

CONSISTENT NONPARAMETRIC TEST ON NONLINEAR REGRESSION  
MODELS WITH NEAR-INTEGRATED COVARIATES

by

Li Wu

A dissertation submitted to the faculty of  
The University of North Carolina at Charlotte  
in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in  
Applied Mathematics

Charlotte

2013

Approved by:

---

Dr. Jiancheng Jiang

---

Dr. Zongwu Cai

---

Dr. Weihua Zhou

---

Dr. Claudia Flowers

©2013  
Li Wu  
ALL RIGHTS RESERVED

## ABSTRACT

LI WU. Consistent Nonparametric Test on Nonlinear Regression Models With Near-Integrated Covariates . (Under the direction of DR. ZONGWU CAI)

In this paper, a  $L^2$  type nonparametric test is developed to test a specific nonlinear parametric regression model with near-integrated regressors. The asymptotic distributions of the proposed test statistic under both null and alternative hypotheses are established. The finite sample performance is also examined by conducting Monte Carlo simulation. The test statistic is applied to testing the linear prediction model of asset return and the predictability of asset return is shown at last.

## ACKNOWLEDGMENTS

I'm very grateful to my advisor Dr. Zongwu Cai in the first place. He introduced me into the field of statistics. His passion in research and teaching conveyed to me and inspired me to explore further in statistics. His strict and high requirements urged my progress and encouraged me to reach my full potential in research. During the process of writing this dissertation, Dr. Cai contributed lots of time and effort in reviewing and correcting my thesis. Without his persistent and patient guidance this dissertation would not be possible.

I'm also indebted to Dr. Jiancheng Jiang and Dr. Yanqing Sun. Not only that the effective training I received in their classes sharpened my professional skills, their interperation of statistic approaches and theorem made those rigid, complicated formula and methods interesting, readily understood and inspiring to me. They also spared no effort in answering my questions after class. Their valuable suggestions and generous help lit my way up in research.

In addition, I'm very thankful to my committee members, Dr. Weihua Zhou and Dr. Flowers. I truely appreciate their precious time and genuine advice, which significantly improved my dissertation and benefited my future study.

I would also like to place on record my sincere gratitude to graduate student coordinators in our department, Dr. Shaozhong Deng and Dr. Joe Avrin. They always put their feet in students' shoes and did their best to help. I can't expect more from a coordinator.

At last, I would like to express my gratitude to my family. Their love, support and

faith in me is the main momentum that got me through all difficulties and frustrations on my way pursuing a PhD.

## TABLE OF CONTENTS

|  |      |
|--|------|
| LIST OF FIGURES  | viii |
| LIST OF TABLES   | ix   |
| CHAPTER 1: INTRODUCTION  | 1    |
| 1.1 Nonlinear Cointegration Model  | 2    |
| 1.2 Estimation of Nonlinear Cointegration Model                          | 2    |
| 1.3 Cointegration Tests  | 6    |
| 1.4 Overview   | 8    |
| CHAPTER 2: Nonlinear Least Square Estimation                             | 9    |
| 2.1 The Model and Preliminary Results                                    | 9    |
| 2.2 Consistency  | 12   |
| 2.3 Asymptotics for Nonlinear Regression with Near-Intergrated Processes | 13   |
| CHAPTER 3: Test Statistic  | 16   |
| 3.1 Construction of Test Statistics                                      | 16   |
| 3.2 Assumptions and Asymptotic Results                                   | 18   |
| CHAPTER 4: Monte Carlo Simulations                                       | 22   |
| CHAPTER 5: Empirical Study   | 26   |
| 5.1 Review of Tests of Predictability of Asset return                    | 26   |
| 5.2 Description of Data and Model  | 28   |
| 5.3 Nonparametric Test of Predictability                                 | 29   |
| CHAPTER 6: Conclusion  | 32   |

|   |    |
|---|----|
| APPENDIX A: Technical Results for Chapter 2 | 34 |
| APPENDIX B: Technical Results for Chapter 3 | 37 |
| REFERENCES                                  | 48 |

## LIST OF FIGURES

|  |    |
|--|----|
| FIGURE 1: Time Series Plot for Log Dividend-Price Ratio. | 29 |
| FIGURE 2: Time Series Plot for Log Earning-Price Ratio.  | 30 |



## LIST OF TABLES

|   |    |
|---|----|
| TABLE 1: Estimated Sizes: Varying Smoothing Parameters    | 23 |
| TABLE 2: Estimated Powers: Varying Smoothing Parameters   | 23 |
| TABLE 3: Estimated Sizes : Varying Integration Parameters | 24 |
| TABLE 4: Estimated Powers: Varying Integration Parameters | 24 |
| TABLE 5: Estimated Powers: Varying models                 | 24 |
| TABLE 6: Estimated Powers: Varying $a_1$                  | 24 |
| TABLE 7: Estimated Autoregression Parameter               | 29 |
| TABLE 8: Test Statistics and Confidence Intervals         | 31 |

## CHAPTER 1: INTRODUCTION

Nonlinear cointegration models are important in a wide range of applications in economics (e.g. [Granger, 1995]). In this paper, a test statistic is introduced to test model specification of a nonlinear parametric model with near-integrated regressors.

Nonlinear Least Square method is applied for parameter estimation for the specific parametric model. The asymptotic theorem of NLS estimates with unit root process was introduced in [Park and Phillips, 2008] and [Kasparis, 2010]. The extension of the existing limit theorem to near-integrated process is straightforward. The major works in [Wang and Phillips, 2009b, Wang and Phillips, 2009a] of limit theorem of sample covariances of nonstationary time series and integrable functions of such time series that involve a bandwidth sequence are referred to in deriving asymptotic distributions of the proposed test statistic.

The construction of our test statistic closely relates to the work in [Sun et al., 2011], in which a test of time-varying coefficients is proposed with null hypothesis of constant coefficients. This paper goes further than the above one in two aspects. First, the null hypothesis is a specific nonlinear parametric model involving a constant as a special case. Next, the functional parameter is assumed to be a nonlinear transformation of near-integrated processes instead of stationary processes, deriving of asymptotic theory of which is much more challenging.

## 1.1 Nonlinear Cointegration Model

The belief that many economic and financial time series are highly persistent and nonlinearly related is widely held. Nonlinear dynamic relationships that has been discussed by economic theorists include, for instance, the correlation between cost and production functions , hysteresis and boundary effect, exchange rate and fundamentals, and inflation and economic growth. Working on modeling the relationships among highly persistent time series, two major questions are faced by econometricians and statisticians: how to specify nonlinear models and how to test the goodness of fit of a specified nonlinear model. This paper will focus on the latter one.

The nonlinear cointegration considered in this paper is modeled as:

$$y_t = f(z_t) + u_t \quad (1)$$

where  $z_t$  (a scalar) is an integrated series  $I(1)$  or nearly integrated series  $NI(1)$ ,  $u_t$  a stationary process, and  $f(\cdot)$  an unknown functional. The null hypothesis of interest in this paper is a specified parametric nonlinear functional:

$$H_0 : \quad \Pr(f(z_t) = g(z_t, \theta)) = 1 \text{ for some } \theta \in \Theta, \quad (2)$$

where  $\Theta$  is the parameter set. The contiguous alternatives are written as follows,

$$H_{1n} : \quad f(z_t) = g(z_t, \theta) + n^{-\gamma}G(z_t) \quad (3)$$

where  $\gamma < \frac{1}{10}$ . That is to test if the function  $f(\cdot)$  in (1) is of the parametric form  $g(z, \theta)$ .

## 1.2 Estimation of Nonlinear Cointegration Model

The nonlinear cointegration model is estimated using parametric and non parametrical technique respectively under the null and alternative hypothesis.

### 1.2.1 Nonlinear Least Square Estimator

The asymptotic theory of linear regression in the context of stationary or weakly dependent processes has been originally developed by [Hansen, 1992], in which strong laws of large number and central limit theory are applied straightly to stationary and ergodic measurable functions. Then, a mechanism for doing asymptotic analysis for linear systems of integrated time series was introduced by [Phillips, 1986], [Phillips, 1987], and [Phillips and Durlauf, 1986]. They applied weak convergence in function spaces, continuous mapping theorem, and weak convergence of martingales in deriving asymptotic distributions.

The development of limit distribution theory for a nonlinear model with high persistent time series has been hamstrung for a long time until the work of [Park and Phillips, 2008], where a new machinery was introduced to analyze the asymptotic behavior of sample moments of nonlinear functions of nonstationary data. The key notion of the new method is to transport the sample function into a spatial function, which is also the basis of later works of Phillips regarding nonparametric regression of nonstationary time series. In particular, they dealt with sample sum by replacing it with a spatial sum and then treating it as a location problem. Our analysis in this paper employs this technique, too.

The following nonlinear regression model for  $y_t$  was considered in [Park and Phillips, 2008],

$$y_t = f(z_t, \theta_0) + u_t$$

$$z_t = z_{t-1} + v_t$$

where  $f : R \times R^m \rightarrow R$  is known, regressor  $z_t$  an integrated process, regression error  $u_t$  a martingale difference sequence, and  $\theta_0$  an  $m$ -dimensional true parameter vector.

They estimated  $\theta_0$  by nonlinear least squares (NLS). That is to choose  $\hat{\theta}_n$  by minimizing the function below,

$$Q_n(\theta) = \sum_{t=1}^n (y_t - f(z_t, \theta))^2$$

Thus, the NLS estimator  $\hat{\theta}_n$  was defined by

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

Under some regularity conditions and assumptions on function  $f$ , they showed the consistency and limit distribution of NLS estimator,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \left( L(1, 0) \int_{-\infty}^{\infty} \dot{f}(s, \theta_0) \dot{f}(s, \theta_0)' ds \right)^{-1/2} W(1)$$

where  $L(1, 0)$  is the local time of the limit data generating process  $v_t$  and  $W(1)$  is a Brownian motion independent of  $L$ .

A similar limit theory of NLS estimator with near integrated (NI(1)) regressors is given in this paper. The only difference of the limit distribution between I(1) and NI(1) time series lies in the local time function. The local time for integrated regressor is the local time of a limit Brownian motion. As in near integrated situation, it's the local time of an O-U process.

### 1.2.2 Nonparametric Cointegration Estimator

In nonparametric estimation, joint dependence between the regressor and the dependent variable is the main complication leading to bias in conventional kernel estimates. It is shown in [Wang and Phillips, 2009a, Wang and Phillips, 2009b] that in functional cointegrating regressions with integrated or near integrated regressors, simple nonparametric estimation of a structural nonparametric cointegrating regression is consistent and the limit distribution is mixed normal.

The nonlinear structural model of cointegration is

$$y_t = f(z_t) + u_t,$$

where  $u_t$  is a zero mean stationary error,  $z_t$  an integrated or near integrated regressor, and  $f$  the unknown function to estimate. Then, the Nadaraya-Watson kernel estimator of  $y_t$  is given by

$$\hat{f}(z) = \frac{\sum_{t=1}^n y_t K_h(z_t - z)}{\sum_{t=1}^n K_h(z_t - z)},$$

where  $K_h(s) = (1/h)K(s/h)$  is a nonnegative kernel function, and  $h$  the bandwidth function, such that  $h \rightarrow 0$  as  $n \rightarrow \infty$ .

Imposing some assumptions, it's proved in [Wang and Phillips, 2009b] that the limit behavior of  $\hat{f}(x)$  is

$$\hat{f}(z) \xrightarrow{p} f(z)$$

when  $nh^2 \rightarrow \infty$  and  $h \rightarrow 0$ . In addition, if  $h$  satisfies that  $nh^2 \rightarrow \infty$  and  $nh^{2(1+2\gamma)} \rightarrow 0$  as  $n \rightarrow \infty$ , the limit distribution of the Nadaraya-Watson kernel estimator is shown

as

$$\left( h \sum_{t=1}^n K_h(z_t - z) \right)^{1/2} \left( \hat{f}(z) - f(z) \right) \xrightarrow{d} N(0, \sigma^2)$$

where  $0 < \gamma \leq 1$ , for sufficiently small  $h$ ,  $|f(hy + z) - f(z)| \leq h^\gamma f_1(y, z)$  for any  $y \in R$  and  $\int_{-\infty}^{\infty} K(s) f_1(s, z) ds < \infty$ , and  $\sigma^2 = E(u_{m_0}^2) \int_{-\infty}^{\infty} K^2(s) ds \int_{-\infty}^{\infty} K(z) dz$ . Notice that they defined  $u_t = 0$  for  $1 \leq t \leq m_0 - 1$ .

It is also proven in [Wang and Phillips, 2009b] that the localized version of sum of squared residuals is a consistent estimate of the error variance  $Eu_{m_0}^2$  with stricter assumptions imposed,

$$\hat{\sigma}_n^2 \xrightarrow{p} Eu_{m_0}^2.$$

for any  $h$  satisfying  $nh^2 \rightarrow \infty$  and  $h \rightarrow 0$  as  $n \rightarrow \infty$ , where

$$\hat{\sigma}_n^2 = \frac{\sum_{t=1}^n [y_t - \hat{f}(z)]^2 K_h(z_t - z)}{\sum_{t=1}^n K_h(z_t - z)}.$$

### 1.3 Cointegration Tests

Tests for a linear cointegrating model has been developed since [Hansen, 1992], that tested parameter stability. Recently, a modified RESET test was introduced by [Hong and Phillips, 2010] to test the existence of linear cointegration. In empirical studies, the RESET test statistic was applied to check the traditional linear cointegration specification in purchasing power parity (PPP) model. A linearity test of cointegrating smooth transition regressions is proposed in [Gao et al., 2009]. They tested the null hypothesis of a linear cointegration model:  $y = \beta_0 + x_t \beta_1 + u_t$  against the alternative hypothesis of a nonlinear cointegration regression system:  $y_t = g(x_t) + u_t$ , where regressor  $x_t$  is a unit root process independent of error  $u_t$ . Based on the work of

[Gao et al., 2009], a similar problem was investigated in [Wang and Phillips, 2011]. They allow regressor  $x_t$  to be more general and not necessarily independent of  $u_t$ . The problem of testing a linear cointegration model against a nonlinear cointegration model was considered by [Choi and Saikkonen, 2004]. The smooth transition regression model developed in [Choi and Saikkonen, 2004] is:  $y_t = x_t' \alpha + \beta' x_t g(x_{ts} - c) + u_t$ , where  $x_t$  is a  $p$  dimensional random walk vector, and  $x_{ts}$  denotes the  $s^{th}$  component of  $x_t$ . The model reduces to a linear cointegration model under the null hypothesis of  $\beta = 0$ .

A semiparametric varying coefficient model was studied in [Sun et al., 2011]. That model was first learned by [Cai et al., 2009] and [Xiao, 2009]:

$$y_t = X_t' \theta(z_t) + u_t, \quad (4)$$

where  $X_t$  is a  $d$ -dimensional non stationary regressor,  $z_t$  and  $u_t$  stationary variables, and  $\theta(\cdot)$  a  $d \times 1$  vector of unknown smooth functions. They tested the parameter constancy

$$H_0 : Pr(\theta(z_t) = \theta_0) = 1, \text{ for some } \theta_0 \in B,$$

against

$$H_1 : Pr(\theta(z_t) \neq \theta) > 0, \text{ for any } \theta \in B.$$

The model studied in this paper differs from all the above ones in that we test a nonlinear cointegration model instead of a linear one. Compared with the varying coefficient model investigated by [Sun et al., 2011], our model could be taken as a varying coefficient model with one dimensional  $X_t = 1$ , and nonstationary  $z_t$ . The combination of nonlinearity and cointegration makes the analysis of limit theory very complicated.



## 1.4 Overview

The rest of the paper is organized as follows. Chapter 2 develops the asymptotic theory of least square estimate of nonlinear regression with near-integrated process. Chapter 3 describes our test statistic and shows asymptotic results of the test statistics under null and alternative hypothesis respectively. In Chapter 4, Monte Carlo simulations are performed to examine the finite sample performance of the proposed tests. We test the predictability of asset return from a linear model using our test statistics in Chapter 5. Chapter 6 concludes the paper. All the mathematical proofs are relegated to Appendices.

The notation is conventional throughout the paper. We offer a summary of notation here for convenience sake. (i)  $\xrightarrow{d}$  stands for convergence in distribution,  $\xrightarrow{p}$  for convergence in probability, and “ $\Rightarrow$ ” for weak convergence with respect to the Skorohod metric, as defined in [Billingsley, 2009]. (ii)  $O_e(a_n)$  denotes a probability order of  $a_n$ , where  $a_n$  is a non-stochastic positive sequence; i.e.  $O_e(a_n) = O_p(a_n)$ . (iii) We define  $L^r$ -norm of a matrix  $X$  by  $\|X\|_r = \left(\sum_{ij} E|X_{ij}|^r\right)^{1/r}$ , where  $X_{ij}$  is the  $(i, j)$ th element of  $X$ . (iv)  $A \stackrel{def}{=} B$  is used to define  $A$  by a previously defined quantity  $B$ , and  $A \equiv B$  is used to assign a new notation  $B$  to  $A$ . (V)  $[a]$  denotes the smallest integer that is greater than  $a$  for  $a > 0$ . (vi) we use  $\mathcal{F}_{nt} = \sigma\{z_i, u_i : 1 \leq i \leq t \leq n\}$  to denote the smallest  $\sigma$ -field containing past history of  $\{z_t, u_t\}$  for all  $n$ .

## CHAPTER 2: NONLINEAR LEAST SQUARE ESTIMATION

### 2.1 The Model and Preliminary Results

We consider the nonlinear regression model for  $y_t$  under  $H_0$

$$y_t = g(z_t, \theta_0) + u_t \quad (5)$$

where  $g : R \times R \rightarrow R$  is known and  $\theta_0$  is the true parameter that lies in the parameter set  $\Theta$ . This section concentrates on nonlinear least square estimation of (5). Let

$$Q_n(\theta) = \sum_{t=0}^n (y_t - g(z_t, \theta))^2, \quad (6)$$

then, the NLS estimator  $\hat{\theta}_n$  is as follows,

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} Q_n(\theta). \quad (7)$$

It is assumed throughout the paper that  $\hat{\theta}_n$  exists and is unique for all  $n$ , and  $\theta_0$  is an interior point of  $\Theta$ , where  $\Theta$  is assumed to be compact and convex. This is standard for NLS regression.  $\hat{\sigma}_n = (1/n) \sum_{t=1}^n \hat{u}_t^2$  is an error variance estimate, where  $\hat{u}_t = y_t - g(z_t, \hat{\theta}_n)$ .

We start by writing  $z_t$  as

$$z_t = \rho z_{t-1} + \eta_t, \quad (8)$$

and initializing it with  $z_0 = 0$  to avoid unnecessary complication in our development of limit theory as in [Park and Phillips, 2008]. Then, define the stochastic processes  $U_n$  and  $V_n$  respectively by

$$U_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} u_t \quad \text{and} \quad V_n(r) = \frac{1}{\sqrt{n}} \sum_{t=1}^{[nr]} \eta_t$$

where  $[s]$  denotes the largest integer less than  $s$ .

**Assumption 2.1:** (a)  $(U_n, V_n) \xrightarrow{d} (U, V_c)$ , where  $U$  is a Brownian motion and  $V_c$  is an O-U process driven by a standard Brownian Motion over  $[0, 1]$  with variance  $\sigma_\eta = \lim_{n \rightarrow \infty} \text{Var}(n^{-1/2} \sum_{t=1}^n \eta_t)$ . (b)  $(u_t, \mathcal{F}_{nt})$  is a martingale difference sequence with  $E(u_t^2 | \mathcal{F}_{n,t-1}) = \sigma^2$  a.s. for all  $t$  and  $\sup_{1 \leq t \leq n} E(|u_t|^q | \mathcal{F}_{n,t-1}) < \infty$  a.s. for some  $q > 2$ .

Assumption 2.1 is routinely imposed on NLS regression with nonstationary processes as in [Park and Phillips, 2008]. Assumption (a) is well known to be satisfied for a wide variety of data generating processes like mildly heterogeneous time series and stationary processes. Condition (b) is essential to the limit distribution theory. But if it's relaxed to allow serial correlation in errors and cross correlation between regressors and errors, the consistency of the least squared estimator still holds.

From Skorohod representation theorem, there exists a common probability space  $(\Omega, \mathcal{F}, \mathcal{P})$  supporting both  $(U, V_c)$  and  $(U_n^0, V_n^0)$  such that

$$(U_n^0, V_n^0) =_d (U_n, V_n) \text{ and } (U_n^0, V_n^0) \rightarrow (U, V_c) \text{ a.s.} \quad (9)$$

Then, there's no loss in generality by assuming  $(U_n, V_n) = (U_n^0, V_n^0)$  throughout this paper.

More restrictive conditions on process  $z_t$  required to develop the asymptotic theory for nonlinear regression are introduced in the following.

**Assumption 2.2** Let  $\eta_t = \varphi(L)\varepsilon_t = \sum_{k=0}^{\infty} \varphi_k \varepsilon_{t-k}$  with  $\varphi(1) \neq 0$ . Assume that  $\sum_{k=0}^{\infty} k|\varphi_k| < \infty$ ,  $\{\varepsilon_t\}$  is i.i.d with  $E|\varepsilon_t|^p < \infty$  for some  $p > 4$ , and the characteristic function  $c(\lambda)$  of  $\{\varepsilon_t\}$  satisfying  $\lim_{\lambda \rightarrow \infty} \lambda^r c(\lambda) = 0$  for some  $r > 0$ .

Assumption 2.2 is satisfied by all invertible Gaussian ARMA models and implies that  $V_n^0 \xrightarrow{d} V_c$ .

In the subsequent development of the asymptotic theory for nonlinear regression of near-integrated time series, the local time of the O-U process is used repeatedly. So, let's recall the definition of local time. The process  $\{L_M(t, s), t \geq 0, s \in \mathcal{R}\}$  is called the local time of a measurable process  $\{M(t), t \geq 0\}$  if,

$$\int_0^t T[M(s)]ds = \int_{-\infty}^{\infty} T(s)L_M(t, s)ds, \text{ all } t \in \mathcal{R} \quad (10)$$

for any locally integrable function  $T(x)$ . Intuitively,  $L_M(t, s)$  is a spatial density recording the sojourn time of process  $\{L_M(t, s), t \geq 0\}$  at the spatial point  $s$  over the time interval  $[0, t]$ . More discussions and applications of local time are provided by [Geman and Horowitz, 1980], [Revuz and Yor, 2004], [Park and Phillips, 2008] and [Phillips, 2009].

Next, some regularity conditions for nonlinear transformation are required to develop the asymptotics. Here, our focus is only on *I-regular* functions as defined in [Park and Phillips, 2008].

**Definition 2.1** A function  $F$  is said to be *I-regular* on a compact set  $\Pi$  if

(a) for each  $\pi_0 \in \Pi$ , there exists a neighborhood  $N_0$  of  $\pi_0$  and a bounded integrable function  $T : \mathcal{R} \rightarrow \mathcal{R}$  such that for all  $\pi \in N_0$ ,  $\|F(x, \pi) - F(x, \pi_0)\| \leq \|\pi - \pi_0\|T(x)$ , and

(b) for some constant  $c > 0$  and  $k > 6/(p-2)$  with  $p > 4$  given in Assumption 2.2,  $\|F(x, \pi) - F(y, \pi)\| \leq c|x - y|^k$  for all  $\pi \in \Pi$ , on each  $S_i$  of their common support

$$S = \bigcup_{i=1}^m S_i \subset \mathcal{R}.$$

Condition (a) requires  $F(x, \cdot)$  be continuous on  $\Pi$  for all  $x \in \mathcal{R}$  as in standard nonlinear regression theory. Condition (b) requires that all functions in the family are sufficiently smooth piecewise on their common support independent of  $\pi$ .

Theorem 2.1.1. Suppose Assumption 2.2 holds. If  $F$  is I-regular on a compact set  $\Pi$ , then

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n F(z_t, \pi) \xrightarrow{p} \left( \int_{-\infty}^{\infty} F(s, \pi) ds \right) L_{V_c}(1, 0)$$

uniformly in  $\pi \in \Pi$ , as  $n \rightarrow \infty$ . Moreover,

$$\frac{1}{\sqrt[4]{n}} \sum_{t=1}^n F(z_t, \pi) u_t \xrightarrow{d} \left( L_{V_c}(1, 0) \int_{-\infty}^{\infty} F(s, \pi) F(s, \pi) ds \right)^{1/2} W(1)$$

as  $n \rightarrow \infty$ .

The sample mean and sample covariance asymptotics are exactly like those in [Park and Phillips, 2008]. But  $L$  here is the local time of the limit O-U process  $V_c$  due to the near-integrated data generating process.

## 2.2 Consistency

To prove the consistency of the NLS estimator  $\hat{\theta}_n$  defined in (6), a sufficient consistency condition is given following [Park and Phillips, 2008]. Define  $D_n(\theta, \theta_0) = Q_n(\theta) - Q_n(\theta_0)$ . Then, the condition is written as follows.

CN1: For some normalizing sequence  $\nu_n$ ,  $\nu_n^{-1} D_n(\theta, \theta_0) \xrightarrow{p} D(\theta, \theta_0)$  uniformly in  $\theta$ , where  $D(\cdot, \theta_0)$  is continuous and has unique minimum  $\theta_0$  a.s.

The above condition is sufficient to guarantee that  $\hat{\theta}_n \xrightarrow{p} \theta_0$ , referring to the work by [Jennrich, 1969].

Theorem 2.2.1. Under Assumption 2.2, CN1 holds if for all  $\theta \neq \theta_0$ ,  $\int_{-\infty}^{\infty} (g(s, \theta) - g(s, \theta_0))^2 ds > 0$ , with  $\theta_0$  being I-regular on  $\Pi$ . Then, we have

$$D(\theta, \theta_0) = \left( \int_{-\infty}^{\infty} (g(s, \theta) - g(s, \theta_0))^2 ds \right) L_{V_c}(1, 0)$$

with  $\nu_n = \sqrt{n}$ .

All bounded integrable functions that are piecewise smooth satisfy the conditions in Theorem 2.1.

**Corollary 2.1:** Let the assumptions in Theorem 2.1 hold. Then  $\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$ , as  $n \rightarrow \infty$ .

This corollary shows the consistency of the error variance estimator  $\hat{\sigma}_n^2$ , which follows from Theorem 3.2 in [Park and Phillips, 2008]

### 2.3 Asymptotics for Nonlinear Regression with Near-Intergrated Processes

In this section, we derive the asymptotic distribution of the NLS estimator  $\hat{\theta}_n$  defined in (6) under stronger assumptions on differentiability of the regression function.

Let's start by the following definitions,

$$\dot{g} = \left( \frac{\partial g}{\partial \theta_i} \right), \ddot{g} = \left( \frac{\partial^2 g}{\partial \theta_i^2} \right), \ddot{\ddot{g}} = \left( \frac{\partial^3 g}{\partial \theta_i^3} \right)$$

to be the first, second and third derivatives of  $g$  with respect to  $\theta$ , and let  $\dot{Q}_n$  and  $\ddot{Q}_n$

be the first and second derivatives of  $Q_n$  with respect to  $\theta$ . Therefore,

$$\begin{aligned} \dot{Q}_n(\theta) &= \frac{\partial Q_n}{\partial \theta} = - \sum_{t=1}^n \dot{g}(x_t, \theta)(y_t - g(x_t, \theta)), \\ \ddot{Q}_n(\theta) &= \frac{\partial^2 Q_n}{\partial \theta^2} = \sum_{t=1}^n \dot{g}(x_t, \theta)^2 - \sum_{t=1}^n \ddot{\ddot{g}}(x_t, \theta)(y_t - g(x_t, \theta)), \end{aligned}$$

by ignoring a constant. The asymptotic distribution of  $\hat{\theta}_n$  is naturally established from the first order Taylor expansion of  $\dot{Q}_n$ ,

$$\dot{Q}_n(\hat{\theta}_n) = \dot{Q}_n(\theta_0) + \ddot{Q}_n(\theta_n)(\hat{\theta}_n - \theta_0), \quad (11)$$

where  $\theta_n$  lies in between  $\hat{\theta}_n$  and  $\theta_0$ . Suppose that  $\hat{\theta}_n$  is an interior solution to the minimization problem (6). Then, it follows that  $\dot{Q}_n(\hat{\theta}_n) = 0$ .

From Theorem 1, normalized by an appropriately chosen sequence  $\nu_n$ ,  $\nu_n^{-1}\dot{Q}_n(\theta_0) \xrightarrow{d} \dot{Q}(\theta_0)$  for some random vector  $\dot{Q}(\theta_0)$ . Also, let

$$\ddot{Q}_n^0 = \sum_{t=1}^n \dot{g}(z_t, \theta_0) \dot{g}(z_t, \theta_0).$$

We have  $\nu_n^{-2}\ddot{Q}_n^0(\theta_0) \xrightarrow{p} \ddot{Q}(\theta_0)$  for some random matrix  $\ddot{Q}(\theta_0)$  by Theorem 1. Thus, with suitable assumptions imposed, we may expect that

$$\nu_n(\hat{\theta}_n - \theta_0) = -(\nu_n^{-2}\ddot{Q}_n(\theta_n))^{-1}\nu_n^{-1}\dot{Q}_n(\theta_0) \quad (12)$$

$$= -(\nu_n^{-2}\ddot{Q}_n^0(\theta_0))^{-1}\nu_n^{-1}\dot{Q}_n(\theta_0) + o_p(1) \quad (13)$$

$$\xrightarrow{d} -\ddot{Q}(\theta_0)^{-1}\dot{Q}(\theta_0) \quad (14)$$

$$(15)$$

as  $n \rightarrow \infty$ .

A set of sufficient conditions leading to (12) are listed below for reference.

AD1:  $\nu_n^{-1}\dot{Q}_n(\theta_0) \xrightarrow{d} \dot{Q}(\theta_0)$  as  $n \rightarrow \infty$ .

AD2:  $\dot{Q}_n(\hat{\theta}_n) = 0$  with probability approaching to one as  $n \rightarrow \infty$ .

AD3:  $\nu_n^{-2}(\ddot{Q}_n(\theta_n) - \ddot{Q}_n^0(\theta_0)) \xrightarrow{p} 0$  as  $n \rightarrow \infty$ .

AD4:  $\ddot{Q}(\theta_0) > 0$  a.s.

AD5:  $\nu_n^{-2}\ddot{Q}_n(\theta_0) = \nu_n^{-2}\ddot{Q}_n^0(\theta_0) + o_p(1)$  for large  $n$ .

AD6:  $\nu_n^{-2}\ddot{Q}_n(\theta_0) \xrightarrow{p} \ddot{Q}(\theta_0)$  as  $n \rightarrow \infty$ .

Under standard asymptotic conditions in nonlinear regression AD1-AD6, it's easy to see that (12) follows from (11).

Theorem 2.3.1. Let Assumption 2.2 holds. Assume  $g$  satisfies conditions in Theorem 2.2,  $\dot{g}$  and  $\ddot{g}$  are I-regular on  $\Theta$ , and  $\int_{-\infty}^{\infty} \dot{g}(s, \theta_0) \dot{g}(s, \theta_0) ds > 0$ . Then we have

$$\sqrt[4]{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} \left( L_{V_c}(1, 0) \int_{-\infty}^{\infty} \dot{g}(s, \theta_0) \dot{g}(s, \theta_0) ds \right)^{-1/2} W(1)$$

as  $n \rightarrow \infty$ . Here,  $V_c$  is defined in (9)  $W(r)$  is a Brownian Motion satisfying

$$\limsup_{r \rightarrow 0^+} \frac{W(r)}{\sqrt{2r \log \log \frac{1}{r}}} = 1.$$

The NLS estimator converges at the rate of  $\sqrt[4]{n}$ , and has a mixed Gaussian limiting distribution with I-regular regression functions. The technology applied in this section follows immediately from [Park and Phillips, 2008].



## CHAPTER 3: TEST STATISTIC

This Chapter constructs the test statistic and derives its asymptotics based on theorems given above. The work in this Chapter follows [Sun et al., 2011].

### 3.1 Construction of Test Statistics

We construct a  $L^2$ -type test statistic as in [Sun et al., 2011],

$$\int [\hat{f}_n(z) - g(z, \hat{\theta}_n)]^2 dz,$$

where  $K_t(z) = K((Z_t - z)/h)$ .

$$\hat{f}_n(z) = \left[ \sum_{t=1}^n K_t(z) \right]^{-1} \sum_{t=1}^n y_t K_t(z)$$

is the NW kernel estimator of nonlinear functional  $f(z)$ , and  $g(z, \hat{\theta}_n)$  is the NLS estimator of  $g(z, \theta)$ . We modify the test statistic by multiplying a weighting matrix

$D_n(z) = \sum_{t=1}^T K_t(z)$  to get rid of the random denominator,

$$\int [D_n(z)(\hat{f}_n(z) - g(z, \hat{\theta}_n))]^2 dz = \sum_{t=1}^n \sum_{s=1}^n \hat{u}_t \hat{u}_s \int K_t(z) K_s(z) dz, \quad (16)$$

where  $\hat{u}_t = y_t - g(z, \hat{\theta}_n)$  is the residual from the parametric model. Then, a convolution

kernel is defined,

$$\bar{K}_{ts} \stackrel{def}{=} \int K_t(z) K_s(z) dz = \begin{cases} h \int K^2(z) dz & \text{if } t = s; \\ h \int K(v) K((Z_s - Z_t)/h + v) dv & \text{if } t \neq s. \end{cases}$$

When  $t \neq s$ ,  $\bar{K}_{ts} = \int K_t(z) K_s(z) dz$  can be regarded as a local weight function.

Therefore, our final test statistic is obtained by removing the global center with  $t = s$

and replacing  $\bar{K}_{ts}$  with  $K_{ts} \equiv K((Z_t - Z_s)/h)$  as in [Sun et al., 2011], where  $K(\cdot)$  is

a kernel function.

$$\hat{I}_n = \frac{1}{n^2 h} \sum_{t=1}^n \sum_{s \neq t}^n \hat{u}_t \hat{u}_s K_{ts}. \quad (17)$$

$\hat{I}_n$  is a second-order U-statistic similar to the test statistic proposed in [Li et al., 2002] and [Sun et al., 2011]. Model (4) was studied by [Li et al., 2002] assuming that both  $x_t$  and  $z_t$  are stationary variables, and the test statistic  $\tilde{I}_n = \frac{1}{n^3 h} \sum_{t=1}^n \sum_{s \neq t}^n X_t^T X_s \hat{u}_t \hat{u}_s K_{ts}$  was constructed. With all variables stationary, it is shown that  $\tilde{I}_n$  converges to  $E\{[E(X_t u_t | z_t)]^2 f(z_t)\} \geq 0$  with proper scale of  $n$  and  $h$ . It's apparent to see that  $\tilde{I}_n$  is a one-sided test statistic. The setting of [Li et al., 2002] was changed in [Sun et al., 2011] by assuming  $X_t$  to be an I(1) process. Law of large numbers applied by [Li et al., 2002] is not applicable when non stationary variables are included. Therefore, Martingale Central Limit Theory was adopted by [Sun et al., 2011] to develop the asymptotic theory of  $\tilde{I}_n$ . It's proved that  $\tilde{I}_n$  is also a one-sided test statistic that approaches a positive random variable under alternatives. In this paper, the kernel function is based on NI(1) random variables rather than stationary variables. The fact that the nonstationary variable is set into a function form significantly complicates the proof of the limit theory. By applying Martingale Central Limit Theory, continuous mapping theorem and the definition of local time, we derive the limit distribution of  $\hat{I}_n$  under both null and alternative hypothesis.  $\hat{I}_n$  is shown to be one sided unsurprisingly.

### 3.2 Assumptions and Asymptotic Results

Assumptions are imposed below for developing asymptotic theories. We start by giving a stronger assumption on  $\{z_t\}$ .

**Assumption 4.1 :** (i) On a suitable probability space, there exists a stochastic process  $V_c(\cdot)$  having a continuous local time such that for some  $\theta_* = (1/2) - 1/(2 + \delta_*)$  and  $\lambda_* > 0$  (a function of  $\delta_*$ ) with  $0 < \delta_* \leq 2$

$$\sup_{0 \leq r \leq 1} \|V_n(r) - V_c(r)\| = O_{a.s.}(n^{-\theta_*} \log^{\lambda_*}(n)), \quad (18)$$

where  $\|x\|$  is the Euclidean norm of  $x$  and  $O_{a.s.}(\cdot)$  denotes almost surely convergence.

(ii) Furthermore,

$$\sup_{r \in [0,1]} \|V_n(r)\| = O_{a.s.}(\sqrt{\log \log n}). \quad (19)$$

**Remark 1:** Apparently, Assumption 4.1 is stronger than Assumption 2.1, since strong approximation in (18) usually requires stronger assumptions than weak convergence as in Assumption 2.1. Theorem 4.1 of [Shao and Lu, 1987] establishes a sufficient condition for Assumption 4.1 to hold. It states that, for a stationary  $\beta$ -mixing sequence  $\{\eta_t\}$  satisfying, for some  $\gamma_* > 2 + \delta_*$ ,

$$E|\eta_t|^{\gamma_*} < \infty, \quad \text{and} \quad \sum_{n=1}^{\infty} \beta_n^{1/(2+\delta_*)-1/\gamma_*} < \infty, \quad (20)$$

where  $\beta_n$  are the mixing coefficients of  $\{\eta_t\}$ , Assumption 4.1 holds true.

Both the weak convergence in Assumption 2.1 and the strong approximation result in (18) are commonly made assumptions in econometrics literature, as Assumptions in [Kasparis, 2008], [Kasparis and Phillips, 2012], and [Wang and Phillips, 2009a].

**Remark 2:** The almost sure assumptions in (18) and (19) can be replaced by  $O_p(\cdot)$ . By the Strassen's functional law of iterative logarithm for a NI(1) process (see

[Rio, 1995]), (19) can be derived.

Now, we work on the limiting distribution of  $\hat{I}_n$  with additional assumptions imposed. First of all, a useful notation is defined.

$$\Omega_n(\eta) = \{(l, k) : \eta n \leq k \leq (1 - \eta)n, k + \eta n \leq l \leq n\},$$

where  $0 < \eta < 1$ , following [Wang and Phillips, 2009a].

(A1)  $\dot{g}(z, \theta)$  is continuously twice differentiable with respect to  $\theta$ .  $\dot{g}(z, \theta)$  and its partial derivative functions with respect to  $\theta$  (up to second order) are all uniformly continuous and bounded. Moreover,  $\int_{-\infty}^{\infty} \dot{g}^2(z, \theta) dz < \infty$ .

(A2) For all  $0 \leq k < l \leq n$ ,  $n \geq 1$ , there exist a sequence of constants  $d_{l,k,n}$  such that

(a) for some  $m_0 > 0$  and  $C > 0$ ,  $\inf_{(l,k) \in \Omega_n(\eta)} d_{l,k,n} \geq \eta^{m_0}/C$  as  $n \rightarrow \infty$ ,

$$\lim_{\eta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{l=(1-\eta)n}^n (d_{l,0,n})^{-1} = 0, \quad (21)$$

$$\lim_{\eta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{1}{n} \max_{0 \leq k \leq (1-\eta)n} \sum_{l=(k+1)}^{k+\eta n} (d_{l,k,n})^{-1} = 0, \quad (22)$$

$$\lim_{n \rightarrow \infty} \sup \frac{1}{n} \max_{0 \leq k \leq n-1} \sum_{l=(k+1)}^n (d_{l,k,n})^{-1} < \infty, \quad (23)$$

$$(24)$$

(b)  $z_{k,n}$  are adapted to  $F_{k,n}$  and, conditional on  $F_{k,n}$ ,  $(z_{l,n} - z_{k,n})/d_{l,k,n}$  has a density  $h_{l,k,n}(x)$  which is uniformly bounded by a constant  $K$  and

$$\lim_{\delta \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{(l,k) \in \Omega_n[\delta^{1/(2m_0)}]} \sup_{|u| \leq \delta} |h_{l,k,n}(u) - h_{l,k,n}(0)| = 0. \quad (25)$$

(A3)  $\{u_t\}$  is an i.i.d. sequence and is independent of  $\{Z_t\}$ . Also,  $E(u_t) = 0$ ,

$$E(u_t^2) = \sigma_u^2 < \infty \text{ and } E(u_t^4) = \mu_4 < \infty.$$

(A4) The kernel function  $K(u)$  is a differentiable symmetric (around zero) probability density function on interval  $[-1, 1]$ . Also, we denote  $\nu_2(K) = \int K^2(u) du$ ,

$$\sup_u K(u) < \infty \text{ and } \sup_u K'(u) < \infty.$$

(A5)  $\{\eta_t\}$  is a strictly stationary, absolutely regular (or  $\beta$ -mixing) sequence satisfying (20).

(A6)  $h(\log \log n)^3 \rightarrow 0$ ,  $nh^2 \rightarrow \infty$ , and  $h \rightarrow 0$  as  $n \rightarrow \infty$ .

(A7)  $\sup_{1 \leq t \leq n} \left\| \hat{f}_n(z_t) - f(z_t) \right\| = o_p(n^{-1/2})$ .

**Remark 3:** (A3) can be relaxed to  $E(u_t|z_t, \mathcal{F}_{n,t-1}) = 0$ ,  $E(u_t^2|z_t, \mathcal{F}_{n,t-1}) = \sigma_u^2$  and  $E(u_t^4|z_t, \mathcal{F}_{n,t-1}) < \infty$  for all  $t$ , which requires a lengthier proof.

**Remark 4:** The bounded support of the kernel function in (A4) is not necessary. Kernel functions with unbounded support, such as Gaussian kernel, is allowed at the cost of a lengthier proof. (A7) is used to simplify the proof of consistency of the estimated asymptotic variance of the test statistic.

Before presenting the asymptotic results of our test statistic, we define a measurable process  $L_{V_c}(r, r, 0)$  as the local time of measurable process  $\{V_c(t) - V_c(s), t \geq 0, s \geq 0\}$ ,

$$\int_0^r \int_0^r T[(V_c(t) - V_c(s))] ds dt = \int_{-\infty}^{\infty} T(x) L_{V_c}(r, r, 0) dx, \text{ all } r \in \mathcal{R} \quad (26)$$

where  $T(x)$  denotes a locally integrable function.

Now, the asymptotic properties of our test statistic are stated in the following theorem with proofs delayed to Appendix B.

Theorem 3.2.1. Under Assumptions A1\*-A7, we obtain (i) under  $H_0$ ,

$$J_n = n^{\frac{5}{4}} h^{\frac{1}{2}} \hat{I}_n \xrightarrow{d} MN(0, \Sigma), \quad (27)$$

where  $MN(0, \Sigma)$  is a mixed normal distribution with zero mean and conditional

variance as

$$\Sigma = \frac{1}{2} \sigma_u^4 \mu_2(K) E[L_{V_c}(r, r, 0)] \quad (28)$$

In addition, if Assumption A8 also holds, a consistent estimator of  $\Sigma$  is given by

$$\widehat{\Sigma} = \frac{2}{n^{\frac{3}{2}}h} \sum_{t=2}^n \sum_{s=1}^{t-1} \tilde{u}_t^2 \tilde{u}_s^2 K_{ts}^2 \xrightarrow{p} \Sigma \quad (29)$$

where  $\tilde{u}_t = Y_t - \hat{f}^{(-t)}(Z_t)$  is the nonparametric residual of the leave-one-out estimator  $\hat{f}^{(-t)}(Z_t)$  for all  $t$ ;

(ii) under  $H_1$ , the test statistic  $J_n$  diverges to  $+\infty$  at the rate of  $h^{-1}$ . Hence, we have

$$\Pr[J_n > B_n] \rightarrow 1 \text{ as } n \rightarrow \infty,$$

where  $B_n$  is a non-stochastic sequence with  $B_n = o(h^{-1})$ . Therefore, the statistic  $J_n$  is a consistent test.

Theorem 3.2.1 shows that  $J_n$  as the leading term of  $n^{5/4}h^{1/2}\hat{I}_n$  converges in distribution to a positive random variable under the alternative hypothesis. That indicates that the test is one-sided. It follows that the nonlinear parametric functional form in null hypothesis is rejected when  $J_n$  is greater than the  $(1 - \alpha)100\%$ th percentile  $z_\alpha$  of a standard normal distribution.

## CHAPTER 4: MONTE CARLO SIMULATIONS

Monte Carlo simulations are performed in this chapter to examine the finite sample performance of the proposed nonparametric test. The test statistic is given by

$$J_n = n^{\frac{5}{4}} h^{\frac{1}{2}} \hat{I}_n \quad (30)$$

where

$$\hat{I}_n = \frac{1}{n^2 h} \sum_{t=1}^n \sum_{s=1}^n \hat{u}_t \hat{u}_s K_{ts} \quad (31)$$

as proposed in Chapter 3.

The data generating process (*DGP*) under  $H_0$  is assumed to be:

$$y_t = \theta z_t^2 + u_t \quad (32)$$

$$z_t = \rho z_t + \eta_t = \left(1 - \frac{c}{n}\right) z_{t-1} + \eta_t \quad (33)$$

where  $u_t$  is an i.i.d random variable satisfying  $N(0, \sigma_u^2)$ ,  $\eta_t$  an i.i.d standard normal random variable, and  $z_t$  a NI(1) process that's independent of  $u_t$ . It's clear that  $z_t$  becomes I(1) process if  $\rho = 1$  or  $c = 0$ . Thus, in model (32), we see that  $y_t$  is a nonlinear function of nonstationary random variable  $z_t$ .

For alternative hypothesis, two different settings are investigated. We use  $DGP_1$  and  $DGP_2$  to indicate two data generating processes constructed under  $H_a$ :

$$DGP_1 : y_t = \theta z_t^2 + a_1 n^{-1/10} z_t + u_t,$$

$$DGP_2 : y_t = \theta z_t^2 + a_2 n^{-1/10} z_t^3 + u_t$$

Table 1: Estimated Sizes: Varying Smoothing Parameters

|     | $d = .8$ |      |      | $d = 1$ |      |      | $d = 1.2$ |      |      |
|-----|----------|------|------|---------|------|------|-----------|------|------|
| $n$ | 1%       | 5%   | 10%  | 1%      | 5%   | 10%  | 1%        | 5%   | 10%  |
| 100 | .014     | .061 | .113 | .008    | .036 | .100 | .014      | .053 | .101 |
| 200 | .016     | .076 | .124 | .013    | .055 | .106 | .010      | .061 | .128 |
| 400 | .014     | .057 | .118 | .024    | .061 | .106 | .014      | .047 | .100 |
| 600 | .015     | .058 | .101 | .010    | .055 | .101 | .014      | .058 | .105 |

Table 2: Estimated Powers: Varying Smoothing Parameters

|     | $d = .8$ |      |      | $d = 1$ |      |      | $d = 1.2$ |      |      |
|-----|----------|------|------|---------|------|------|-----------|------|------|
| $n$ | 1%       | 5%   | 10%  | 1%      | 5%   | 10%  | 1%        | 5%   | 10%  |
| 100 | .846     | .914 | .939 | .832    | .900 | .952 | .811      | .900 | .928 |
| 200 | .996     | .998 | .999 | .997    | 1    | 1    | .995      | .997 | 1    |
| 400 | 1        | 1    | 1    | 1       | 1    | 1    | 1         | 1    | 1    |
| 600 | 1        | 1    | 1    | 1       | 1    | 1    | 1         | 1    | 1    |

The replication time of Monte Carlo simulation is  $m = 1000$ . Sample sizes are  $n = 100$ ,  $n = 200$ ,  $n = 400$  and  $n = 600$ . Gaussian kernel function is used with bandwidth  $h = dn^{-1/10}$ . First, we let near-integration parameter  $c = 2$  and choose different values of  $d$  to check the effect of different amount of smoothing. The results are listed in Table 1 and Table 2. Then, we compare tests under 3 settings of near-integration parameter  $c$  with  $c = 0$ ,  $c = 2$  and  $c = 20$ , and  $d$  is fixed to be 1. Table 3 and Table 4 give the results of the above comparison. Estimated powers above are calculated based on  $DGP_1$  with  $a_1 = 0.5$ . Estimated powers against  $DGP_j$  are reported in Table 5, where  $c$  and  $d$  are both set to be 1, and  $a_1 = a_2 = 0.5$ . We report estimated powers against  $DGP_1$  according to different settings of  $a_1$  in Table 6, where  $c = d = 1$ .

From Table 1 and 2, we don't see significant effect on test sizes and powers from the smoothing parameter. Table 2 shows that even the sample sizes are small, the proposed test statistic reject the null hypothesis effectively under  $H_a$ .





Table 3 and 4 offer the estimated test sizes and powers when the integration parameter varies. It's obvious that for  $c = 20$ , our test has less power against the alternative than the other cases with  $c = 0$  and  $c = 2$ , when sample size is pretty small. As sample size increases, the test has power for all 3 settings of  $c$ . This indicates that the test is more powerful if regressors are closer to an I(1) process rather than the stationary process, especially when sample size is limited.

The test has power against both generating processes  $DGP_1$  and  $DGP_2$  as presented in Table 5.

We see from table 6 that the proposed test statistic is sensitive to parameter  $a_1$  in alternative hypothesis. The greater the value of  $a_1$  is, the better we can detect the alternative hypothesis. We also see that for sample sizes large enough, our test is equivalently powerful to all values of  $a_1$ .

The finite sample performance of the proposed test statistic was demonstrated by Monte Carlo simulations implemented above. Then, we'll apply it to testing predictability of asset return in the following chapter.

## CHAPTER 5: EMPIRICAL STUDY

### 5.1 Review of Tests of Predictability of Asset return

Monte Carlo simulations conducted in the previous chapter illustrate finite sample performance of our test. In this chapter, we apply the proposed test statistic to testing the predictability of asset return.

Whether asset returns can be predicted by financial variables like dividend-to-price ratio and earning-to-price ratios has been a hot topic for last two decades. Conventional tests of predictability of asset return could lead to invalid inference due to the high persistency of financial variables. The large sample theory of traditional t-statistic is shown to be a poor approximation to the finite sample distribution of test statistic based on a persistent predictor variable (see [Elliott and Stock, 1994]; [Gregory Mankiw and Shapiro, 1986]; and [Stambaugh, 1999]), since the asymptotic theory for t-statistic is established on the assumption that the predictor is a process with autoregressive root less than 1. Hence, the strong evidence for the predictability of asset returns provided by traditional t-test is not reliable.

Later on, new methods are developed to address the problem caused by high persistence of financial variable. Extending work of [Richardson and Stock, 1989] and [Cavanagh et al., ], [Torous et al., 2004] shows that returns are predictable at short horizons but not at long horizons. No evidence for predictability of stock return

was found in [Lanne, 2002] by testing the stationarity of long-horizon returns, while predictability with some ratios was verified in [Lewellen, 2004].

An unifying understanding of various test procedures mentioned above refers to [Campbell and Yogo, 2006]. They used theory of uniformly most powerful (UMP) test as a benchmark to compare different methods. In addition, a new Bonferroni test was proposed by [Campbell and Yogo, 2006] based on the theory of UMP test.

The test procedure proposed by this paper differs from that in [Campbell and Yogo, 2006] in that we test a specific parametric model against a nonparametric model. In the context of testing predictability of asset returns, we check the linear regression model with high persistent financial predictor. Then, the null hypothesis is

$$H_0 : \quad \Pr(r_t = \theta_0 + \theta_1 z_{t-1} + u_t) = 1 \text{ for any } \theta_0 \in \Theta_0 \text{ and } \theta_1 \in \Theta_1,$$

where  $r_t$  denotes asset return, and  $z_t$  financial variable. The alternative hypothesis is

$$H_1 : \quad \Pr(r_t = \theta_0 + \theta_1 z_{t-1} + u_t) = 0 \text{ for any } \theta_0 \in \Theta_0 \text{ and } \theta_1 \in \Theta_1,$$

The work of [Campbell and Yogo, 2006] focused on testing whether the value of parameter in linear prediction model  $r_t = \theta_0 + \theta_1 z_{t-1} + u_t$  equals zero or not. The hypotheses are stated as

$$H_0 : \quad \theta_1 = 0,$$

and

$$H_1 : \quad \theta_1 \neq 0,$$

It's clear to see that the rejection of null hypothesis indicates no linear predictability of asset return in our test procedure, while in [Campbell and Yogo, 2006], the rejection of null hypothesis provides evidence for predictability of asset return.

## 5.2 Description of Data and Model

In this section, the nonparametric test of the linear prediction model of asset return is implemented on monthly NYSE/AMEX value-weighted index data (1926-2002) from the Center for Research in Security Prices (CRSP), referring to the data used by [Campbell and Yogo, 2006]. Dividend-price ratio and earnings-price ratio are used to predict excess stock returns separately, where dividend-price ratio is defined by the ratio of past year dividends over current price, and earnings-price ratio by dividing moving average of earnings over previous ten years by current stock price. Monthly earnings are constructed by linear extrapolation using data from S&P 500 as in [Schiller, 2000], since no earnings available from CRSP. Excess returns are computed as stock returns subtracting risk-free returns. The one-month T-bill rate from CRSP Indices database is used as monthly risk-free return.

The regression model we consider is

$$r_t = \alpha + \beta x_{t-1} + u_t, \quad (34)$$

$$x_t = \gamma + \rho x_{t-1} + e_t, \quad (35)$$

where  $r_t$  denotes the excess stock return at time  $t$ , and  $x_{t-1}$  the financial predictor at time  $t-1$ . The financial variables used to predict excess return are log dividend-price ratio and log earning-price ratio.

Fig.1 and Fig.2 provide time series plots of monthly log dividend-price ratio and monthly log earnings-to price ratio from 1926 to 2002. Both ratios appear persistent, especially at the end of the sample period. We estimate  $\rho$  by least square method and construct the confidence intervals toward log dividend-price ratio and log earning-

Table 7: Estimated Autoregression Parameter

|     | $\rho$ | 95% CI for $\rho$ |
|-----|--------|-------------------|
| ldp | .9895  | (.9796, .9994)    |
| lep | .9885  | (.9786, .9985)    |

price ratio in Table 7. It's apparent that log dividend-price ratio and log earning-price ratio are both near integrated time series with autoregression coefficient close to 1.

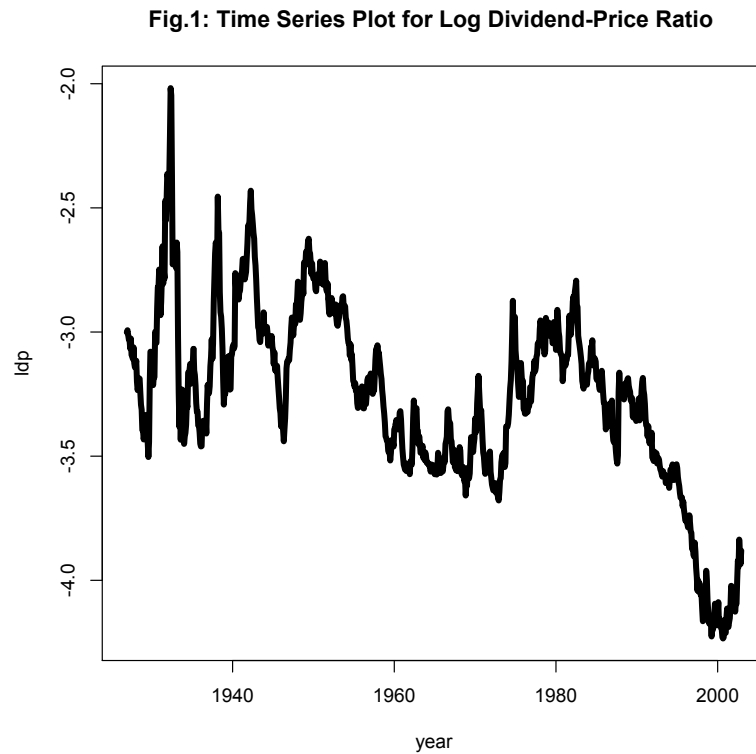


Figure 1: Time Series Plot for Log Dividend-Price Ratio.

### 5.3 Nonparametric Test of Predictability

We show in previous section that the predictors are near integrated processes. So the proposed nonparametric test statistic can be applied to testing the predictability of stock return. The test statistic is defined as (30) in Chapter 4.

To get the estimated critical values, we perform nonparametric wild bootstrap to

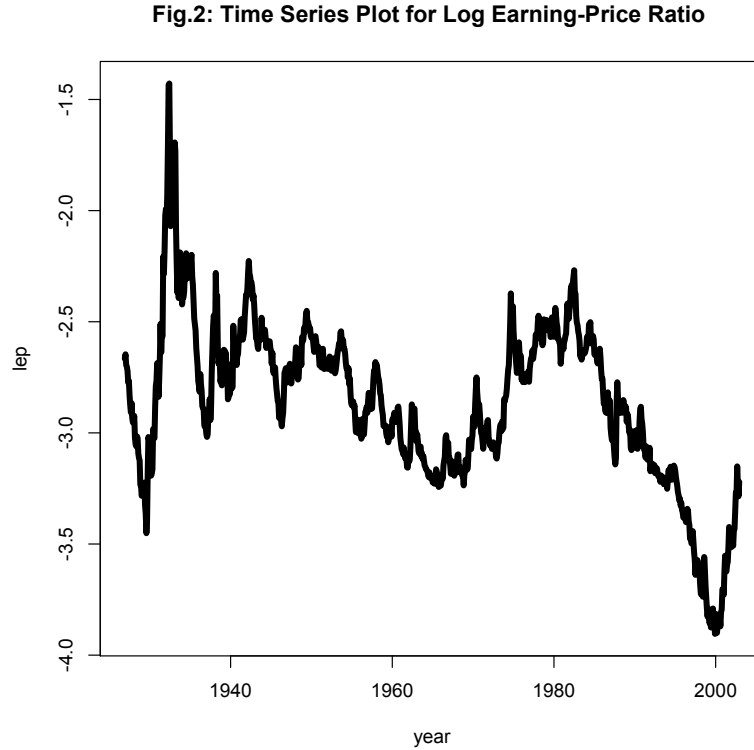


Figure 2: Time Series Plot for Log Earning-Price Ratio.

do residual resampling. The procedure is described as below,

1. Generate bootstrap residuals  $u^*$  from multiplying nonparametric residuals  $\tilde{u}$  by standard normal random variable  $\epsilon$ .
2. The resampled response variable  $r_t^*$  is calculated as

$$r_t^* = \hat{\alpha} + \hat{\beta}x_{t-1} + u_t^*,$$

where  $\hat{\alpha}$  and  $\hat{\beta}$  are linear least square estimates from the original data.

3. Compute test statistic  $J_n$  by using bootstrap response observations  $r_t^*$  and  $x_t$ .

We repeat the above procedure for 400 times to construct the confidence interval for  $J_n$ . The result of our empirical study is provided in Table 8

The values of test statistic  $J_n$  based on log dividend-price ratio and log earning-price

Table 8: Test Statistics and Confidence Intervals

|     | $J_n$  | 90% CI           | 95% CI           | 99% CI           |
|-----|--------|------------------|------------------|------------------|
| ldp | -.0259 | (-.0374, -.0229) | (-.0390, -.0206) | (-.0415, -.0142) |
| lep | -.0277 | (-.0360, -.0207) | (-.0375, -.0186) | (-.0394, -.0155) |

ratio are both inside all of the confidence intervals reported in Table 8. Therefore, our nonparametric test fail to reject the linear prediction model for stock return. This could be viewed as evidence for predictability of asset return from a linear model of financial variables also.



## CHAPTER 6: CONCLUSION

We propose a  $L^2$  type nonparametric test statistic to test the nonlinear parametric model with near integrated regressors in this dissertation. The construction of test statistic is based on [Sun et al., 2011], where the limit distribution of the test statistic is derived under the null hypothesis of a linear function of nonstationary time series. We extend the method to testing a model of nonlinear function of a near-integrated process. The contribution of this dissertation is to provide the asymptotics of a  $L^2$  type test statistic with a nonlinear function of near-integrated process included. The asymptotic distribution under the null hypothesis of a nonlinear function is mixed normal, similar to testing a linear model as in [Sun et al., 2011]. Since We test the null against contiguous alternatives, the convergence rate for alternative models is derived to be less than or equal to  $n^{-\frac{1}{10}}$  to make it detectable, when the rate for bandwidth is set to be optimal  $h = n^{-\frac{1}{10}}$ .

Monte Carlo simulation demonstrates the finite sample performance of the test statistic. It shows that even the sample sizes are quite small, like 100 and 200, the proposed test has power against the alternative, and the power increases rapidly as the sample size increases. Table 2 shows that the test isn't sensitive to the selection of smoothing parameter. But it is noticeably more powerful if the regressor is closer to a unit root process than to a stationary process seen from Table 4. We also see

that the power of the test is significantly sensitive to parameter  $a_1$  in the alternative model. The power is positively related with  $a_1$ .

In empirical studies, the test is applied to testing the linear prediction model of asset return. The high persistence of financial variables used to predict asset return is shown. Since traditional test procedures are not appropriate in case of high persistent predictors, the strong evidence for predictability from traditional tests are not reliable due to over-rejection (see [Campbell and Yogo, 2006]). Thus, the nonparametric test proposed here is performed and evidence for linear predictability is shown. The linear prediction model of stock return with log dividend price ratio and earning price ratio as predictors respectively is verified.

In short, a nonparametric test procedure, that can be used for detecting nonstationary nonlinear parametric model, is developed in this dissertation. The linear prediction model of asset return is evidently supported by this method.

## Appendix A: TECHNICAL RESULTS FOR CHAPTER 2

The proof of limit theory of NLS estimator follows the procedure applied by [Park and Phillips, 2008]. We start by defining regular functions as follows (see [Park and Phillips, 2008] ):

Definition A.1. A transformation  $T$  on  $\mathcal{R}$  is said to be regular if and only if

- (a). it is continuous in a neighborhood of infinity, and
- (b). given any compact set  $K \subset \mathcal{R}$ , for each  $\epsilon > 0$  there exists continuous functions

$$\underline{T}_\epsilon, \bar{T}_\epsilon, \text{ and } \delta_\epsilon > 0 \text{ such that } \underline{T}_\epsilon(x) \leq T(y) \leq \bar{T}_\epsilon(x) \text{ for all } |x - y| < \delta_\epsilon > 0 \text{ on } K, \text{ and such that } \int_K (\underline{T}_\epsilon - \bar{T}_\epsilon)(x) dx \rightarrow 0 \text{ as } \epsilon \rightarrow 0.$$

The so called regularity conditions are defined also.

Definition A.2.  $F$  is regular on  $\Pi$  if

- (a).  $F(\cdot, \pi)$  is regular for all  $\pi \in \Pi$ , and
- (b). for all  $x \in \mathcal{R}$ ,  $F(x, \cdot)$  is discontinuous in a neighborhood of  $x$ .

The regularity conditions (a) is a sufficient condition that ensures the existence of both sample mean and sample covariance asymptotics for  $F(\cdot, \pi)$  for each  $\pi \in \Pi$ . Condition (b) guarantees that there's a neighborhood  $N_0$  of any  $\pi_0 \in \Pi$  such that  $\sup_{\pi \in N_0} F(\cdot, \pi)$  and  $\inf_{\pi \in N_0} F(\cdot, \pi)$  are regular. These results are shown in the following lemmas.

Next, we provide some useful lemmas:

Lemma A.3. If  $T_1$  and  $T_2$  are regular transformations, then so are  $T_1 \pm T_2$  and  $T_1 T_2$ .

Lemma A.4. Suppose that Assumption 2.1 holds. If  $T$  is regular, then

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n T\left(\frac{z_t}{\sqrt{n}}\right) &\rightarrow_{a.s.} \int_0^1 T(V_c(r)) dr, \\ \frac{1}{n} \sum_{t=1}^n T\left(\frac{z_t}{\sqrt{n}}\right) u_t &\xrightarrow{d} \int_0^1 T(V_c(r)) dU(r), \end{aligned}$$

as  $n \rightarrow \infty$

Lemma A.5. (a) If  $F(\cdot, \pi)$  is a regular family on  $\Pi$ , then for each  $\pi_0 \in \Pi$ , there exists a neighborhood  $N_0$  of  $\pi_0$  such that  $\sup_{\pi \in \Pi} F(\cdot, \pi)$  and  $\inf_{\pi \in \Pi} F(\cdot, \pi)$  are regular for all  $N \subset N_0$ .

(b) If  $F$  is regular on a compact set  $\Pi$ , then  $\sup_{\pi \in \Pi} |F(\cdot, \pi)|$  is locally bounded.

Lemma A.6. (a) Let Assumption 2.1 hold. If  $F$  is regular on a compact set  $\Pi$ , then for large  $n$ ,  $n^{-1} \sum_{t=1}^n F(z_t/\sqrt{n}, \pi) u_t = o_p(1)$  uniformly in  $\pi \in \Pi$ .

(b) Let Assumption 2.2 hold. If  $F$  is I-regular on a compact set  $\Pi$ , then for large  $n$ ,  $n^{-1/2} \sum_{t=1}^n$

$F(z_t, \pi) u_t = o_p(1)$  uniformly in  $\pi \in \Pi$ .

Lemma A.7. (a) If  $F$  is regular on a compact set  $\Pi$ , then  $\int_0^1 F(V_c(r), \cdot) dr$  is continuous *a.s.* on  $\Pi$ .

(b) If  $F$  is I-regular on a compact set  $\Pi$ ,  $\int_{-\infty}^{\infty} F(s, \cdot) ds$  is continuous on  $\Pi$ .

Lemma A.8. Let Assumptions 2.1 hold. Then  $U_n^0$  introduced in (8) can be represented by

$$U_n^0\left(\frac{t}{n}\right) = U\left(\frac{\tau_{nt}}{n}\right)$$

with an increasing sequence of stopping times  $\tau_{nt}$  in  $(\Sigma, \mathcal{F}, P)$  with  $\tau_{n0} = 0$  such that

$$\sup_{1 \leq t \leq n} \left| \frac{\tau_{nt} - t}{n^\delta} \right| \rightarrow_{a.s.} 0$$

as  $n \rightarrow \infty$  for any  $\delta > \max(1/2, 2/q)$  where  $q$  is the moment exponent given in Assumption 2.1.

Lemma A.9. (See Theorem 3.1 in [Park and Phillips, 2008]) Let Assumptions 2.1 hold. If  $F$  is regular on a compact set  $\Pi$ , then

$$\frac{1}{n} \sum_{t=1}^n F\left(\frac{z_t}{\sqrt{n}}, \pi\right) \rightarrow_{a,s} \int_0^1 F(V_c(r), \pi) dr$$

uniformly in  $\pi \in \Pi$ , as  $n \rightarrow \infty$ . Moreover, if  $F(\cdot, \pi)$  is regular, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n F\left(\frac{z_t}{\sqrt{n}}, \pi\right) u_t \rightarrow_{a,s} \int_0^1 F(V_c(r), \pi) dU(r)$$

as  $n \rightarrow \infty$ .

See Appendix A of [Park and Phillips, 2008] for proofs of lemmas. Now, we use lemmas given above to prove Theorem 2.1.1.

**Proof of Theorem 2.1.1:** See proof of Theorem 3.2 of [Park and Phillips, 2008].

**Proof of Theorem 2.2.1:** See proof of Theorem 4.1 of [Park and Phillips, 2008].

**Proof of Theorem 2.3.1:** See proof of Theorem 5.1 of [Park and Phillips, 2008].

## Appendix B: TECHNICAL RESULTS FOR CHAPTER 3

Throughout this section we will use the notation that  $A_n \approx B_n$  to denote that  $B_n$  is the leading term of  $A_n$ , i.e.,  $A_n = B_n + (s.o.)$ , where  $(s.o.)$  denotes terms having probability order smaller than that of  $B_n$ . In addition, we use  $A_n \sim B_n$  to denote  $A_n$  and  $B_n$  having the same stochastic order. Also, we let  $M$  denote a generic constant, which may take different values at different places.

**Proof of Theorem 3.2.1 (i):** Under  $H_0$ ,  $\hat{u}_t = y_t - g(z_t, \hat{\theta}_n) = u_t - (g(z_t, \hat{\theta}_n) - g(z_t, \theta_0))$ , where  $\theta_0$  is the true parameter to be estimated. We decompose  $\hat{I}_n$  in (17) as

$$\begin{aligned} \hat{I}_n &= \frac{1}{n^2 h} \sum_{t=1}^n \sum_{s \neq t} [u_t u_s + (g(z_t, \hat{\theta}_n) - g(z_t, \theta_0))(g(z_s, \hat{\theta}_n) - g(z_s, \theta_0)) \\ &\quad - 2 u_t (g(z_s, \hat{\theta}_n) - g(z_s, \theta_0))] K_{ts} \\ &\equiv I_{1n} + G_{2n} - 2G_{3n}, \end{aligned}$$

where

$$I_{1n} = \frac{1}{n^2 h} \sum_{t=1}^n \sum_{s \neq t} u_t u_s K_{ts}, \quad (36)$$

$$G_{2n} = \frac{1}{n^2 h} \sum_{t=1}^n (g(z_t, \hat{\theta}_n) - g(z_t, \theta_0)) \sum_{s \neq t} (g(z_s, \hat{\theta}_n) - g(z_s, \theta_0)) K_{ts}, \quad (37)$$

and

$$G_{3n} = \frac{1}{n^2 h} \sum_{t=1}^n u_t \sum_{s \neq t} (g(z_s, \hat{\theta}_n) - g(z_s, \theta_0)) K_{ts}. \quad (38)$$

Lemma B.11 below shows that, under  $H_0$ ,  $n^{5/4} h^{1/2} I_{1n} \xrightarrow{d} MN(0, \Sigma)$ . Also, Lemmas

B.14 and B.15 show that  $G_{2n} = O_p(n^{-\frac{3}{2}}h)$  and  $G_{3n} = O_p(n^{-\frac{5}{4}})$  under  $H_0$ . These results lead to  $n^{5/4}h^{1/2}\hat{I}_n = n^{5/4}h^{1/2}I_{1n} + o_p(1) \xrightarrow{d} MN(0, \Sigma)$  relying on Assumption A7. Finally, Lemma B.15 gives that  $\hat{\Sigma} \xrightarrow{p} \Sigma$ , which completes the proof of Theorem 3.2.1 (i) (under  $H_0$ ).

Now, we give a lemma to show the asymptotic distribution of a sample moment useful in subsequent proofs.

Lemma B.10. Under Assumption 2.1 and Assumptions A1-A7, for  $d_n = \frac{\sqrt{n}}{h}$  and  $r \in [0, 1]$ , we have

$$\frac{d_n}{n^2} \sum_{t=1}^{[nr]} \sum_{s=1}^t K_{ts} \Rightarrow \frac{1}{2} L_{V_c}(r, r, 0) \quad (39)$$

as  $n \rightarrow \infty$ . where  $L_{V_c}(r, r, 0)$  is defined by (26).

**Proof of Lemma B.10:** Let

$$L_{n,\epsilon}^{(r)} = \frac{d_n}{n^2} \sum_{t=1}^{[nr]} \sum_{s=1}^t \int_{-1}^1 K[d_n(z_{t,n} - z_{s,n} + x\epsilon)] \phi(x) dx,$$

where  $z_{t,n} = \frac{z_t}{\sqrt{n}}$ ,

$$\phi_\epsilon(x) = \frac{1}{\epsilon\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2\epsilon^2}\right\},$$

and

$$\phi(x) = \phi_1(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\}.$$

Then, for each  $\epsilon > 0$ , we have

$$L_{n,\epsilon}^{(r)} - \left(\int_{-1}^1 K(u) du\right) \frac{d_n}{n^2} \sum_{t=1}^{[nr]} \sum_{s=1}^t \phi_\epsilon(z_{t,n} - z_{s,n}) = o_p(1) \quad (40)$$

uniformly in  $r \in [0, 1]$ ,  $z_{t,n}$  and  $z_{s,n}$  as  $n \rightarrow \infty$  and  $d_n \rightarrow \infty$ . Since  $\int_{-1}^1 K(u) du = 1$ ,

it becomes

$$L_{n,\epsilon}^{(r)} - \frac{d_n}{n^2} \sum_{t=1}^{[nr]} \sum_{s=1}^t \phi_\epsilon(z_{t,n} - z_{s,n}) = o_p(1)$$

The proof of (40) refers to the proof of Lemma B in [Phillips, 2009].

Next, it follows from the continuous mapping theorem that, for  $\forall \epsilon > 0$  and any  $r \in [0, 1]$ ,

$$\frac{d_n}{n^2} \sum_{t=1}^{[nr]} \sum_{s=1}^t \phi_\epsilon(z_{t,n} - z_{s,n}) \xrightarrow{d} \int_0^r \int_0^t \phi_\epsilon(V_c(t) - V_c(s)) ds dt \quad (41)$$

By recalling the definition of local time of a measurable process, as  $n \rightarrow 0$ , we get

$$\int_0^r \int_0^t \phi_\epsilon(V_c(t) - V_c(s)) ds dt = \frac{1}{2} \int_{-\infty}^{\infty} \phi(x) L_{V_c}(r, r, \epsilon x) dx = \frac{1}{2} L_{V_c}(r, r, 0) + o_{a.s.}(1)$$

where  $\{L_{V_c}(r, r, \epsilon x), 0 \leq r \leq 1, s \in \mathcal{R}\}$  satisfies the following equation,

$$\int_0^r \int_0^r \phi_\epsilon(V_c(t) - V_c(s)) ds dt = \int_{-\infty}^{\infty} \phi(x) L_{V_c}(r, r, \epsilon x) dx$$

Then, write

$$L_n^{(r)} = \frac{d_n}{n^2} \sum_{t=2}^{[nr]} \sum_{s=1}^{t-1} K_{ts}$$

Lemma B.10 follows if we prove that

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{0 \leq r \leq 1} E|L_n^{(r)} - L_{n,\epsilon}^{(r)}| = 0 \quad (42)$$

The proof of (42) is similar to proof of Theorem 2.1 in [Wang and Phillips, 2009a].

Lemma B.11. Under Assumptions A1-A7, we obtain  $n^{5/4}h^{1/2}I_{1n} \xrightarrow{d} MN(0, \Sigma)$ , where  $MN(0, \Sigma)$  is a mixed normal with mean zero and conditional variance  $\Sigma$  given in (28).

**Proof of Lemma B.11:** Denote  $Z_{nt} = n^{-3/4}h^{-1/2}u_t \sum_{s=1}^{t-1} u_s K_{ts}$ . It follows that  $n^{5/4}h^{1/2}I_{1n} = 2 \sum_{t=2}^n Z_{nt}$ . Let  $\mathcal{F}_{nt} = \sigma\{\eta_i, u_i : 1 \leq i \leq t \leq n\}$  be the smallest  $\sigma$ -field containing the past history of  $\{\eta_t, u_t\}$  for all  $n$  and  $E_t(Z)$  denote  $E(Z|\mathcal{F}_{nt})$  for short. It is easy to see that  $\{Z_{nt}; \mathcal{F}_{nt}\}$  is a martingale difference process by showing  $E_{t-1}(Z_{nt}) = 0$  given  $E(u_t|Z_t, \mathcal{F}_{n,t-1}) = 0$  for all  $t$ . Therefore, central limit theorem for a martingale difference (Theorem 3.2 of [Hall and Heyde, 1980]) is applied to establish our results. We verify that the two conditions of the central limit theorem



for martingale difference are satisfied.

$$\sum_{t=2}^n E_{t-1} [Z_{nt}^2 I(|Z_{nt}| > \xi_1)] \xrightarrow{p} 0 \quad \text{for all } \xi_1 > 0 \quad (43)$$

and

$$V_n^2 = \sum_{t=2}^n E_{t-1} (Z_{nt}^2) \xrightarrow{p} \Sigma/2, \quad (44)$$

where  $\Sigma$  is given in (28) and  $I(A)$  is the indicator function of event  $A$ . We start by

checking (44). Define  $a_{t-1,s} = E_{t-1}(u_s^2 K_{ts}^2) - E(u_s^2 K_{ts}^2)$ . Then,  $V_n^2$  is decomposed as

$$\begin{aligned} V_n^2 &= \sum_{t=2}^n E_{t-1} (Z_{nt}^2) = n^{-3/2} h^{-1} \sum_{t=2}^n E_{t-1} \left[ \left( u_t \sum_{s=1}^{t-1} u_s K_{ts} \right)^2 \right] \\ &= \sigma_u^2 n^{-3/2} h^{-1} \sum_{t=2}^n \sum_{s_1=1}^{t-1} \sum_{s_2=1}^{t-1} u_{s_1} u_{s_2} E_{t-1}(K_{ts_1} K_{ts_2}) \\ &= \sigma_u^2 n^{-3/2} h^{-1} \sum_{t=2}^n \sum_{s=1}^{t-1} E(u_s^2 K_{ts}^2) + \sigma_u^2 n^{-3/2} h^{-1} \sum_{t=2}^n \sum_{s=1}^{t-1} a_{t-1,s} \\ &\quad + 2\sigma_u^2 n^{-3/2} h^{-1} \sum_{t=3}^n \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{s_1-1} u_{s_1} u_{s_2} E_{t-1}(K_{ts_1} K_{ts_2}) \\ &= B_{1n} + B_{2n} + 2B_{3n}. \end{aligned}$$

The probability limits of  $B_{1n}$ ,  $B_{2n}$  and  $B_{3n}$  are derived respectively with  $B_{1n} =$

$$\sigma_u^4 n^{-2} \sum_{t=2}^n \sum_{s=1}^{t-1} E(K_{ts}^2), \quad B_{2n} \xrightarrow{p} o(1), \quad \text{and } B_{3n} \xrightarrow{p} o(1).$$

by lemma B.10, we have

$$\begin{aligned} B_{1n} &= \sigma_u^2 n^{-3/2} h^{-1} \sum_{t=2}^n \sum_{s=1}^{t-1} E(u_s^2 K_{ts}^2) \\ &= \sigma_u^4 E \left[ n^{-3/2} h^{-1} \sum_{t=2}^n \sum_{s=1}^{t-1} K_{ts}^2 \right] \\ &= \sigma_u^4 \nu_2(K) E[L_{V_c}(r, r, 0)] \end{aligned}$$

as  $n \rightarrow \infty$ . Notice that  $\nu_2(K) = \int K^2(u) du$ .

Next, we consider  $B_{2n}$ . To show that  $B_{2n} = o_p(1)$ , we specify some useful notations.

For any small  $\delta \in (0, 1)$ , set  $N = \lceil 1/\delta \rceil$ ,  $s_k = \lfloor kn/N \rfloor + 1$ ,  $s_k^* = s_{k+1} - 1$ ,  $N_t^* = \lfloor (N-1)(t-1)/n \rfloor$  and  $s_k^{**} = \min\{s_k^*, t-1\}$ . Then ,

$$\begin{aligned} B_{2n} &= \sigma_u^2 n^{-3/2} h^{-1} \sum_{t=2}^n \sum_{s=1}^{t-1} a_{t-1,s} \\ &\leq \left| \sigma_u^2 n^{-3/2} h^{-1} \sum_{t=2}^n \sum_{s=1}^{t-1} a_{t-1,s} \right| \\ &\leq \sigma_u^2 n^{-3/2} h^{-1} \sum_{t=2}^n \sum_{k=0}^{N_t^*} \left| \sum_{s=s_k}^{s_k^{**}} a_{t-1,s} \right| \end{aligned}$$

Also, it's easy to see that  $d_n E|a_{t-1,s}| = O_p(1)$  where  $d_n = \frac{\sqrt{n}}{h}$ . Then,

$$\begin{aligned} E \left[ n^{-3/2} h^{-1} \sum_{t=2}^n \sum_{k=0}^{N_t^*} \left| \sum_{s=s_k}^{s_k^{**}} a_{t-1,s} \right| \right] &\leq n^{-3/2} h^{-1} \sum_{t=2}^n N_n^* \sup_{s+n\delta < t} E \left| \sum_{i=s}^{s+\delta n} a_{t-1,i} \right| \\ &\leq n^{-1} \sum_{t=2}^n \sup_{s+n\delta < t} E \left| \frac{d_n}{\delta n} \sum_{i=s}^{s+\delta n} a_{t-1,i} \right| \\ &= M(\delta n)^{-1/2} = o_p(1) \end{aligned}$$

as  $n \rightarrow \infty$ . This implies that  $B_{2n} = o(1)$ . Apply the same method to  $B_{3n}$ . It follows that  $B_{3n} = o(1)$ .

Finally, we prove that (43) holds. For all  $\xi_2 > 0$ ,

$$\begin{aligned}
& \Pr \left\{ \sum_{t=2}^n E_{t-1} [Z_{nt}^2 I(|Z_{nt}| > \xi_1)] > \xi_2 \right\} \\
&= \Pr \left\{ \sum_{t=2}^n E_{t-1} \left[ Z_{nt}^2 I \left( \frac{|Z_{nt}|}{\xi_1} > 1 \right) \right] > \xi_2 \right\} \\
&\leq \Pr \left\{ \xi_1^{-2} \sum_{t=2}^n E_{t-1} (Z_{nt}^4) > \xi_2 \right\} \\
&\leq \xi_1^{-2} \xi_2^{-1} \sum_{t=2}^n E(Z_{nt}^4),
\end{aligned}$$

where the last inequality follows from Markov inequality. Condition (43) holds if

$\sum_{t=2}^n E(Z_{nt}^4) \rightarrow 0$  as  $n \rightarrow \infty$ . Simple calculations give

$$\begin{aligned}
\sum_{t=2}^n E(Z_{nt}^4) &= n^{-4} \sum_{t=2}^n E(u_t \sum_{s=1}^{t-1} u_s K_{ts})^4 \\
&= \mu_4^2 n^{-4} \sum_{t=2}^n \sum_{s=1}^{t-1} E(K_{ts}^4) + 2\mu_4 \sigma_u^4 n^{-4} \sum_{t=3}^n \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{s_1-1} E(K_{ts_1}^2 K_{ts_2}^2) \\
&= o(1),
\end{aligned}$$

where in the above we have used (A3) and (A5). This completes the proof of the Lemma B.11.

To prove the convergence of  $G_{2n}$  and  $G_{3n}$ , lemma B.12 and lemma B.13 are provided.

Lemma B.12. Let

$$\begin{aligned}
L_{n,\epsilon}^{(r)} &= \frac{c_n^2}{n^2 h} \sum_{t=2}^{[nr]} \sum_{s=1}^{t-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dot{g}[c_n(z_{t,n} + x_1\epsilon)] \dot{g}[c_n(z_{s,n} + x_2\epsilon)] \\
&\quad K[c_n(z_{t,n} - z_{s,n} + x_1\epsilon - x_2\epsilon)] \phi(x_1) \phi(x_2) dx_1 dx_2
\end{aligned}$$

$$M_{n,\epsilon}^{(r)} = \tau \frac{1}{n^2} \sum_{t=2}^{[nr]} \sum_{s=1}^{t-1} \phi_\epsilon(z_{t,n}) \phi_\epsilon(z_{s,n})$$

where  $c_n = \sqrt{n}$ ,  $\tau = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dot{g}(a) \dot{g}(b) K(a-b) da db$ ,  $\phi_\epsilon(z) = \frac{1}{\epsilon \sqrt{2\pi}} \exp\left\{-\frac{z^2}{2\epsilon}\right\}$ , and  $\phi(x) = \phi_1(x)$ . Suppose Assumptions 4.1, (A1)-(A6) hold. Then, for any  $r \in [0, 1]$

and  $\epsilon > 0$ ,

$$L_{n,\epsilon}^{(r)} - M_{n,\epsilon}^{(r)} = o_p(1)$$

**Proof:** The proof refers to Lemma B of [Phillips, 2009] Write

$$\begin{aligned} L_{n,\epsilon}^{(r)} &= \frac{c_n^2}{n^2} \sum_{t=2}^{[nr]} \sum_{s=1}^{t-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dot{g}[c_n(z_{t,n} + x_1\epsilon)] \dot{g}[c_n(z_{s,n} + x_2\epsilon)] \\ &\quad K\left[\frac{c_n}{h}(z_{t,n} - z_{s,n} + x_1\epsilon - x_2\epsilon)\right] \phi(x_1) \phi(x_2) dx_1 dx_2 \\ &= \frac{c_n^2}{n^2 h} \sum_{t=2}^{[nr]} \sum_{s=1}^{t-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dot{g}(c_n a) \dot{g}(c_n b) K[d_n(a-b)] \phi_\epsilon(a - z_{t,n}) \phi_\epsilon(b - z_{s,n}) da db \\ &= \frac{c_n^2}{2n^2 h} \sum_{t=1}^{[nr]} \sum_{s=1}^{[nr]} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dot{g}(c_n a) \dot{g}(c_n b) K[d_n(a-b)] \phi_\epsilon(a - z_{t,n}) \phi_\epsilon(b - z_{s,n}) da db \\ &\quad + s.o. \end{aligned}$$

Then, similar to the proof of Lemma B in [Phillips, 2009], it is readily seen that as  $n \rightarrow \infty$ ,

$$\sup_r |L_{n,\epsilon}^{(r)} - M_{n,\epsilon}^{(r)}| \rightarrow 0.$$

Lemma B.12 follows.

Lemma B.13. Let  $L_{V_c}(r, s)$  be a continuous local time process for measurable process  $V_c(t)$  satisfying the following equation,

$$\int_0^r \phi_\epsilon(V_C(t)) dt = \int_{-\infty}^{\infty} \phi_\epsilon(s) L_{V_c}(r, s) ds \quad (45)$$

Suppose Assumptions 4.1, (A1)-(A6) hold. Then, for  $c_n = \sqrt{n}$  and  $r \in [0, 1]$ ,

$$\frac{c_n^2}{n^2 h} \sum_{t=2}^{[nr]} \sum_{s=1}^{t-1} \dot{g}(c_n z_{t,n}) \dot{g}(c_n z_{s,n}) K(c_n(z_{t,n} - z_{s,n})) \xrightarrow{d} \frac{1}{2} \tau L_{V_c}^2(r, 0)$$

**Proof:** The proof refers to Theorem 2.1 of [Wang and Phillips, 2009a]. Write

$$\begin{aligned} L_n^{(r)} &= \frac{c_n^2}{n^2 h} \sum_{t=2}^{[nr]} \sum_{s=1}^{t-1} \dot{g}(c_n z_{t,n}) \dot{g}(c_n z_{s,n}) K\left(\frac{c_n}{h}(z_{t,n} - z_{s,n})\right) \\ L_{n,\epsilon}^{(r)} &= \frac{c_n^2}{n^2 h} \sum_{t=2}^{[nr]} \sum_{s=1}^{t-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dot{g}(c_n(z_{t,n} + x_1\epsilon)) \dot{g}(c_n(z_{s,n} + x_2\epsilon)) K[c_n(z_{t,n} - z_{s,n} \\ &\quad + x_1\epsilon - x_2\epsilon)] \phi(x_1) \phi(x_2) dx_1 dx_2 \end{aligned}$$

where  $\phi(x) = \phi_1(x)$  with  $\phi_\epsilon(x) = (1/\epsilon\sqrt{2\pi} \exp\{-x^2/2\epsilon^2\})$ .

Then, by lemma B.12, We have

$$L_{n,\epsilon}^{(r)} - \frac{\tau}{n^2} \sum_{t=2}^{[nr]} \sum_{s=1}^{t-1} \phi_\epsilon(z_{t,n}) \phi_\epsilon(z_{s,n}) = o_p(1)$$

uniformly in  $r \in [0, 1]$ . Next, we just need to show

$$\lim_{\epsilon \rightarrow 0} \lim_{n \rightarrow \infty} \sup_{0 \leq r \leq 1} E|L_n^{(r)} - L_{n,\epsilon}^{(r)}| = 0 \quad (46)$$

It follows from the continuous mapping theorem that, for  $\forall \epsilon > 0$  and  $\forall r \in [0, 1]$ ,

$$\begin{aligned} &\frac{1}{n^2} \sum_{t=2}^{[nr]} \sum_{s=1}^{t-1} \phi_\epsilon(z_{t,n}) \phi_\epsilon(z_{s,n}) \\ &= \frac{1}{2} \int_0^r \int_0^r \phi_\epsilon(z_{[tn],n}) \phi_\epsilon(z_{[sn],n}) ds dt + s.o. \\ &\xrightarrow{d} \frac{1}{2} L_{V_c}^2(r, 0) \end{aligned}$$

Then, we prove (46). Write  $Y_{t,s,n} = \dot{g}[c_n z_{t,n}] \dot{g}[c_n z_{s,n}] K[c_n(z_{t,n} - z_{s,n})] - \dot{g}[c_n(z_{t,n} + x_1\epsilon)] \dot{g}[c_n(z_{s,n} + x_2\epsilon)] K[c_n(z_{t,n} - z_{s,n} + x_1\epsilon - x_2\epsilon)]$ . Next, it's easy to see that

$$\sup_{0 \leq r \leq 1} E|L_n - L_{n,\epsilon}| \leq \frac{1}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{c_n^2}{n^2 h} \sup_{0 \leq r \leq 1} E \left| \sum_{t=1}^{[nr]} \sum_{s=1}^{[nr]} Y_{t,s,n}(x_1, x_2) \right| \phi(x_1) \phi(x_2) dx_1 dx_2 \quad (47)$$

Because  $z_{t,n}/d_{t,0,n}$  has a density  $h_{t,0,n}(x)$  that is bounded by a constant and the kernel function  $K(\cdot)$  is also bounded, we have

$$\begin{aligned}
\frac{c_n^2}{h} E|Y_{t,s,n}| &\leq \frac{Ac_n^2}{2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\dot{g}[c_n(d_{t,0,n} + x_1\epsilon)]\dot{g}[c_n(d_{s,0,n} + x_2\epsilon)] - \dot{g}[c_n d_{t,0,n} z_1]\dot{g}[c_n d_{s,0,n} z_2]| \\
&\quad h_{t,0,n}(z_1)h_{s,0,n}(z_2) dz_1 dz_2 \\
&\leq \frac{A}{2d_{t,0,n}d_{s,0,n}} \int_{-\infty}^{\infty} |\dot{g}(z_1 + c_n x_1\epsilon) - \dot{g}(z_1)| dz_1 \int_{-\infty}^{\infty} |\dot{g}(z_2 + c_n x_2\epsilon) - \dot{g}(z_2)| dz_2 \\
&\leq A \left[ \int_{-\infty}^{\infty} |\dot{g}(z)| dz / d_{t,0,n} \right]^2
\end{aligned}$$

Then, it follows that

$$\frac{c_n^2}{2n^2 h} \sup_{0 \leq r \leq 1} E \left| \sum_{t=1}^{[nr]} \sum_{s=1}^{[nr]} Y_{t,s,n}(x_1, x_2) \right| \leq A_1 \frac{1}{n^2} \sum_{t=1}^n \sum_{s=1}^n \frac{1}{d_{t,0,n} d_{s,0,n}} < \infty \quad (48)$$

This, together with (47) and the dominated convergence theorem, implies that, to prove (46), it suffices to show that, for fixed  $x_1$  and  $x_2$ ,

$$\Lambda_n(\epsilon) = \frac{c_n^2}{n^2 h} \sup_{0 \leq r \leq 1} E \left[ \sum_{t=1}^{[nr]} \sum_{s=1}^{[nr]} Y_{t,s,n}(x_1, x_2) \right]^2 \rightarrow 0 \quad (49)$$

Refer to Proof of Theorem 2.1 in [Wang and Phillips, 2009a], we can see that (49)

is true. Now, the result is stated.

Lemma B.14. Under Assumptions given in Theorem 3.2.1, under  $H_0$ , we obtain  $G_{2n} = O_p(n^{-\frac{3}{2}})$  and  $G_{3n} = O_p(n^{-\frac{3}{2}})$ , where  $G_{2n}$  and  $G_{3n}$  are defined in (37) and (38), respectively.

**Proof:** By Taylor expansion,  $g(z_t, \hat{\theta})$  is written as

$$g(z_t, \hat{\theta}) = g(z_t, \theta) + \dot{g}(z_t, \theta)(\hat{\theta}_n - \theta) + s.o.$$

Also, note that the convergence rate of  $\hat{\theta}_n$  is  $n^{1/4}$  according to Theorem 2.3.1. Then,

lemma B.12 and lemma B.13 are applied to get the following result

$$\frac{c_n^2}{n^2 h} \sum_{t=2}^n \sum_{s=1}^{t-1} \dot{g}(z_t, \theta) \dot{g}(z_s, \theta) K \left( \frac{z_t - z_s}{h} \right) \xrightarrow{d} \frac{1}{2} \tau \mathbf{L}_{V_c}^2(1, 0)$$

Hence, we have

$$\begin{aligned}
G_{2n} &\sim \frac{1}{n^2 h} \sum_{t=1}^n \sum_{s \neq t}^n (g(z_t, \hat{\theta}_n) - g(z_t, \theta)) (g(z_s, \hat{\theta}_n) - g(z_s, \theta)) K_{ts} \\
&= \frac{2}{n} \frac{c_n^2}{n^2 h} \sum_{t=2}^n \sum_{s=1}^{t-1} \dot{g}(z_t, \theta) (\hat{\theta}_n - \theta) \dot{g}(z_s, \theta) (\hat{\theta}_n - \theta) K_{ts} + s.o. \\
&= \frac{1}{n} (\hat{\theta}_n - \theta)^2 \frac{c_n^2}{n^2 h} \sum_{t=2}^n \sum_{s=1}^{t-1} \dot{g}(z_t, \theta) \dot{g}(z_s, \theta) K_{ts} + s.o. \\
&= O_p(n^{-3/2})
\end{aligned}$$

Next, let

$$G_{3n} \sim \frac{1}{n^2 h} \sum_{t=1}^n \sum_{s \neq t}^n u_t (g(z_s, \hat{\theta}_n) - g(z_s, \theta)) K_{ts} \equiv A_n$$

In a similar way to dealing with  $G_{2n}$ ,

$$\begin{aligned}
A_n^2 &= \frac{4\sigma_u^2}{n^4 h^2} \sum_{t=3}^n \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{s_1-1} \left( g(z_{s_1}, \hat{\theta}_n) - g(z_{s_1}, \theta) \right) \left( g(z_{s_2}, \hat{\theta}_n) - g(z_{s_2}, \theta) \right) K_{ts_1} K_{ts_2} \\
&= \frac{\sigma_u^2}{n^{5/2}} (\hat{\theta}_n - \theta)^2 \frac{c_n^3}{n^3 h^2} \sum_{t=3}^n \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{s_1-1} \dot{g}(z_{s_1}, \theta) \dot{g}(z_{s_2}, \theta) K_{ts_1} K_{ts_2} + s.o. \\
&= O_p(n^{-3})
\end{aligned}$$

So,  $G_{3n} = O_p(n^{-3/2})$ . This completes the proof of Lemma B.14.

Lemma B.15. Under Assumptions given in Theorem 3.2.1, we obtain

$$\widehat{\Sigma} = \frac{1}{n^{\frac{3}{2}} h} \sum_{t=1}^n \sum_{s \neq t}^n \tilde{u}_s^2 \tilde{u}_t^2 K_{ts}^2 \xrightarrow{p} \Sigma,$$

where  $\tilde{u}_t = Y_t - \hat{f}^{(-t)}(Z_t)$  is the nonparametric residual and  $\Sigma$  is defined in (28).

**Proof:** Note that  $\tilde{u}_t = Y_t - \hat{f}^{(-t)}(Z_t) = u_t - [\hat{f}^{(-t)}(Z_t) - f(Z_t)]$ . By Assumption A8

we know that we can replace  $\tilde{u}_t$  by  $u_t$  to obtain the leading term of  $\widehat{\Sigma}$ . Following the

proof in Lemma B.11, we obtain

$$\widehat{\Sigma} = \frac{1}{n^{\frac{3}{2}}h} \sum_{t=1}^n \sum_{s \neq t} \tilde{u}_s^2 \tilde{u}_t^2 K_{ts}^2 + o_p(1) = \frac{1}{n^{\frac{3}{2}}h} \sigma_u^4 \sum_{t=1}^n \sum_{s \neq t} E(K_{ts}^2) + o_p(1) \xrightarrow{p} \Sigma.$$

**Remark:** Here we emphasize that it is important to use the nonparametric residual in computing  $\widehat{\Sigma}$ . If the nonparametric residual  $\tilde{u}_t$  is replaced by the parametric residual  $\hat{u}_t = Y_t - g(Z_t, \hat{\theta}) = u_t - [g(Z_t, \hat{\theta}) - f(Z_t)]$ , then under  $H_1$ ,  $\hat{u}_t = u_t + O_p(1)$ , and Lemma B.15 does not hold and the resulting test may have only trivial power even as  $n \rightarrow \infty$ .

**Proof of Theorem 3.2.1 (ii):** Under  $H_1$ ,  $f_n(z_t) = g(z_t, \theta_0) + n^{-\gamma}G(z_t) + u_t$  and  $I_{1n}$  is the same as that defined under  $H_0$ . Hence,  $I_{1n} = O_p(n^{-\frac{5}{4}}h^{-\frac{1}{2}})$ .

Now, we consider  $G_{2n}$ .

$$\begin{aligned} G_{2n} &\sim \frac{1}{n^2h} \sum_{t=1}^n \sum_{s \neq t} \left[ g(z_t, \hat{\theta}_n) - f_n(z_t) \right] \left[ g(z_s, \hat{\theta}_n) - f_n(z_s) \right] K_{ts} \\ &= \frac{1}{n^{1+2\gamma}} \frac{d_n}{n^2} \sum_{t=1}^n \sum_{s \neq t} G(z_t), G(z_s) K_{ts} + s.o. \\ &= O_p(n^{-(1+2\gamma)}) \end{aligned}$$

Finally, we deal with  $G_{3n}$  in a similar way as  $G_{2n}$ ,

$$G_{3n} \sim \frac{1}{n^2h} \sum_{t=1}^n \sum_{s \neq t} u_t (g(z_s, \hat{\theta}_n) - f_n(z_s)) K_{ts} \equiv B_n$$



$$\begin{aligned}
B_n^2 &= \frac{2\sigma_u^2}{n^4 h^2} \sum_{t=3}^n \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{t-1} \left( g(z_{s_1}, \hat{\theta}_n) - f(z_{s_1}) \right) \left( g(z_{s_2}, \hat{\theta}_n) - f(z_{s_2}) \right) K_{ts_1} K_{ts_2} \\
&= \frac{\sigma_u^2}{n^{2+2\gamma}} \frac{d_n^2}{n^3} \sum_{t=3}^n \sum_{s_1=2}^{t-1} \sum_{s_2=1}^{t-1} G(z_{s_1}) G(z_{s_2}) K_{ts_1} K_{ts_2} + s.o. \\
&= O_p(n^{-(2+2\gamma)})
\end{aligned}$$

Therefore,  $G_{2n} = O_p(n^{-(1+2\gamma)})$  and  $G_{3n} = O_p(n^{-(1+\gamma)})$  under  $H_1$ . Since  $\gamma > 0$ ,  $G_{2n}$  is the leading term. Then, the test has power if

$$n^{\frac{5}{4}} h^{\frac{1}{2}} O_p(n^{-(1+2\gamma)}) \geq O_p(1)$$

is satisfied.

Suppose bandwidth  $h = an^{-\delta}$ , where  $a$  and  $\delta$  are constant, we get  $\gamma \leq \frac{1}{8} - \frac{\delta}{4}$  by solving inequality B. If the rate for  $h$  is set to be  $n^{-\frac{1}{10}}$ , the optimal rate for bandwidth in nonparametric nonstationary regression,  $\gamma \leq \frac{1}{10}$  is required for the test to have power.

This concludes the proof of Theorem 3.2.1 (ii).

## REFERENCES

- [Billingsley, 2009] Billingsley, P. (2009). *Convergence of probability measures*, volume 493. Wiley-Interscience.
- [Cai et al., 2009] Cai, Z., Li, Q., and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *Journal of Econometrics*, 148(2):101–113.
- [Campbell and Yogo, 2006] Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of financial economics*, 81(1):27–60.
- [Cavanagh et al., ] Cavanagh, C. L., Elliott, G., and James, H. Stock, 1995, inference in models with nearly integrated regressors. *Econometric theory*, 11(5):1131–1147.
- [Choi and Saikkonen, 2004] Choi, I. and Saikkonen, P. (2004). Testing linearity in cointegrating smooth transition regressions. *The Econometrics Journal*, 7(2):341–365.
- [Elliott and Stock, 1994] Elliott, G. and Stock, J. (1994). Inference in models with nearly integrated regressors. *Econometric theory*, 11:1131–1147.
- [Gao et al., 2009] Gao, J., King, M., Lu, Z., and Tjøstheim, D. (2009). Specification testing in nonlinear and nonstationary time series autoregression. *The Annals of Statistics*, 37(6B):3893–3928.
- [Geman and Horowitz, 1980] Geman, D. and Horowitz, J. (1980). Occupation densities. *The Annals of Probability*, 8(1):1–67.
- [Granger, 1995] Granger, C. W. (1995). Modelling nonlinear relationships between extended-memory variables. *Econometrica: Journal of the Econometric Society*, pages 265–279.
- [Gregory Mankiw and Shapiro, 1986] Gregory Mankiw, N. and Shapiro, M. D. (1986). Do we reject too often?: Small sample properties of tests of rational expectations models. *Economics Letters*, 20(2):139–145.
- [Hall and Heyde, 1980] Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application*. Academic press New York.
- [Hansen, 1992] Hansen, B. E. (1992). Convergence to stochastic integrals for dependent heterogeneous processes. *Econometric Theory*, 8:489–489.
- [Hong and Phillips, 2010] Hong, S. H. and Phillips, P. C. (2010). Testing linearity in cointegrating relations with an application to purchasing power parity. *Journal of Business & Economic Statistics*, 28(1):96–114.
- [Jennrich, 1969] Jennrich, R. I. (1969). Asymptotic properties of non-linear least squares estimators. *The Annals of Mathematical Statistics*, pages 633–643.

- [Kasparis, 2008] Kasparis, I. (2008). Detection of functional form misspecification in cointegrating relations. *Econometric Theory*, 24(5):1373.
- [Kasparis, 2010] Kasparis, I. (2010). Functional form misspecification in regressions with a unit root. *Econometric Theory*, 27(2):285.
- [Kasparis and Phillips, 2012] Kasparis, I. and Phillips, P. C. (2012). Dynamic misspecification in nonparametric cointegrating regression. *Journal of Econometrics*.
- [Lanne, 2002] Lanne, M. (2002). Testing the predictability of stock returns. *Review of Economics and Statistics*, 84(3):407–415.
- [Lewellen, 2004] Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics*, 74(2):209–235.
- [Li et al., 2002] Li, Q., Huang, C. J., Li, D., and Fu, T.-T. (2002). Semiparametric smooth coefficient models. *Journal of Business & Economic Statistics*, 20(3):412–422.
- [Park and Phillips, 2008] Park, J. Y. and Phillips, P. C. (2008). Nonlinear regressions with integrated time series. *Econometrica*, 69(1):117–161.
- [Phillips, 1986] Phillips, P. C. (1986). Understanding spurious regressions in econometrics. *Journal of econometrics*, 33(3):311–340.
- [Phillips, 1987] Phillips, P. C. (1987). Time series regression with a unit root. *Econometrica: Journal of the Econometric Society*, pages 277–301.
- [Phillips, 2009] Phillips, P. C. (2009). Local limit theory and spurious nonparametric regression. *Econometric Theory*, 25(6):1466.
- [Phillips and Durlauf, 1986] Phillips, P. C. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53(4):473–495.
- [Revuz and Yor, 2004] Revuz, D. and Yor, M. (2004). *Continuous martingales and Brownian motion*, volume 293. Springer.
- [Richardson and Stock, 1989] Richardson, M. and Stock, J. H. (1989). Drawing inferences from statistics based on multiyear asset returns. *Journal of Financial Economics*, 25(2):323–348.
- [Rio, 1995] Rio, E. (1995). The functional law of the iterated logarithm for stationary strongly mixing sequences. *The Annals of Probability*, 23(3):1188–1203.
- [Schiller, 2000] Schiller, R. J. (2000). The irrational exuberance.
- [Shao and Lu, 1987] Shao, Q. and Lu, C. (1987). Strong approximation for partial sums of weakly dependent random variables. *Scientia Sinica*, A(15):576–587.

- [Stambaugh, 1999] Stambaugh, R. F. (1999). Predictive regressions. *Journal of Financial Economics*, 54(3):375–421.
- [Sun et al., 2011] Sun, Y., Cai, Z., and Li, Q. (2011). Consistent nonparametric test on parametric smooth coefficient models with integrated time series.
- [Torous et al., 2004] Torous, W., Valkanov, R., and Yan, S. (2004). On predicting stock returns with nearly integrated explanatory variables\*. *The Journal of Business*, 77(4):937–966.
- [Wang and Phillips, 2011] Wang, Q. and Phillips, P. (2011). Specification testing for nonlinear cointegrating regression.
- [Wang and Phillips, 2009a] Wang, Q. and Phillips, P. C. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25(3):710.
- [Wang and Phillips, 2009b] Wang, Q. and Phillips, P. C. (2009b). Structural nonparametric cointegrating regression. *Econometrica*, 77(6):1901–1948.
- [Xiao, 2009] Xiao, Z. (2009). Functional-coefficient cointegration models. *Journal of Econometrics*, 152(2):81–92.