TESTING PREDICTABILITY OF ASSET RETURNS

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

2014

Approved by:

____

Dr. Jiancheng Jiang

____

Dr. Zongwu Cai

____

Dr. Weihua Zhou

____

Dr. Claudia Flowers

ABSTRACT

LI WU. Testing predictability of asset returns. (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

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Nonlinear cointegration models are important in a wide range of applications in economics (e.g. [1]). 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 [2] and [3]. The extension of the existing limit theorem to near-integrated process is straightforward. The major works in [4, 5] 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 [6], 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:

$$y_t = f(z_t) + u_t \tag{1}$$

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:

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

where Θ is the parameter set. The contiguous alternatives are written as follows,

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

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 [7] , 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 [8], [9], and [10]. 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 [2], 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 [2],

$$
\begin{aligned}
y_t &= f(z_t, \theta_0) + u_t \\
z_t &= z_{t-1} + v_t
\end{aligned}
$$

where $f : R \times R^m \to R$ is known, regressor z_t an integrated process, regression error u_t a martingale difference sequence, and θ_0 an m-dimensional true parameter vector.

They estimated θ_0 by nonlinear least squares (NLS). That is to choose $\hat{\theta}_n$ by minimizing the function below,

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

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,

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

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 [5, 4] 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 $h \to 0$ as $n \to \infty$.

Imposing some assumptions, it's proved in [4] that the limit behavior of $\hat{f}(x)$ is

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

when $nh^2 \to \infty$ and $h \to 0$. In addition, if h satisfies that $nh^2 \to \infty$ and $nh^{2(1+2\gamma)} \to 0$ as $n \to \infty$, 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 [4] 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 $nh^2 \to \infty$ and $h \to 0$ as $n \to \infty$, 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 [7], that tested parameter stability. Recently, a modified RESET test was introduced by [11] 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 [12]. 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 [12], a similar problem was investigated in [13]. 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 considerred by [14]. The smooth transition regression model developed in [14] is: $y_t = x_t'\alpha + \beta' x_t g(x_{ts} - c) + u_t$, where x_t is a p dimentional 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 [6]. That model was first learned by [15] and [16]:

$$y_t = X_t'\theta(z_t) + u_t, \tag{4}$$

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 cointegraion model instead of a linear one. Compared with the varying coefficient model investigated by [6], our model could be taken as a varying coefficient model with one diminutional $X_t = 1$, 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 aymptotic 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 [17]. (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) $A \overset{def}{=} B$ 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 σ-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 regresstion model for y_t under H_0

$$y_t = g(z_t, \theta_0) + u_t \tag{5}$$

where $g : R \times R \to R$ is known and θ_0 is the true parameter that lies in the parameter set Θ. This section concentrates on nonlinear least square estimation of (5). Let

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

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

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

It is assumed throughout the paper that $\hat{\theta}_n$ exists and is unique for all n, and θ_0 is an interior point of Θ, where Θ 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

$$z_t = \rho z_{t-1} + \eta_t, \tag{8}$$

and initializing it with $z_0 = 0$ to avoid unnecessary complication in our development of limit theory as in [2]. Then, define the stochastic processes U_n and V_n respectively by

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

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 $\sigma_\eta = \lim_{n\to\infty} \mathrm{Var}(n^{-1/2}\sum_{t=1}^{n}\eta_t)$. (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

$sup_{1\le t\le n} E(|u_t|^q|\mathcal{F}_{n,t-1}) < \infty$ a.s. for some $q > 2$.

Assumption 2.1 is routinely imposed on NLS regression with nonstationary processes as in [2]. 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

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

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 $\lim_{\lambda\to\infty}\lambda^r c(\lambda) = 0$ 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 proses 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,

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

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 $[0,1]$. More discussions and applications of local time are provided by [18], [19], [2] and [20].

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 [2].

Definition 2.1: A function F is said to be *I-regular* on a compact set Π if

(a) for each $\pi_0 \in \Pi$, there exists a neighborhood N_0 of π_0 and a bounded integrable function $T : \mathcal{R} \to \mathcal{R}$ 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 Π 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 π.

Theorem 2.1.1. Suppose Assumption 2.2 holds. If F is I-regular on a compact set Π, 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 $n \to \infty$. 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 $n \to \infty$.

The sample mean and sample covariance asymptotics are exactly like those in [2]. 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 [2]. 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 ν_n, $\nu_n^{-1} D_n(\theta,\theta_0) \xrightarrow{p} D(\theta,\theta_0)$ uniformly in θ, where $D(\cdot,\theta_0)$ is continuous and has unique minimum θ_0 a.s.

The above condition is sufficient to guarantee that $\hat{\theta}_n \xrightarrow{p} \theta_0$, referring to the work by [21].

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 θ_0 being I-regular on Π. 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 $n \to \infty$.

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

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,

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

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

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

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

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

where θ_n lies in between $\hat{\theta}_n$ and θ_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 ν_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

$$\begin{aligned} \nu_n(\hat{\theta}_n - \theta_0) &= -(\nu_n^{-2}\ddot{Q}_n(\theta_n))^{-1}\nu_n^{-1}\dot{Q}_n(\theta_0) && (12) \\ &= -(\nu_n^{-2}\ddot{Q}_n^0(\theta_0))^{-1}\nu_n^{-1}\dot{Q}_n(\theta_0) + o_p(1) && (13) \\ &\xrightarrow{d} -\ddot{Q}(\theta_0)^{-1}\dot{Q}(\theta_0) && (14) \\ & && (15) \end{aligned}$$

as $n \to \infty$.

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 $n \to \infty$.

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

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

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 $n \to \infty$.

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

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

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 [2].

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 [6].

3.1 Construction of Test Statistics

We construct a L^2 -type test statistic as in [6],

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

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,

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

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

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

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 [6], where $K(\cdot)$ is a kernel function.

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

$\hat{I}_n$ is a second-order U-statistic similar to the test statistic proposed in [22] and [6]. Model (4) was studied by [22] 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 $E\left\{[E(X_t u_t | z_t)^2] f(z_t)\right\} \geq 0$ 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 [22] was changed in [6] by assuming X_t to be an I(1) process. Law of large numbers applied by [22] is not applicable when non stationary variables are included. Therefore, Martingale Central Limit Theory was adopted by [6] 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 δ_*) with $0 < \delta_* \le 2$

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

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

(ii) Furthermore,

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

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 [23] establishes a sufficient condition for Assumption 4.1 to hold. It states that, for a stationary β-mixing sequence $\{\eta_t\}$ satisfying, for some $\gamma_* > 2+\delta_*$,

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

where β_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 [24], [25], and [5].

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 [26]), (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.

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

where $0 < \eta < 1$, following [5].

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

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

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

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

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

$$\limsup_{n\to\infty}\frac{1}{n}\max_{0\le k\le n-1}\sum_{l=(k+1)}^{n}(d_{l,k,n})^{-1} < \infty, \tag{23}$$

(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

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{(l,k)\in\Omega_n[\delta^{1/(2m_0)}]}\sup_{|u|\le\delta}|h_{l,k,n}(u)-h_{l,k,n}(0)|=0. \tag{24}$$

(A3) $\{u_t\}$ is an i.i.d. sequence and is independent of $\{Z_t\}$. Also, $E(u_t)=0$, $E(u_t^2)=\sigma_u^2<\infty$ 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 $\nu_2(K)=\int K^2(u)\,du$, $\sup_u K(u)<\infty$ and $\sup_u K'(u)<\infty$.

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

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

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

(A8) $\int\int G(z_t)G(z_s)K_{ts}dz_tdz_s\ne 0$ and $\int\int\int G(z_{s1})G(z_{s2})K_{ts_1}K_{ts_2}dz_tdz_{s_1}dz_{s_2}\ne 0$.

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

$E\left(u_t^4|z_t, \mathcal{F}_{n,t-1}\right) < \infty$ for all t, which requires a lengther 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 mearurable process $L_{V_c}(r,r,0)$ as the local time of measurable process $\{V_c(t)-V_c(s), t \geq 0, s \geq 0\}$,

$$\int_0^r \int_0^r T[(V_c(t) - V_c(s))]dsdt = \int_{-\infty}^{\infty} T(x)L_{V_c}(r,r,0)dx, \text{ all } r \in \mathcal{R} \tag{25}$$

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-A8, we obtain (i) under H_0,

$$J_n = n^{\frac{5}{4}} h^{\frac{1}{2}} \widehat{I}_n \xrightarrow{d} MN(0, \Sigma), \tag{26}$$

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

$$\Sigma = \frac{1}{2}\sigma_u^4 \mu_2(K) E[L_{V_c}(r,r,0)] \tag{27}$$

In addition, if Assumption A7 also holds, a consistent estimator of Σ is given by

$$\widehat{\Sigma} = \frac{2}{n^{\frac{3}{2}}h} \sum_{t=2}^{n} \sum_{s=1}^{t-1} \tilde{u}_t^2 \tilde{u}_s^2 K_{ts}^2 \xrightarrow{p} \Sigma \tag{28}$$

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

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

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

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_α 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

$$J_n = n^{\frac{5}{4}} h^{\frac{1}{2}} \hat{I}_n \tag{29}$$

where

$$\hat{I}_n = \frac{1}{n^2 h} \sum_{t=1}^{n} \sum_{s=1}^{n} \hat{u}_t \hat{u}_s K_{ts} \tag{30}$$

as proposed in Chapter 3.

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

$$y_t = \theta z_t^2 + u_t \tag{31}$$

$$z_t = \rho z_t + \eta_t = \left(1 - \frac{c}{n}\right) z_{t-1} + \eta_t \tag{32}$$

where u_t is an i.i.d random variable satisfying $N(0, \sigma_u^2)$, η_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 (31), 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:

$$\text{DGP}_1 \; : \; y_t = \theta z_t^2 + a_1 n^{-1/10} z_t + u_t,$$

$$\text{DGP}_2 \; : \; y_t = \theta z_t^2 + a_2 n^{-1/10} z_t^3 + u_t$$

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 $n = 600$. Gussian 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: Estimated sizes : varying integration parameters

	$c=0$			$c=2$			$c=20$		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	.018	.050	.109	.008	.036	.100	.013	.054	.107
200	.015	.060	.107	.013	.055	.106	.010	.043	.087
400	.015	.050	.102	.024	.061	.106	.013	.056	.100
600	.012	.057	.097	.010	.055	.101	.021	.070	.114

Table 4: Estimated powers: varying integration parameters

	$c=0$			$c=2$			$c=20$		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	.902	.953	.973	.832	.900	.952	.579	.774	.849
200	.997	1	1	.997	1	1	.975	.994	.996
400	1	1	1	1	1	1	1	1	1
600	1	1	1	1	1	1	1	1	1

Table 5: Estimated powers: varying models

	DGP_1			DGP_2		
n	1%	5%	10%	1%	5%	10%
100	.856	.919	.952	.829	.907	.944
200	.998	1	1	.997	1	1
400	1	1	1	1	1	1
600	1	1	1	1	1	1

Table 6: Estimated powers: varying a_1

	$a_1=0.25$			$a_1=0.5$			$a_1=1$		
n	1%	5%	10%	1%	5%	10%	1%	5%	10%
100	.328	.478	.561	.856	.919	.952	.999	1	1
200	.754	.860	.904	.998	1	1	1	1	1
400	.995	.998	.999	1	1	1	1	1	1
600	1	1	1	1	1	1	1	1	1

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 [27]; [28]; and [29]), 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 [30] and [31], [32] shows that returns are predictable at short horizons but not at long horizons. No evidence for predictability of stock return was found in [33] by testing the stationarity of long-horizon returns, while predictability with some ratios was verified in [34].

A unifying understanding of various test procedures mentioned above refers to [35]. 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 [35] based on the theory of UMP test.

The test procedure proposed by this paper differs from that in [35] 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

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

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

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

The work of [35] 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

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

and

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

It's clear to see that the rejection of null hypothesis indicates no linear predictability of asset return in our test procedure, while in [35], 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 [35]. 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 [36] , 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

$$r_t = \alpha + \beta x_{t-1} + u_t, \tag{33}$$

$$x_t = \gamma + \rho x_{t-1} + e_t, \tag{34}$$

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 ρ by least square method and construct the confidence intervals toward log dividend-price ratio and log earning-

Table 7: Estimated autoregression parameter

	ρ	95% CI for ρ
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.

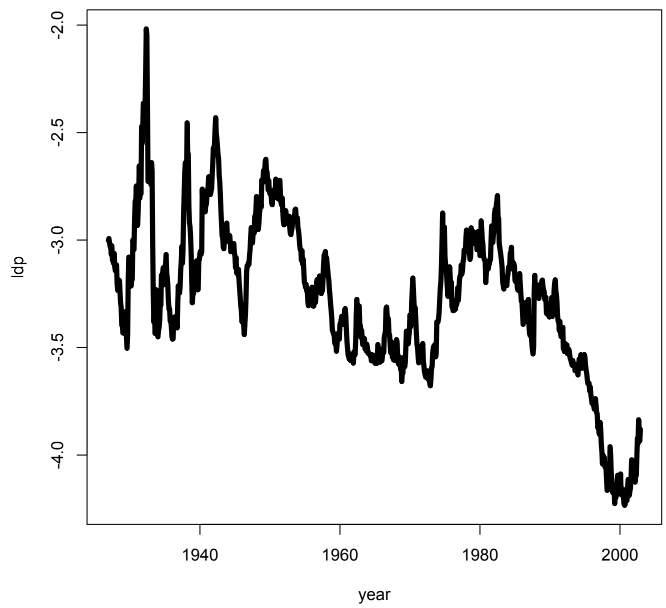

Figure 1: Time series plot for log dividend-price ratio.

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 (29) in Chapter 4.

To get the estimated critical values, we perform nonparametric wild bootstrap to do residual resampling. The procedure is described as below,

1. Generate bootstrap residuals u^* from multiplying nonparametric residuals $\tilde{u}$ by

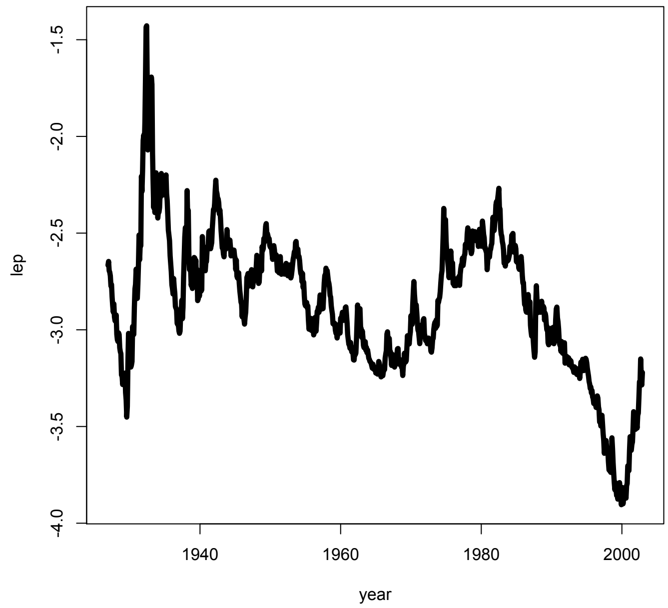

Figure 2: Time series plot for log earning-price ratio.

standard normal random variable ϵ.

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 calculate the P-value for J_n. The result of our empirical study is provided in Table 8

Table 8: Test statistics and p-values

	J_n	p-value
ldp	-.0259	0.145
lep	-.0277	0.35

P-values of test statistic J_n based on log dividend-price ratio and log earning-price ratio are both greater than 10% as reported in Table 8. Therefore, our nonparametric test fail to reject the linear prediction model for stock return if the significance level is no more than 10%. 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 [6], 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 [6]. 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 increses. Table 2 shows that the test isn't sensitive to the selection of smoothing parameter. But it is noticably 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 lieanr 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 [35]). 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.

REFERENCES

[1] C. W. Granger, "Modelling nonlinear relationships between extended-memory variables," *Econometrica: Journal of the Econometric Society*, pp. 265–279, 1995.

[2] J. Y. Park and P. C. Phillips, "Nonlinear regressions with integrated time series," *Econometrica*, vol. 69, no. 1, pp. 117–161, 2008.

[3] I. Kasparis, "Functional form misspecification in regressions with a unit root," *Econometric Theory*, vol. 27, no. 2, p. 285, 2010.

[4] Q. Wang and P. C. Phillips, "Structural nonparametric cointegrating regression," *Econometrica*, vol. 77, no. 6, pp. 1901–1948, 2009.

[5] Q. Wang and P. C. Phillips, "Asymptotic theory for local time density estimation and nonparametric cointegrating regression," *Econometric Theory*, vol. 25, no. 3, p. 710, 2009.

[6] Y. Sun, Z. Cai, and Q. Li, "Consistent nonparametric test on parametric smooth coefficient models with integrated time series," 2011.

[7] B. E. Hansen, "Convergence to stochastic integrals for dependent heterogeneous processes," *Econometric Theory*, vol. 8, pp. 489–489, 1992.

[8] P. C. Phillips, "Understanding spurious regressions in econometrics," *Journal of econometrics*, vol. 33, no. 3, pp. 311–340, 1986.

[9] P. C. Phillips, "Time series regression with a unit root," *Econometrica: Journal of the Econometric Society*, pp. 277–301, 1987.

[10] P. C. Phillips and S. N. Durlauf, "Multiple time series regression with integrated processes," *The Review of Economic Studies*, vol. 53, no. 4, pp. 473–495, 1986.

[11] S. H. Hong and P. C. Phillips, "Testing linearity in cointegrating relations with an application to purchasing power parity," *Journal of Business & Economic Statistics*, vol. 28, no. 1, pp. 96–114, 2010.

[12] J. Gao, M. King, Z. Lu, and D. Tjøstheim, "Specification testing in nonlinear and nonstationary time series autoregression," *The Annals of Statistics*, vol. 37, no. 6B, pp. 3893–3928, 2009.

[13] Q. Wang and P. Phillips, "Specification testing for nonlinear cointegrating regression," 2011.

[14] I. Choi and P. Saikkonen, "Testing linearity in cointegrating smooth transition regressions," *The Econometrics Journal*, vol. 7, no. 2, pp. 341–365, 2004.

[15] Z. Cai, Q. Li, and J. Y. Park, "Functional-coefficient models for nonstationary time series data," *Journal of Econometrics*, vol. 148, no. 2, pp. 101–113, 2009.

[16] Z. Xiao, "Functional-coefficient cointegration models," *Journal of Econometrics*, vol. 152, no. 2, pp. 81–92, 2009.

[17] P. Billingsley, *Convergence of probability measures*, vol. 493. Wiley-Interscience, 2009.

[18] D. Geman and J. Horowitz, "Occupation densities," *The Annals of Probability*, vol. 8, no. 1, pp. 1–67, 1980.

[19] D. Revuz and M. Yor, *Continuous martingales and Brownian motion*, vol. 293. Springer, 2004.

[20] P. C. Phillips, "Local limit theory and spurious nonparametric regression," *Econometric Theory*, vol. 25, no. 6, p. 1466, 2009.

[21] R. I. Jennrich, "Asymptotic properties of non-linear least squares estimators," *The Annals of Mathematical Statistics*, pp. 633–643, 1969.

[22] Q. Li, C. J. Huang, D. Li, and T.-T. Fu, "Semiparametric smooth coefficient models," *Journal of Business & Economic Statistics*, vol. 20, no. 3, pp. 412–422, 2002.

[23] Q. Shao and C. Lu, "Strong approximation for partial sums of weakly dependent random variables," *Scientia Sinica*, vol. A, no. 15, pp. 576–587, 1987.

[24] I. Kasparis, "Detection of functional form misspecification in cointegrating relations," *Econometric Theory*, vol. 24, no. 5, p. 1373, 2008.

[25] I. Kasparis and P. C. Phillips, "Dynamic misspecification in nonparametric cointegrating regression," *Journal of Econometrics*, 2012.

[26] E. Rio, "The functional law of the iterated logarithm for stationary strongly mixing sequences," *The Annals of Probability*, vol. 23, no. 3, pp. 1188–1203, 1995.

[27] G. Elliott and J. Stock, "Inference in models with nearly integrated regressors," *Econometric theory*, vol. 11, pp. 1131–1147, 1994.

[28] N. Gregory Mankiw and M. D. Shapiro, "Do we reject too often?: Small sample properties of tests of rational expectations models," *Economics Letters*, vol. 20, no. 2, pp. 139–145, 1986.

[29] R. F. Stambaugh, "Predictive regressions," *Journal of Financial Economics*, vol. 54, no. 3, pp. 375–421, 1999.

[30] M. Richardson and J. H. Stock, "Drawing inferences from statistics based on multiyear asset returns," *Journal of Financial Economics*, vol. 25, no. 2, pp. 323–348, 1989.

[31] C. L. Cavanagh, G. Elliott, and H. James, "Stock, 1995, inference in models with nearly integrated regressors," *Econometric theory*, vol. 11, no. 5, pp. 1131–1147.

[32] W. Torous, R. Valkanov, and S. Yan, "On predicting stock returns with nearly integrated explanatory variables*," *The Journal of Business*, vol. 77, no. 4, pp. 937–966, 2004.

[33] M. Lanne, "Testing the predictability of stock returns," *Review of Economics and Statistics*, vol. 84, no. 3, pp. 407–415, 2002.

[34] J. Lewellen, "Predicting returns with financial ratios," *Journal of Financial Economics*, vol. 74, no. 2, pp. 209–235, 2004.

[35] J. Y. Campbell and M. Yogo, "Efficient tests of stock return predictability," *Journal of financial economics*, vol. 81, no. 1, pp. 27–60, 2006.

[36] R. J. Schiller, "The irrational exuberance," 2000.

[37] P. Hall and C. C. Heyde, *Martingale limit theory and its application.* Academic press New York, 1980.

APPENDIX A: TECHNICAL RESULTS FOR CHAPTER 2

The proof of limit theory of NLS estimator follows the procedure applied by [2]. We start by defining regular functions as follows (see [2]):

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

The so called regularity conditions are defined also.

Definition A.2. F is regular on Π 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

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

$$\frac{1}{n}\sum_{t=1}^{n} T\left(\frac{z_t}{\sqrt{n}}\right) u_t \xrightarrow{d} \int_0^1 T(V_c(r))dU(r),$$

as $n \to \infty$

Lemma A.5. (a) If $F(\cdot, \pi)$ is a regular family on Π, then for each $\pi_0 \in \Pi$, there exits a neighborhood N_0 of π_0 such that $\sup_{\pi\in\Pi} F(\cdot, \pi)$ and $\inf_{\pi\in\Pi} F(\cdot, \pi)$ are relular for all $N \subset N_0$.

(b) If F is regular on a compact set Π, 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 Π, 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 Π, 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 Π, then $\int_0^1 F(V_c(r), \cdot)dr$ is continuous $a.s.$ on Π.

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

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 τ_{nt} in $(\Sigma, \mathcal{F}, P)$ with $\tau_{n0} = 0$ such that

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

as $n \to \infty$ 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 [2])Let Assumptions 2.1 hold. If F is regular on a compact set Π, then

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

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

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

as $n \to \infty$.

See Appendix A of [2] 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 [2].

Proof of Theorem 2.2.1: See proof of Theorem 4.1 of [2].

Proof of Theorem 2.3.1: See proof of Theorem 5.1 of [2].

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 θ_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

$$I_{1n} = \frac{1}{n^2 h} \sum_{t=1}^{n} \sum_{s \neq t} u_t u_s\, K_{ts}, \tag{35}$$

$$G_{2n} = \frac{1}{n^2 h} \sum_{t=1}^{n} (g(z_t, \hat{\theta}_n) - g(z_t, \theta_0)) \sum_{s \neq t} (g(z_s, \hat{\theta}_n) - g(z_s, \theta_0))\, K_{ts}, \tag{36}$$

and

$$G_{3n} = \frac{1}{n^2 h} \sum_{t=1}^{n} u_t \sum_{s \neq t} (g(z_s, \hat{\theta}_n) - g(z_s, \theta_0))\, K_{ts}. \tag{37}$$

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 $\widehat{\Sigma} \xrightarrow{p} \Sigma$, 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

$$\frac{d_n}{n^2} \sum_{t=1}^{[nr]} \sum_{s=1}^{t} K_{ts} \Rightarrow \frac{1}{2} L_{V_c}(r,r,0) \tag{38}$$

as $n \to \infty$. where $L_{V_c}(r,r,0)$ is defined by (25).

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

$$L_{n,\epsilon}^{(r)} - \left(\int_{-1}^{1} K(u)du\right) \frac{d_n}{n^2} \sum_{t=1}^{[nr]} \sum_{s=1}^{t} \phi_\epsilon(z_{t,n} - z_{s,n}) = o_p(1) \tag{39}$$

uniformly in $r \in [0,1]$, $z_{t,n}$ and $z_{s,n}$ as $n \to \infty$ and $d_n \to \infty$. 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 (39) refers to the proof of Lemma B in [20].

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

$$\frac{d_n}{n^2} \sum_{t=1}^{[nr]} \sum_{s=1}^{t} \phi_\epsilon(z_{t,n} - z_{s,n}) \xrightarrow{d} \int_0^r \int_0^t \phi_\epsilon(V_c(t) - V_c(s))dsdt \tag{40}$$

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

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

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

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

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

$$\lim_{\epsilon \to 0} \lim_{n \to \infty} \sup_{0 \le r \le 1} E|L_n^{(r)} - L_{n,\epsilon}^{(r)}| = 0 \tag{41}$$

The proof of (41) is similar to proof of Theorem 2.1 in [5].

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 Σ given in (27).

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 $\mathcal{F}_{nt} = \sigma\{\eta_i, u_i : 1 \le i \le t \le n\}$ be the smallest σ-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 [37] is applied to establish our results. We verify that the two conditions of the central limit theorem for martingale difference are satisfied.

$$\sum_{t=2}^{n} E_{t-1}\left[Z_{nt}^2 I(|Z_{nt}| > \xi_1)\right] \xrightarrow{p} 0 \quad \text{for all} \quad \xi_1 > 0 \tag{42}$$

and

$$V_n^2 = \sum_{t=2}^{n} E_{t-1}\left(Z_{nt}^2\right) \xrightarrow{p} \Sigma/2, \tag{43}$$

where Σ is given in (27) and $I(A)$ is the indicator function of event A. We start by checking (43). 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

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

The probability limits of B_{1n}, B_{2n} and B_{3n} are derived respectively with $B_{1n} = \sigma_u^4 n^{-2}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E(K_{ts}^2)$, $B_{2n} \xrightarrow{p} o(1)$, 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 $n \to \infty$. 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 $N = [1/\delta]$, $s_k = [kn/N] + 1$, $s_k^* = s_{k+1} - 1$, $N_t^* =$

$[(N-1)(t-1)/n]$ 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 $n \to \infty$. 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 (42) holds. For all $\xi_2 > 0$,

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

where the last inequality follows from Markov inequality. Condition (42) holds if

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

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

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

$$L_{n,\epsilon}^{(r)} = \frac{c_n^2}{n^2 h} \sum_{t=2}^{[nr]}\sum_{s=1}^{t-1} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \dot{g}[c_n(z_{t,n} + x_1\epsilon)]\dot{g}[c_n(z_{s,n} + x_2\epsilon)]$$

$$K[c_n(z_{t,n} - z_{s,n} + x_1\epsilon - x_2\epsilon)]\phi(x_1)\phi(x_2)dx_1dx_2$$

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

where $c_n = \sqrt{n}$, $\tau = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \dot{g}(a)\dot{g}(b)K(a-b)dadb$, $\phi_\epsilon(z) = \frac{1}{\epsilon\sqrt{2\pi}} \exp\left\{-\frac{z^2}{2\epsilon}\right\}$, 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 [20] Write

$$\begin{aligned} L_{n,\epsilon}^{(r)} &= \frac{c_n^2}{n^2}\sum_{t=2}^{[nr]}\sum_{s=1}^{t-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dot{g}[c_n(z_{t,n}+x_1\epsilon)]\dot{g}[c_n(z_{s,n}+x_2\epsilon)] \\ & K[\frac{c_n}{h}(z_{t,n}-z_{s,n}+x_1\epsilon-x_2\epsilon)]\phi(x_1)\phi(x_2)dx_1dx_2 \\ &= \frac{c_n^2}{n^2h}\sum_{t=2}^{[nr]}\sum_{s=1}^{t-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dot{g}(c_na)\dot{g}(c_nb)K[d_n(a-b)]\phi_\epsilon(a-z_{t,n})\phi_\epsilon(b-z_{s,n})dadb \\ &= \frac{c_n^2}{2n^2h}\sum_{t=1}^{[nr]}\sum_{s=1}^{[nr]}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dot{g}(c_na)\dot{g}(c_nb)K[d_n(a-b)]\phi_\epsilon(a-z_{t,n})\phi_\epsilon(b-z_{s,n})dadb \\ & +s.o. \end{aligned}$$

Then, similar to the proof of Lemma B in [20], it is readily seen that as $n\to\infty$,

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

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,

$$\int_0^r \phi_\epsilon(V_C(t))dt = \int_{-\infty}^{\infty}\phi_\epsilon(s)L_{V_c}(r,s)ds \tag{44}$$

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

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

Proof: The proof refers to Theorem 2.1 of [5]. Write

$$\begin{aligned} L_n^{(r)} &= \frac{c_n^2}{n^2h}\sum_{t=2}^{[nr]}\sum_{s=1}^{t-1}\dot{g}(c_nz_{t,n})\dot{g}(c_nz_{s,n})K(\frac{c_n}{h}(z_{t,n}-z_{s,n})) \\ L_{n,\epsilon}^{(r)} &= \frac{c_n^2}{n^2h}\sum_{t=2}^{[nr]}\sum_{s=1}^{t-1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\dot{g}(c_n(z_{t,n}+x_1\epsilon))\dot{g}(c_n(z_{s,n}+x_2\epsilon))K[c_n(z_{t,n}-z_{s,n} \\ & +x_1\epsilon-x_2\epsilon)]\phi(x_1)\phi(x_2)dx_1dx_2 \end{aligned}$$

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

$$\lim_{\epsilon\to 0}\lim_{n\to\infty}\sup_{0\le r\le 1} E|L_n^{(r)} - L_{n,\epsilon}^{(r)}| = 0 \tag{45}$$

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

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

Then, we prove (45). 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

$$\sup_{0\le r\le 1} E|L_n - L_{n,\epsilon}| \le \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{c_n^2}{n^2 h}\sup_{0\le r\le 1} E\left|\sum_{t=1}^{[nr]}\sum_{s=1}^{[nr]} Y_{t,s,n}(x_1,x_2)\right|\phi(x_1)\phi(x_2)dx_1dx_2 \tag{46}$$

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

$$\begin{aligned} \frac{c_n^2}{h}E|Y_{t,s,n}| \;\le\; & \frac{Ac_n^2}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}|\dot{g}[c_n(d_{t,0,n} + x_1\epsilon)]\dot{g}[c_n(d_{s,0,n} + x_2\epsilon)] - \dot{g}[c_n d_{t,0,n}z_1]\dot{g}[c_n d_{s,0,n}z_2]| \\ & h_{t,0,n}(z_1)h_{s,0,n}(z_2)dz_1dz_2 \\ \le\; & \frac{A}{2d_{t,0,n}d_{s,0,n}}\int_{-\infty}^{\infty}|\dot{g}(z_1 + c_n x_1\epsilon) - \dot{g}(z_1)|dz_1\int_{-\infty}^{\infty}|\dot{g}(z_2 + c_n x_2\epsilon) - \dot{g}(z_2)|dz_2 \\ \le\; & A\left[\int_{-\infty}^{\infty}|\dot{g}(z)|dz/d_{t,0,n}\right]^2 \end{aligned}$$

Then, it follows that

$$\frac{c_n^2}{2n^2 h}\sup_{0\le r\le 1} E|\sum_{t=1}^{[nr]}\sum_{s=1}^{[nr]} Y_{t,s,n}(x_1,x_2)| \le A_1\frac{1}{n^2}\sum_{t=1}^{n}\sum_{s=1}^{n}\frac{1}{d_{t,0,n}d_{s,0,n}} < \infty \tag{47}$$

This, together with (46) and the dominated convergence theorem, implies that , to prove (45), it suffices to show that, for fixed x_1 and x_2,

$$\Lambda_n(\epsilon) = \frac{c_n^2}{n^2 h} \sup_{0\leq r\leq 1} E\left[\sum_{t=1}^{[nr]}\sum_{s=1}^{[nr]} Y_{t,s,n}(x_1,x_2)\right]^2 \to 0 \tag{48}$$

Refer to Proof of Theoren 2.1 in [5], we can see that (48) is true. Now, the result is stated.

Lemma B.14. Under Assumptions given in Theorem 3.2.1, under H_0, we obtain $G_{2n} = O_p\left(n^{-\frac{3}{2}}\right)$ and $G_{3n} = O_p(n^{-\frac{3}{2}})$, where G_{2n} and G_{3n} are defined in (36) and (37), 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

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

Hence, we have

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

Next, let

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

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

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

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

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

where $\tilde{u}_t=Y_t-\hat{f}^{(-t)}(Z_t)$ is the nonparametric residual and Σ is defined in (27).

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

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

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 $n\to\infty$.

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}.

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

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

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

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

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 δ 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).