delta_{n2}w_n\|_{2,(0,1)} \leq & \left\|\frac{1}{n}\sum_{s=[nu]+1}^{n}\left(1+\log\frac{s}{n}\right)+u\log u\right\|_{2,(0,1)} \\ & +\left\|\frac{1}{n}\sum_{s=[nu]+1}^{n}\frac{L(s)-\bar{L}}{L(n)\varepsilon(n)}+\frac{[nu]}{n}G([nu],n)\right\|_{2,(0,1)}+\left\|\frac{[nu]}{n}G([nu],n)-u\log u\right\|_{2,(0,1)}.\end{aligned} \tag{3.27}$$

The first and second terms of the upper bound in (3.27) converge to zero by Lemmas 6 and 7, respectively. For the third term, we have

$$\begin{aligned}&\left\|\frac{[nu]}{n}G([nu],n)-u\log u\right\|_{2,(0,1)} \\ &\leq \left\|\frac{[nu]}{n}\right\|_{2,(0,1)}\|G([nu],n)-\log u\|_{2,(0,1)}+\left\|\frac{[nu]}{n}-u\right\|_{2,(0,1)}\|\log u\|_{2,(0,1)} \\ &\leq \|G([nu],n)-\log u\|_{2,(0,1)}+2\left\|\frac{[nu]}{n}-u\right\|_{2,(0,1)}.\end{aligned} \tag{3.28}$$

The first term in the right-hand side of (3.28) converges to zero because of the L_2-closeness of the G-function to the logarithmic function; see Mynbaev (2011, 4.4.1). The second term also converges to zero since $[nu]/n$ converges to u uniformly in $u \in [0,1]$. In consequence, we obtain $\|\Delta_{n2}w_n\|_{2,(0,1)} \to 0$ by (3.27) and the result follows. ■

Proof of Theorem 3 The marginal limiting distribution of $\hat{\beta}_n$ is clear from Lemma 3. The remaining claim is the distribution of $\hat{\alpha}_n$ and their joint behavior. It is easy to see that

$$\begin{aligned}\frac{\varepsilon(n)}{n^{1/2}}(\hat{\alpha}_n-\alpha) &= \frac{\varepsilon(n)}{n^{3/2}}\sum_{t=1}^{n}u_t-\frac{\varepsilon(n)}{n^{1/2}}\frac{\bar{L}\sum_{t=1}^{n}L(t)u_t-\bar{L}^2\sum_{t=1}^{n}u_t}{\sum_{t=1}^{n}(L(t)-\bar{L})^2} \\ &= O_p(\varepsilon(n))-\frac{\varepsilon(n)\bar{L}}{n^{1/2}}(\hat{\beta}_n-\beta) \\ &= o_p(1)-\frac{L(n)\varepsilon(n)}{n^{1/2}}(\hat{\beta}_n-\beta)[1+O(\varepsilon(n))] \\ &= -\frac{L(n)\varepsilon(n)}{n^{1/2}}(\hat{\beta}_n-\beta)+o_p(1).\end{aligned}$$

Thus, symmetry on the origin of the limiting normal distribution of the last term implies the desired result. ■

Proof of Lemma 4 (i) From (3.7) and Theorem 3, applying the FCLT, together with the continuous mapping theorem, yields

$$\begin{aligned}\frac{1}{n}\hat{u}_n^2 &= \left\{\frac{u_n}{n^{1/2}} - \left(\frac{1}{n}\sum_{t=1}^{n}\frac{u_t}{n^{1/2}}\right) - \frac{L(n)-\bar{L}}{n^{1/2}}(\hat{\beta}_n-\beta)\right\}^2 \\ &= \left\{\frac{u_n}{n^{1/2}} - \left(\frac{1}{n}\sum_{t=1}^{n}\frac{u_t}{n^{1/2}}\right) - \frac{L(n)-L(n)+L(n)\varepsilon(n)[1+O(\varepsilon(n))]}{n^{1/2}}(\hat{\beta}_n-\beta)\right\}^2 \\ &\xrightarrow{d} \sigma_L^2\left\{W(1) - \int_0^1 W(r)dr - \int_0^1 (1+\log r)W(r)dr\right\}^2,\end{aligned}$$

as stated. ■

(ii) By a similar manner to the proof of (i), we obtain

$$\begin{aligned}\frac{1}{n}\hat{u}_1^2 &= \left\{\frac{u_1}{n^{1/2}} - \left(\frac{1}{n}\sum_{t=1}^{n}\frac{u_t}{n^{1/2}}\right) - \frac{L(1)-\bar{L}}{n^{1/2}}(\hat{\beta}_n-\beta)\right\}^2 \\ &= \left\{O_p\left(\frac{1}{n^{1/2}}\right) - O_p(1) - \frac{L(1)-L(n)+L(n)\varepsilon(n)[1+O(\varepsilon(n))]}{n^{1/2}}(\hat{\beta}_n-\beta)\right\}^2.\end{aligned} \tag{3.29}$$

Since (3.29) multiplied by $\varepsilon(n)^2$ is $O_p(1)$, we obtain

$$\begin{aligned}\frac{\varepsilon(n)^2}{n}\hat{u}_1^2 &= \left\{o_p(1) + \frac{L(n)\varepsilon(n)}{n^{1/2}}(\hat{\beta}_n-\beta)(1+o_p(1))\right\}^2 \\ &\xrightarrow{d} \sigma_L^2\left\{\int_0^1 (1+\log r)W(r)dr\right\}^2.\end{aligned} \tag{3.30}$$

Hence, (3.30) gives the conclusion. ■

(iii) We have

$$\begin{aligned}\frac{1}{n}\sum_{t=2}^{n}(\hat{u}_t-\hat{u}_{t-1})^2 &= \frac{1}{n}\sum_{t=2}^{n}\left(v_t - \Delta L(t)(\hat{\beta}_n-\beta)\right)^2 \\ &\le \frac{2}{n}\sum_{t=2}^{n}v_t^2 + 2\left(\frac{\hat{\beta}_n-\beta}{n^{1/2}}\right)^2\sum_{t=2}^{n}(\Delta L(t))^2.\end{aligned} \tag{3.31}$$

The first term converges to $2\sigma_S^2$ in probability. For the second term, we have

$$\left(\frac{\hat{\beta}_n-\beta}{n^{1/2}}\right)^2\sum_{t=2}^{n}(\Delta L(t))^2 = O_p\left(\frac{1}{L(n)^2\varepsilon(n)^2}\right)o\left(L(n)^2\right) = o_p\left(\frac{1}{\varepsilon(n)^2}\right) \tag{3.32}$$

by Theorem 3 and Lemma 8. We therefore conclude that $\varepsilon(n)^2$ times (3.31) is $o_p(1)$. ■

(iv) If we use $\sum_{t=1}^{n} L(t) = nL(n)[1+O(\varepsilon(n))]$, $\sum_{t=1}^{n} L(t)^2 = nL(n)^2[1+O(\varepsilon(n))]$ and Lemma 3 (i), it follows that

$$\begin{aligned}
\frac{1}{n^2}\sum_{t=1}^{n}\hat{u}_t^2 &= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{u_t}{n^{1/2}} - \left(\frac{1}{n}\sum_{s=1}^{n}\frac{u_s}{n^{1/2}}\right) - \frac{\hat{\beta}_n-\beta}{n^{1/2}}(L(t)-\bar{L})\right)^2 \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{u_t}{n^{1/2}}\right)^2 + \left(\frac{1}{n}\sum_{s=1}^{n}\frac{u_s}{n^{1/2}}\right)^2 + \left(\frac{\hat{\beta}_n-\beta}{n^{1/2}}\right)^2 \frac{1}{n}\sum_{t=1}^{n}(L(t)-\bar{L})^2 \\
&\quad -2\left(\frac{1}{n}\sum_{t=1}^{n}\frac{u_t}{n^{1/2}}\right)^2 - \frac{L(n)\varepsilon(n)(\hat{\beta}_n-\beta)}{n^{1/2}}\frac{2}{n^{3/2}L(n)\varepsilon(n)}\sum_{t=1}^{n}(L(t)-\bar{L})u_t \\
&= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{u_t}{n^{1/2}}\right)^2 - \left(\frac{1}{n}\sum_{s=1}^{n}\frac{u_s}{n^{1/2}}\right)^2 - \left(\frac{L(n)\varepsilon(n)(\hat{\beta}_n-\beta)}{n^{1/2}}\right)^2[1+O(\varepsilon(n))] \\
&\xrightarrow{d} \sigma_L^2\int_0^1 W(r)^2dr - \sigma_L^2\left(\int_0^1 W(r)dr\right)^2 - \sigma_L^2\left\{\int_0^1(1+\log r)W(r)dr\right\}^2,
\end{aligned}$$

which completes the proof. ■

Proof of Theorem 4 and Corollary 1 It is clear from Lemma 4. ■

Proof of Theorem 5 The result follows from Lemma 3. ■

Proof of Lemma 5 (i) Note that

$$\hat{u}_t = u_t - (\hat{\beta}_n-\beta)L(t)$$

under model (3.8). Following from Theorem 5, we obtain

$$\frac{1}{n}\hat{u}_n^2 = \left(\frac{1}{n^{1/2}}u_n - \frac{L(n)}{n^{1/2}}\left(\hat{\beta}_n-\beta\right)\right)^2 \xrightarrow{d} \sigma_L^2\left(W(1)-\int_0^1 W(r)dr\right)^2,$$

which gives the proof. ■

(ii) It follows from $u_1 = O_p(1)$ and Theorem 5 that

$$\frac{1}{n}\hat{u}_1^2 = \left(\frac{1}{n^{1/2}}u_1 - \frac{L(1)}{n^{1/2}}\left(\hat{\beta}_n - \beta\right)\right)^2 = \left(O_p\left(\frac{1}{n^{1/2}}\right) - O_p\left(\frac{1}{L(n)}\right)\right)^2.$$

Hence, this converges to zero in probability. ■

(iii) Using Theorem 5, together with Lemma 8, we have

$$\begin{aligned}
\frac{1}{n}\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2 &= \frac{1}{n}\sum_{t=2}^{n}\left(v_t - \Delta L(t)(\hat{\beta}_n - \beta)\right)^2 \\
&= \frac{1}{n}\sum_{t=2}^{n}v_t^2 + \left(\frac{\hat{\beta}_n - \beta}{n^{1/2}}\right)^2\sum_{t=2}^{n}(\Delta L(t))^2 - \frac{2(\hat{\beta}_n - \beta)}{n}\sum_{t=2}^{n}v_t\Delta L(t) \qquad (3.33) \\
&= \sigma_S^2 + o_p(1) + O_p\left(\frac{1}{n^{1/2}L(n)}\right)\sum_{t=2}^{n}v_t\Delta L(t).
\end{aligned}$$

Thus, the proof is completed if the last term in (3.33) is $o_p(1)$. From the Schwarz inequality and Lemma 8, we get

$$\begin{aligned}
\left|\sum_{t=2}^{n}v_t\Delta L(t)\right| &\leq \left(\sum_{t=2}^{n}v_t^2\right)^{1/2}\left(\sum_{t=2}^{n}(\Delta L(t))^2\right)^{1/2} \\
&= n^{1/2}\left(\sigma_S^2 + o_p(1)\right)^{1/2}\left(o_p\left(L(n)^2\right)\right)^{1/2} = o_p\left(n^{1/2}L(n)\right).
\end{aligned}$$

Therefore, the last term in (3.33) is $o_p(1)$ and the result follows. ■

(iv) It follows from Lemma 3 (i) and Theorem 5 that

$$\begin{aligned}
\frac{1}{n^2}\sum_{t=1}^{n}\hat{u}_t^2 = \frac{1}{n}\sum_{t=1}^{n}\left(\frac{u_t}{n^{1/2}}\right)^2 &- 2\frac{L(n)}{n^{1/2}}\left(\hat{\beta}_n - \beta\right)\frac{1}{n^{3/2}L(n)}\sum_{t=1}^{n}L(t)u_t \\
&+ \frac{L(n)^2}{n}\left(\hat{\beta}_n - \beta\right)^2[1 + O(\varepsilon(n))] \\
\xrightarrow{d} \sigma_L^2\int_0^1 W(r)^2dr &- 2\sigma_L^2\left(\int_0^1 W(r)dr\right)^2 + \sigma_L^2\left(\int_0^1 W(r)dr\right)^2.
\end{aligned}$$

Collecting terms gives the result. ■

Proof of Theorem 6 and Corollary 2 It is clear from Lemma 5. ■

Proof of Theorem 7 We consider the following situation

True DGP : $y_t = \alpha + \beta L(t) + u_t,$

Regression : Regress y_t on only $L(t)$ to get $\hat{\beta}_n$.

Then, the OLS estimator is

$$\begin{aligned}\hat{\beta}_n = \frac{\sum_{t=1}^n L(t) y_t}{\sum_{t=1}^n L(t)^2} &= \alpha \frac{\sum_{t=1}^n L(t)}{\sum_{t=1}^n L(t)^2} + \beta + \frac{\sum_{t=1}^n L(t) u_t}{\sum_{t=1}^n L(t)^2} \\ &= \alpha \frac{nL(n)[1+O(\varepsilon(n))]}{nL(n)^2[1+O(\varepsilon(n))]} + \beta + O_p\left(\frac{n^{1/2}}{L(n)}\right) \\ &= \alpha \frac{1+o(1)}{L(n)[1+o(1)]} + \beta + O_p\left(\frac{n^{1/2}}{L(n)}\right)\end{aligned}$$

by Lemma 3 and Theorem 3. Collecting terms and scaling by $L(n)/n^{1/2}$, we have $n^{-1/2}L(n)(\hat{\beta}_n - \beta) = O(n^{-1/2}) + O_p(1)$. Thus, it leads to the same limiting result as in Theorem 5. The conclusions for the other statistics in this situation are also derived in the same way via Lemma 5. ■

Proof of Theorem 8 It suffices to prove the consistency of the estimated long-run variance $\hat{\sigma}_L^2$. First, it can be written as

$$\begin{aligned}\frac{1}{n}\sum_{t=j+1}^n \Delta\hat{u}_t \Delta\hat{u}_{t-j} = \frac{1}{n}\sum_{t=j+1}^n v_t v_{t-j} - \frac{\hat{\beta}_n - \beta}{n}\sum_{t=j+1}^n v_t \Delta L(t-j) \\ - \frac{\hat{\beta}_n - \beta}{n}\sum_{t=j+1}^n v_{t-j}\Delta L(t) + \frac{(\hat{\beta}_n - \beta)^2}{n}\sum_{t=j+1}^n \Delta L(t)\Delta L(t-j).\end{aligned} \tag{3.34}$$

Then, the last three terms of (3.34) are $o_p(1)$ from the Schwarz inequality and Lemma 8 as in the proof of Lemma 5 (iii). Therefore, we have

$$\frac{1}{n}\sum_{t=j+1}^n \Delta\hat{u}_t \Delta\hat{u}_{t-j} - \frac{1}{n}\sum_{t=j+1}^n v_t v_{t-j} = o_p(1).$$

Combining the fact that $\hat{\sigma}_S^2 - \sigma_S^2 = o_p(1)$, we conclude that $\hat{\sigma}_L^2$ is consistent for σ_L^2 from Theorem 2 in Newey and West (1987). ■

Table 3.1: $\varepsilon(x)$ and $\eta(x)$ associated with some $L(x)$. $\gamma > 0$.

$L(x)$	$\varepsilon(x)$	$\eta(x)$
$\log^{\gamma} x$	$\gamma/\log x$	$-1/\log x$
$1/\log^{\gamma} x$	$-\gamma/\log x$	$-1/\log x$
$\log\log x$	$1/(\log x \log\log x)$	$-1/(\log x \log\log x) - 1/\log x$
$1/\log\log x$	$-1/(\log x \log\log x)$	$-1/(\log x \log\log x) - 1/\log x$

Table 3.2: Percentage points of the limiting distributions.

Statistic	Probability (%) 1.0	2.5	5.0	10.0	90.0	95.0	97.5	99.0
$n\varepsilon(n)^2(\hat{\rho}_n - 1)$	-2.13	-1.69	-1.38	-1.05	-0.01	-0.00	-0.00	-0.00
$\varepsilon(n)^2 t_{\hat{\rho}_n}$	-0.63	-0.50	-0.41	-0.31	-0.00	-0.00	-0.00	-0.00

Table 3.3: Percentage points of the limiting distributions.

Statistic	Probability (%) 1.0	2.5	5.0	10.0	90.0	95.0	97.5	99.0
$n(\hat{\rho}_n - 1)$ based on (3.8)	-19.69	-15.88	-12.91	-10.02	-0.15	0.45	0.95	1.45
$t_{\hat{\rho}_n}$ based on (3.8)	-3.07	-2.75	-2.48	-2.17	-0.08	0.27	0.55	0.91

Table 3.4: Empirical size of tests with $\alpha = 0$ $(\%)$.

Statistic	# observations 100	200	300	500
$n(\hat{\rho}_n - 1)$ based on (3.8)	8.74	8.46	7.99	7.55
$t_{\hat{\rho}_n}$ based on (3.8)	10.66	9.40	8.86	7.97

Table 3.5: Empirical size of tests with $\alpha = 1$ $(\%)$.

Statistic	# observations 100	200	300	500
$n(\hat{\rho}_n - 1)$ based on (3.8)	8.78	8.59	8.05	7.57
$t_{\hat{\rho}_n}$ based on (3.8)	11.03	9.56	8.91	8.16

Table 3.6: Empirical size of tests with $\alpha = 5$ $(\%)$.

Statistic	# observations 100	200	300	500
$n(\hat{\rho}_n - 1)$ based on (3.8)	10.63	10.43	9.27	8.54
$t_{\hat{\rho}_n}$ based on (3.8)	21.37	15.15	12.23	10.41

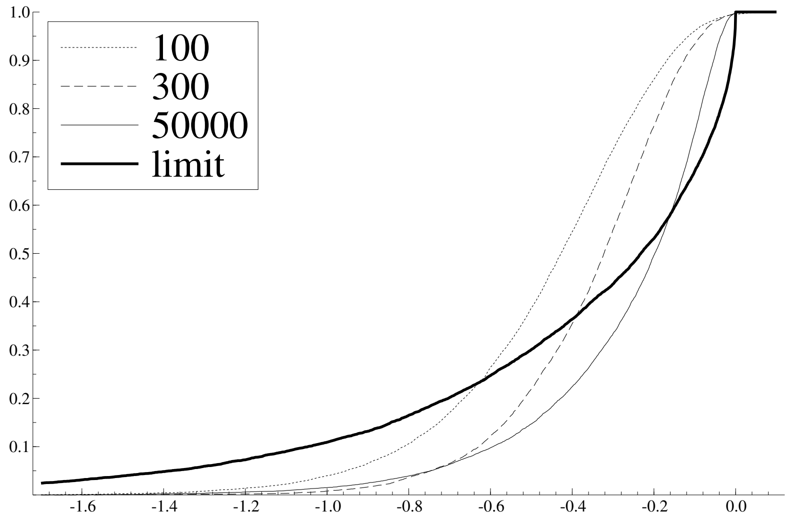


Figure 3.1: CDFs of $n\varepsilon(n)^2(\hat{\rho}_n - 1)$.

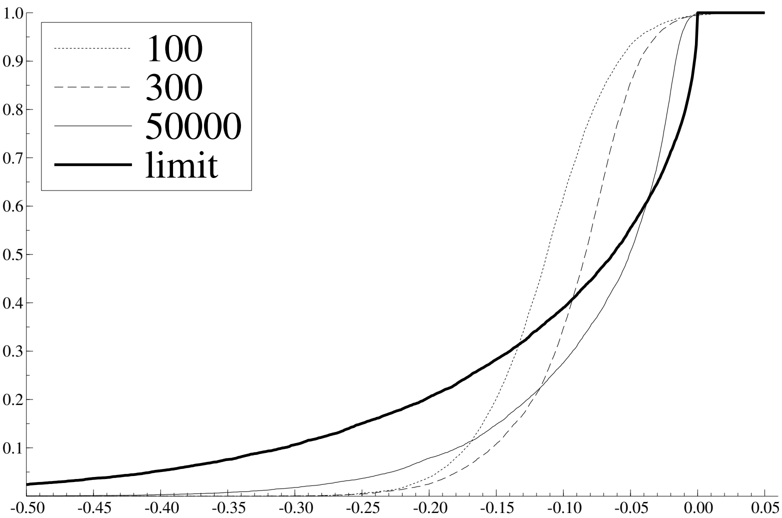


Figure 3.2: CDFs of $\varepsilon(n)^2 t_{\hat{\rho}_n}$.

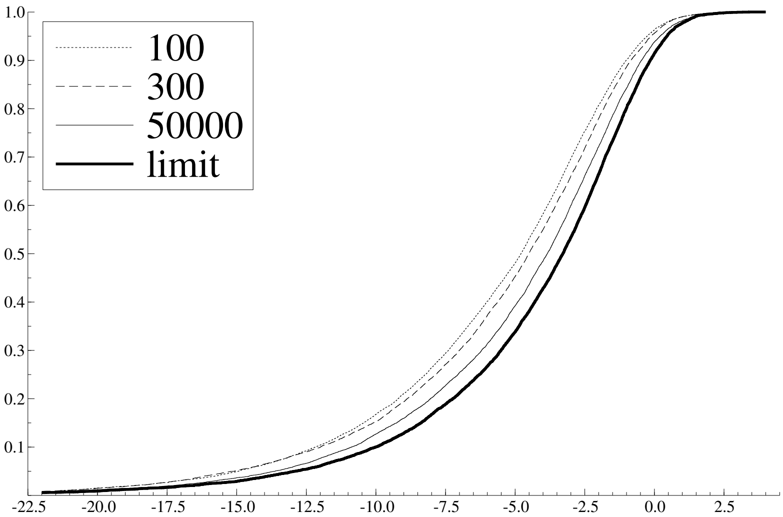


Figure 3.3: CDF of $n(\hat{\rho}_n - 1)$ based on (3.8) with $\alpha = 0$.

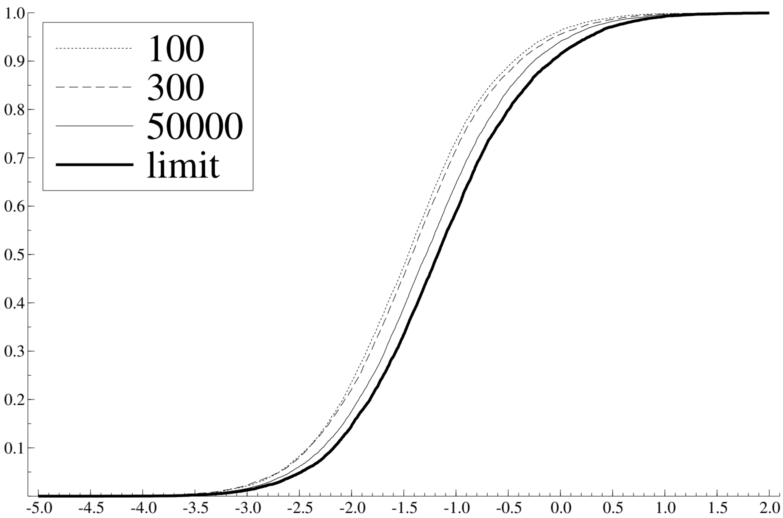


Figure 3.4: CDF of $t_{\hat{\rho}_n}$ based on (3.8) with $\alpha = 0$.

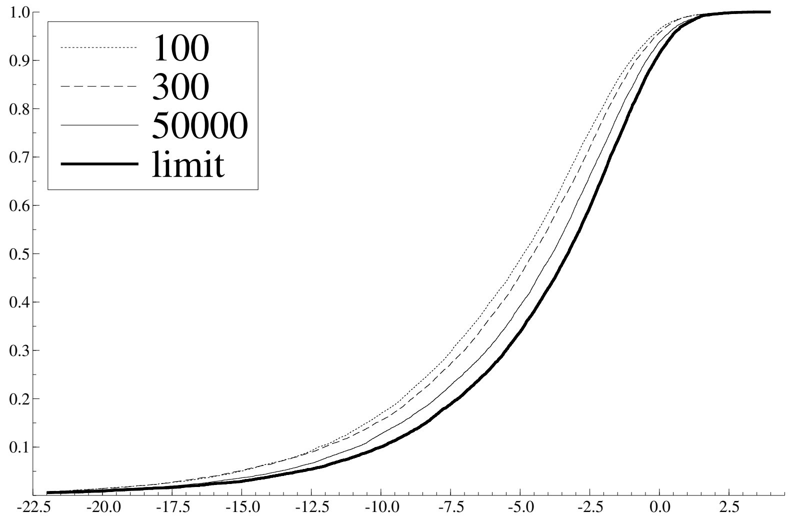


Figure 3.5: CDF of $n(\hat{\rho}_n - 1)$ based on (3.8) with $\alpha = 1$.

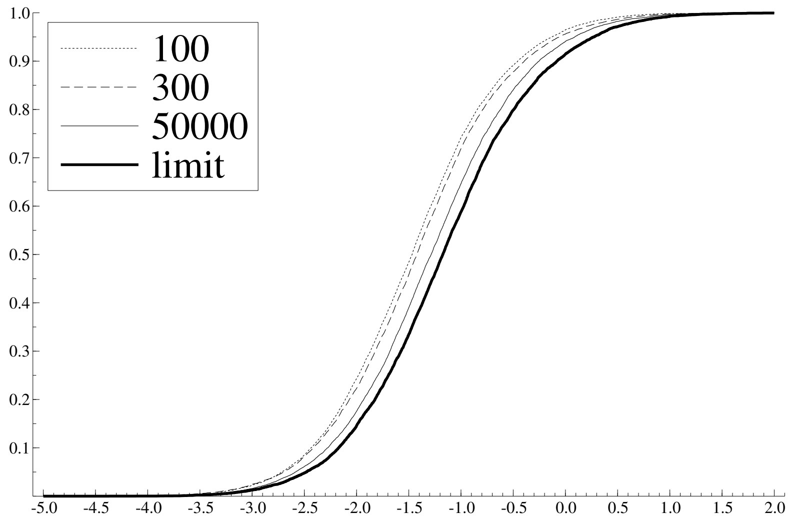


Figure 3.6: CDF of $t_{\hat{\rho}_n}$ based on (3.8) with $\alpha = 1$.

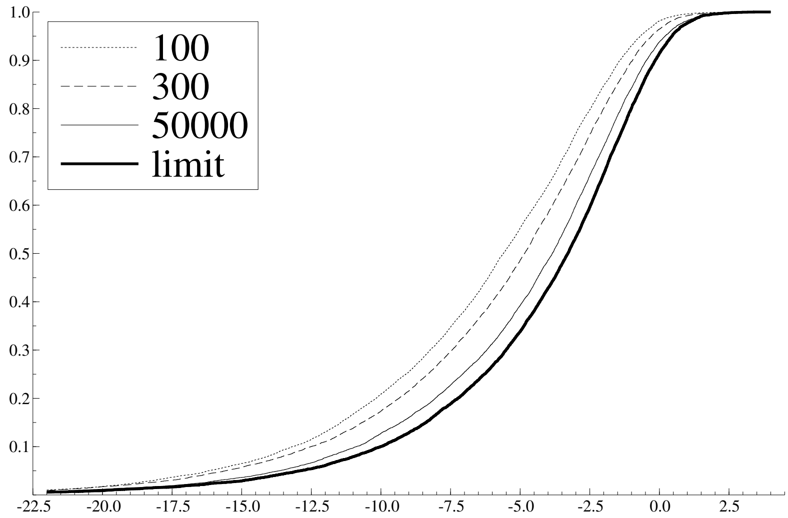


Figure 3.7: CDF of $n(\hat{\rho}_n - 1)$ based on (3.8) with $\alpha = 5$.

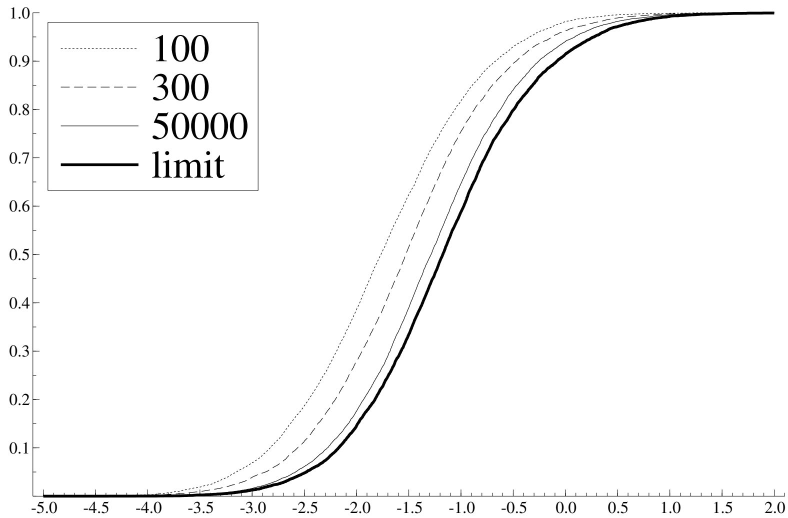


Figure 3.8: CDF of $t_{\hat{\rho}_n}$ based on (3.8) with $\alpha = 5$.

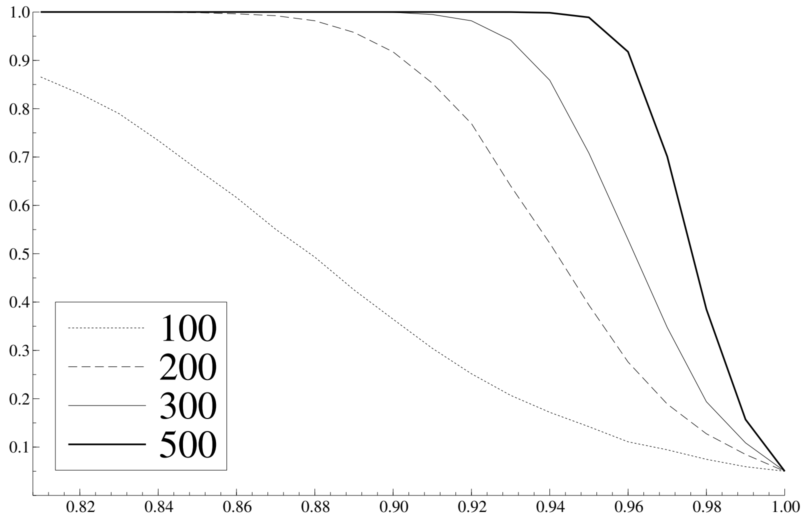


Figure 3.9: Power of $n(\hat{\rho}_n - 1)$ based on (3.8) with $\alpha = 0$.

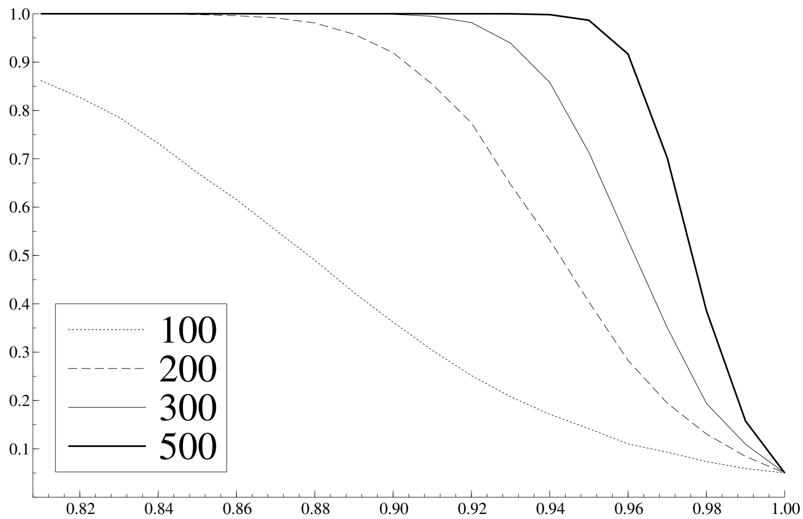


Figure 3.10: Power of $t_{\hat{\rho}_n}$ based on (3.8) with $\alpha = 0$.

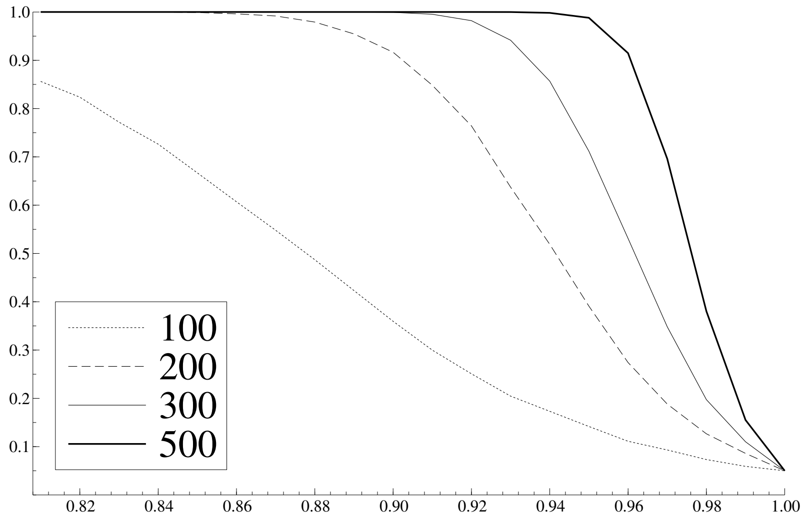


Figure 3.11: Power of $n(\hat{\rho}_n - 1)$ based on (3.8) with $\alpha = 1$.

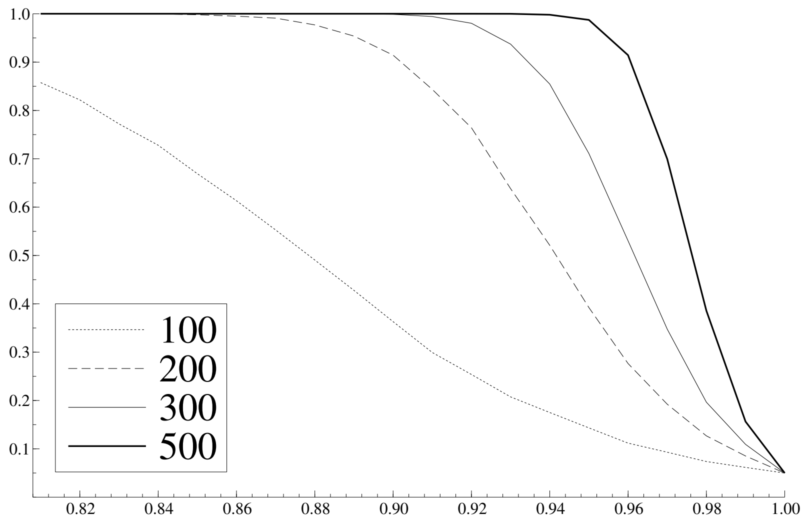


Figure 3.12: Power of $t_{\hat{\rho}_n}$ based on (3.8) with $\alpha = 1$.

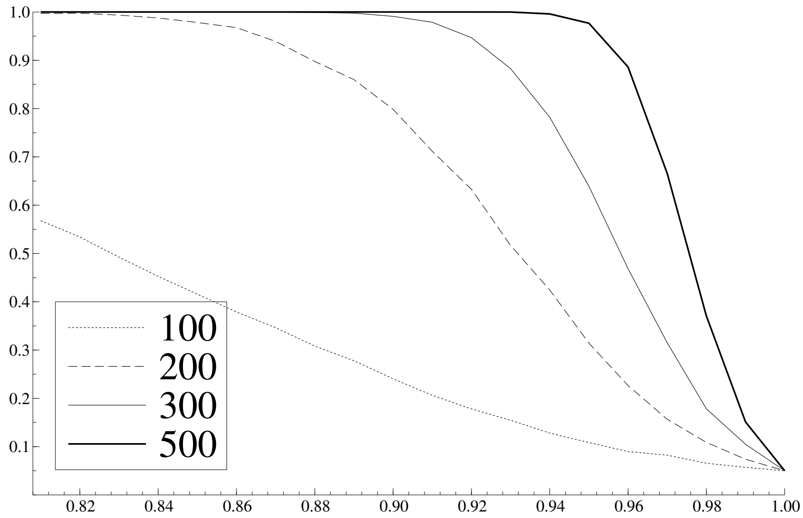


Figure 3.13: Power of $n(\hat{\rho}_n - 1)$ based on (3.8) with $\alpha = 5$.

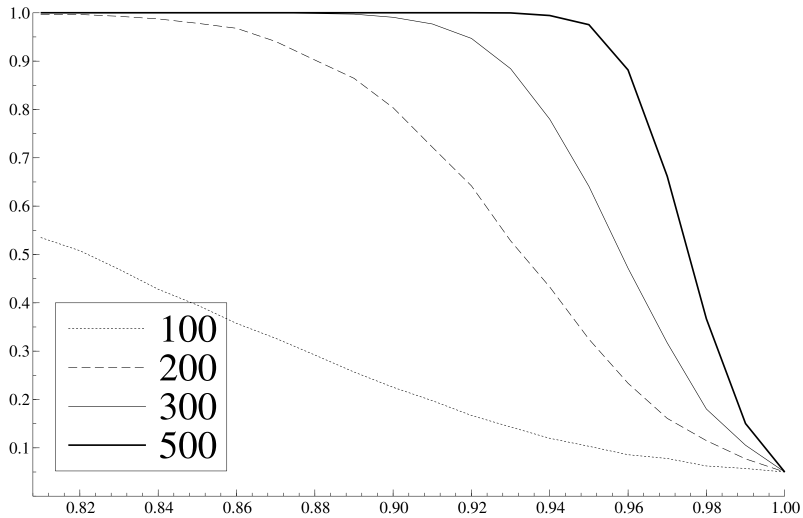


Figure 3.14: Power of $t_{\hat{\rho}_n}$ based on (3.8) with $\alpha = 5$.

Chapter 4

Nonstationary Nonlinear Quantile Regression

This chapter studies estimation and inference for nonlinear regression models with integrated time series by quantile regression method. The derivatives of regression function are specified as asymptotically homogeneous function, which has been analyzed by nonlinear least squares in Park and Phillips (2001, *Econometrica*) (PP). In this chapter, we derive the limiting distributions of the nonlinear quantile regression (NQR) estimator when it is close to the true parameter. We find that the estimator does not converge weakly to a mixed normal distribution as in PP. In that case, a fully-modified type NQR estimator is proposed. The class of integrable regression function derivatives are considered only when the model is linear in parameter. Finally, we observe from simulations that the NQR estimators are desirable when distributions of regression errors possess fat tails.

4.1 Introduction

Since the seminal works of Park and Phillips (1999, 2001) (PP hereafter), the literature on the analysis of unit root nonstationary nonlinear time series has been highly developed.

PP (2001) derived limiting distributions of nonlinear least squares (NLS) estimators for two cases of regression function derivatives: integrated and asymptotically homogeneous functions. Along these lines, many related works have appeared in the last ten years. Park and Phillips (2000) analyzed the maximum likelihood estimation of a nonstationary binary choice and Chang and Park (2003) extended it to index models. Moon (2004) and Guerre and Moon (2006) dealt with the maximum score and semiparametric estimation in reference to it. Nonstationary nonlinear heteroskedasticity was analyzed in Park (2002) and Chung and Park (2007). Moon and Schorfheide (2002) proposed the limit distribution of minimum distance estimators. Phillips et al. (2004) presented the instrumental variable estimation. Marmer (2008) studied nonstationary nonlinearity from a forecasting perspective. Recently, the misspecification problem was addressed in Kasparis (2010, 2011).

The quantile regression method has become widely used in several aspects of theoretical and applied works since the novel work of Koenker and Basset (1978). There are many articles on parametric quantile regression and least absolute deviation (LAD) estimation in the time series context, including works by Knight (1989, 1991), Phillips (1995), Herce (1996), Koenker and Xiao (2004) and Xiao (2009). Studies on nonlinear quantile regressions include Powel (1984, 1986), Weiss (1991), Jurečková, J. and B. Procházka (1994), Wang (1995), Koenker and Park (1996), Mukherjee (1999), Oberhofer and Haupt (2006) and Chen et al. (2009). Recently, Honda (2013) studied the nonparametric LAD estimation of the nonlinear model with an integrated covariate. However, to the best of the author's knowledge, there is no work on the quantile regression of nonstationary nonlinear models like PP's. Given this background, this chapter explores the theory of nonstationary nonlinear quantile regression, which is located on the intersection of these two bodies of literatures.

It is well-known that inference based on an order statistic is generally more robust than that based on a sample mean in the presence of fat-tailed behavior; see van der Vaart (2000, chapter 14.1) and Koenker (2005, chapter 3.5.1), for instance. Similarly, compared to the NLS estimator, the LAD estimator, which is the same as 50%-quantile regression estimator, may be expected to provide robust inference when regression errors have fat-tailed distribution. Since many time series are believed to possess fat-tailed distributions, the quantile regression-based method may be desirable in terms of robustness. Furthermore, in financial econometrics, quantile regression applies to risk estimation measures called conditional value-at-risks by focusing on quantiles of the lower tail; see Chernozhukov and Umantsev (2001) and Engle and Manganelli (2004), for instance. The latter article proposed a conditional VaR as a regression quantile of a nonlinear model. In this way, the establishment of a foundation of statistical analysis for nonstationary nonlinear models by quantile regression will be a further step in studies on time series econometrics.

The first contribution of the chapter is to derive the asymptotic distribution of the nonlinear quantile regression (NQR) estimator when the regression function derivative is specified as an asymptotically homogeneous function and the estimator is close to the true parameter. We then find that the estimator is not asymptotically mixed normal as in PP (2001). Therefore, the second contribution is to propose the fully-modified-type NQR estimator for that case to recover the mixed normality. The estimator makes standard inference, such as the Wald test, available. When the regression function derivative is given by an integrable function, we restrict the nonlinear model to be linear in parameters. Finally the third contribution is the observation from simulations that our suggested NQR estimators are desirable relative to the NLS estimators in terms of estimation accuracy and powers of tests when distributions of regression errors possess fat tails.

This chapter is organized as follows. The model, assumptions and preliminary results for nonstationary NQR estimators are introduced in Section 4.2. The main analytical results are stated in Sections 4.3 to 4.5. Section 3 derives the limiting distributions of the estimators for the class of asymptotically homogeneous regression functions. In Section 4.4, we construct the fully-modified-type NQR estimator and provide an asymptotic test on parameter restrictions based on it. Section 4.5 studies linear-in-parameter models with the classes of regression functions, including the class of integrable ones, and derive the limiting distributions again. Section 4.6 deals with simulation studies to determine the performances of suggested NQR estimators in comparison to the NLS estimators. In Section 4.7, we conclude.

For a matrix $A = (a_{ij})$ and a vector $x = (x_i)$, we introduce some notations. The modulus $|\cdot|$ is taken element by element such that $|A| = (|a_{ij}|)$ and $|x| = (|x_i|)$. The norm $\|\cdot\|$ is defined as the maximum of the moduli such that $\|A\| = \max_{ij}|a_{ij}|$ and $\|x\| = \max_i |x_i|$. For a function g, which is a scalar or vector, the norm $\|g\|_p$ is defined as $(E\|g(x)\|^p)^{1/p}$.

4.2 Preliminaries

4.2.1 The model and estimator

Suppose that a scalar-valued random variable y_t is generated from the following nonlinear model

$$y_t = \alpha_0 + g(x_t, \beta_0) + u_t \tag{4.1}$$

for $t = 1, \ldots, n$, where $g : \mathbb{R} \times \mathbb{R}^\ell \to \mathbb{R}$ is a known regression function specified later and the error term u_t is a zero-mean stationary process. Hereafter, we simply denote $g(x_t, \beta)$ as $g_t(\beta)$. The covariate x_t is a univariate $I(1)$ time series defined by

$$x_t = x_{t-1} + v_t, \tag{4.2}$$

where $x_0 = O_p(1)$ is allowed, but we set $x_0 = 0$ for simplicity. The innovation $\{v_t\}$ is assumed to be stationary and mean zero. We then let $\mathscr{F}_{t-1}$ denote an increasing filtration generated by $\{u_{t-j}, j \geq 1; v_{t-k}, k \geq 0\}$. Furthermore, let $E_{t-1}[\ \cdot\ ]$ denote the conditional expectation with respect to the filtration $\mathscr{F}_{t-1}$.

The $(1+\ell)$-dimensional true parameter vector $\theta_0 = (\alpha_0, \beta_0')'$ is assumed to lie in the parameter set $\Theta = A \times B$, where $A \subset \mathbb{R}$ and $B \subset \mathbb{R}^\ell$. PP (2001) analyzed the asymptotic properties of the NLS estimator of a model like (4.1), whereas we investigate the asymptotic behavior of the estimator based on the NQR.

The NQR estimator $\hat{\theta}_n$ of θ_0 can be obtained by solving the minimization problem

$$\hat{\theta}_n = \arg\min_{\theta \in \Theta} \sum_{t=1}^{n} \rho_\tau(y_t - \alpha - g_t(\beta)) \tag{4.3}$$

where $\rho_\tau(u) = u(\tau - 1(u < 0))$ is referred to as a *check function*. In the special case $\tau = 1/2$, the NQR estimator agrees with the nonlinear LAD (NLAD) estimator. Here, we set $\psi_\tau(u) = \tau - 1(u < 0)$.

Let F and f denote the distribution function and density function of u_t, respectively. Let $\alpha_0(\tau) = \alpha_0 + F^{-1}(\tau)$ and define the new parameter vector $\theta_0(\tau) = (\alpha_0(\tau), \beta_0')'$. We may also rewrite the error term so that

$$u_{t\tau} = y_t - \alpha_0(\tau) - g_t(\beta_0) = u_t - F^{-1}(\tau).$$

Notice that $E\psi_\tau(u_{t\tau}) = 0$ and $Q_{u_{t\tau}}(\tau) = 0$, where $Q_{u_t}(\tau)$ is the τth quantile of u_t. Furthermore we need some assumptions on the distribution of u_t.

Assumption 1 (a) The distribution function of $\{u_t\}$, F, is strictly increasing and has a continuous density function, $f \in (0, \infty)$, defined on $\{F \in (0,1)\}$.

(b) The density function f satisfies $f(F^{-1}(\tau)) > 0$ and $\max_{u \in \mathbb{R}} f(u) \leq K$ for some $K < \infty$.

(c) The conditional distribution function $F_{t-1}(u) = P(u_t < u | \mathscr{F}_{t-1})$ has its derivative f_{t-1} a.s. with $E[f_{t-1}^r] < \infty$ for some $r > 1$.

Remark 1 If Assumption 1 (a) is satisfied, then the inverse function of F, F^{-1}, is well-defined. Assumption 1 (b) is required for a technical reasons. The conditional density is asymptotically associated with the unconditional one under Assumption 1 (c). This assumption is utilized in many researches, such as Knight (1989), Herce (1996), Koenker and Xiao (2006) and Xiao (2009).

For the error terms $u_{t\tau}$ and v_t, we construct two partial sum processes

$$U_n^{\psi}(\tau,r) = n^{-1/2}\sum_{t=1}^{[nr]}\psi_\tau(u_{t\tau}) \quad \text{and} \quad V_n(r) = n^{-1/2}\sum_{t=0}^{[nr]} v_{t+1}. \tag{4.4}$$

To develop the asymptotic theory, we should impose some assumptions on the processes (4.4).

Assumption 2 $\{\psi_\tau(u_{t\tau}), \mathscr{F}_t\}$ is a martingale difference sequence.

To determine the asymptotic behavior of V_n, either Assumption 3 or 4 below is required in response to the class of a transformation g.

Assumption 3 $\{x_t\}$ is adapted to the filtration $\mathscr{F}_{t-1}$ and, for all $r \in [0,1]$, the vector $(U_n^{\psi}(\tau,r), V_n(r))$ converges weakly to a two-dimensional vector Brownian motion $(U^{\psi}(\tau,r), V(r))$ with the covariance matrix

$$r\Omega(\tau) = r\begin{bmatrix} \omega_\psi(\tau)^2 & \omega_{\psi v}(\tau) \\ \omega_{v\psi}(\tau) & \omega_v^2 \end{bmatrix}.$$

Remark 2 We may easily confirm that the assumption of PP on errors is satisfied under Assumptions 2 and 3 for an H-regular g introduced later. Under Assumption 2, it follows that $U_n^{\psi}(\tau,r) \to_d U^{\psi}(\tau,r)$, where $U^{\psi}(\tau,r)$ is viewed as a Brownian motion with variance $r\omega_\psi(\tau)^2 = r\tau(1-\tau)$ for a fixed τ. Thus, for each fixed pair (τ,r), $U^{\psi}(\tau,r)$ is distributed as $N(0, r\omega_\psi(\tau)^2)$. Assumption 3 also requires V_n to converge weakly jointly with U_n^{ψ} to a vector Brownian motion; this requirement is standard in time series analysis.

For a certain class of a transformation g, a stronger assumption on $\{v_t\}$ is needed to derive the asymptotic results, but is utilized only in Section 4.5.

Assumption 4 Suppose Assumption 3 is satisfied. In addition, we assume:
$v_t = C(L)\varepsilon_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j}$ with $C(1) \neq 0$ and $\sum_{j=0}^{\infty} j|c_j| < \infty$. The innovation $\{\varepsilon_t\}$ is a sequence of i.i.d. random variables with mean zero and $E|\varepsilon_t|^p < \infty$ for some $p > 8$, the distribution of which is absolutely continuous with respect to the Lebesgue measure and has the characteristic function $\varphi(\lambda)$ with $\lambda^\delta \varphi(\lambda) \to 0$ as $\lambda \to \infty$ for some $\delta > 0$.

This requirement is somewhat restrictive and not necessary for some results. However, this still holds for many processes including all invertible Gaussian ARMA models.

We introduce an important concept called the *local time* of the Brownian motion V at x up to time t. This is defined by

$$L(t,x) = \lim_{\varepsilon \searrow 0} \frac{1}{2\varepsilon} \int_0^t 1\{|V(r) - x| \leq \varepsilon\} dr. \tag{4.5}$$

This random variable measures the time spent by the Brownian motion V in the neighborhood of x up to time t. It is known that the local time is a.s. continuous in t and x. For further information, see PP (1999, 2001) and Chung and Williams (1990).

4.2.2 The classes of transformation

We define the classes of regression functions as in PP. Our subsequent analysis is based on the classification of the regression functions into H- and I-regular functions. Here, we review these classes briefly; their precise definitions are listed in Appendix A.

An H-regular function $(x,\beta) \mapsto g(x,\beta)$ (see Definition 2) is guided by a *regular* function $(x,\beta) \mapsto h(x,\beta)$ (see Definition 1) and is defined by

$$g(\lambda x,\beta) = \kappa(\lambda)h(x,\beta) + R(x,\lambda,\beta), \tag{4.6}$$

where the function κ is said to be the *asymptotic order* of g and the *regular* function h in (4.6) is called the *limit homogeneous function* of g. The last term, $R(x,\lambda,\beta)$, is a remainder. Polynomial functions, distribution-like functions and logarithmic functions are included in this H-regular class; see also PP (2001, p.135 and p.140). From the definition of an H-regular function g and the property of a *regular* function h, we have an intuition that

$$\kappa(n^{1/2})^{-1} g(x_t,\beta) \approx h(n^{-1/2}x_t,\beta) \approx h(V(r),\beta)$$

uniformly in $\beta \in B$ for a sufficiently large n with an $I(1)$ covariate x_t. Thus, the asymptotic properties of an H-regular transformation come to those of a regular function h; see PP

(1999, 2001) for details. We use the brief notation κ_n instead of $\kappa(n^{1/2})$ hereafter. Further, we write a limit homogeneous function $h_t(\beta) = h(n^{-1/2}x_t, \beta)$ like $g_t(\beta) = g(x_t, \beta)$.

An I-regular function $(x, \beta) \mapsto g(x, \beta)$ (see Definition 3) is characterized as a bounded and integrable function with respect to x with sufficient smoothness in β. An example is an exponential function of the form $\exp(-\beta x^2)$ for $\beta > 0$.

After this, the argument τ is always fixed, so we omit the argument τ unless otherwise confused in the remaining sections. The τth quantile of u_t, $F^{-1}(\tau)$ is written by F^{-1} and the Brownian motion $U_n^{\psi}(\tau, r)$ is simply denoted by $U_n^{\psi}(r)$.

4.3 Local Asymptotic Behavior

In this section, we derive the asymptotic distribution of NQR estimator $\hat{\theta}_n$. We restrict our attention to the class of H-regular functions and investigate the local behavior of the NQR estimator. The proof is completed by Huber's (1967) method with some modifications; a similar proof can be found in Powell (1984, 1986) and Weiss (1991), for example. Note that the class of I-regular functions cannot be treated in the same framework due to the irregular convergence rate $n^{1/4}$; see the proof in Appendix for detail. This class will, however, be dealt with in the later section with a restricted model specification.

In the proof, we will need to apply the mean value theorem twice to a regression function g with respect to the parameter vector β. Hence $g(x, \cdot)$ is always supposed to be twice continuously differentiable. Define notation of the first and second order derivatives as

$$\dot{g}(x, \beta) = \frac{\partial g}{\partial \beta}(x, \beta) \quad \text{and} \quad \ddot{G}(x, \beta) = \frac{\partial^2 g}{\partial \beta \partial \beta'}(x, \beta),$$

and we further write $\ddot{g} = \mathrm{vec}(\ddot{G})$. Corresponding to the ℓ-dimensional vector $\dot{g}$ and ℓ^2-dimensional vector $\ddot{g}$, the asymptotic order matrices $\dot{\kappa}_n$ $(\ell \times \ell)$ and $\ddot{\kappa}_n$ $(\ell^2 \times \ell^2)$ and the vector of the limit homogeneous functions $\dot{h}$ and $\ddot{h}$ are introduced when $\dot{g}$ and $\ddot{g}$ are H-regular. We further let $\tilde{g} = (1, \dot{g}')'$, $\tilde{h} = (1, \dot{h}')'$ and $\tilde{\kappa}_n = \mathrm{diag}(1, \dot{\kappa}_n)$.

To obtain the limiting distribution of the NQR estimator $\hat{\theta}_n$, we need to suppose an additional assumption on the parameter vector θ.

Assumption 5 The parameter vector $\theta = (\alpha, \beta')' \in \mathbb{R} \times \mathbb{R}^\ell$ is specified as $\theta = \theta_0(\tau) + n^{-1/2}\tilde{\kappa}_n'^{-1}\pi$, where π lies in a compact set $\Pi \subset \mathbb{R} \times \mathbb{R}^\ell$.

We may regard the parameter θ as a function of π under Assumption 5 and, if necessary, denote $\theta = \theta(\pi)$ with $\theta(\pi) = \theta_0(\tau) + n^{-1/2}\tilde{\kappa}_n^{-1}\pi$ for notational convenience. Note that $\theta(0) = \theta_0(\tau)$.

Define the derivative from the right of the objective function (4.3) as

$$\begin{aligned} z_t(\theta) &= \tilde{g}_t(\beta)\psi_\tau(y_t - \alpha - g_t(\beta)) \\ &= \tilde{g}_t(\beta)\psi_\tau(u_{t\tau} - [\alpha - \alpha_0(\tau) + g_t(\beta) - g_t(\beta_0)]). \end{aligned} \tag{4.7}$$

Utilizing the function (4.7), we may derive the limiting distribution by considering the "first order condition"

$$n^{-1/2}\tilde{\kappa}_n^{-1}\sum_{t=1}^{n} z_t(\hat{\theta}_n) = o_p(1). \tag{4.8}$$

This estimating equation leads to the Bahadur representation of the NQR estimator $\hat{\theta}_n$. The following lemma ensures that (4.8) holds and then the main result is obtained in the next theorem.

Lemma 1 *Let $\dot{g}$ and $\ddot{g}$ be H-regular on B. If Assumptions 1–3 and 5 are assumed, then the estimating equation (4.8) holds with the NQR estimator $\hat{\theta}_n$.*

Theorem 1 *Let $\dot{g}$ and $\ddot{g}$ be H-regular on B, and suppose that Assumptions 1–3 and 5 hold. Furthermore, the identifiability condition*

$$\int_{-\delta}^{\delta} \dot{h}(x, \beta_0)\dot{h}(x, \beta_0)'dx > 0 \quad \textit{for all} \quad \delta > 0 \tag{4.9}$$

and $\left\|\dot{\kappa}_n^{-1}\right\|^2 \|\ddot{\kappa}_n\| < \infty$ are assumed to hold. Then we have

$$\begin{aligned} & n^{1/2}\tilde{\kappa}_n'(\hat{\theta}_n(\tau) - \theta_0(\tau)) \\ & \xrightarrow{d} \frac{1}{f(F^{-1}(\tau))}\left[\int_0^1 \tilde{h}(V(r), \beta_0)\tilde{h}(V(r), \beta_0)'dr\right]^{-1}\int_0^1 \tilde{h}(V(r), \beta_0)dU^\Psi(r). \end{aligned}$$

Remark 3 (a) The same identifiability condition in (4.9) is employed in Theorem 5.2 of PP (2001), whereas the condition on the asymptotic order is different from that in Theorem 5.2 of PP (2001); that assumes $\|(\dot{\kappa}_n \otimes \dot{\kappa}_n)^{-1} \kappa_n \ddot{\kappa}_n\| < \infty$.

(b) The limiting distribution in Theorem 1 is not usually (mixed) normal because of the possibly nonzero correlation between U^ψ and V. Therefore standard inferences are not applicable in this case.

(c) The limiting random variable in Theorem 1 resembles that of the NLS estimator derived in Theorem 5.2 in PP (2001), but the critical difference lies in the coefficient. The NQR estimator depends on the variance of U^ψ, ω_ψ^2, which is the function of the predetermined τ and the functional form of the density f of $\{u_t\}$, while the PP's NLS estimator depends on the variance of $\{u_t\}$. This is typical in investigating the relative efficiency of two statistics, such as sign versus t-test; see van der Vaart (2000, chapter 14.1).

4.4 Inferences

4.4.1 Fully-modified estimation

By Assumption 3 (or 4), we see that $\sum_0^\infty E v_t \psi_\tau(u_{t+j})$ is zero, but $\sum_{-\infty}^\infty E v_t \psi_\tau(u_{t+j})$ and $\sum_0^\infty E v_t v_{t+j}$ are not. In this subsection, we eliminate the endogeneity bias by means of *fully modified* (FM) type estimation developed by Phillips and Hansen (1990) in linear cointegrating models. Along this line, Phillips (1995) and Xiao (2009) suggested the FM type quantile regression estimator for linear time series. In the literature of nonlinear regressions with integrated time series, Chang et.al. (2001) worked out the so-called *efficient nonstationary nonlinear least squares* (EN-NLS) estimator, which is parametrically constructed through AR approximation of $\Delta x_t = v_t$. In the present chapter, however, we suggest the FM-NQR estimator based on the results of Phillips and Hansen (1990) and de Jong (2002). To utilize the result of de Jong (2002), we write $\nabla \dot{h}(x,\beta) = \partial \dot{h}(x,\beta)/\partial x$ and introduce additional assumptions on an H-regular transformation and the weak dependence structure of v_t.

Assumption 6 (a) v_t is *near-epoch dependent in L_2-norm* (L_2-NED) of size -1 on w_t, where w_t is an α-mixing array of size $-2p/(p-2)$ or v_t is L_2-NED of size -1 on w_t, where w_t is a ϕ-mixing array of size $-p/(p-1)$;

(b) for all $\beta \in B$, $\nabla \dot{h}(x,\beta)$ is continuous in $x \in \mathbb{R}$;

(c) for any sequence δ_n such that $\delta_n \to 0$ as $n \to \infty$, and for any $\beta \in B$ and $c > 0$,

$$\limsup_{n\to\infty} \sup_{|x|\le c} \sup_{y:|x-y|\le\delta_n} \left\|\dot{h}(x,\beta) - \dot{h}(y,\beta)\right\|^2 = 0.$$

Remark 4 (a) On Assumption 6 (a), we say that v_t is L_2-NED of size -1 on w_t if v_t satisfies

$$\|v_t - E[v_t|w_{t-m},\ldots,w_{t+m}]\|_2 \le \Psi(m) = O(m^{-\varphi})$$

for some $\varphi > 1$. For more information on the concepts of mixing and NED sequences, see Davidson (1994).

(b) Assumption 6 (b) and (c) provide the smoothness of a function; (c) is referred to as *asymptotic uniform equicontinuity*, which is seen in Davidson (1994, p. 90), for example.

The idea of the FM-NQR is based on the decomposition of the integral appearing in the limiting variable of $n^{1/2}\tilde{\kappa}_n'(\hat{\theta}_n(\tau) - \theta_0(\tau))$ in Theorem 1 so that

$$\begin{aligned}&\int_0^1 \tilde{h}(V(r),\beta_0)dU^{\psi}(r)\\&= \left[\int_0^1 \dot{h}(V(r),\beta_0)dU^{\psi}(r) - \frac{\omega_{\psi v}}{\omega_v^2}\int_0^1 \tilde{h}(V(r),\beta_0)dV(r)\right] + \frac{\omega_{\psi v}}{\omega_v^2}\int_0^1 \tilde{h}(V(r),\beta_0)dV(r).\end{aligned}$$

Taking note of the integrals in the brackets, we may define a new Gaussian process

$$U^{\psi+}(r) = U^{\psi}(r) - \frac{\omega_{\psi v}}{\omega_v^2}V(r).$$

Then, $U^{\psi+}$ is easily found to be uncorrelated with V and, hence, independent of V. Therefore, we can construct a new estimator of β_0 that has the limiting variable integrated with respect to the integrator $dU^{\Psi+}$ instead of dU^{Ψ}. For this purpose, it is necessary to form two statistics S_n and T_n satisfying

$$S_n \xrightarrow{p} \int_0^1 \tilde{h}(V(r),\beta_0)\tilde{h}(V(r),\beta_0)'dr \quad \text{and} \quad T_n \xrightarrow{d} \int_0^1 \tilde{h}(V(r),\beta_0)dV(r)$$

and arrange consistent estimators $1/\widehat{f(F^{-1}(\tau))}$, $\hat{\omega}_{\psi v}$, $\hat{\omega}_v^2$ and $\hat{\lambda}$. It should be remarked that estimation of the *sparsity function*, $1/f(F^{-1}(\tau))$, is a frequent problem in quantile regression. A portion of the method is introduced in Appendix. For further details, see Koenker (2005, section 3.4.1) and Koenker and Xiao (2004). Note, moreover, that λ is the one-sided long-run variance λ defined by $\sum_{j=0}^{\infty} E v_t v_{t+j}$. Using these statistics, we may accommodate the FM-NQR estimator

$$\hat{\theta}_n^+(\tau) = \hat{\theta}_n(\tau) - \frac{n^{-1/2}\tilde{\kappa}_n'^{-1}}{\widehat{f(F^{-1}(\tau))}} \frac{\hat{\omega}_{\psi v}}{\hat{\omega}_v^2} S_n^{-1} T_n, \tag{4.10}$$

where

$$S_n = n^{-1}\sum_{t=1}^{n} \tilde{h}(n^{-1/2}x_t,\hat{\beta}_n(\tau))\tilde{h}(n^{-1/2}x_t,\hat{\beta}_n(\tau))',$$
$$T_n = n^{-1/2}\sum_{t=1}^{n} \left\{ \tilde{h}(n^{-1/2}x_t,\hat{\beta}_n(\tau))\Delta x_t - n^{-1/2}\hat{\lambda}\nabla\tilde{h}(n^{-1/2}x_t,\hat{\beta}_n(\tau)) \right\}.$$

Theorem 2 *If all Assumptions in Theorem 1 and Assumption 6 are satisfied, then we have*

$$n^{1/2}\tilde{\kappa}_n'(\hat{\theta}_n^+(\tau) - \theta_0(\tau))$$
$$\xrightarrow{d} \frac{1}{f(F^{-1}(\tau))}\left[\int_0^1 \tilde{h}(V(r),\beta_0)\tilde{h}(V(r),\beta_0)'dr\right]^{-1}\int_0^1 \tilde{h}(V(r),\beta_0)dU^{\Psi+}(r).$$

Remark 5 (a) The mixed normality of the limiting random variable in Theorem 2 is brought to light immediately because of the construction of $U^{\Psi+}$. Indeed, it may be expressed as

$$MN\left(0, \frac{\omega_\psi^2 - \omega_{\psi v}^2/\omega_v^2}{f(F^{-1}(\tau))^2}\left[\int_0^1 \dot{h}(V(r),\beta_0)\dot{h}(V(r),\beta_0)'dr\right]^{-1}\right). \tag{4.11}$$

(b) The consistent estimators of the nuisance parameters are obtained by a standard non-parametric method as in Xiao (2009). That is, if we let the sample variance and covariance denote

$$C_v^2(j) = \frac{1}{n}\sum_{t=1}^{n-j} v_t v_{t+j} \quad \text{and} \quad C_{\psi v}(j) = \frac{1}{n}\sum_{t=1}^{n-j} v_t \psi_\tau(\hat{u}_{t+j,\tau}),$$

then we may utilize

$$\hat{\omega}_v^2 = \sum_{j=-M}^{M} k\left(\frac{j}{M}\right) C_v^2(j), \ \ \hat{\omega}_{\psi v} = \sum_{j=-M}^{M} k\left(\frac{j}{M}\right) C_{\psi v}(j), \ \ \hat{\lambda} = \sum_{j=0}^{M} k\left(\frac{j}{M}\right) C_v^2(j).$$

Here, $k(\cdot)$ is a lag window on $[-1,1]$ with $k(0)=1$ and M is a bandwidth satisfying $M\to\infty$ and $M/n\to 0$. A typical order is $M=O(n^{1/3})$.

4.4.2 Testing for parameter restrictions

We consider testing linear restrictions on the parameter vector according to FM-NQR estimator proposed in the previous subsection.

A null hypothesis is supposed to be of the simple form

$$H_0 : R\theta_0(\tau) = q,$$

where R denotes an $(m\times(1+\ell))$ matrix and q denotes an m-dimensional vector. By Theorem 2 and (4.11), we have, under H_0,

$$\left(\frac{\omega_\psi^2-\omega_{\psi v}^2/\omega_v^2}{f(F^{-1}(\tau))^2} R\left[\int_0^1 \tilde{h}(V(r),\beta_0)\tilde{h}(V(r),\beta_0)' dr\right]^{-1} R'\right)^{-1/2} \left[RD_n\hat{\theta}_n^+(\tau)-q\right]$$

$$\xrightarrow{d} N(0,I_m),$$

where $D_n = n^{1/2}\tilde{\kappa}_n'$ and $N(0,I_m)$ is an m-dimensional standard normal random variable. Therefore, for each τ, we may construct the Wald statistic $W_n(\tau)$ below:

$$W_n(\tau) = \frac{f(\widehat{F^{-1}(\tau)})^2}{\omega_\psi^2-\hat{\omega}_{\psi v}^2/\hat{\omega}_v^2}\left[RD_n\hat{\theta}_n^+(\tau)-q\right]'\left(RS_n^{-1}R'\right)^{-1}\left[RD_n\hat{\theta}_n^+(\tau)-q\right],$$

where S_n is defined in Theorem 2. The asymptotic behavior of $W_n(\tau)$ is summarized as follows:

Theorem 3 *Suppose all Assumptions of Theorem 2 are satisfied. Then, under the null hypothesis H_0, we have*

$$W_n(\tau) \xrightarrow{d} \chi_m^2,$$

where χ_m^2 is a chi-square random variable with m degrees of freedom.

4.5 Asymptotic Behavior of Linear-in-parameter Models

We have considered the NQR estimator of the nonlinear model only in the case of H-regular $\dot{g}$ and $\ddot{g}$. In this section, we then investigate the so-called linear-in-parameter model obtained by confining model (4.1) to

$$g(x_t, \beta_0) = \beta_0 g(x_t). \tag{4.12}$$

The regression function g is either I-regular or H-regular and write $g_t = g(x_t)$. The parameter β_0 is allowed ℓ-dimensional, but is assumed $\ell = 1$ for the sake of brevity. Because the model is linear in parameter, the asymptotics of the quantile regression estimator $\hat{\beta}_n$ can be derived even if g is an I-regular function as well as an H-regular one. The proof is quite different from that of Theorem 1; Theorems 4 and 5 below are shown after the fashion of the proofs for linear models achieved by Knight (1989, 1991), Herce (1996), Koenker and Xiao (2004) and Xiao (2009), for example. To make the discussion simple, we employ the following assumption.

Assumption 7 $\{u_t\}$ is a sequence of independent random variables.

Define the localized parameter $\pi = D_n(\theta - \theta_0)$ in a compact parameter set $\Pi \subset \mathbb{R} \times \mathbb{R}$. The scaling coefficient D_n is given by either $D_n^I = \text{diag}(n^{1/2}, n^{1/4})$ for an I-regular model or $D_n^H = n^{1/2}\tilde{\kappa}_n$ with $\tilde{\kappa}_n = \text{diag}(1, \kappa_n)$ for an H-regular model. Then θ is considered as a function of π so that $\theta = \theta(\pi) = \beta_0 + D_n^{-1}\pi$. It may be understood that $\hat{\theta}_n$ is the minimizer of the re-parameterized objective function $M_n(\theta) = M_n(\theta(\pi))$ defined by

$$M_n(\theta(\pi)) = \sum_{t=1}^{n} \left[\rho_\tau\left(u_{t\tau} - \pi' D_n^{-1}\tilde{g}_t\right) - \rho_\tau(u_{t\tau})\right].$$

First, we consider model (4.1) under restriction (4.12) with I-regular regression function derivative $\dot{g}$. The limiting distribution of the NQR estimator $\hat{\theta}_n$ is summarized as follows.

Theorem 4 *Let g be I-regular on B, and suppose that Assumptions 1, 4 and 7 hold. Furthermore, the identifiability condition*

$$\int_{-\infty}^{\infty} g(x)^2 dx > 0$$

is assumed to be satisfied. Then we have

$$D_n^I(\hat{\theta}_n(\tau) - \theta_0(\tau)) \xrightarrow{d} \frac{1}{f(F^{-1}(\tau))} \begin{bmatrix} U^{\psi}(1) \\ \left(L(1,0)\int_{-\infty}^{\infty} g(x)^2 dx\right)^{-1/2} W(1), \end{bmatrix}$$

where $W(r)$ is the Brownian motion independent of $(U^{\psi}(r), V(r))$.

Remark 6 The Brownian motion $W(r)$ in Theorem 4 has variance $r\tau(1-\tau)$, which is the same as $r\omega_{\psi}^2$, the variance of $U^{\psi}(r)$; see Theorem 3.2 of PP (2001) for detail. This implies that $\hat{\alpha}_n$ and $\hat{\beta}_n$ are asymptotically independent; consequently, the limiting joint distribution is mixed normal of the form

$$MN\left(0,\ \frac{\omega_{\psi}^2}{f(F^{-1}(\tau))^2}\begin{bmatrix} 1 & 0 \\ 0 & L(1,0)\int_{-\infty}^{\infty} g(x)^2 dx \end{bmatrix}^{-1}\right). \tag{4.13}$$

Hence, standard inferences are applicable in an asymptotic sense.

Finally, we derive the asymptotic distribution again for the H-regular case to pursue the completeness.

Theorem 5 *Let g be H-regular on B, and suppose that Assumptions 1, 3 and 7 hold. Furthermore, the identifiability condition*

$$\int_{-\delta}^{\delta} h(x)^2 dx > 0$$

for all $\delta > 0$ is assumed to be satisfied. Then we have

$$n^{1/2}\tilde{\kappa}_n(\hat{\theta}_n(\tau) - \theta_0(\tau)) \xrightarrow{d} \frac{1}{f(F^{-1}(\tau))}\left[\int_0^1 h(V(r))^2 dr\right]^{-1} \int_0^1 h(V(r)) dU^{\psi}(r).$$

4.6 Simulations

We investigate the finite sample performances of the NQR estimators with $\tau = 0.5$ by comparison to NLS estimators by simulations. Note that the NQR estimator here is equivalent to the nonlinear lease absolute deviation (NLAD) estimator. In the following subsections, we will consider estimation accuracy first and then observe the performance of tests under fat-tailed behavior. All the computations are implemented by R version 2.14.1. The NLAD estimators are obtained by the interior point algorithm through the `nlrq` command in the `quantreg` package. For more information on computational aspects, see Koenker (2005, Appendix A) and Koenker and Park (1994).

Throughout this section, the distributions of the regression error sequence $\{u_t\}$ for $t = 1,\ldots,n$ are supposed to be standard normal (SN) or student's t with k-degrees of freedom (t_k) for $k = 3,4$. The integrated covariate $\{x_t\}$ for $t = 1,\ldots,n$ is assumed to be driven by the error sequence $\{v_t\}$, which is specified by SN or student's t_9. The sample size is either $n = 250$ or 500 in all the experiments, and the number of replications is 500 for each computation. To clarify the objective and pursue simplicity, $\{u_t\}$ and $\{v_t\}$ are assumed to be i.i.d. and mutually independent.

Two linear-in-parameter models and one nonlinear model are used as examples of I- and H-regular regression functions. The first has an I-regular regression function called an exponential-type model, which is given by

$$y_t = \beta_{01} \exp\left(-cx_t^2\right) + u_t, \tag{4.14}$$

where $\beta_{01} \in \mathbb{R}$ is a parameter of interest and c is a known positive constant. The true value of β_{01} is set to 2. We should pay attention to the value of this type of regression function in practice. If exponential functions like (4.14) are used, the functional value tends to be close to zero very quickly for a relatively large value of $|x_t|$. As a result, numerical optimization will become unstable. In our settings, the constant c is set to 0.01 to stabilize a variation of cx_t^2 in (4.14). The second is a smooth transition model, which is an example of H-regular functions, and is given by

$$y_t = \frac{\beta_{02} \exp(x_t)}{1 + \exp(x_t)} + u_t, \tag{4.15}$$

where $\beta_{02} \in \mathbb{R}$ is unknown and to be estimated, and the true value of β_{02} is set to 2. The asymptotic order and asymptotically homogeneous function associated with this regression function derivative are calculated as $\dot{\kappa}_n = 1$ and $\dot{h}(x, \beta_{02}) = 1(x \geq 0)$, respectively. The third is a reformulated version of the regression function $g(x, \beta_{03}) = (x + \beta_{03})^2$ investigated in Example 4.1(c) of PP(2001), which is an example of nonlinear H-regular functions, and is given by

$$y_t = 2\beta_{03} x_t + \beta_{03}^2 + u_t, \tag{4.16}$$

where $\beta_{03} \in \mathbb{R}$ is unknown and to be estimated, and the true value of β_{03} is set to 2. The asymptotic order and asymptotically homogeneous function associated with this regression function derivative are calculated as $\dot{\kappa}_n = n^{1/2}$ and $\dot{h}(x, \beta_{03}) = 2x$, respectively.

4.6.1 Estimation

For each model, we observe tolerances of the NLS and NLAD estimators for fat-tailed errors by comparing their root mean squared errors (RMSEs). The results are shown in Tables 4.1 to 4.3.

From Table 4.1, which indicates the RMSEs for I-regular model (4.14), we may observe that the NLAD estimators are relatively superior to the NLS estimators when the regression error sequence $\{u_t\}$ has a fat-tailed distribution although the NLS estimators have an advantage in case of a Gaussian error sequence $\{u_t\}$. Meanwhile, the fat-tailedness of $\{v_t\}$ is not very effective in this experiment. From Tables 4.2 and 4.3, which indicate the RMSEs for H-regular models (4.15) and (4.16), we may find the same tendency as in Table 4.1. As a result, it may seem desirable to rely on the NLAD estimation relative to the NLS method if Gaussianity for errors is unlikely to be satisfied.

4.6.2 Testing

We examine the empirical sizes and size-adjusted powers of t-tests based on the estimated parameters $\hat{\beta}_{01}$, $\hat{\beta}_{02}$ and $\hat{\beta}_{03}$ in models (4.14), (4.15) and (4.16), respectively. The nominal size we desire is assumed to be 0.05. Tables 4.4 to 4.6 report the empirical sizes and Figures 4.1 to 4.12 display the size-adjusted powers under several error distributions.

For the NLAD estimation, we need to estimate a nuisance sparsity function $s(\tau)$ defined as $1/f(F^{-1}(\tau))$. There are several ways to obtain the estimator $\hat{s}_n(\tau)$, but we adopt a residual-based method here; see Appendix B. By Appendix B, we write (B) and (HS) for the NLAD estimates based on Bofinger (1975) and Hall and Sheather (1988), respectively, in Tables 4.4 to 4.6. Further, (0) in these tables means that their values are computed with the true values of the nuisance parameters.

From Tables 4.4 to 4.6, we may observe that the sizes are at least almost good although there are tendencies for under-rejections to arise to some degree in some NLAD estimates and for over-rejections to occur slightly in some NLS estimates. We finally compare the size-adjusted powers of the tests. In the figures, the values of coordinates on the horizontal axis show the difference between the parameter estimate and parameter value under the null hypothesis. Figures from 4.1 to 4.4 show the powers of the tests when I-regular model (4.14) is used. We note that the tests based on the NLAD estimators are superior to those based on the NLS estimators for fat-tailed regression errors in terms of the powers, and this is also the case with the RMSEs. Figures 4.5 to 4.12 indicate the powers of the tests when I-regular model (4.15) is used and we may find the same propensity.

4.7 Conclusion

The first contribution of this chapter is that we have investigated the asymptotic behavior of the NQR estimators when the regression function derivatives are given by H-regular functions and the covariate is given by an integrated time series. That is, we have obtained the limit distributions of the estimator when it is close to the true value. For the second contribution, we have proposed the FM-NQR estimator for an H-regular case since the derived NQR estimator corresponding to an H-regular function does not converge weakly to a mixed normal distribution. Due to this estimator, standard inference becomes available in consequence. As the third contribution, we have derived the limiting distribution of the NQR estimator for I-regular functions as well as H-regular ones when the model is linear in a parameter. Finally, we observe from simulations that our suggested NQR estimators are desirable relative to the NLS estimators in terms of estimation accuracy and powers of tests when distributions of regression errors possess fat tails.

4.8 Appendix

4.8.1 Functional classes

Definition 1 A function h is called *regular* on B if it satisfies

(a) for all $\beta \in B$, $h(\cdot, \beta)$ is continuous in a neighborhood of infinity,

(b) for any $\beta \in B$ and compact subset K of $\mathbb{R}$, there exist for each $\varepsilon > 0$ continuous functions $\underline{h}_\varepsilon$, $\overline{h}_\varepsilon$ and a constant $\delta_\varepsilon > 0$ such that $\underline{h}_\varepsilon(x, \beta) \leq h(y, \beta) \leq \overline{h}_\varepsilon(x, \beta)$ for all $|x - y| < \delta_\varepsilon$ on K, and such that $\int_K (\underline{h}_\varepsilon(x, \beta) - \overline{h}_\varepsilon(x, \beta))dx \to 0$ as $\varepsilon \to 0$, and

(c) for all $x \in \mathbb{R}$, $h(x, \cdot)$ is equicontinuous in a neighborhood of x.

Definition 2 A function g is called *H-regular* on B if it satisfies

(a) $g(\lambda x, \beta) = \kappa(\lambda)h(x, \beta) + R(x, \lambda, \beta)$ with h being regular on B,

and either

(b-i) $|R(x, \lambda, \beta)| \leq a(\lambda, \beta)P(x, \beta)$ with $\limsup_{\lambda \to \infty} \sup_{\beta \in B} \|\kappa(\lambda)^{-1} a(\lambda, \beta)\| = 0$, or (b-ii) $|R(x,\lambda,\beta)| \leq b(\lambda,\beta)P(x,\beta)Q(\lambda x,\beta)$ with $\limsup_{\lambda\to\infty}\sup_{\beta\in B}\|\kappa(\lambda)^{-1}b(\lambda,\beta)\| < \infty$,

where

(c) $\sup_{\beta\in B}P(\cdot,\beta)$ is locally bounded such that for some $c>0$, $\sup_{\beta\in B}P(x,\beta) = O(e^{c|x|})$ as $|x|\to\infty$, and $\sup_{\beta\in B}Q(\cdot,\beta)$ is bounded with $\sup_{\beta\in B}Q(x,\beta) = o(1)$ as $|x|\to\infty$.

Definition 3 A function g is called *I-regular* on B if it satisfies both

(a) for each $\beta_0 \in B$, there exist a neighborhood B_0 of β_0 and bounded integrable function $T:\mathbb{R}\to\mathbb{R}$ satisfying $|g(x,\beta)-g(x,\beta_0)| \leq T(x)\|\beta-\beta_0\|$ for all $\beta\in B_0$ and $\sup_{\beta\in B_0}|g(\cdot,\beta)|$ is integrable, and

(b) for some $c>0$ and $k>6/(p-2)$ with $p>4$ given in Assumption 4, $|g(x,\beta)-g(y,\beta)| \leq c|x-y|^k$ for all $\beta\in B$.

4.8.2 Estimation of sparsity functions

If we let $F_n^{-1}(t) = \hat{u}_{(i)}$ for $t\in[(i-1)/n, i/n)$ with ith smallest regression residual $\hat{u}_{(i)}$, $\hat{s}_n(\tau)$ may be obtained by

$$\hat{s}_n(\tau) = \frac{F_n^{-1}(\tau+h_n) - F_n^{-1}(\tau-h_n)}{2h_n},$$

where h_n is a bandwidth. Let Φ and ϕ denote the distribution function and density function of a standard normal random variable. The first way to estimate the bandwidth h_n is due to Bofinger (1975), which suggests computing

$$h_n = \left[\frac{4.5\phi(\Phi^{-1}(\tau))^4}{n(2\Phi^{-1}(\tau)^2+1)^2}\right]^{1/5}. \tag{4.17}$$

Another one is according to Hall and Sheather (1988), which recommends that

$$h_n = \left[\frac{1.5z_\alpha^2\phi(\Phi^{-1}(\tau))^2}{2n\Phi^{-1}(\tau)^2+1}\right]^{1/3}, \tag{4.18}$$

where z_α satisfies $\Phi(z_a) = 1-\alpha/2$, and α denotes the size of the test, 0.05, in this case. For more discussion on sparsity estimation, see Koenker (2005, chapter 4.10.1).

4.8.3 Useful lemmas

Remember that $z_t(\theta) = \tilde{g}_t(\beta)\psi_\tau(y_t - \alpha - g_t(\beta))$. Define the following values

$$\lambda_n(\theta) = \frac{1}{n}\sum_{t=1}^{n} E_{t-1} z_t(\theta),$$

$$\mu_t(\theta(\pi), d) = \sup_{\eta:\|\pi-\eta\|\le d} \left\| \tilde{\kappa}_n^{-1}\{z_t(\theta(\pi)) - z_t(\theta(\eta))\}\right\|,$$

where $\theta(\pi) = \theta_0 + n^{1/2}\tilde{\kappa}_n'\pi$ and π lies in the compact set $\Pi \subset \mathbb{R}^{1+\ell}$; see Assumption 5. More specifically, we may write $\theta(\pi) = (\alpha(\pi), \beta(\pi)')'$ with $\alpha(\pi) = \alpha_0(\tau) + n^{1/2}\pi_1$ and $\beta(\pi) = \beta_0 + n^{1/2}\dot{\kappa}_n'\pi_2$.

Lemma 2 *Let all the conditions in Theorem 1 hold. Then, for a sufficiently large n, there exist positive numbers a, b, c, and d_0 such that*

(i) $\|\tilde{\kappa}_n^{-1}\lambda_n(\theta(\pi))\| \ge n^{-1/2}a\|\pi\|$ *for* $\|\pi\| \le d_0$;

(ii) $E_{t-1}\mu_t(\theta(\pi), d) \le n^{-1/2}bd$ *for* $\|\pi\| + d \le d_0,\ d \ge 0$;

(iii) $E_{t-1}\mu_t(\theta(\pi), d)^2 \le n^{-1/2}cd$ *for* $\|\pi\| + d \le d_0,\ d \ge 0$.

Proof of Lemma 2 First we show (i). Since Assumption 2 implies $\tau = F_{t-1}(F^{-1}(\tau))$, by the definition of λ_n and applying the mean value theorem twice, we have

$$\begin{aligned}
\lambda_n(\theta) &= \frac{1}{n}\sum_{t=1}^{n} E_{t-1} z_t(\theta) \\
&= \frac{1}{n}\sum_{t=1}^{n} \tilde{g}_t(\beta)\{F_{t-1}(F^{-1}(\tau)) - F_{t-1}(F^{-1}(\tau) + \alpha - \alpha_0(\tau) + g_t(\beta) - g_t(\beta_0))\} \\
&= -\frac{1}{n}\sum_{t=1}^{n} \tilde{g}_t(\beta) f_{t-1}(\gamma_t(\theta))\{\alpha - \alpha_0(\tau) + g_t(\beta) - g_t(\beta_0)\} \\
&= -\frac{1}{n}\sum_{t=1}^{n} f_{t-1}(\gamma_t(\theta))\tilde{g}_t(\beta)\tilde{g}_t(\bar{\beta})'(\theta - \theta_0(\tau)),
\end{aligned} \tag{4.19}$$

where $\gamma_t(\theta)$ is a point between $F^{-1}(\tau)$ and $F^{-1}(\tau) + \alpha - \alpha_0(\tau) + g_t(\beta) - g_t(\beta_0)$, and $\bar{\beta}$ is a point between β and β_0. We now utilize Assumption 5. Under the assumption, $\gamma_t(\theta(\pi)) \to_p F^{-1}(\tau)$ uniformly in π and t. Thus, $\gamma_t(\theta)$ is asymptotically replaced with

$F^{-1}(\tau)$ without loss of generality. By Assumption 1 (c), furthermore, $f_{t-1}(F^{-1}(\tau))$ is asymptotically replaced with $f(F^{-1}(\tau))$; see Xiao (2009) for example. In consequence, noting that $\theta = \theta(\pi)$, we have

$$\lambda_n(\theta(\pi)) = -\frac{f(F^{-1}(\tau))}{n} \sum_{t=1}^{n} \tilde{g}_t(\beta(\pi))\tilde{g}_t(\bar{\beta}(\pi))'(\theta(\pi) - \theta_0(\tau))(1 + o_p(1)) \tag{4.20}$$

Therefore, it follows that

$$\begin{aligned} &\|\tilde{\kappa}_n^{-1}\lambda_n(\theta(\pi))\| \\ &= \frac{f(F^{-1}(\tau))}{n} \left\| \sum_{t=1}^{n} \tilde{\kappa}_n^{-1}\tilde{g}_t(\theta(\pi))\tilde{g}_t(\bar{\theta}(\pi))' \frac{\tilde{\kappa}_n'^{-1}}{n^{1/2}} \pi(1 + o_p(1)) \right\| \\ &= \frac{f(F^{-1}(\tau))}{n^{1/2}} \left\| \int_{-\infty}^{\infty} L(1,s)\tilde{h}(s,\beta_0)\tilde{h}(s,\beta_0)' ds\pi(1 + o_p(1)) \right\| \end{aligned} \tag{4.21}$$

uniformly in $\pi \in \Pi$ by PP (2001). Because the integration in (4.21) is eventually positive by the identifiability condition, the last value is bounded below by $n^{-1/2}a\|\pi\|$ for some $a > 0$.

Next, we prove (*ii*). By the definition of μ_t, the triangle inequality and the definition of $\psi_\tau(\cdot)$, we have

$$\begin{aligned} &\mu_t(\theta(\pi), d) \\ &= \sup_{\eta:\|\pi-\eta\|\le d} \left\| \tilde{\kappa}_n^{-1}\{z_t(\theta(\pi)) - z_t(\theta(\eta))\} \right\| \\ &\le \sup_{\eta} \left\| \tilde{\kappa}_n^{-1} \{\tilde{g}_t(\beta(\pi)) - \tilde{g}_t(\beta(\eta))\} \right. \\ &\qquad\qquad \left. \times\ \psi_\tau(u_{t\tau} - \alpha(\eta) + \alpha_0(\tau) - g_t(\beta(\eta)) + g_t(\beta_0)) \right\| \\ &+ \sup_{\eta} \left\| \tilde{\kappa}_n^{-1}\tilde{g}_t(\beta(\pi)) \{\psi_\tau(u_{t\tau} - \alpha(\pi) + \alpha_0(\tau) - g_t(\beta(\pi)) + g_t(\beta_0)) \right. \\ &\qquad\qquad \left. -\ \psi_\tau(u_{t\tau} - \alpha(\eta) + \alpha_0(\tau) - g_t(\beta(\eta)) + g_t(\beta_0))\} \right\| \\ &\le \sup_{\eta} \left\| \tilde{\kappa}_n^{-1} \{\tilde{g}_t(\beta(\pi)) - \tilde{g}_t(\beta(\eta))\} \right\| \\ &+ \left\| \tilde{\kappa}_n^{-1}\tilde{g}_t(\beta(\pi)) \right\| \sup_{\eta} \left| 1(u_{t\tau} < \alpha(\pi) - \alpha_0(\tau) + g_t(\beta(\pi)) - g_t(\beta_0)) \right. \\ &\qquad\qquad \left. -1(u_{t\tau} < \alpha(\eta) - \alpha_0(\tau) + g_t(\beta(\eta)) - g_t(\beta_0)) \right|. \end{aligned} \tag{4.22}$$

We focus on the second term of the last expression in (4.22). By monotonicity of the indicator functions, the supremum part is equal to either $1(u_{t\tau} < \sup_\eta[\alpha(\eta) + g_t(\beta(\eta))] -$

$\alpha_0(\tau) - g_t(\beta_0)) - 1(u_{t\tau} < \alpha(\pi) + g_t(\beta(\pi)) - \alpha_0(\tau) - g_t(\beta_0))$ or $1(u_{t\tau} < \alpha(\pi) + g_t(\beta(\pi)) - \alpha_0(\tau) - g_t(\beta_0)) - 1(u_{t\tau} < \inf_\eta[\alpha(\eta) + g_t(\beta(\eta))] - \alpha_0(\tau) - g_t(\beta_0))$. We then take conditional expectations on $\mu_t(\theta(\pi), d)$. If we apply the mean value theorem to the difference $E_{t-1}1(u_{t\tau} < \cdot) - E_{t-1}1(u_{t\tau} < \cdot)$ in either case, we have the expression so that

$$\begin{aligned} E_{t-1}\mu_t(\theta(\pi), d) \leq \sup_\eta \left\| \tilde{\kappa}_n^{-1} \{\tilde{g}_t(\beta(\pi)) - \tilde{g}_t(\beta(\eta))\} \right\| \\ + \left\| \tilde{\kappa}_n^{-1} \tilde{g}_t(\beta(\pi)) \right\| \bar{f} \sup_\eta |\alpha(\pi) - \alpha(\eta) + g_t(\beta(\pi)) - g_t(\beta(\eta))|, \end{aligned} \tag{4.23}$$

where $\bar{f}$ is the upper bound of the density f. Applying the mean value theorem to $\tilde{g}(\cdot)$ and $g(\cdot)$ in (4.23) and using the property of H-regular functions and $\|\tilde{\kappa}_n^{-1}\|^2 \|\ddot{\kappa}_n\| < \infty$, we have

$$\begin{aligned} E_{t-1}\mu_t(\theta(\pi), d) &\leq \left\| \tilde{\kappa}_n^{-1} \right\| \sup_\eta \|\ddot{g}_t(\beta(\eta))\| \sup_\eta \|\theta(\pi) - \theta(\eta)\| \\ &\quad + \left\| \tilde{\kappa}_n^{-1} \tilde{g}_t(\beta(\pi)) \right\| \bar{f} \sup_\eta \|\tilde{g}_t(\bar{\beta}(\eta))'(\theta(\pi) - \theta(\eta))\| \\ &\leq n^{-1/2} \|\tilde{\kappa}_n^{-1}\|^2 \|\ddot{\kappa}_n\| \sup_\eta \|\ddot{h}_t(\beta(\eta)) + o_p(1)\| \sup_\eta \|\pi - \eta\| \\ &\quad + n^{-1/2} \bar{f} \sup_\eta \left\| \tilde{h}_t(\beta(\eta)) + o_p(1) \right\|^2 \sup_\eta \|\pi - \eta\| \\ &\leq n^{-1/2} bd, \end{aligned} \tag{4.24}$$

where the last inequality in (4.23) follows from the local boundedness of the regular functions $\dot{h}$ and $\ddot{h}$ (Lemma A3 (b) of PP(2001), for a sufficiently large n and some $b > 0$.

Finally we show (*iii*). From (4.22), we have the bound

$$\begin{aligned} &\mu_t(\theta(\pi), d)^2 \\ &\leq 2 \sup_\eta \left\| \tilde{\kappa}_n^{-1} \{\tilde{g}_t(\beta(\pi)) - \tilde{g}_t(\beta(\eta))\} \right\|^2 \\ &+ 2 \left\| \tilde{\kappa}_n^{-1} \tilde{g}_t(\beta(\pi)) \right\|^2 \sup_\eta |1(u_{t\tau} < \alpha(\pi) - \alpha_0(\tau) + g_t(\beta(\pi)) - g_t(\beta_0)) \\ &\qquad - 1(u_{t\tau} < \alpha(\eta) - \alpha_0(\tau) + g_t(\beta(\eta)) - g_t(\beta_0))|. \end{aligned} \tag{4.25}$$

Thus the same manipulation as in (4.23) and (4.24) yields

$$
\begin{aligned}
&E_{t-1}\mu_t(\theta(\pi),d)^2 \\
&\leq 2n^{-1}\|\tilde{\kappa}_n^{-1}\|^4\|\ddot{\kappa}_n\|^2 \sup_\eta \|\ddot{h}_t(\beta(\eta))+o_p(1)\|^2 \sup_\eta \|\pi-\eta\|^2 \\
&+2n^{-1/2}\bar{f}^2 \sup_\eta \left\|\tilde{h}_t(\beta(\eta))+o_p(1)\right\|^3 \sup_\eta \|\pi-\eta\| \\
&\leq n^{-1/2}cd
\end{aligned}
\tag{4.26}
$$

for a sufficiently large n and some $c > 0$. ■

Lemma 3 *Define*

$$
Q_n(\theta(\eta),\theta(\pi)) = \frac{\left\|\tilde{\kappa}_n^{-1}\sum_{t=1}^n\{z_t(\theta(\eta))-z_t(\theta(\pi))-\lambda_n(\theta(\eta))+\lambda_n(\theta(\pi))\}\right\|}{n^{1/2}+n\left\|\tilde{\kappa}_n^{-1}\lambda_n(\theta(\eta))\right\|}. \tag{4.27}
$$

Suppose all the conditions in Theorem 1. Then, for a sufficiently large n, we have

$$
\sup_{\|\eta\|\leq d_0} Q_n(\theta(\eta),\theta(0)) = o_p(1). \tag{4.28}
$$

Proof of Lemma 3 The idea of the proof is to subdivide the cube $\|\eta\| \leq d_0$ into a slowly increasing number of smaller cubes and bound $Q_n(\theta(\eta),\theta(0))$ in probability on each of those smaller cubes. This way is identical to Huber (1967, pp. 227–230) except that:

- a fixed number γ (see eq. (36) in Huber (1967)) is now chosen arbitrary from $(0,1/2)$ in order to control the convergence rate in response to the localized parameter;
- terms of the form $\lambda(\cdot)$, $nE\mu_t(\cdot,\cdot)$ and so on in Huber's notation are replaced by those of the form λ_n, $\sum E_{t-1}\mu_t(\cdot,\cdot)$ and so on (i.e., "averaged" counterparts and conditional expectations are used) without affecting the validity of the argument if n is sufficiently large; see, for example, Powell (1984) and Weiss (1991) for similar discussions.

In consequence, the proof is completed by the proof in Huber (1967, pp. 227–230) because (N-3) in Huber (1967, p. 227) is satisfied from Lemma 2 and the other conditions (N-1) and

(N-2) in Huber (1967, pp. 226–227) hold in this case. ■

Remark on Lemma 3 The result does not hold when $\dot{g}$ and $\ddot{g}$ are I-regular because of the irregular convergence rate, $n^{1/4}$, at least utilizing Q_n defined in (4.27). We are then required to rearrange Q_n, in such a way as to make (4.28) hold, but it does not seem easy. A completely different way might be needed.

Lemma 4 *Define $\zeta_n = -D_n^{-1}\sum_{t=1}^n \tilde{g}_t \psi_\tau(u_{t\tau})$, where g is either I- or H-regular.*

(a) Suppose that all Assumptions in Theorem 4 are satisfied. Then, ζ_n converges weakly to ζ^I, where

$$\zeta^I = -\begin{bmatrix} U^\Psi(1) \\ \left(L(1,0)\int_{-\infty}^{\infty} g(s)^2 ds\right)^{1/2} W(1) \end{bmatrix}.$$

(b) Suppose that all Assumptions in Theorem 5 are satisfied. Then, ζ_n converges weakly to ζ^H, where

$$\zeta^H = -\begin{bmatrix} U^\Psi(1) \\ \int_0^1 h(V(r))dU^\Psi(r) \end{bmatrix}.$$

Proof of Lemma 4 The first element of ζ_n, which is identical in (a) and (b), converges weakly to $U^\Psi(1)$ by a central limit theorem for martingale difference sequences. For the second element of ζ_n in (a) and (b), the result is immediately obtained from Theorems 3.2 and 3.3 of PP (2001), respectively. That is, in the case of (a), we have

$$n^{-1/4}\sum_{t=1}^n g_t \psi_\tau(u_{t\tau}) \xrightarrow{d} \left(L(1,0)\int_{-\infty}^{\infty} g(s)^2 ds\right)^{1/2} W(1), \tag{4.29}$$

where $W(r)$ is a Brownian motion independent of $(U^\Psi(r), V(r))$ and has variance $r\omega_\psi^2$. Since joint convergence is ensured from Lemma 5 in Chang et al. (2001), this leads to result (a).

Similarly, in the case of (b), we have

$$n^{-1/2}\kappa_n^{-1}\sum_{t=1}^{n} g_t\psi_\tau(u_{t\tau}) \xrightarrow{d} \int_0^1 h(V(r))dU^{\psi}(r). \tag{4.30}$$

This in turn leads to result (b) since joint convergence is again ensured from Lemma 5 in Chang et al. (2001). ■

Lemma 5 *Define*

$$\Xi_n(\pi) = \sum_{t=1}^{n}\left(\pi' D_n^{-1}\tilde{g}_t - u_{t\tau}\right)\left[1\left\{\pi' D_n^{-1}\tilde{g}_t > u_{t\tau} > 0\right\} - 1\left\{\pi' D_n^{-1}\tilde{g}_t < u_{t\tau} < 0\right\}\right],$$

where g is either I- or H-regular.

(a) Suppose that all Assumptions in Theorem 4 are satisfied. Then, $\Xi_n(\pi)$ converges in probability to $\pi'\Xi^I\pi/2$, where

$$\Xi^I = f(F^{-1}(\tau))\begin{bmatrix} 1 & 0 \\ 0 & L(1,0)\int_{-\infty}^{\infty} g(s)^2 ds\theta^2 \end{bmatrix}.$$

(b) Suppose that all Assumptions in Theorem 5 are satisfied. Then, $\Xi_n(\pi)$ converges in probability to $\pi'\Xi^H\pi/2$, where

$$\Xi^H = f(F^{-1}(\tau))\begin{bmatrix} 1 & \int_0^1 h(V(r))dr \\ \int_0^1 h(V(r))dr & \int_0^1 h(V(r))^2 dr \end{bmatrix}.$$

Proof of Lemma 5 Denote $V_n = \sum_{t=1}^{n} V_{tn}$ with $V_{tn} = \left(\pi' D_n^{-1}\tilde{g}_t - u_{t\tau}\right) 1\left\{\pi' D_n^{-1}\tilde{g}_t > u_{t\tau} > 0\right\}$. We consider the truncation of V_{tn} by an indicator function I_t, which depends on the class of g. In the I-regular case, we set $I_t = I_t^I$, where

$$I_t^I = 1\left(0 < g_t \leq m\right) \tag{4.31}$$

for some $m > 0$. In the H-regular case, we set $I_t = I_t^H$, where

$$I_t^H = 1\left(0 < \kappa_n^{-1} g_t \le m\right) = 1\left(0 < h\left(n^{-1/2} x_t\right) + o_p(1) \le m\right) \tag{4.32}$$

for some $m > 0$. Define V_{nm} and V_{tnm} as

$$V_{nm} = \sum_{t=1}^{n} V_{tnm} = \sum_{t=1}^{n} V_{tn} I_t. \tag{4.33}$$

In addition, we define its conditional expectation as $\mu_{tnm} = E_{t-1} V_{tnm}$ and its summation as $\mu_{nm} = \sum_{t=1}^{n} \mu_{tnm}$.

Using the notations above, we will derive the probability limit of V_n through the following steps. First, we compute the limiting variable of μ_{nm} as $n \to \infty$ and $m \to \infty$. Next, we check the asymptotic equivalence of V_{nm} and μ_{nm}. Finally, we verify that the effect of truncation by I_t is asymptotically negligible.

For the first step, we have, from Assumption 7,

$$\begin{aligned}
\mu_{nm} &= \sum_{t=1}^{n} E_{t-1}\left[\left(\pi' D_n^{-1} \tilde{g}_t - u_{t\tau}\right) 1\left(\pi' D_n^{-1} \tilde{g}_t > u_{t\tau} > 0\right) I_t\right] \\
&= \sum_{t=1}^{n} \int_{F^{-1}(\tau)}^{\pi' D_n^{-1} \tilde{g}_t + F^{-1}(\tau)} \left(\pi' D_n^{-1} \tilde{g}_t + F^{-1}(\tau) - u\right) f(u) du I_t \\
&= \sum_{t=1}^{n} \int_{F^{-1}(\tau)}^{\pi' D_n^{-1} \tilde{g}_t + F^{-1}(\tau)} \int_{u}^{\pi' D_n^{-1} \tilde{g}_t + F^{-1}(\tau)} dv f(u) du I_t \\
&= \sum_{t=1}^{n} \int_{F^{-1}(\tau)}^{\pi' D_n^{-1} \tilde{g}_t + F^{-1}(\tau)} \int_{F^{-1}(\tau)}^{v} f(u) du dv I_t \\
&= \sum_{t=1}^{n} \int_{F^{-1}(\tau)}^{\pi' D_n^{-1} \tilde{g}_t + F^{-1}(\tau)} (v - F^{-1}(\tau)) \frac{F(v) - F(F^{-1}(\tau))}{v - F^{-1}(\tau)} dv I_t,
\end{aligned}$$

where the fourth equality follows from Fubini's theorem. Under Assumption 7, we then get

$$\begin{aligned}
\mu_{nm} &= f(F^{-1}(\tau)) \sum_{t=1}^{n} \int_{F^{-1}(\tau)}^{\pi' D_n^{-1} \tilde{g}_t + F^{-1}(\tau)} v dv I_t \\
&= \frac{f(F^{-1}(\tau))}{2} \pi' D_n^{-1} \sum_{t=1}^{n} \tilde{g}_t \tilde{g}_t' I_t D_n^{-1} \pi.
\end{aligned} \tag{4.34}$$

We analyze μ_{nm} in (4.34) with explicit functional form of g. Hereafter, we use the notation $\cdot^I$ and $\cdot^H$ for several variables in the same way as I_t^I and I_t^H.

When g is I-regular, we remember that $D_n = D_n^I = \mathrm{diag}(n^{1/2}, n^{1/4})$ and $I_t = I_t^I$ was given in (4.31). Because the product of I-regular functions is again I-regular (Lemma A6 (b) of PP (2001)), μ_{nm} in (4.34) deduces

$$\begin{aligned}\mu_{nm} = \mu_{nm}^I &= \frac{f(F^{-1}(\tau))}{2} \sum_{t=1}^{n} \pi' \begin{bmatrix} n^{-1} & n^{-3/4} g_t \\ n^{-3/4} g_t & n^{-1/2} g_t^2 \end{bmatrix} I_t^I \pi \\ &\xrightarrow{p} \frac{f(F^{-1}(\tau))}{2} \pi' \begin{bmatrix} \xi_{11m}^I & 0 \\ 0 & \xi_{22m}^I \end{bmatrix} \pi =: \mu_m^I \end{aligned} \tag{4.35}$$

as $n \to \infty$, where

$$\xi_{11m}^I = \int_0^1 1\left(0 < g(V(r)) \leq m\right) dr,$$
$$\xi_{22m}^I = L(1,0) \int_{-\infty}^{\infty} g(s)^2 1\left(0 < g(s) \leq m\right) ds.$$

Letting $m \to \infty$, we then obtain

$$\mu_m^I \xrightarrow{p} \frac{f(F^{-1}(\tau))}{2} \pi' \begin{bmatrix} \xi_{11}^I & 0 \\ 0 & \xi_{22}^I \end{bmatrix} \pi =: \mu^I, \tag{4.36}$$

where

$$\xi_{11}^I = \int_0^1 1\left(0 < g(V(r))\right) dr,$$
$$\xi_{22}^I = L(1,0) \int_{-\infty}^{\infty} g(s)^2 1\left(0 < g(s)\right) ds.$$

For the next step, we show the asymptotic equivalence of V_{nm}^I and μ_{nm}^I for a sufficiently large n. Note that $\{V_{tnm}^I - \mu_{tnm}^I\}$ forms a martingale difference sequence. Since g is I-regular, we see that

$$\pi' D_n^{I-1} \tilde{g}_t I_t^I \xrightarrow{p} 0 \tag{4.37}$$

uniformly in $t = 1, \ldots, n$. Then, it follows that

$$\begin{aligned}\sum_{t=1}^{n} E_{t-1} V_{tnm}^{I2} &= \sum_{t=1}^{n} E_{t-1} \left[\left(\pi' D_n^{I-1} \tilde{g}_t - u_{t\tau}\right)^2 1\left(\pi' D_n^{I-1} \tilde{g}_t > u_{t\tau} > 0\right) I_t^I \right] \\ &\leq \sum_{t=1}^{n} E_{t-1} \left[\pi' D_n^{I-1} \tilde{g}_t I_t^I V_{tnm}^I\right] \leq \max_{1 \leq t \leq n} \left\{\pi' D_n^{I-1} \tilde{g}_t I_t^I\right\} O_p(1) \xrightarrow{p} 0. \end{aligned} \tag{4.38}$$

Therefore, from the arguments of Pollard (1984, p. 171), we conclude that $V_{nm}^I \to_p \mu_m^I$ as $n \to \infty$.

We finally prove $V_n^I \to_p \mu^I$. This statement is true provided that the approximation error brought by truncation at m is asymptotically negligible. That is, it suffices to show that, for any $\varepsilon > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} P\left(|V_n - V_{nm}| \geq \varepsilon\right) = 0 \tag{4.39}$$

for $V_n = V_n^I$ and $V_{nm} = V_{nm}^I$. However, we may observe that

$$\begin{aligned} &\limsup_{n\to\infty} P\left(|V_n^I - V_{nm}^I| \geq \varepsilon\right) \leq \limsup_{n\to\infty} P\left(|V_n^I - V_{nm}^I| > 0\right) \\ &= \limsup_{n\to\infty} P\left(\sum_{t=1}^n \left(\pi' D_n^{I-1} \tilde{g}_t - u_{t\tau}\right) 1\left(\pi' D_n^{I-1} \tilde{g}_t > u_{t\tau} > 0\right) 1\left(g_t > m\right) > 0\right) \\ &\leq \limsup_{n\to\infty} P\left(\bigcup_{t=1}^n \{g_t > m\}\right) = \limsup_{n\to\infty} P\left(\max_{1\leq t\leq n} g_t > m\right) \leq P\left(\max_{x\in\mathbb{R}} g(x) > m\right). \end{aligned} \tag{4.40}$$

Thus, (4.39) holds by letting $m \to \infty$. In consequence, we obtain $V_n^I \to_p \mu^I$. Because the above argument is also true for $-(\pi' D_n^{I-1} \tilde{g}_t - u_{t\tau}) 1\left\{\pi' D_n^{I-1} \tilde{g}_t < u_{t\tau} < 0\right\}$, result (a) follows.

When g is H-regular, we remember that $D_n = D_n^H = n^{1/2}\tilde{\kappa}_n$ and that $I_t = I_t^H$ was given in (4.32). Then, μ_{nm} in (4.34) deduces

$$\begin{aligned} \mu_{nm} = \mu_{nm}^H &= \frac{f(F^{-1}(\tau))}{2} \sum_{t=1}^n \pi' \begin{bmatrix} n^{-1} & n^{-1}\kappa_n^{-1} g_t \\ n^{-1}\kappa_n^{-1} g_t & n^{-1}\kappa_n^{-2} g_t^2 \end{bmatrix} I_t^H \pi \\ &\xrightarrow{p} \frac{f(F^{-1}(\tau))}{2} \pi' \begin{bmatrix} \xi_{11m}^H & \xi_{12m}^H \\ \xi_{21m}^H & \xi_{22m}^H \end{bmatrix} \pi =: \mu_m^H \end{aligned} \tag{4.41}$$

as $n \to \infty$, where

$$\begin{aligned} \xi_{11m}^H &= \int_0^1 1\left(0 < h(V(r)) \leq m\right) dr, \\ \xi_{12m}^H &= \int_0^1 h(V(r)) 1\left(0 < h(V(r)) \leq m\right) dr = \xi_{21m}^H, \\ \xi_{22m}^H &= \int_0^1 h(V(r))^2 1\left(0 < h(V(r)) \leq m\right) dr. \end{aligned}$$

Letting $m \to \infty$, we obtain

$$\mu_m^H \xrightarrow{p} \frac{f(F^{-1}(\tau))}{2}\pi' \begin{bmatrix} \xi_{11}^H & \xi_{12}^H \\ \xi_{21}^H & \xi_{22}^H \end{bmatrix} \pi =: \mu^H, \tag{4.42}$$

where

$$\xi_{11}^H = \int 1\,(0 < h(V(r)))\,dr,$$
$$\xi_{12}^H = \int_0^1 h(V(r))1\,(0 < h(V(r)))\,dr = \xi_{21}^H,$$
$$\xi_{22}^H = \int_0^1 h(V(r))^2 1\,(0 < h(V(r)))\,dr.$$

Next, we show the asymptotic equivalence of V_{nm}^H and μ_{nm}^H as in the I-regular case. Since g is H-regular, we have (4.37) again uniformly in t. Thus, from the same argument as in (4.38), we observe that

$$\sum_{t=1}^n E_{t-1}\left[V_{tnm}^{H2}\right] \leq \max_{1\leq t\leq n}\left\{\pi' D_n^{H-1}\tilde{g}_t 1\left(0 < \kappa_n^{-1}\tilde{g}_t \leq m\right)\right\} O_p(1) \xrightarrow{p} 0. \tag{4.43}$$

Therefore, from the arguments of Pollard (1984, p. 171), we conclude that $V_{nm}^H - \mu_{nm}^H \to_p 0$ as $n \to \infty$.

We finally prove $V_{nm}^H \to_p \mu^H$, but (4.39) holds for $V_n = V_n^H$ and $V_{nm} = V_{nm}^H$ by a similar calculation as in (4.40). That is, for any $\varepsilon > 0$, we have

$$\begin{aligned}
&\limsup_{n\to\infty} P(|V_n^H - V_{nm}^H| \geq \varepsilon) \leq \limsup_{n\to\infty} P(|V_n^H - V_{nm}^H| > 0) \\
&= \limsup_{n\to\infty} P\left(\sum_{t=1}^n \left(\pi' D_n^{H-1}\tilde{g}_t - u_{t\tau}\right) 1\left(\pi' D_n^{H-1}\tilde{g}_t > u_{t\tau} > 0\right) 1\left(\kappa_n^{-1} g_t > m\right) > 0\right) \\
&\leq \limsup_{n\to\infty} P\left(\bigcup_{t=1}^n \left\{\kappa_n^{-1} g_t > m\right\}\right) \leq \limsup_{n\to\infty} P\left(\max_{1\leq t\leq n}\left\{h\left(n^{-1/2}x_t\right) + o_p(1)\right\} > m\right) \\
&\leq P\left(\max_{x\in\mathscr{K}} h(x) > m\right),
\end{aligned}$$

where $\mathscr{K} = [x_{\min} - 1, x_{\max} + 1]$ with $x_{\min(\max)} = \min(\max)_{0\leq r\leq 1} V(r)$ in the last equation (see Lemma A3 (b) of PP (2001)). Thus, (4.39) holds by letting $m \to \infty$. Therefore $V_n^H \to_p \mu^H$. Because the above argument is also true for $-(\pi' D_n^{H-1}\tilde{g}_t - u_{t\tau})1\left\{\pi' D_n^{H-1}\tilde{g}_t < u_{t\tau} < 0\right\}$, the result (b) follows. ■

4.8.4 Proofs of the main results

Proof of Lemma 1 This is the same as Weiss (1991). See also Ruppert and Carroll (1980). ■

Proof of Theorem 1 Because we can write

$$\sum_{t=1}^{n} z_t(\hat{\theta}_n) = \sum_{t=1}^{n} \{z_t(\hat{\theta}_n) - z_t(\theta_0) - \lambda_n(\hat{\theta}_n)\} + \sum_{t=1}^{n} \{z_t(\theta_0) + \lambda_n(\hat{\theta}_n)\}, \tag{4.44}$$

we have, from Lemmas 3 and 4.8,

$$\frac{\|\tilde{\kappa}_n^{-1} \sum_{t=1}^{n} \{z_t(\theta_0) + \lambda_n(\hat{\theta}_n)\}\|}{n^{1/2} + n\|\tilde{\kappa}_n^{-1}\lambda_n(\hat{\theta}_n)\|} \leq \sup_{\|\pi\| \leq d_0} Q_n(\theta(\pi), \theta_0) + o_p(1) = o_p(1) \tag{4.45}$$

Since $n^{1/2}\|\tilde{\kappa}_n^{-1}\lambda_n(\hat{\theta}_n)\|$ is $O_p(1)$ from the proof of Lemma 2 (*i*), it follows from (4.44) and (4.45) that

$$n^{-1/2}\tilde{\kappa}_n^{-1} \sum_{t=1}^{n} z_t(\theta_0) + n^{1/2}\tilde{\kappa}_n^{-1}\lambda_n(\hat{\theta}_n) = o_p(1). \tag{4.46}$$

By definitions of z_t and λ_n, (4.46) is equivalent to

$$n^{-1/2}\tilde{\kappa}_n^{-1} \sum_{t=1}^{n} \tilde{g}_t(\beta_0)\psi_\tau(u_{t\tau}) + n^{-1/2}\tilde{\kappa}_n^{-1} \sum_{t=1}^{n} \tilde{g}_t(\hat{\beta}_n)E_{t-1}\psi_\tau(y_t - \hat{\alpha}_n - g_t(\hat{\beta}_n)) = o_p(1). \tag{4.47}$$

By the proof of Lemma 2 (*i*) again, we consequently have

$$\begin{aligned} &n^{1/2}\tilde{\kappa}_n'(\hat{\theta}_n - \theta_0) \\ &= \left[\frac{f(F^{-1}(\tau))}{n} \sum_{t=1}^{n} \tilde{h}_t(\beta_0)\tilde{h}_t(\beta_0)'\right]^{-1} \left[-n^{-1/2} \sum_{t=1}^{n} \tilde{h}_t(\beta_0)\psi_\tau(u_{t\tau}) + o_p(1)\right]. \end{aligned} \tag{4.48}$$

Thus the result follows immediately by the limit theory of Park and Phillips (2001) and Lemma 4 . ■

Proof of Theorem 2 First, we observe that

$$S_n = n^{-1}\sum_{t=1}^{n}\tilde{h}(x_t/\sqrt{n},\hat{\beta}_n)\tilde{h}(x_t/\sqrt{n},\hat{\beta}_n)' \xrightarrow{p} \int_0^1 \tilde{h}(V(r),\beta_0)\tilde{h}(V(r),\beta_0)'dr \tag{4.49}$$

and, by de Jong (2002),

$$\begin{aligned} T_n &= n^{-1/2}\sum_{t=1}^{n}\left\{\tilde{h}(n^{-1/2}x_t,\hat{\beta}_n)\Delta x_t - n^{-1/2}\hat{\lambda}\nabla\tilde{h}(n^{-1/2}x_t,\hat{\beta}_n)\right\} \\ &\xrightarrow{p} \int_0^1 \tilde{h}(V(r),\beta_0)dV(r) + \lambda\int_0^1 \nabla\tilde{h}(V(r),\beta_0)dr - \lambda\int_0^1 \nabla\tilde{h}(V(r),\beta_0)dr \\ &= \int_0^1 \tilde{h}(V(r),\beta_0)dV(r), \end{aligned} \tag{4.50}$$

where λ in the first integral of (4.50) is the one-sided long-run variance. Applying Theorem 1, and utilizing (4.49) and (4.50), we see that

$$\begin{aligned} & n^{1/2}\tilde{\kappa}_n'(\hat{\theta}_n^+ - \theta_0) \\ &= n^{1/2}\tilde{\kappa}_n'(\hat{\theta}_n - \theta_0) - \frac{1}{\widehat{f(F^{-1}(\tau))}}\frac{\hat{\omega}_{\psi v}}{\hat{\omega}_v^2}S_n^{-1}T_n \\ &\xrightarrow{d} \frac{1}{f(F^{-1}(\tau))}\left[\int_0^1 \tilde{h}(V(r),\beta_0)\tilde{h}(V(r),\beta_0)'dr\right]^{-1}\int_0^1 \tilde{h}(V(r),\beta_0)dU^{\Psi}(r) \\ &\quad - \frac{1}{f(F^{-1}(\tau))}\frac{\omega_{\psi v}}{\omega_v^2}\left[\int_0^1 \tilde{h}(V(r),\beta_0)\tilde{h}(V(r),\beta_0)'dr\right]^{-1}\left[\int_0^1 \tilde{h}(V(r),\beta_0)dV(r)\right]. \end{aligned}$$

If we let

$$U^{\Psi+}(r) = U^{\Psi}(r) - \frac{\omega_{\psi v}}{\omega_v^2}V(r),$$

then simple algebra gives the desired limiting variable

$$\frac{1}{f(F^{-1}(\tau))}\left[\int_0^1 \tilde{h}(V(r),\beta_0)\tilde{h}(V(r),\beta_0)'dr\right]^{-1}\int_0^1 \tilde{h}(V(r),\beta_0)dU^{\Psi+}(r),$$

which completes the proof. ■

Proof of Theorem 3 The result is immediately achieved by Theorem 2. ■

Proofs of Theorems 4 and 5 If we apply the identity

$$\rho_\tau(u-\gamma)-\rho_\tau(u)=-\gamma\psi_\tau(u)+(\gamma-u)\{1(\gamma>u>0)-1(\gamma<u<0)\}$$

to $M_n(\theta)$, it is written as

$$M_n(\theta)=\pi'\zeta+\Xi_n(\pi), \tag{4.51}$$

where

$$\zeta_n=-D_n^{-1}\sum_{t=1}^{n}\tilde{g}_t\psi_\tau(u_t),$$
$$\Xi_n(\pi)=\sum_{t=1}^{n}\left(\pi'D_n^{-1}\tilde{g}_t-u_{t\tau}\right)\left[1\left\{\pi'D_n^{-1}\tilde{g}_t>u_{t\tau}>0\right\}-1\left\{\pi'D_n^{-1}\tilde{g}_t<u_{t\tau}<0\right\}\right].$$

From Lemma 4, ζ_n converges weakly to ζ, where $\zeta=\zeta^I$ for I- and $\zeta=\zeta^H$ for H-regular cases, respectively. According to Lemma 5, $\Xi_n(\pi)$ converges to $\pi'\Xi\pi/2$ in probability, where $\Xi=\Xi^I$ for I- and $\Xi=\Xi^H$ for H-regular cases, respectively. Thus, (4.51) converges weakly to the quadratic form so that

$$M_n(\pi)\xrightarrow{d}-\pi'\zeta+\frac{1}{2}\pi'\Xi\pi=:M(\pi)$$

for both I- and H-regular cases. Note that $M_n(\pi)$ and $M(\pi)$ are minimized at $\hat{\pi}=D_n(\hat{\theta}-\theta_0)$ and $\Xi^{-1}\zeta$, respectively. By the convexity lemma of Pollard (1991) and arguments of Knight (1989), the results are achieved. ■

Table 4.1: The RMSEs for estimating β_{10} in I-regular model (4.14).

		$n = 250$			$n = 500$		
u_t	v_t	NLS	NLAD	ratio	NLS	NLAD	ratio
SN	SN	0.1226	0.1439	0.85	0.1046	0.1222	0.86
SN	t_9	0.1232	0.1432	0.86	0.1091	0.1234	0.88
t_3	SN	0.1948	0.1506	1.29	0.1707	0.1317	1.30
t_3	t_9	0.2050	0.1653	1.24	0.1831	0.1449	1.26

Table 4.2: The RMSEs for estimating β_{20} in H-regular model (4.15).

		$n = 250$			$n = 500$		
u_t	v_t	NLS	NLAD	ratio	NLS	NLAD	ratio
SN	SN	0.3529	0.3974	0.89	0.3212	0.3665	0.88
SN	t_9	0.4970	0.5560	0.89	0.4602	0.5165	0.89
t_3	SN	0.5387	0.4282	1.26	0.4999	0.3915	1.28
t_3	t_9	0.7800	0.6467	1.21	0.6623	0.5397	1.23

Table 4.3: The RMSEs for estimating β_{30} in H-regular model (4.16).

		$n = 250$			$n = 500$		
u_t	v_t	NLS	NLAD	ratio	NLS	NLAD	ratio
SN	SN	0.0047	0.0053	0.88	0.0024	0.0029	0.85
SN	t_9	0.0043	0.0048	0.90	0.0020	0.0023	0.85
t_3	SN	0.0085	0.0064	1.33	0.0040	0.0032	1.27
t_3	t_9	0.0062	0.0050	1.23	0.0033	0.0026	1.27

Table 4.4: The empirical sizes of test when the model is (4.14).

		$n = 250$				
u_t	v_t	NLS (0)	NLS	NLAD (0)	NLAD (B)	NLAD (HS)
SN	SN	0.066	0.066	0.042	0.026	0.036
SN	t_9	0.054	0.060	0.042	0.038	0.042
t_3	SN	0.056	0.044	0.058	0.032	0.040
t_3	t_9	0.054	0.056	0.036	0.026	0.036
		$n = 500$				
u_t	v_t	NLS (0)	NLS	NLAD (0)	NLAD (B)	NLAD (HS)
SN	SN	0.046	0.046	0.032	0.022	0.024
SN	t_9	0.064	0.070	0.058	0.056	0.058
t_3	SN	0.058	0.054	0.060	0.048	0.048
t_3	t_9	0.060	0.056	0.046	0.044	0.050

Table 4.5: The empirical sizes of test when the model is (4.15).

		$n = 250$				
u_t	v_t	NLS (0)	NLS	NLAD (0)	NLAD (B)	NLAD (HS)
SN	SN	0.072	0.070	0.052	0.046	0.054
SN	t_9	0.064	0.064	0.050	0.044	0.058
t_3	SN	0.068	0.072	0.058	0.050	0.058
t_3	t_9	0.070	0.066	0.076	0.062	0.070

		$n = 500$				
u_t	v_t	NLS (0)	NLS	NLAD (0)	NLAD (B)	NLAD (HS)
SN	SN	0.078	0.080	0.054	0.046	0.060
SN	t_9	0.056	0.058	0.032	0.026	0.034
t_3	SN	0.058	0.072	0.068	0.058	0.064
t_3	t_9	0.060	0.060	0.058	0.036	0.056

Table 4.6: The empirical sizes of test when the model is (4.16).

		$n = 250$				
u_t	v_t	NLS (0)	NLS	NLAD (0)	NLAD (B)	NLAD (HS)
SN	SN	0.056	0.052	0.044	0.032	0.036
SN	t_9	0.058	0.068	0.030	0.028	0.032
t_3	SN	0.238	0.054	0.068	0.040	0.050
t_3	t_9	0.202	0.032	0.054	0.024	0.026
		$n = 500$				
u_t	v_t	NLS (0)	NLS	NLAD (0)	NLAD (B)	NLAD (HS)
SN	SN	0.062	0.064	0.056	0.050	0.058
SN	t_9	0.052	0.052	0.032	0.032	0.038
t_3	SN	0.244	0.060	0.064	0.042	0.046
t_3	t_9	0.230	0.042	0.048	0.022	0.022

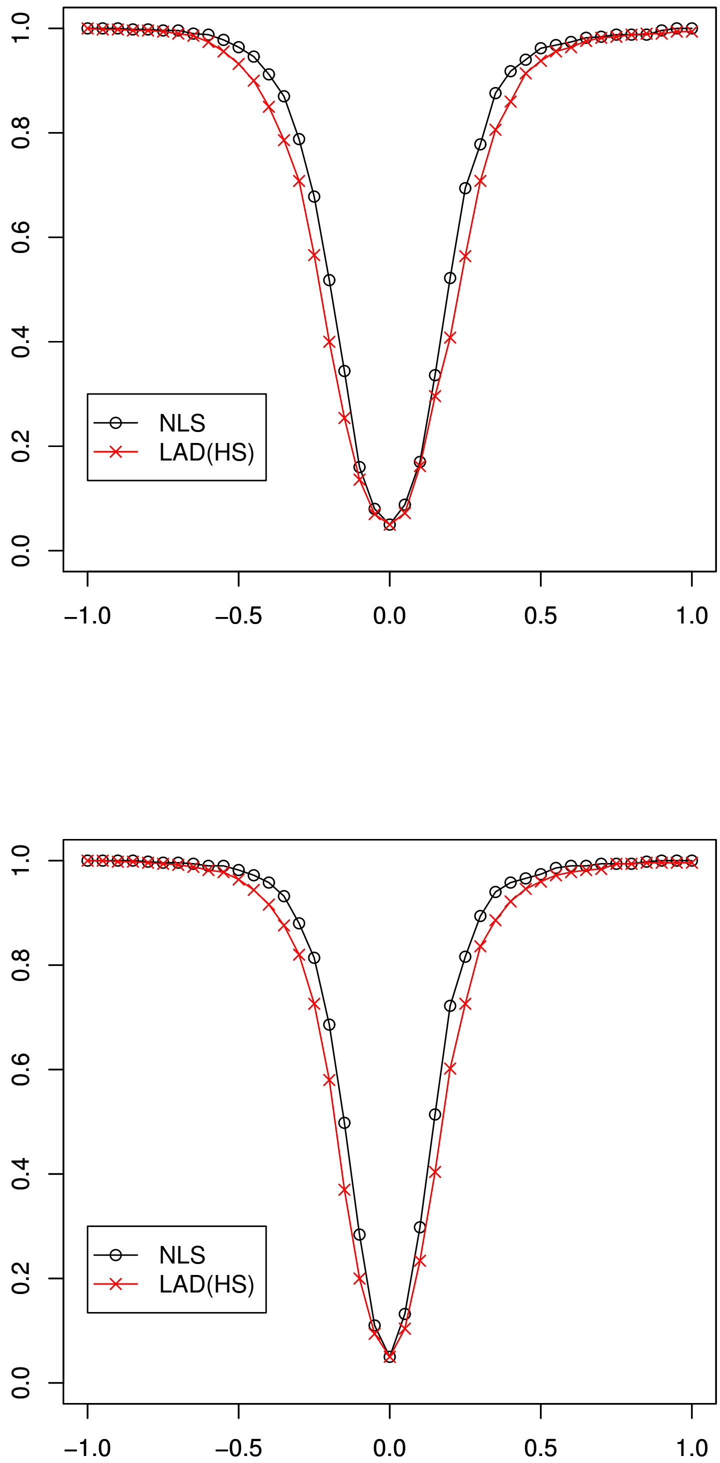


Figure 4.1: The powers of test when the model is (4.14) with $u_t \sim SN$ and $v_t \sim SN$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.

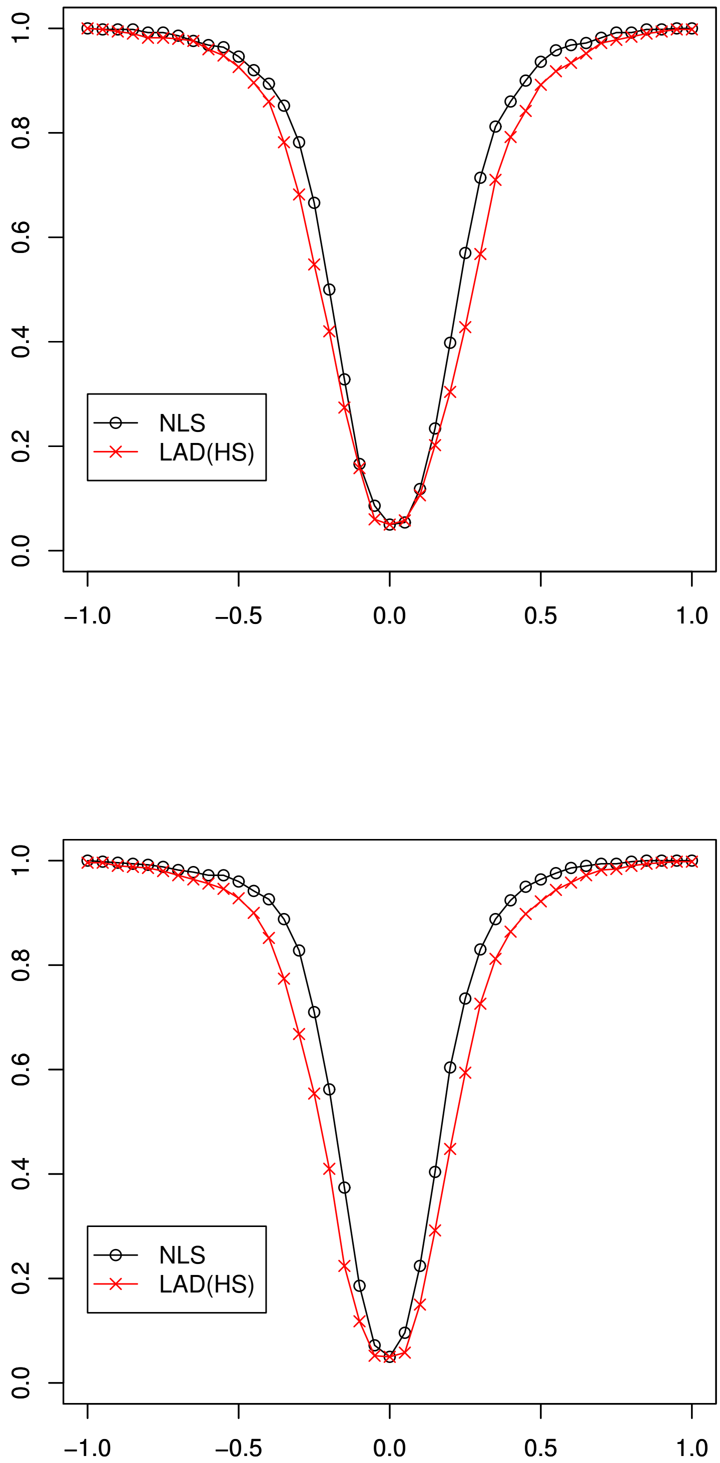
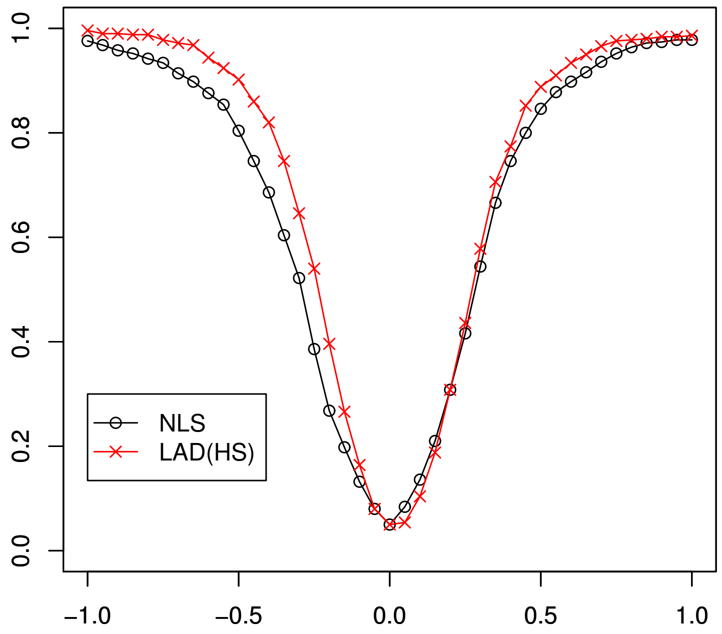


Figure 4.2: The powers of test when the model is (4.14) with $u_t \sim SN$ and $v_t \sim t_9$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.



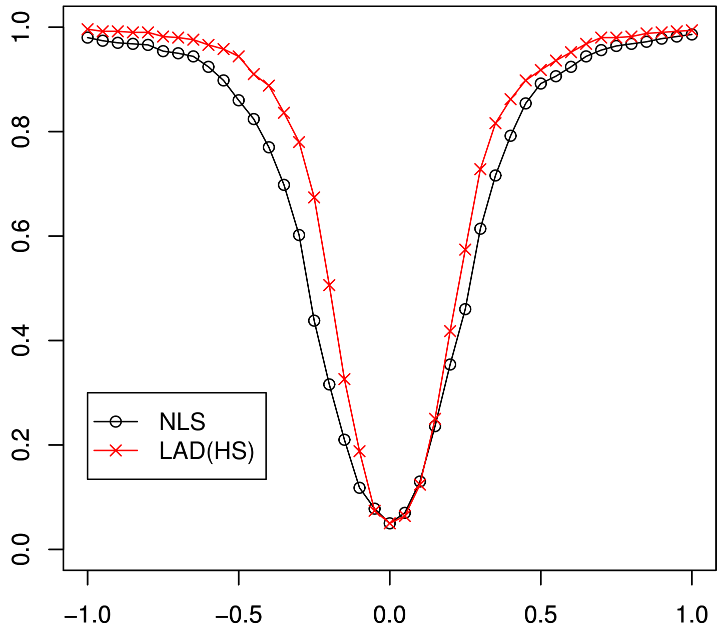
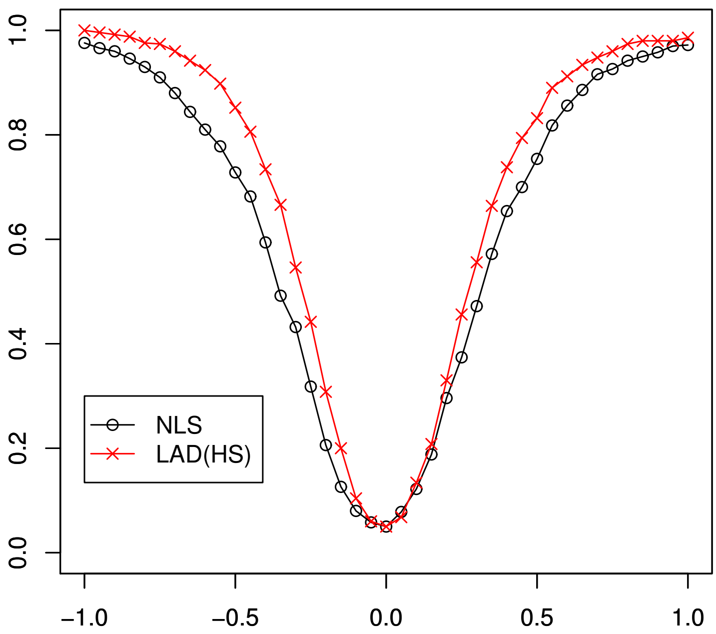


Figure 4.3: The powers of test when the model is (4.14) with $u_t \sim t_3$ and $v_t \sim SN$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.



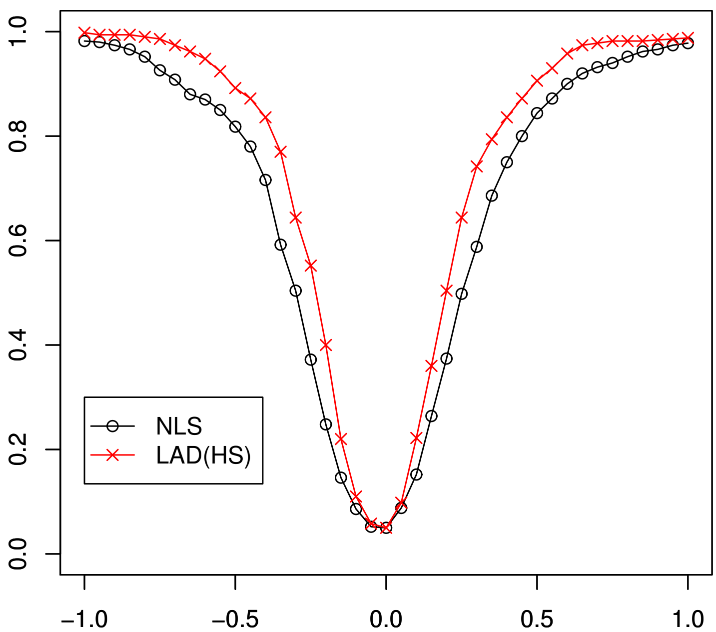


Figure 4.4: The powers of test when the model is (4.14) with $u_t \sim t_3$ and $v_t \sim t_9$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.

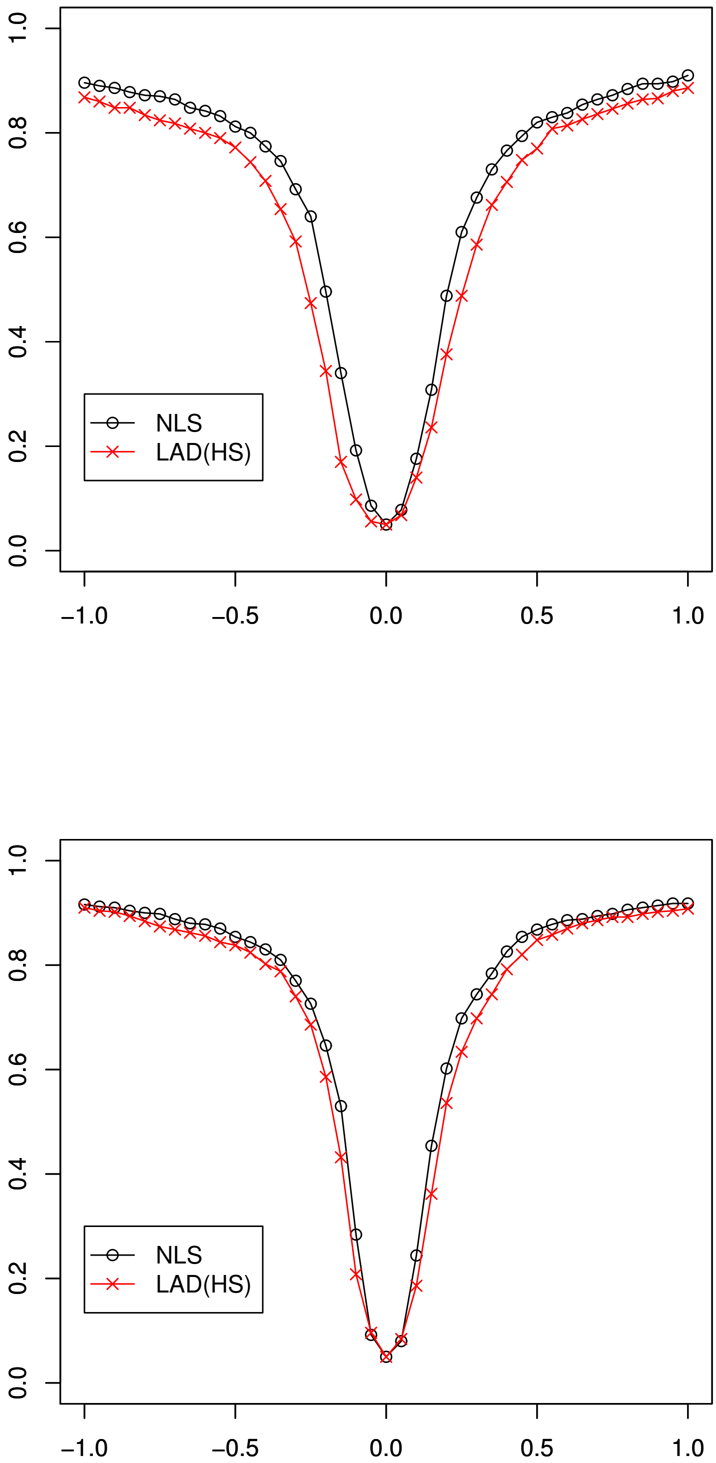


Figure 4.5: The powers of test when the model is (4.15) with $u_t \sim SN$ and $v_t \sim SN$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.

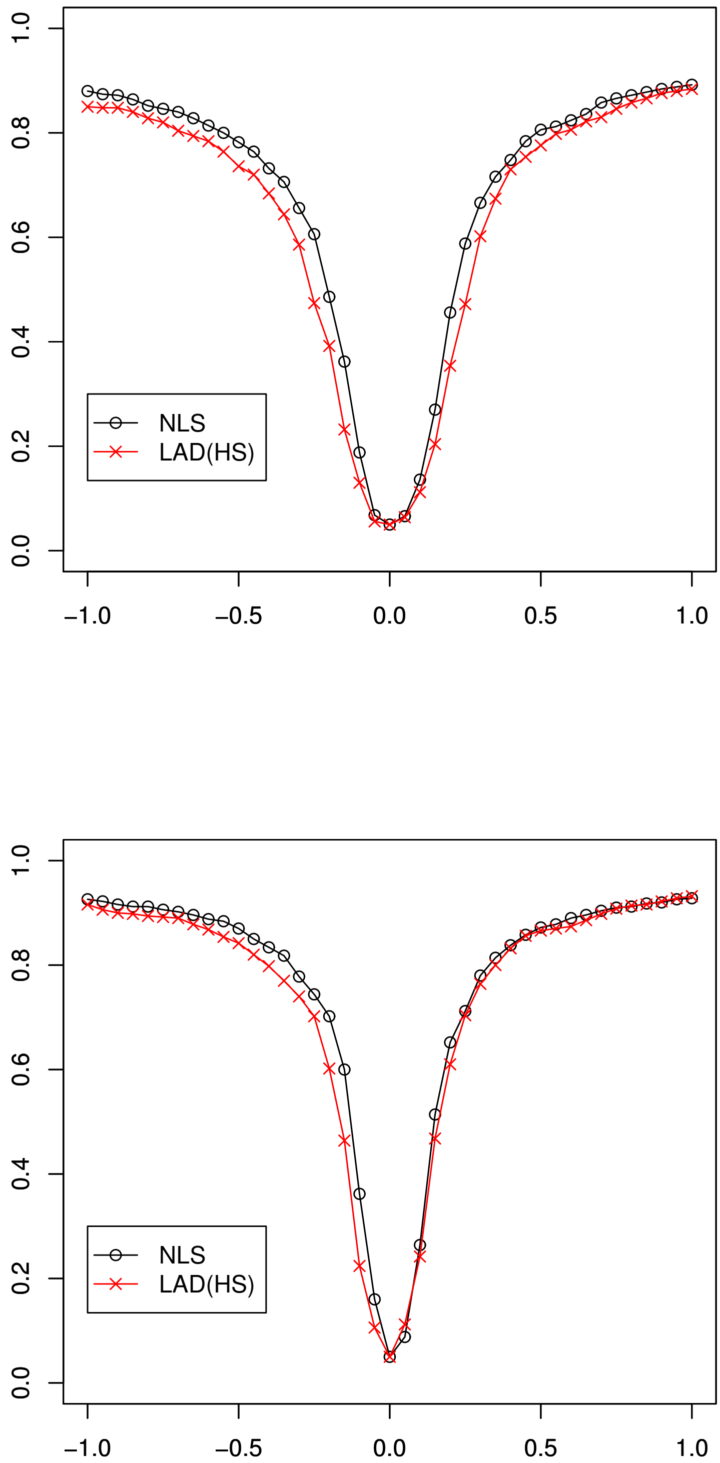
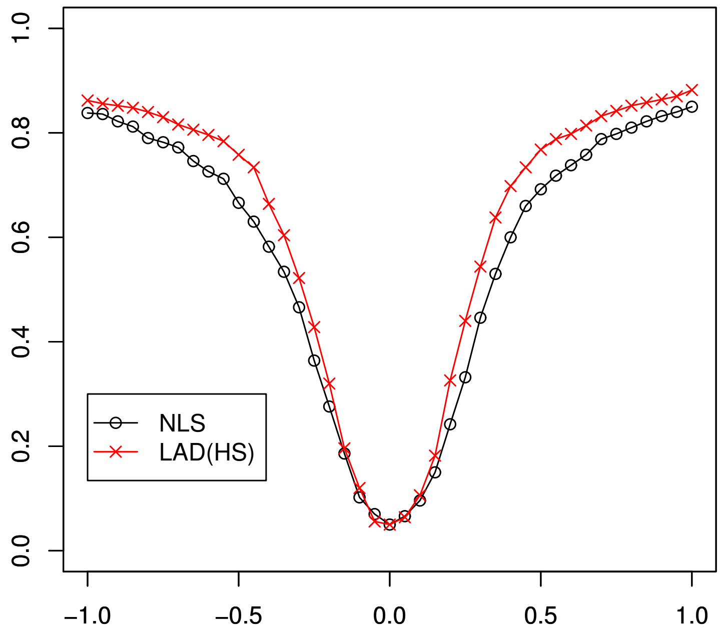


Figure 4.6: The powers of test when the model is (4.15) with $u_t \sim SN$ and $v_t \sim t_9$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.



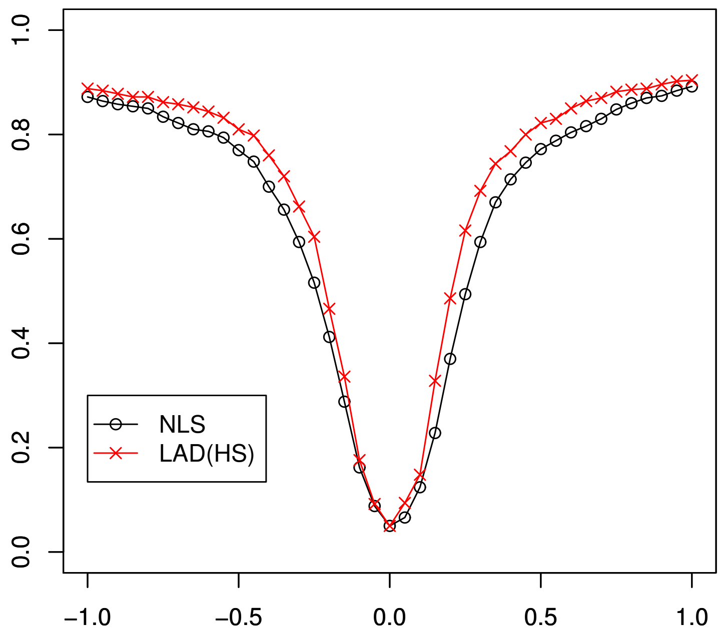
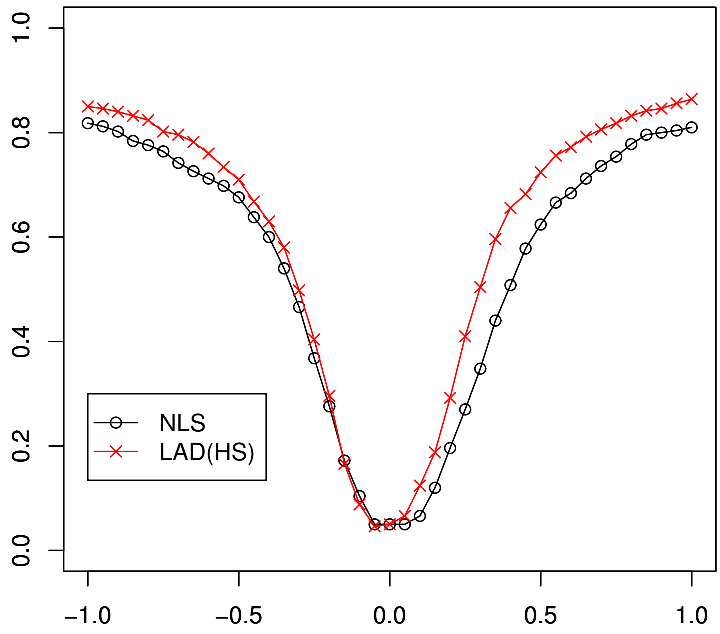


Figure 4.7: The powers of test when the model is (4.15) with $u_t \sim t_3$ and $v_t \sim SN$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.



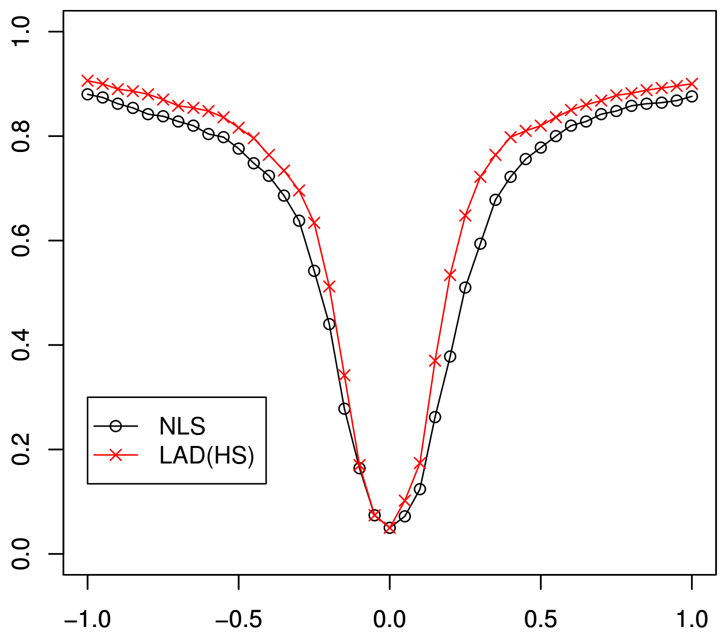


Figure 4.8: The powers of test when the model is (4.15) with $u_t \sim t_3$ and $v_t \sim t_9$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.

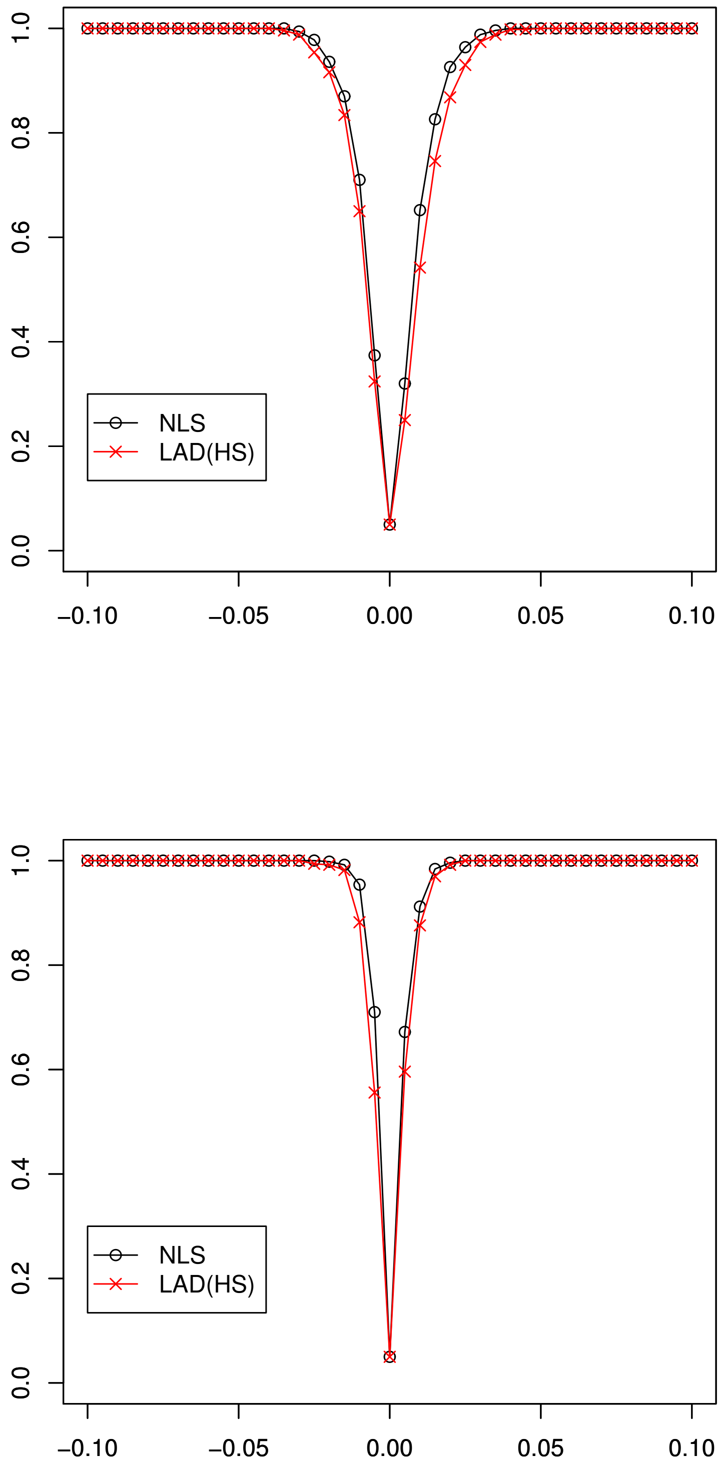
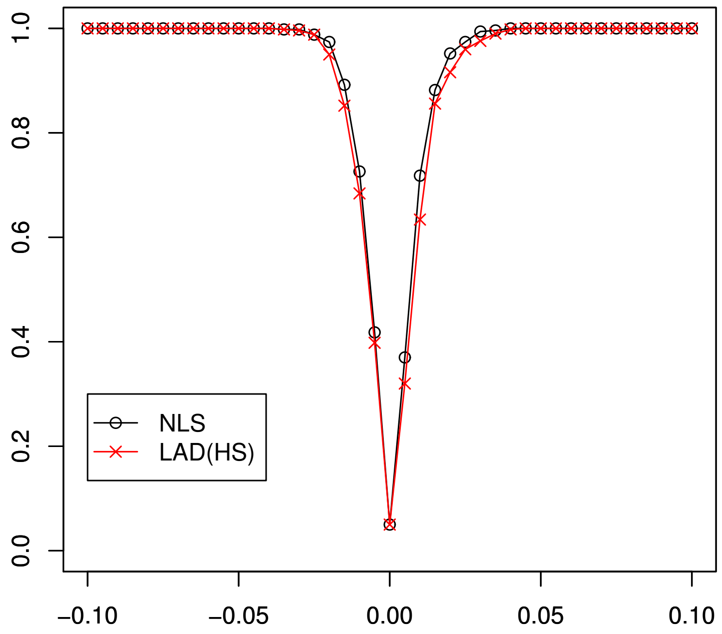


Figure 4.9: The powers of test when the model is (4.16) with $u_t \sim SN$ and $v_t \sim SN$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.



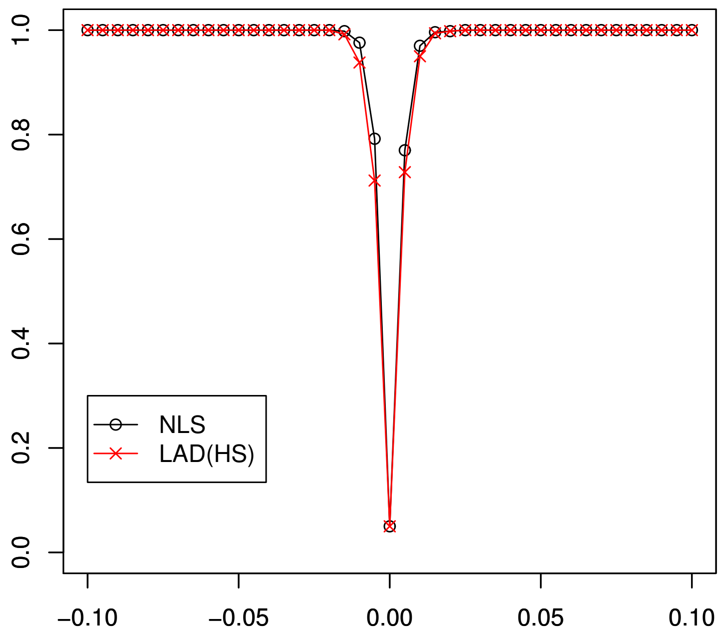
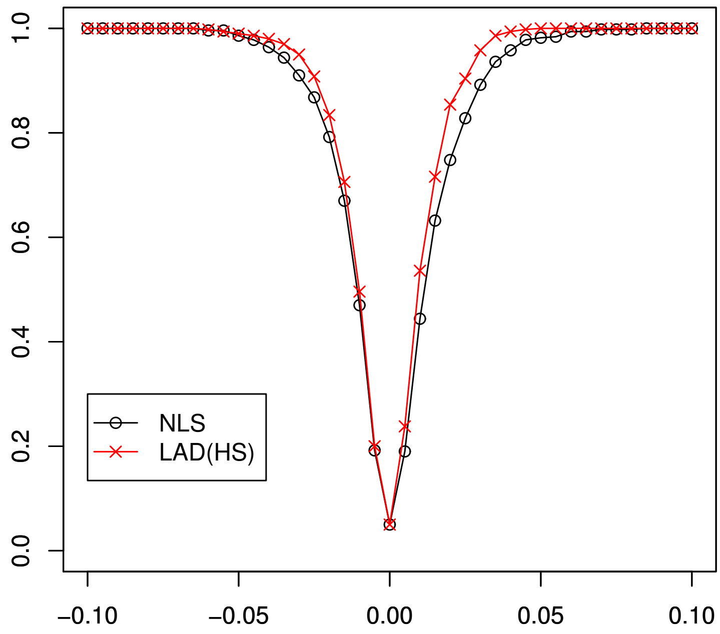


Figure 4.10: The powers of test when the model is (4.16) with $u_t \sim SN$ and $v_t \sim t_9$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.



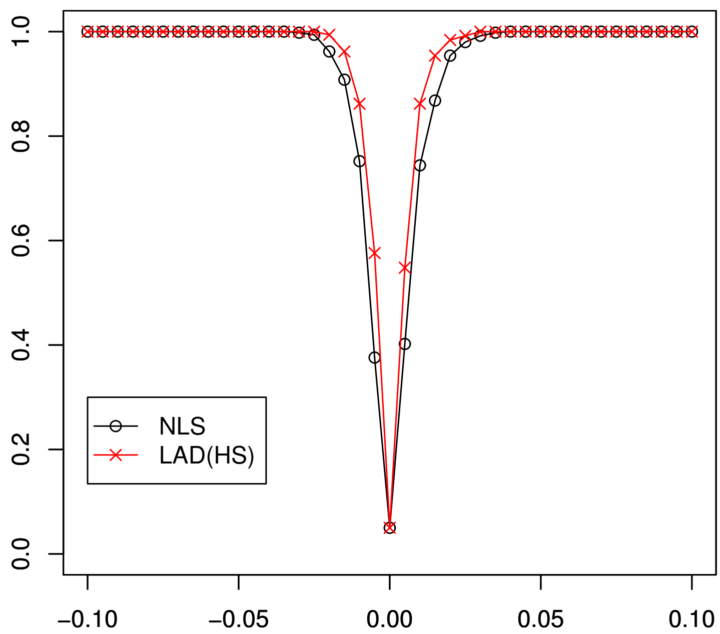


Figure 4.11: The powers of test when the model is (4.16) with $u_t \sim t_3$ and $v_t \sim SN$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.

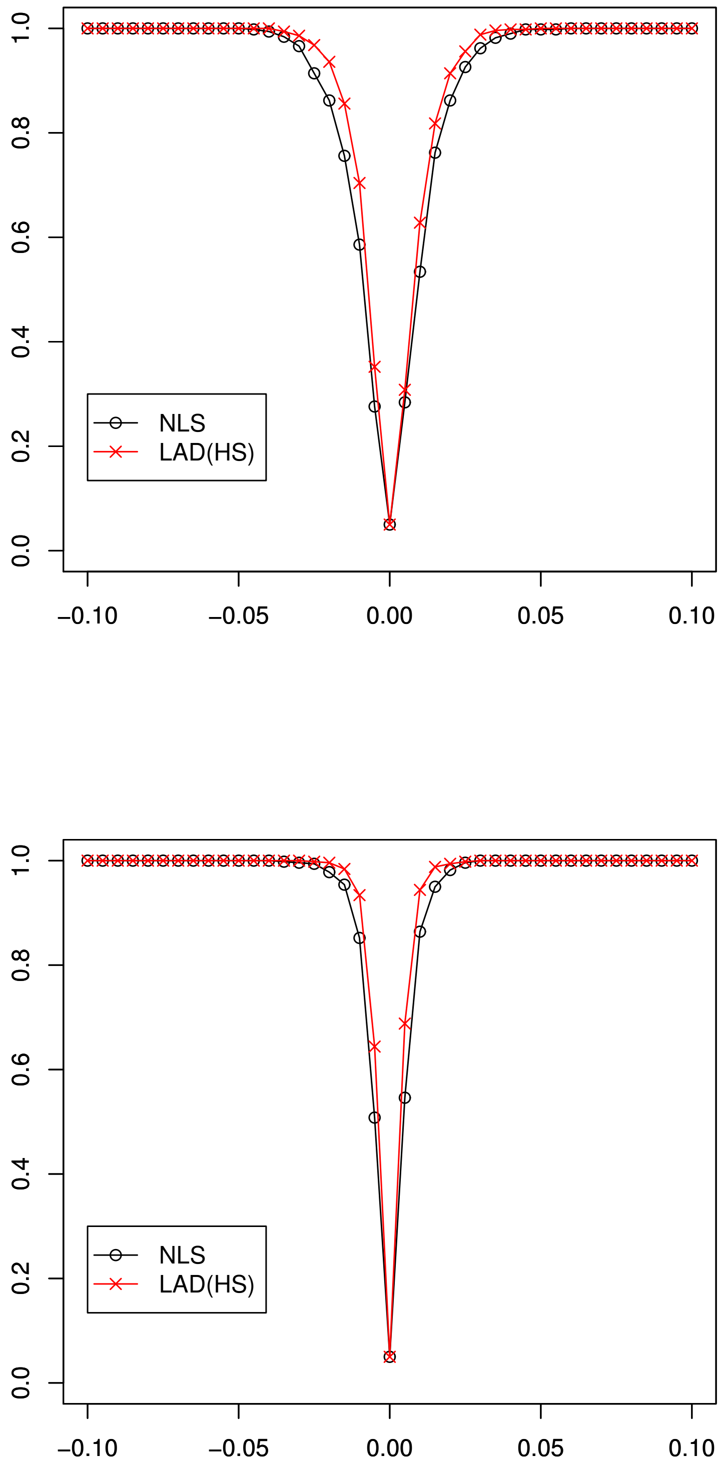


Figure 4.12: The powers of test when the model is (4.16) with $u_t \sim t_3$ and $v_t \sim t_9$. The sample sizes are 250 and 500 for the top and bottom pictures, respectively.

Bibliography

[1] Anderson, T. W. (1971), *The Statistical Analysis of Time Series*, Wiley.

[2] Andrews, D. W. K. and C. J. McDermott (1995), "Nonlinear econometric models with deterministically trending variables," *Review of Economic Studies*, **62**, 343–360.

[3] Barro, R. J. and X. Sala-i-Martin (2004), *Economic Growth*, 2nd ed. Prentice Hall.

[4] Bingham, N. H., C. M. Goldie and J. L. Teugels (1987), *Regular Variation*, Cambridge University Press.

[5] Bofinger, E. (1975), "Estimation of a density function using order statistics," *Australian Journal of Statistics*, **17**, 1–7.

[6] Brockwell, P. J. and R. A. Davis (1991), *Time Series: Theory and Methods*, 2nd ed. Springer.

[7] Canjels, E. and M. W. Watson (1997), "Estimating deterministic trends in the presence of serially correlated errors," *Review of Economics and Statistics*, **79**, 184–200.

[8] Chang, Y. and J. Y. Park (2003), "Index models with integrated time series," *Journal of Econometrics* **114**, 73–106.

[9] Chang, Y. and J. Y. Park (2011), "Endogeneity in nonlinear regressions with integrated time series," *Econometric Reviews*, **30**, 51–87.

[10] Chang, Y., J. Y. Park and P. C. B. Phillips (2001), "Nonlinear econometric models with cointegrated and deterministically trending regressors," *Econometrics Journal*, **4**, 1–36.

[11] Chernozhukov, V. and L. Umantsev (2001), "Conditional value-at-risk: Aspects of modeling and estimation," *Empirical Economics*, **26**, 271–292.

[12] Chen, X., R. Koenker and Z. Xiao (2009), "Copula-based nonlinear quantile autoregression," *Econometrics Journal*, **12**, 50–67.

[13] Chung, K. L. and R. J. Williams (1990), *An Introduction to Stochastic Integration*, 2nd. ed., Springer.

[14] Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press.

[15] de Jong, R. M. (2002), "Nonlinear estimators with integrated regressors but without exogeneity," *mimeo*.

[16] de Jong, R. M. and C.-H. Wang (2005), "Further results on the asymptotics for nonlinear transformations of integrated time series," *Econometric Theory*, **21**, 413–430.

[17] Dickey, D. A. and W. A. Fuller (1979), "Distribution of the estimators for autoregressive time series with a unit root," *Journal of the American Statistical Association*, **74**, 427–431.

[18] Doornik, J. A. (2007), *Object-Oriented Matrix Programming Using Ox*, 3rd ed. London: Timberlake Consultants Press and Oxford: www.doornik.com.

[19] Durlauf, S. N. and P. C. B. Phillips (1988), "Trends versus random walks in time series analysis," *Econometrica*, **56**, 1333–1354.

[20] Engle, R. F. and S. Manganelli (2004), "CAViaR: Conditional value at risk by regression quantiles," *Journal of Business & Economic Statistics*, **22**, 367–381.

[21] Fuller, W. (1996), *Introduction to Statistical Time Series*, 2nd ed. Wiley.

[22] Grenander, U. (1954), "On the estimation of regression coefficients in the case of an autocorrelated disturbance," *Annals of Mathematical Statistics*, **25**, 252–272.

[23] Grenander, U. and M. Rosenblatt (1957), *Statistical Analysis of Stationary Time Series,* Wiley.

[24] Guerre, E. and H. R. Moon (2006), "A study of a semiparametric binary choice model with integrated covariates," *Econometric Theory*, **22**, 721–742.

[25] Hall, P and S. Sheather (1988), "On the distribution of a studentized quantile," *Journal of Royal Statistical Society, Series B*, **50**, 381–391.

[26] Hamilton, J. D. (1994), *Time Series Analysis*, Princeton.

[27] Herce, M. A. (1996), "Asymptotic theory of LAD estimation in a unit root process with finite variance errors," *Econometric Theory*, **12**, 129–153.

[28] Honda, T. (2013), "Nonparametric LAD cointegrating regression," *Journal of Multivariate Analysis*, **117**, 150–162.

[29] Huber, P. J. (1967), "The behavior of maximum likelihood estimates under nonstandard conditions," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1: Statistics, 221–233.

[30] Hurvich, C. M., R. Deo and J. Brodsky (1998), "The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series," *Journal of Time Series Analysis*, **19**, 19–46.

[31] Jurečková, J. and B. Procházka (1994), "Regression quantiles and trimmed least squares estimator in nonlinear regression model," *Nonparametric Statistics*, **3**, 201–222.

[32] Knight, K. (1989), “Limit theory for autoregressive-parameter estimates in an infinite-variance random walk,” *Canadian Journal of Statistics*, **3**, 261–278.

[33] Knight, K. (1991), “Limit theory for M-estimates in an integrated infinite variance process,” *Econometric Theory*, **7**, 200–212.

[34] Knight, K. (1998), “Limiting distributions for L_1 regression estimators using general conditions,” *Annals of Statistics*, **26**, 755–770.

[35] Koenker, R. (2005), *Quantile Regression*, Cambridge University Press.

[36] Koenker, R. and G. Bassett (1978), “Regression quantiles,” *Econometrica*, **46**, 33–50.

[37] Koenker, R. and B. J. Park (1994), “An interior point algorithm for nonlinear quantile regression,” *Journal of Econometrics*, **71**, 265–283.

[38] Koenker, R. and Z. Xiao (2004), “Unit root quantile autoregression inference,” *Journal of the American Statistical Association*, **99**, 775–787.

[39] Krämer, W. (1986), “Least squares regression when the independent variable follows an ARIMA process,” *Journal of the American Statistical Association*, **81**, 150–154.

[40] Krämer, W. and U. Hassler (1998), “Limiting efficiency OLS vs. GLS when the regressors are fractionally integrated,” *Economics Letters*, **60**, 285–290.

[41] Moon, H. R. (2004), “Maximum score estimation of a nonstationary binary choice model,” *Journal of Econometrics*, **122**, 385–403.

[42] Mukherjee, K. (1999), “Asymptotics of quantiles and rank scores in nonlinear time series,” *Journal of Time Series Analysis*, **20**, 173–192.

[43] Mynbaev, K. T. (2009), “Central limit theorems for weighted sums of linear processes: L_p-approximability versus Brownian motion,” *Econometric Theory*, **25**, 748–763.

[44] Mynbaev, K. T. (2011a), *Short-Memory Linear Processes and Econometric Applications*, Wiley.

[45] Mynbaev, K. T. (2011b), "Regressions with asymptotically collinear regressors," *Econometrics Journal*, **14**, 304–320.

[46] Newey, W. and K. West (1987), "A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica*, **55**, 703–705.

[47] Oberhofer W. and H. Haupt (2006), "Asymptotic theory for nonlinear quantile regression under weak dependence," *mimeo*.

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

[49] Park, J. Y. and P. C. B. Phillips (2000), "Nonstationary binary choice," *Econometrica*, **68**, 1249–1280.

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

[51] Perron, P. and S. Ng (1996), "Useful modifications to some unit root tests with dependent errors and their local asymptotic properties," *Review of Economic Studies*, **63**, 435–63.

[52] Perron, P. and T. Yabu (2009), "Estimating deterministic trends with an integrated or stationary noise component," *Journal of Econometrics*, **151**, 56–69.

[53] Phillips, P. C. B. (1987), "Time series regression with a unit root," *Econometrica*, **55**, 277–301.

[54] Phillips, P. C. B. (1995), "Robust nonstationary regression," *Econometric Theory*, **11**, 912–951.

[55] Phillips, P. C. B. (1999), "Discrete Fourier transforms of fractional processes," Cowles Foundation discussion paper no. 1243, Yale University.

[56] Phillips, P. C. B. (2007), "Regression with slowly varying regressors and nonlinear trends," *Econometric Theory*, **23**, 557–614.

[57] Phillips, P. C. B. and B. E. Hansen (1990), "Statistical inference in instrumental variables regression with I(1) processes," *Review of Economic Studies*, **57**, 99–125.

[58] Phillips, P. C. B. and J. Y. Park (1988), "Asymptotic equivalence of ordinary least squares and generalized least squares in regressions with integrated regressors," *Journal of the American Statistical Association*, **83**, 111–115.

[59] Phillips, P. C. B. and P. Perron (1988), "Testing for a unit root in time series regression," *Biometrika*, **75**, 335–346.

[60] Phillips, P. C. B. and D. Sul (2007), "Transition modeling and econometric convergence tests," *Econometrica*, **75**, 1771–1855.

[61] Phillips, P. C. B. and D. Sul (2009), "Economic transition and growth," *Journal of Applied Econometrics*, **24**, 1153–1185.

[62] Phillips, P. C. B. and Y. Sun (2003), "Regression with an evaporating logarithmic trend," *Econometric Theory*, **19**, 692–701.

[63] Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer.

[64] Pollard, D. (1991), "Asymptotics for least absolute deviation regression estimators," *Econometric Theory*, **7**, 186–199.

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

[66] Powell, J. L. (1984), "Least absolute deviations estimation for the censored regression model," *Journal of Econometrics*, **25**, 303–325.

[67] Powell, J. L. (1986), "Censored regression quantiles," *Journal of Econometrics*, **32**, 143–155.

[68] Robinson, P. M. (1995), "Log-periodogram regression of time series with long range dependence." *Annals of Statistics*, **23**, 1048–1072.

[69] Rosenblatt, M. (1956), "On the estimation of regression coefficients of a vector-valued time series with a stationary residual," *Annals of Mathematical Statistics*, **27**, 99–121.

[70] Ruppert, D. and R. J. Carroll (1980) "Trimmed Least Squares Estimation in the Linear Model," *Journal of the American Statistical Association*, **75**, 828–838.

[71] Shaman, P. (1975), "An approximate inverse for the covariance matrix of moving average and autoregressive processes," *Annals of Statistics*, **3**, 532–538.

[72] Shin, D. W. and M. S. Oh (2002), "Asymptotic efficiency of the ordinary least squares estimator for regressions with unstable regressors," *Econometric Theory*, **18**, 1121–1138.

[73] Tanaka, K. (1996), *Time Series Analysis: Nonstationary and Noninvertible Distribution Theory*, Wiley.

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

[75] van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

[76] Vogelsang, T. (1998), "Trend function hypothesis testing in the presence of serial correlation," *Econometrica*, **66**, 123–148.

[77] Wang J. (1995), "Asymptotic normality of L_1-estimators in nonlinear regression," *Journal of Multivariate Analysis*, **54**, 227–238.

[78] Weiss A. A. (1991), "Estimating nonlinear dynamic models using least absolute error estimation," *Econometric Theory*, **7**, 46–68.

[79] Wooldridge, J. M. (1994), "Estimation and inference for dependent processes," Ch. 10 in R. F. Engle and D. McFadden (eds.), *Handbook of Econometrics, Vol. 4*, North-Holland, 2639–2738.

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

[81] Xiao Z. (2009), "Quantile cointegrating regression," *Journal of Econometrics*, **150**, 248–260.

[82] Zygmund, A. (2003), *Trigonometric Series*, Cambridge.