ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

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

Hermite Series Estimation in Nonlinear Cointegrating Models

Biqing Cai and Jiti Gao

August 2013

Working Paper 17/13

Hermite Series Estimation in Nonlinear Cointegrating Models

Biqing Cai and Jiti Gao*

August 5, 2013

Abstract

This paper discusses nonparametric series estimation of integrable cointegration models using Hermite functions. We establish the uniform consistency and asymptotic normality of the series estimator. The Monte Carlo simulation results show that the performance of the estimator is numerically satisfactory. We then apply the estimator to estimate the stock return predictive function. The out–of–sample evaluation results suggest that dividend yield has nonlinear predictive power for stock returns while book–to–market ratio and earning–price ratio have little predictive power.

Key words: Cointegration, Hermite Functions, Return Predictability, Series Estimator, Unit Root

JEL Classification Numbers: C14; C22; G17.

*Corresponding author: Jiti Gao, Department of Econometrics and Business Statistics, Monash University, Caulfield East, Victoria 3145, Australia. Email:jiti.gao@monash.edu.

1 Introduction

Since Engle and Granger (1987), the concept of cointegration has become popular in economics because cointegration relationships are often used to describe economic variables which share some common stochastic trends or have long–run equilibrium relationships. However, the idea that every small deviation from the long–run equilibrium will lead instantaneously to error correction mechanisms is implicit in the definition. Nonetheless, as argued by Blake and Fomby (1997), the presence of fixed costs of adjustment may prevent economic agents from adjusting continuously, thus the movement towards the long–run equilibrium need not occur in every period such that linear cointegration may fail. Also, there is consensus in econometrics that nonlinearity is now the norm, rather than the exception (as discussed in Granger 1995; Gao 2007; Teräsvirta, Tjøstheim and Granger 2010, for example). Misspecifying a linear cointegration model may lead to non-finding of cointegration.

Recently, nonlinear cointegration models have become a hot topic in econometrics. Park and Phillips (1999) discuss asymptotics for nonlinear transformation of unit root process and Park and Phillips (2001) for nonlinear regression with a unit root process. Furthermore, asymptotic properties for nonparametric estimation for nonlinear cointegration models have been derived by Wang and Phillips (2009a; 2009b). Meanwhile, Karlsen and Tjøtheim (2001) and Karlsen et al. (2007) also derive some limit theory for nonparametric estimation of nonlinear cointegration based on different assumptions on the data generating process and different mathematical techniques. Chen, Gao and Li (2012) consider estimation issues in a partially linear model with nonstationary regressors. Gao and Phillips (2013) consider semiparametric estimation in triangular system equations with nonstationarity and endogeneity.

In addition to the kernel–based estimation proposed in the literature, the series estimation method is an alternative to the kernel–based method. When the data are either independent and identically distributed or stationary, estimation theories based on series estimation methods have been discussed in Andrews (1991), Newey (1997) and Gao (2007) for example. However, as far as we know, when the data is unit root nonstationary, there are only a couple of studies based on series estimation. Dong and Gao (2011; 2012) were among the first considering series expansion for nonstationary data. Dong and Gao (2011) discuss series expansion for Lévy processes which can be considered as an orthogonal series expansion based on time varying probability densities. In contrast to their papers, we propose using Hermite series expansion which is orthogonal with respect to Lebesgue density without specifying the distribution of the innovation to unit root process. Thus, we allow for much more general data generating assumptions. It is well known that the series estimation has some advantages over the kernel–based estimation. For example, it's easy to impose some types of restrictions, such as additive separability. In addition, it is computationally convenient.

In this paper, we propose using a Hermite series estimation method for a class of nonparametric cointegration models where the mean function is integrable. Meanwhile, we establish an asymptotic distribution theory for a matrix of partial sums of nonlinear nonstationary time series in Theorem 4 below. Such a result is of general interest and is applicable to deal with inverses of matrices of unit root nonstationary time series. As a consequence, we are able to establish asymptotic results for the nonparametric series estimator itself rather than a transformed version as in Dong and Gao (2011). For instances, we establish some uniform consistency results and an asymptotic normality for a series–based estimator with a rate of convergence of an order of $\sqrt{\sqrt{T}\, p^{-1}}$, where p is the truncation parameter involved in the series approximation and T is the sample size. This rate is equivalent to that of $\sqrt{\sqrt{T}h}$ based on the kernel method when we choose $h = p^{-1}$ with h being a bandwidth parameter.

We then apply the proposed estimation method to estimate the stock return predictive function. The time series properties of stock returns and financial ratios (highly persistent financial ratios and far less persistent stock returns) suggest that a nonlinear integrable model should be more suitable than the simple linear model. Note that the linear model is commonly used in the literature, although the linear model has some model identification and specification problems for modelling a stationary return series by a linear function of unit root nonstationary predictive time series. By contrast, a nonlinear integrable function of a unit root nonstationary time series significantly reduces the nonstationarity of the original unit root time series. Our empirical results support the existence of nonlinear predictive power of the dividend yield, while the predictive power of book–to–market ratio and earning–price ratio is very weak. A detailed analysis is given in Section 5 below.

The organisation of this paper is as follows. In Section 2, we propose the model and discuss the estimation method. In Section 3, we derive the consistency and asymptotic normality of the series estimator. Section 4 separately establishes an asymptotic consistency results for a regression matrix of series estimation that is vital for the establishment of the results given in Section 3. In section 5, we conduct Monte Carlo simulation to evaluate the finite sample performance of the nonparametric series estimator. Section 6 applies the series estimator to estimate the stock return predictive function and compare the out–of–sample performance of an integrable model with the historic mean and linear models. Section 7 concludes the paper.

Throughout this paper, all limits are taken "as $T \to \infty$", $\to_D$ denotes weak convergence, $\to_P$ denotes convergence with probability approaching one, $\to_{a.s.}$ denotes almost sure convergence. $O_P(.)$ means stochastic order same as, $o_P(.)$ means stochastic order less than. For a $n \times m$ matrix A, $\|A\| = tr(A^\tau A) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2$, with a_{ij} being the $ij-th$ element of A, and for a vector a, $\|a\| = (\sum_i a_i^2)^{1/2}$ with a_i being the $i-th$ element of a. The mathematical proofs are collected in two appendices.

2 The Model and Estimation

2.1 Preliminaries of Hermite Functions

In this paper, we use Hermite functions to estimate square integrable functionals of a unit root process. The Hermite functions form a complete basis in $L^2(R)$ and have advantages in the nonlinear nonstationarity setup. Unlike the stationary data case where orthogonality is defined according to the marginal (stationary) distribution, in the nonstationary case, the orthogonality is defined with respect to Lebesgue measure. In this case, Lebesgue measure is the invariant measure. For more information about the invariant measure for a null recurrent process, see Karlsen and Tjøtheim (2001). Thus, Hermite series are naturally orthogonal without fully specifying the data generating process, i.e., we don't need to impose restrictions on the distribution of the unit root process. Also, it's well known that the use of orthogonal basis functions allows us to extract information of different pieces of the unknown function effectively. In this subsection, we introduce some basic properties about Hermite functions.

Let $\{H_i(x)\}_{i=0,1,2,\ldots}$ be the orthogonal Hermite polynomial system with respect to the weight function $\exp(-x^2)$. It's well known that $\{H_i(x)\}_{i=0,1,2,\ldots}$ is a complete orthogonal system in Hilbert Space $L^2(R, \exp(-x^2))$, in which the conventional inner product is used, i.e., $\langle f, g\rangle = \int f(x)g(x)\exp(-x^2)dx$. In addition, the orthogonality of the system can be expressed as follows:

$$\langle H_i(x), H_j(x)\rangle = \sqrt{\pi}2^i i!\delta_{ij}, \tag{2.1}$$

where δ_{ij} is the Kronecker delta function. Put $\varphi(x) = \exp(-\frac{1}{2}x^2)$ and define

$$h_i(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{2^i i!}}H_i(x) \quad \text{and} \quad F_i(x) = h_i(x)\varphi(x), i \geq 0. \tag{2.2}$$

Then, $\{F_i(x)\}_{i=0,1,2,\ldots}$ are the so-called Hermite series or Hermite functions in the literature. $\{F_i(x)\}_{i=0,1,2,\ldots}$ are complete orthonormal in $L^2(R)$, such that any continuous function $f(x)$ in $L^2(R)$ has an expression of the form

$$f(x) = \sum_{i=0}^{\infty}\theta_i F_i(x), \tag{2.3}$$

which is the projection coefficient of $f(x)$ on $F_i(x)$, where $\theta_i = \int f(x)F_i(x)dx$.

The Hermite functions can be listed as follows: $F_0(x) = \frac{1}{\sqrt[4]{\pi}}\exp(-\frac{1}{2}x^2)$; $F_1(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{2}} \times 2x \times \exp(-\frac{1}{2}x^2)$; $F_2(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{8}} \times (4x^2 - 2) \times \exp(-\frac{1}{2}x^2)$; $F_3(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{48}}(8x^3 - 12x) \times \exp(-\frac{1}{2}x^2)$, and so on.

2.2 Model and Assumptions

Consider a nonparametric regression model of the form

$$y_t = f(x_t) + \varepsilon_t, t = 1, 2, \ldots, T, \tag{2.4}$$

where x_t is a scalar unit root process and ε_t is a stationary error term. Letting f be a square integrable function, it admits an expansion of the form:

$$f(x) = \sum_{j=0}^{\infty} \theta_j F_j(x), \tag{2.5}$$

where $\theta_j = \int f(x)F_j(x)dx$, $\{F_i(x)\}_{i=0,1,2,...}$ $j = 0, 1, 2, ...$ are Hermite functions. We then have the following least squares estimator of θ of the form[1]

$$\widehat{\theta} = \left[\sum_{t=1}^{T} Fp(x_t)Fp(x_t)^\tau\right]^{-1} \sum_{t=1}^{T} Fp(x_t)y_t. \tag{2.6}$$

Then, the series estimator of function f taking value at x is $\widehat{f}(x) = Fp^\tau(x)\widehat{\theta}$, where $Fp^\tau(x) = [F_0(x), F_2(x), ..., F_{p-1}(x)]$. We want to derive some asymptotic properties for $\widehat{f}(x) - f(x)$.

Observe that

$$\widehat{f}(x) - f(x) = Fp^\tau(x)[\widehat{\theta} - \theta] + \sum_{j=p}^{\infty} \theta_j F_j(x), \tag{2.7}$$

where $\theta = [\theta_0, \ldots \theta_{p-1}]^\tau$.

To derive asymptotic properties for $\widehat{f}(x) - f(x)$, we need following technical assumptions.

Assumption 1. Let $x_t = \sum_{i=1}^{t} \zeta_i$ for $t = 1, 2, ..., T$ with $x_0 = O_P(1)$, in which $\zeta_t = \sum_{j=0}^{\infty} \pi_j e_{t-j}$, where $\{e_t\}$ is a sequence of independent and identically distributed (i.i.d.) random errors with mean 0, variance 1 and $E(|e_1|^{6+\delta}) < \infty$ for some $\delta > 0$, $\pi_0 = 1, \sum_{j=0}^{\infty} \pi_j \neq 0$ and $\sum_{j=0}^{\infty} j|\pi_j| < \infty$. The characteristic function of e_1 denoted by $\varphi(t)$ satisfies $\int_{-\infty}^{\infty} |\varphi(t)|dt < \infty$.

Assumption 2. Let $\{\varepsilon_t, \mathcal{F}_t\}_{t\geq 1}$ be a martingale difference sequence satisfying $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \sigma^2$ a.s. and $E(\varepsilon_t^4|\mathcal{F}_{t-1}) < \infty$ a.s., for all $t = 1, 2, \ldots, T$, where $\mathcal{F}_t$ is a sequence of σ–fields defined by $\mathcal{F}_t = \sigma(\varepsilon_1, ..., \varepsilon_t, x_1, x_2, ..., x_{t+1})$.

Assumption 3. f is square integrable such that

$$e^{x^2/2}(d^r/dx^r)[e^{-\frac{x^2}{2}} f(x)] = \sum_{i=1}^{r} [r!/i!(i-r)!](-1)^i 2^{-i/2} H_i(x/2^{1/2})(d^{r-i}/dx^{r-i})f(x)$$

exists and is square integrable for some $r \geq 5$.

[1]From the proof of theorem 4, we can see that the inverse of $\sum_{t=1}^{T} Fp(x_t)Fp(x_t)^\tau$ is well defined asymptotically, so we avoid the usage of a generalized inverse here.

Assumption 4. Let the truncation parameter of the Hermite series expansion p satisfy $p = c[T^{\alpha}]$, where $c > 0$ is a constant and $\frac{1}{2(r-1)} < \alpha < \frac{3}{22}$.

Remark on Assumptions 1 and 2: Assumptions 1 and 2 are used widely in such kind of problems, such as, Park and Phillips (2001), Wang and Phillips (2009a). We impose some restricted assumptions on the moment condition of the innovation of unit root process to ensure sufficiently fast rate of convergence when applying strong approximation. Under Assumptions 1 and 2, the joint invariance principle holds. Define

$$W_{1Tr} = \frac{x_{[Tr]}}{\sqrt{T}} \quad \text{and} \quad W_{2Tr} = \frac{1}{\sqrt{T}} \sum_{t=1}^{[Tr]} \varepsilon_t. \tag{2.8}$$

Then, we have $(W_{1Tr}, W_{2Tr}) \to_D (B_1(r), B_2(r))$, where $(B_1(r), B_2(r))$ is a two dimensional Brownian motion.

Also, due to Skorohod-Dudley-Wichura representation, joint strong invariance principle holds such that[2]

$$\left(\frac{W_{1Tr}}{\Psi}, \frac{W_{2Tr}}{\sigma} \right) \to_{a.s.} (W_1(r), W_2(r)) \tag{2.9}$$

in an expanded probability space, where $W_1(r)$ and $W_2(r)$ are individually standard Brownian motion and $(W_1(r), W_2(r))$ is a two dimensional Brownian motion.

In this paper, we will avoid repetitious embedding $\frac{x_{[Tr]}}{\Psi\sqrt{T}}$ and $W_1(r)$ on the expanded space. Because we only care about weak convergence, the convention is justified.

Obviously, $x_t = \Psi S_t + \sum_{k=0}^{t-1} \Psi_k e_{t-k} + \sum_{k=0}^{\infty} (\Psi_{k+1} - \Psi_k) e_{-k}$, with $\Psi_k = \sum_{j=0}^{k} \pi_j$, $\Psi = \sum_{j=0}^{\infty} \pi_j$ and $S_t = \sum_{k=1}^{t} e_k$. It can then be shown that in the same expanded space

$$\sup_{0<r<1} \left| \frac{S_{[Tr]}}{\Psi\sqrt{T}} - W_1(r) \right| = o_P(1). \tag{2.10}$$

That is, the unit root process can be decomposed into a random walk process (a unit root process with i.i.d. innovation) and two asymptotically negligible terms such that the random walk process converge in probability to the same limiting process in the expanded space.

The Brownian motion $\Psi W_1(r)$ admits a local time $L(t, s)$, defined by

$$L(t, s) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_0^t I\{|\Psi W_1(r) - s| < \varepsilon\} dr, \tag{2.11}$$

where $I(A)$ denote the conventional indicator function.

[2]Strong approximation has different expression as in Park and Phillips (2001), we will use this expression because it facilitates our derivation.

Roughly speaking, the local time can be interpreted as a spatial occupation density in s for Brownian motion $\Psi W_1(r)$. The local time is a key tool in studying the intersection of nonlinearity and nonstationarity, e.g., Park and Phillips (1999, 2001), Wang and Phillips (2009a). Phillips (2001) provides some examples where the tool of local time can be used to analyse economic time series which is called "spatial analysis of time series".

Remark on Assumption 3: Assumption 3 requires $f(x)$ to be sufficiently smooth and the tail of $f(x)$ to be sufficiently thin. A sufficient condition for Assumption 3 is L_2 integrality of $|x^i f(x)|dx$ and $|f^i(x)|$ for $i = 0, 1, ..., r$, where $f^i(x)$ denotes $i-th$ order of derivative. Due to the choice of p, here we require a large r to make sure the bias term diminish. The classes of f includes Gaussian functions, Laplace functions and functions with compact support.

Remark on Assumption 4: Assumption 4 restricts the increase of growth rate of the terms of Hermite series to guarantee the convergence of the regression matrix. The condition of $\frac{1}{2(r-1)} < \alpha < \frac{3}{22}$ also requires $r > \frac{14}{3}$.

2.3 Consistency of Series Estimator

In this subsection, we discuss the consistency of the series estimator. According to (2.7), we have

$$\widehat{\theta} - \theta = [\frac{1}{\sqrt{T}} F^\tau F]^{-1} F^\tau (Y - A)/\sqrt{T} + [\frac{1}{\sqrt{T}} F^\tau F]^{-1} F^\tau (A - F\theta)/\sqrt{T}, \tag{2.12}$$

where $F \equiv \begin{pmatrix} F_0(x_1) & ... & F_{p-1}(x_1) \\ ... & & ... \\ F_0(x_T) & ... & F_{p-1}(x_T) \end{pmatrix}$, $Y = (y_1, ..., y_T)^\tau$, and $A = (f(x_1), ..., f(x_T))^\tau$.

The two terms in equation (2.12) can be regarded as the bias and variance terms for nonparametric series regression. We now establish the first theorem of this paper.

Theorem 1. Under Assumptions 1-4, we have as $T \to \infty$

$$\left\| \widehat{\theta} - \theta \right\| = o_P(1). \tag{2.13}$$

Theorem 1 shows that the estimated coefficients converge to the true coefficients. We also have the uniform consistency result for the series estimator $\widehat{f}(x)$.

Corollary 1. Under Assumptions 1-4, we have as $T \to \infty$

$$\sup_x |\widehat{f}(x) - f(x)| = o_P(1). \tag{2.14}$$

Remark: When the data are stationary, polynomials or splines are usually used as basis functions, e.g., in Andrews (1991), Newey (1997), and Gao (2007). In their cases, the uniform consistency is usually based on more restrictive assumptions than those for the point–wise consistency. In our case, due to uniform boundedness of the Hermite series, the uniform consistency requires the same conditions as those for the point–wise consistency.

2.4 Asymptotic Normality

In this subsection, we will establish an asymptotic normality for the series estimator.

Theorem 2. Under Assumptions 1–4, we have as $T \to \infty$

$$\frac{\sqrt[4]{T}}{\Sigma_T}\left(\widehat{f}(x) - f(x)\right) \to_D N(0,1), \tag{2.15}$$

where $\Sigma_T^2 = \frac{\sigma^2 \sum_{i=0}^{p-1} F_i^2(x)}{L(1,0)}$, in which $L(1,0) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\int_0^1 I\{|\Psi W_1(r)| < \varepsilon\}dr$.

Remark: Since $P(L(1,0) > 0) = 1$ and $\sum_{i=0}^{p-1} F_i^2(x) > 0$, Σ_T^2 is well defined. Moreover, $\sum_{i=0}^{p-1} F_i^2(x) = O(p)$ uniformly in x because for any orthogonal polynomials $H_i(x)$ on any compact interval, we have $\sum_{i=0}^{p-1} H_i^2(x) = O(p)$ (see, p. 295 of Alexits 1961). Note that the rate of convergence of the kernel estimator is $T^{1/4}\sqrt{h}$ (see, for example, Theorem 3.1 of Wang and Phillips 2009a), where h is the bandwidth parameter, and the rate of convergence of the series estimator is $T^{1/4}\sqrt{p^{-1}}$. They are equivalent if we replace h by p^{-1}.

Remark: For the purpose of statistical inference, σ^2 can be estimated by $\widehat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} y_t^2$ because of $\frac{1}{T}\sum_{t=1}^{n} y_t^2 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 + o_P(1) \to_P E[\varepsilon_1^2] = \sigma^2$, and $L(1,0)$ can be estimated by $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_0^2(x_t)$, since $\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_0^2(x_t) \to_P L(1,0)\cdot\int F_0^2(x)dx = L(1,0)$ in the expanded probability space.

Combining the above remarks, we have following theorem.

Theorem 3. Under Assumptions 1–4, we have as $T \to \infty$

$$\frac{\sqrt[4]{T}}{\widehat{\Sigma}_T}\left(\widehat{f}(x) - f(x)\right) \to_D N(0,1), \tag{2.16}$$

where $\widehat{\Sigma}_T^2 = \frac{\widehat{\sigma}^2 \sum_{i=0}^{p-1} F_i^2(x)}{\widehat{F}_0}$, in which $\widehat{F}_0 = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_0^2(x_t)$.

Remark: The term $\sum_{i=0}^{p-1} F_i^2(x)/L(1,0)$ can also be estimated by $Fp^\tau(x)[\frac{1}{\sqrt{T}}F^\tau F]^{-1}Fp(x)$, as discussed below.

3 Asymptotic Convergence of $\frac{1}{\sqrt{T}}F^{\tau}F$

As mentioned in the introductory section and seen in the above discussion, the least squares estimator of θ involves an inverse matrix of $\frac{1}{\sqrt{T}}F^{\tau}F$, which causes both theoretical and computational difficulties. In the literature, such difficulties are avoided through using a transformed version of $\widehat{\theta}$ of the form $\widetilde{\theta} = F^{\tau}F \cdot \widehat{\theta}$ (see, for example, Dong and Gao 2011; 2012). As a consequence, it is difficult to obtain a rate of convergence for $\widehat{\theta}$, although a rate of convergence of $\widetilde{\theta}$ is available.

Therefore, we tackle this difficulty by studying the convergence of $\frac{1}{\sqrt{T}}F^{\tau}F$ directly. Our experience suggests that such convergence itself is of general interest and may be applied to significantly simplify the construction of existing estimation and specification procedures, such as those discussed in Dong and Gao (2011; 2012).

We now have following theorem.

Theorem 4. Let Assumption 4 hold. Then, in an expanded probability space, we have as $T \to \infty$

$$\left\| \frac{1}{\sqrt{T}}F^{\tau}F - L(1,0)I \right\| \to_P 0, \tag{3.1}$$

where I is an identity matrix of $p \times p$ order.

Remark: It follows from the definition of F that

$$\left\| \frac{1}{\sqrt{T}}F^{\tau}F - L(1,0)\ I \right\|^2 = \sum_{i=0}^{p-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_i^2(x_t) - L(1,0) \right)^2 + \sum_{i=0}^{p-1} \sum_{j=0,\neq i}^{p-1} \left(\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_i(x_t)F_j(x_t) \right)^2 .$$

Since $p \to \infty$, existing results (see, for example, Wang and Phillips 2009a; 2011) are not applicable. In Appendix B below, we therefore develop some new convergence results with certain rates to complete the proof of Theorem 4.

Remark: From this theorem, we have $\lambda_{\min}(\frac{1}{\sqrt{T}}F^{\tau}F - L(1,0)I) \le \lambda_{\max}(\frac{1}{\sqrt{T}}F^{\tau}F - L(1,0)I) = o_P(1)$, and thus we can also have $L(1,0) + o_P(1) \le \lambda_{\min}(Q) \le \lambda_{\max}(Q) \le L(1,0) + o_P(1)$, where $\lambda_{\max}(Q)$ and $\lambda_{\min}(Q)$ denote the largest and smallest eigenvalues of $Q \equiv \frac{1}{\sqrt{T}}F^{\tau}F$.

Before the proofs of Theorems 1–4 are given in Appendices A and B below, we examine the finite–sample performance of the series estimation in Section 4 below.

4 Simulation Study

In this section, we conduct some simulation studies to assess the finite sample performance of the proposed nonparametric series estimation method for the following nonstationary models:

$$\text{Model1}: \ y_t = \frac{1}{2}\exp(-|x_t|) + \varepsilon_t, t = 1, ...T; \tag{4.1}$$

$$\text{Model2}: \ y_t = \frac{1}{\pi(1+x_t^2)} + \varepsilon_t, t = 1, ...T; \tag{4.2}$$

$$\text{Model3}: \ y_t = \frac{1}{4}I(-2 \leq x_t \leq 2) + \varepsilon_t, t = 1, ...T; \tag{4.3}$$

$$\text{Model4}: \ y_t = x_t + \varepsilon_t, t = 1, ...T; \tag{4.4}$$

where $x_t = x_{t-1} + e_t$ for $t = 1, ..., T$, $x_0 = 0$ and for $t = 1, 2, \cdots, T$, we simulate (e_t, ε_t) by

$$\{e_t, \varepsilon_t\} \sim i.i.d.N\left(0, \begin{pmatrix} 0.1^2 & \rho \times 0.1^2 \\ \rho \times 0.1^2 & 0.1^2 \end{pmatrix}\right),$$

where $\rho = 0$ or 0.9 corresponding to exogeneity or endogeneity respectively. The true regression function of the first model is Laplace density, the second model is the Cauchy density, the third model is indicator function with support on $[-2, 2]$. Models 1 and 3 are integrable models satisfying Assumption 3, while Model 2 is integrable model but doesn't satisfy Assumption 3, because Assumption 3 requires the tails of the integrable functions to be very thin. The fourth model is a linear (not integrable) model, which can't be estimated by the series estimation method proposed in this paper. Thus, it is expected that our method will perform badly in this case.

The sample size is 300, 600 and 1200, the replication number is 5000. The truncation parameters are set to be $[c_1 \times T^{c_2}]$, where $c_2 = \frac{5}{44}$, $c_1 = 1, 2, 3$. The choice of c_2 is relatively less important than the choice of c_1. Because by Assumption 3, c_2 should be smaller than 3/22. when $c_2 = 3/22$, to let $[T^{c_2}]$ larger than 2, T should be larger than 162, and to let $[T^{c_2}]$ larger than 3, T should be larger than 3155 (there are few data sets in economics where T is so large). Thus, with a data of normal size, $[T^{c_2}]$ usually takes only values 1 or 2. Thus, when $[T^{c_2}]$ satisfies Assumption 3, the choice of p is less sensitive to c_2 than c_1, thus we just report the case $c_2 = 5/44$. And from our simulation results, by the root mean squared error (RMSE) criterion, the choice of p should not be large such that c_1 should be at most equal to 2 (when the integrable function is the Cauchy function which is of heavy tail within the integrable functions). So we choose c_1 to be 1, 2 or 3.

The sample bias is defined by:

$$\text{Bias} = \frac{1}{N}\frac{1}{T}\sum_{n=1}^{N}\sum_{t=1}^{T}(\overline{\widehat{f}}_n - f_n(x_t)), \tag{4.5}$$

where $f_n(.)$ denotes the values of f in $n-th$ replication and $\widehat{f}_n(.)$ is the series estimator of the regression function in $n-th$ replication, and $\overline{\widehat{f}}_n = \frac{1}{T}\sum_{t=1}^{T}\widehat{f}_n(x_t)$.

The sample standard deviation is defined by:

$$\text{Std} = \sqrt{\frac{1}{N}\frac{1}{T}\sum_{n=1}^{N}\sum_{t=1}^{T}(\widehat{f}_n(x_t) - \overline{\widehat{f}}_n)^2}, \tag{4.6}$$

where N is the number of replications, and T is the sample size. The sample root mean squared error (RMSE) is defined by:

$$\text{Rmse} = \sqrt{\frac{1}{N}\frac{1}{T}\sum_{n=1}^{N}\sum_{t=1}^{T}(\widehat{f}_n(x_t) - f_n(x_t))^2}. \tag{4.7}$$

The results of the simulation are summarised in following tables:

Table 1: Simulation Results for Bias: ρ=0

c_1	T	Model 1	Model 2	Model 3	Model 4
	300	-0.0021	-0.0045	-0.0054	-0.0550
1	600	-0.0019	-2.5099×10^{-4}	-0.0038	-0.2872
	1200	-0.0018	-1.2862×10^{-4}	-0.0025	-0.6497
	300	-7.4710×10^{-4}	0.0021	-0.0036	0.1023
2	600	-4.6000×10^{-4}	0.0011	-0.0024	-0.1502
	1200	-3.2299×10^{-4}	2.9098×10^{-4}	-0.0018	0.4936
	300	-0.0021	-9.8858×10^{-4}	-0.0031	0.0645
3	600	-2.9926×10^{-4}	-2.3019×10^{-4}	-0.0016	-0.8155
	1200	-4.9918×10^{-5}	-3.6976×10^{-4}	-0.0014	-0.2671

Table 2: Simulation Results for Std: ρ=0

c_1	T	Model 1	Model 2	Model 3	Model 4
	300	0.0875	0.0786	0.0722	0.2421
1	600	0.0745	0.0655	0.0596	0.3790
	1200	0.0583	0.0537	0.0473	0.3004
	300	0.0911	0.0781	0.0734	0.5992
2	600	0.0750	0.0679	0.0612	0.5915
	1200	0.0583	0.0542	0.0480	0.4789
	300	0.0914	0.0820	0.0737	0.8324
3	600	0.0753	0.0689	0.0610	0.7824
	1200	0.0593	0.0563	0.0490	0.6250

Table 3: Simulation Results for Rmse: ρ=0

c_1	T	Model 1	Model 2	Model 3	Model 4
	300	0.0462	0.0592	0.0434	17.1426
1	600	0.0378	0.0440	0.0371	34.4133
	1200	0.0285	0.0348	0.0293	68.9292
	300	0.0512	0.0560	0.0503	17.3494
2	600	0.0416	0.0437	0.0423	34.8402
	1200	0.0308	0.0346	0.0318	68.7829
	300	0.0577	0.0578	0.0572	17.0786
3	600	0.0466	0.0477	0.0465	34.8762
	1200	0.0388	0.0377	0.0362	69.4788

Table 4: Simulation Results for Bias: ρ=0.9

c_1	T	Model 1	Model 2	Model 3	Model 4
	300	-0.0018	-0.0046	-0.0058	-0.2738
1	600	-0.0026	-3.8877$\times10^{-4}$	-0.0044	0.6136
	1200	-0.0016	-7.7728$\times10^{-5}$	-0.0027	0.1744
	300	-0.0014	0.0016	-0.0043	0.1019
2	600	-6.0914$\times10^{-4}$	2.7683$\times10^{-4}$	-0.0025	-0.0778
	1200	-4.8686$\times10^{-4}$	8.4685$\times10^{-5}$	-0.0016	0.3188
	300	-0.0013	2.2917$\times10^{-4}$	-0.0023	-0.0405
3	600	3.8673$\times10^{-4}$	-2.3072$\times10^{-4}$	-0.0019	0.5984
	1200	-1.4246$\times10^{-5}$	-9.7208$\times10^{-5}$	-0.0012	0.1447

Table 5: Simulation Results for Std: ρ=0.9

c_1	T	Model 1	Model 2	Model 3	Model 4
	300	0.0846	0.0749	0.0686	0.2547
1	600	0.0712	0.0617	0.0572	0.3875
	1200	0.0560	0.0497	0.0449	0.3072
	300	0.0885	0.0757	0.0697	0.6120
2	600	0.0714	0.0649	0.0571	0.6111
	1200	0.0565	0.0506	0.0453	0.4891
	300	0.0888	0.0794	0.0711	0.8469
3	600	0.0728	0.0667	0.0577	0.7894
	1200	0.0561	0.0524	0.0457	0.6373

Table 6: Simulation Results for Rmse: ρ=0.9

c_1	T	Model 1	Model 2	Model 3	Model 4
	300	0.0301	0.0484	0.0285	17.1768
1	600	0.0230	0.0335	0.0241	34.6006
	1200	0.0183	0.0251	0.0183	68.8479
	300	0.0327	0.0403	0.0333	17.3540
2	600	0.0262	0.0323	0.0269	34.8270
	1200	0.0197	0.0236	0.0197	68.9452
	300	0.0451	0.0570	0.0387	17.0256
3	600	0.0353	0.0328	0.0291	34.5640
	1200	0.0207	0.0240	0.0214	69.1129

From the above six tables, we can see that for the first three models, the bias, standard deviation and mean square error perform well that is consistent with our theory that Hermite functions can be used to approximate this kind of models, while for the fourth model, the simulation shows that the linear function can't be approximated well by the Hermite functions due to different asymptotic properties of homogeneous and integrable functions. As is well known in the literature, in each case, the bias of the series estimator decreases with larger p, while the standard deviation increases with larger p. From our simulation, we can see that generally the value of p does not need to be large to deliver a satisfactory result. Also, we find that there will be finite sample singularity problem when p is larger. For Models 1 and 3, when $c_1 = 1$, the estimator performs best by RMSE criterion. For Model 2, when $c_1 = 2$, the estimator performs the best. This is due to the fact that the Cauchy density function has heavier tail than the Laplace function and truncated function. Also, the truncation parameter can not be too large to avoid the singularity problem which leading to inaccurate results. We can also see that the existence of endogeneity does not affect the finite sample performance of the series estimator. This suggests that the theory may be extendable to the case where x_t and ε_t are correlated. We leave this as a future topic.

5 Empirical Study: Estimation of Stock Return Predictive Function

In this section, we will apply the estimator proposed here to estimate the stock return predictive function. Whether the stock returns are predictable or not is one of the most important questions in finance. It is not only of interest to the practitioners but also has implications for financial models of risk and returns. Meanwhile, the issue of return predictability is also very controversial because people usually reach different conclusions according to different data sample periods as well as different statistical inferential methods.

In the last few decades, based on in–sample evidence, a consensus has emerged among economists that the equity risk premium seems to vary considerably over the business cycle and perhaps at lower frequencies as well. Campbell (2000) concludes that the stock returns contain significant predictable components. However, the out–of–sample evidence of stock return predictability is much weaker. Goyal and Welch (2003; 2008) find that the predictive ability of a variety of popular financial and economic variables in the literature does not hold up in out–of–sample forecasting exercises. They conclude that "the profession has yet to find some variable that has meaningful and robust empirical equity premium forecasting power." Under the widely held view that predictive models require out–of–sample validation (e.g., Campbell (2008) argues: "The ultimate tests of any predictive model is its out of sample performance."), the findings of Goyal and Welch (2003; 2008) cast doubt on the reliability of return predictability.

In the study of predictive regression, affine functions are now pervasive in the literature, to name only a few, Stambaugh (1999), Lewellen (2004), Campbell and Yogo (2006). However, the linear model has some pitfalls. First of all, the linear model is derived from a simplified present value model, e.g., Campbell and Shiller (1988), which may not be a good approximation in reality. Engstrom (2003) shows that under a reasonable theoretical setting which just imposes the no arbitrage condition, the linear predictive relationship may not hold and structural treatment of risk may lead to nonlinear relationship between return and dividend yield. Secondly, there have been more and more empirical evidence suggesting that the nonlinear model may provide better approximation than the linear model. Lattau and Nieuwerburgh (2008) suggest that after controlling the structural shift in the mean of dividend yield, the evidence of return predictability is much stronger. Also, Gonzalo and Pitarakis (2012) show that the return predictability has different characteristics in different economic regimes. Moreover, Chen and Hong (2010) show that the nonparametric model can outperform the linear model. Thirdly, it is well known that the predicting financial ratios are highly persistent that we can not reject that they contain unit roots. However, the stock return is far less persistent, and thus is often regarded as a stationary process. A linear model will lead to an unbalanced relationship between left–hand–side and right–hand–side variables in predictive regression which is called model inconsistency by Granger (1995).

We propose using the Hermite series–based estimator to estimate the return predictive function. It is consensus in econometrics that the choice of an empirical model might be suggested by economic theory, and often by the requirement that the model is capable of generating the key characteristics of the data at hand. Note that the nonparametric integrable model can be regarded as a time–varying model as discussed in Engstrom (2003) where the parameter is smoothly co–evolving with dividend yield.

5.1 Data Description

In this paper, we will estimate the predictive function of dividend yield, book–to–market ratio and earning–price ratio to stock return. Price and dividends come from Center for Research in Security Prices (CRSP) dataset [3]. We focus on NYSE equal and value–weighted indices to be consistent with prior research. DY is calculated monthly on the value–weighted NYSE index, and it is defined as dividends paid over the prior year divided by current level of index. Thus DY is based on the rolling windows of annual dividends.

The returns data are from January, 1946 to December, 2007 with a total number of 744 data points. Here vwny denotes value–weighted NYSE stock return (nominal return), ewny denotes equal–weighed stock return, evwny denotes excess value–weighted stock return (real return or excess return) which is defined by value-weighted return minus t–bill rate, eewny denotes excess equal–weighted stock return. Dy and log(dy) respectively denote the dividend yield and logarithm of the dividend yield, have 744 data points. Bm denotes book–to–market ratio with 744 data points[4]. Ep denotes earning–price ratio with sample size 536 [5].

The sample mean, standard error, skewness, kurtosis and the first order autocorrelation coefficient are summarised in following table.

---

[3] We thank Professor Lewellen for providing us with his dataset.

[4] Book equity uses data set (http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/) available at Ken French's website.

[5] It is from Compustat, thus only starts from 1963.

Table 7: Summary Statistics

variables	mean	std	skewness	kurtosis	auto
vwny	0.0099	0.0401	-0.3830	4.8756	0.0414
ewny	0.0112	0.0470	-0.1863	6.8156	0.1417
evwny	0.0062	0.0403	-0.4119	4.8167	0.0483
eewny	0.0075	0.0472	-0.2163	6.7615	0.1463
dy	0.0359	0.0128	0.4047	2.5653	0.9904
log(dy)	-3.3948	0.3757	-0.3398	2.4429	0.9927
bm	0.6060	0.2046	0.0703	1.9156	0.9872
log(bm)	-0.5641	0.3676	-0.4561	2.2697	0.9908
ep	0.0657	0.0268	0.9577	2.9629	0.9877
log(ep)	-2.7977	0.3833	0.3150	2.3562	0.9876

From the table, we can find that there are little serial correlations in stock returns. Especially, the value–weighted return series is nearly white noise. At the same time, the forecasting dividend yield, book–to–market ratio and earning–price ratio are highly persistent with the first order autocorrelation close to 1. This suggests that there is no linear cointegration relationship between stock return and its predictors, but can not rule out the possibility that the the predictors can forecast stock return nonlinearly. The sample kurtosis of the return series suggests the stock return has heavy tail as is well known in the literature. The conventional ADF and KPSS tests suggest there is unit root in dividend yield, book–to–market ratio and earning–price ratio, while the stock returns are stationary. The scatter plots of value–weighted stock return v.s. logarithm of dividend yield, logarithm of book–to–market ratio, logarithm of earning–price ratios suggest there are no explicit patterns of forecasting relationship of these financial ratios to stock return (see the graph below for the scatter plot of value–weighted stock return v.s. logarithm of dividend yield).

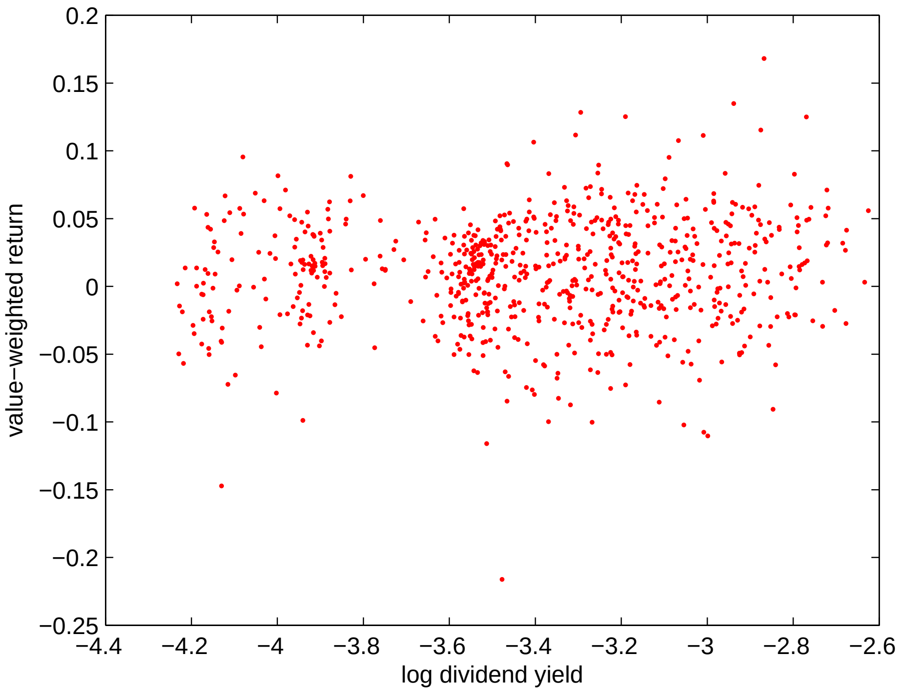
These empirical facts suggest the relationships of stock return and its predictors should not be linear and also the stock return inherently contains a sizable unpredictable component, so the best forecasting models can explain only a relatively small part of stock returns.

5.2 Evaluation Strategy

We estimate the predictive regression relationship using there types of models:

(Model 1) Historic mean model

$$r_{t+1} = \mu + \varepsilon_{t+1}. \tag{4.1}$$



(Model 2) Linear model

$$r_{t+1} = \mu + \beta x_t + \varepsilon_{t+1}. \tag{4.2}$$

(Model 3) Integrable model ,

$$r_{t+1} = \mu + \sum_{i=0}^{p-1} \theta_i F_i(x_t) + \varepsilon_{t+1}, \tag{4.3}$$

where $p = [c_1 T^{5/44}]$ with c_1 being chosen to be 1, 2, 3, x is log dividend yield, or, log book–to–market ratio, or log earning–price ratio. The performances of the estimators are evaluated based on out–of–sample criterion, because the out–of–sample performance provides more convincing evidence as discussed above.

In the literature of economic forecasting, there are basically two methods widely used. The first estimation method uses recursive (or expanding) window. The initial in–sample period is the first $[Tr]$ observations. Then, the estimator in each model is used to estimate next period return using next period predictors. The performance of different models is evaluated by its prediction accuracy through comparing the estimators with the true returns. The next in–sample period is the first $[Tr] + 1$ observations. And then a period ahead forecasting is made. The second estimation method uses a rolling window with the in–sample size always being $[Tr]$, i.e., in each recursion, one observation is abandoned and one new observation is added.

Rolling windows are typically justified by appealing to structure breaks. However, as demonstrated by Pesaran and Timmermann (2007), by criterion of the minimum square forecasting error, the optimal window size is a complicated function of timing and size of breaks. It is not easy to account for the effect of breaks and to choose the window size. The recursive window utilizes all the data when one forecasts next period return, and thus it may increase the precision of in–sample estimation, which in turn, leads to better out–of–sample forecasting. Considering the slow convergence rate of Hermite series estimation, this point is especially important. Although not reported here, the performance of the recursive window method is better than the rolling window for all the choice of r in our experiment for each model. Thus, we just report the results of the recursive window method.

The performance of out–of–sample performance is evaluated by a root mean squared forecasting error (RMSFE) of the form:

$$RMSFE = \sqrt{\frac{1}{T-[Tr]} \sum_{s=[Tr]+1}^{T} (r_s - \widehat{r}_s)^2}. \tag{4.4}$$

5.3 Predict Returns with Dividend Yield

Among instruments thought to capture some of the equity premium's variation in the sense of forecasting stock returns, the lagged dividend price ratio has emerged as a favourite. However, the existence of return predictability by dividend yield is also a very controversial issue. Fama and French (1988) find dividend–price ratio and book–to–market ratio are useful in predicting stock return. Similar results are reported in Campbell and Shiller (1988). Campbell (1991) and Cochrane (1992) attribute large fraction of the variance of dividend price ration to the variation in expected returns. However, as noticed by Mankiw and Shapiro (1986), Nelson and Kim (1993) and Stambaugh (1999), persistence of explaining variables and correlation of explaining variables and error terms lead to over–rejection of the null hypothesis of no predictability in finite sample, thus can be seriously biased towards finding predictability. Stambaugh (1999) finds that after controlling the finite sample bias, there is little evidence of stock return predictability, and the one–sided p–value is 0.15 when stock return is regressed on dividend yield over the period 1952-1996. Contrary to Stambaugh (1999), Lewellen (2004) finds that the evidence of predictability can be strong if we require the explaining variables to be stationary (i.e., the first order autoregressive parameter is less than 1). Using a robust test based on local to unit root framework, Campbell and Yogo (2006) find that the dividend–price ratio predicts returns at annual frequency, but they can not reject the null hypothesis of no predictability at quarterly and monthly frequencies.

The out–of–sample evidence of Goyal and Welch (2003; 2008) suggests that dividend yield can not outperform the historic mean model. However, Campbell and Thompson (2008) show that by imposing the restriction of the sign of regression coefficient, the dividend yield can predict stock returns. Also, Lettau and Nieuwerburgh (2008) suggest the dividend yield can predict stock returns by controlling the structural break in dividend yield. These papers lead us to conclude that the dividend yield should predict stock returns in a nonlinear fashion.

We apply the the evaluation strategy proposed in the last subsection to study the forecasting power of log dividend yield to stock return (nominal and real returns), the results are summarised in following four tables:

Table 8: Forecasting VWNY with Dividend Yield: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	4.0492	4.0583	4.0534	4.0810	4.0908
	1/6	4.0726	4.0689	4.0607	4.0899	4.1052
1946.1	1/4	4.1549	4.1470	4.1394	4.1678	4.1724
To	1/3	4.2260	4.2212	4.2126	4.2429	4.2462
2007.12	1/2	4.0519	4.0623	4.0481	4.0865	4.0841
	5/8	3.9586	3.9894	3.9780	4.0178	4.0182
	2/3	3.8941	3.9298	3.9100	3.9613	3.9597
	$[Tr]$=240	4.2253	4.2164	4.2084	4.2384	4.2411
	$[Tr]$=360	4.2392	4.2414	4.2276	4.2709	4.2723

Table 9: Forecasting EWNY with Dividend Yield: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	4.7287	4.7400	4.7352	4.7616	4.7808
	1/6	4.7844	4.7834	4.7751	4.8034	4.8303
1946.1	1/4	4.9069	4.9001	4.8922	4.9198	4.9356
To	1/3	4.9916	4.9845	4.9764	5.0045	5.0225
2007.12	1/2	4.4265	4.4325	4.4145	4.4574	4.4720
	5/8	4.1610	4.2000	4.1767	4.2182	4.2464
	2/3	4.1426	4.1883	4.1588	4.2109	4.2348
	$[Tr]$=240	4.9932	4.9837	4.9755	5.0044	5.0203
	$[Tr]$=360	4.4057	4.4109	4.3946	4.4341	4.4496

Table 10: Forecasting EVWNY with Dividend Yield: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	4.0698	4.0805	4.0715	4.1000	4.1071
	1/6	4.0937	4.0925	4.0800	4.1102	4.1221
1946.1	1/4	4.1754	4.1714	4.1590	4.1882	4.1896
To	1/3	4.2456	4.2473	4.2331	4.2648	4.2644
2007.12	1/2	4.0566	4.0834	4.0612	4.1013	4.0936
	5/8	3.9472	3.9833	3.9683	4.0123	4.0112
	2/3	3.8872	3.9290	3.9034	3.9597	3.9574
	$[Tr]$=240	4.2471	4.2438	4.2304	4.2615	4.2606
	$[Tr]$=360	4.0223	4.0475	4.0258	4.0655	4.0590

Table 11: Forecasting EEWNY with Dividend Yield: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	4.7481	4.7608	4.7518	4.7814	4.7973
	1/6	4.8041	4.8054	4.7925	4.8242	4.8475
1946.1	1/4	4.9262	4.9225	4.9098	4.9407	4.9528
To	1/3	5.0112	5.0090	4.9955	5.0272	5.0411
2007.12	1/2	4.4367	4.4572	4.4302	4.4790	4.4874
	5/8	4.1650	4.2090	4.1800	4.2293	4.2542
	2/3	4.1511	4.2025	4.1654	4.2259	4.2471
	$[Tr]$=240	5.0148	5.0094	4.9960	5.0283	5.0401
	$[Tr]$=360	4.4152	4.4344	4.4093	4.4548	4.4643

For the value–weighted return, if the initial estimation window is taken to be 20 years (e.g., Goyal and Welch 2008; Chen and Hong 2010; Rapach and Zhou 2012), the integrable model outperforms both the historic mean model and the linear model. If the initial estimation window is taken to be 30 years (e.g., Lattau and Nieuwerburgh 2008), the integrable model also outperforms both the historic mean model and the linear model. Hansen and Timmermann (2012) show that out–of–sample tests of predictive ability have better size properties when the forecast evaluation period is a relatively large proportion of the available sample. Using reasonable in-sample size and larger proportion of the sample in out–of–sample evaluation, i.e., $r = 1/6, 1/4, 1/3, 1/2$, the integrable model performs best when the truncation parameter is chosen to be small. Using different measures of return may lead to somehow different results, however, the conclusion that nonlinear predictive ability of dividend yield seems to be robust to different measures of return.

To be more specific, when $c_1 = 1$, the full–sample choice of p is equal to $[744^{5/44}] = 2$. The estimated regression is:

$$\begin{aligned} \widehat{r}_{t+1} &= 0.0065 + 2.2676 F_0(x_t) + 0.3379 F_1(x_t) \\ &= 0.0065 + (1.7032 + 0.4498\ x_t) \cdot \exp\left(-0.5\ x_t^2\right), \end{aligned} \tag{4.5}$$

where x_t is the log dividend yield, $F_0(x) = \frac{1}{\sqrt{\sqrt{\pi}}} \exp\left(-\frac{1}{2}x^2\right)$, $F_1(x) = \frac{\sqrt{2}}{\sqrt{\sqrt{\pi}}} x \cdot \exp\left(-\frac{1}{2}x^2\right)$, and r_{t+1} is the value–weighted stock return[6].

Model (4.5) shows that there is no support for neither the historic mean nor the simple linear model.

5.4 Predict Return with Book–to–Market Ratio or Earning–Price Ratio

Book–to–market ratio and earning–price ratio are two other commonly used stock return predictors. As shown previously and elsewhere in the literature, the two ratios are also very persistent. Using the annual data over the period of 1926–1991, Kothari and Shanken (1997) show that book–to–market ratio is helpful in predicting stock return. Pontiff and Schall (1998) show that the book-to-market ratio of the Dow Jones Industrial Average predicts market returns and small firm excess returns over the period of 1926-1994. Both papers also find that the predictability is not robust to different sub–samples. Using quarterly data over the period of the first quarter of 1947 to the last quarter of 1994, Lamont (1998) shows that earning price ratio can predict stock return. Generally speaking, the evidence of return predictability of these two ratios are weaker than dividend yield. Lewellen (2004) shows that these two ratios have limited predictive power.

The out–of–sample evidence by Goyal and Welch (2008) suggests that these two variables are not helpful in predicting returns. However, Campbell and Thompson (2008) show that

[6] Of course, we can also write the estimated regression function for the case where r is equal–weighed return or excess returns. To save space, we just report the results for the value–weighted return case.

earning–price ratio can forecast excess return by imposing sign constraints on the regression coefficient using monthly and annual data, while book–to–market can forecast excess return using the annual data. In this subsection, we will compare out–of–sample performance of the historic mean model, the linear model and the nonlinear integrable model using the same strategy as in the last subsection, the results are summarised in the following eight tables.

Table 12: Forecasting VWNY with Book–to–Market Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	4.0492	4.0641	4.0819	4.0869	4.1240
	1/6	4.0726	4.0832	4.1045	4.1069	4.1428
1946.1	1/4	4.1549	4.1600	4.1828	4.1835	4.2141
To	1/3	4.2260	4.2340	4.2566	4.2518	4.2750
2007.12	1/2	4.0519	4.0731	4.1006	4.1037	4.1420
	5/8	3.9586	4.0061	4.0365	4.0671	4.2542
	2/3	3.8941	3.9398	3.9581	4.0052	4.0423
	$[Tr]$=240	4.2253	4.2257	4.2491	4.2471	4.2710
	$[Tr]$=360	4.0174	4.0376	4.0649	4.0677	4.1044

Table 13: Forecasting EWNY with Book–to–Market Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	4.7287	4.7431	4.7646	4.7596	4.8069
	1/6	4.7844	4.7960	4.8210	4.8123	4.8588
1946.1	1/4	4.9069	4.9103	4.9373	4.9256	4.9650
To	1/3	4.9916	4.9906	5.0191	5.0012	5.0233
2007.12	1/2	4.4265	4.4337	4.4724	4.4474	4.4825
	5/8	4.1610	4.2007	4.2453	4.2424	4.2754
	2/3	4.1426	4.1825	4.2162	4.2197	4.2551
	$[Tr]$=240	4.9932	4.9881	5.0187	5.0057	5.0280
	$[Tr]$=360	4.4057	4.4119	4.4495	4.4289	4.4609

Table 14: Forecasting EVWNY with Book–to–Market Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	4.0698	4.0848	4.1029	4.1019	4.1388
	1/6	4.0937	4.1049	4.1263	4.1223	4.1579
1946.1	1/4	4.1754	4.1819	4.2047	4.1985	4.2289
To	1/3	4.2456	4.2582	4.2805	4.2685	4.2912
2007.12	1/2	4.0566	4.0904	4.1161	4.1106	4.1487
	5/8	3.9472	3.9996	4.0317	4.0541	4.0936
	2/3	3.8872	3.9388	3.9590	3.9965	4.0334
	$[Tr]$=240	4.2471	4.2502	4.2732	4.2643	4.2878
	$[Tr]$=360	4.0223	4.0549	4.0806	4.0741	4.1110

Table 15: Forecasting EEWNY with Book–to–Market Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	4.7481	4.7626	4.7845	4.7732	4.8208
	1/6	4.8041	4.8162	4.8416	4.8262	4.8728
1946.1	1/4	4.9262	4.9308	4.9580	4.9392	4.9789
To	1/3	5.0112	5.0133	5.0413	5.0161	5.0382
2007.12	1/2	4.4367	4.4546	4.4914	4.4575	4.4923
	5/8	4.1650	4.2096	4.2563	4.2422	4.2757
	2/3	4.1511	4.1969	4.2338	4.2245	4.2608
	$[Tr]$=240	5.0148	5.0112	5.0413	5.0211	5.0435
	$[Tr]$=360	4.4152	4.4317	4.4676	4.4375	4.4698

Table 16: Forecasting VWNY with Earning–Price Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	4.2694	4.2995	4.3543	4.8438	8.6921
	1/6	4.2179	4.2416	4.2953	4.6907	8.5831
1963.5	1/4	4.2512	4.2765	4.3366	4.9392	4.9789
To	1/3	4.0969	4.0838	4.0919	4.1446	4.2798
2007.12	1/2	3.9929	3.9975	4.0213	4.0222	4.0501
	5/8	3.5305	3.5496	3.5677	3.5727	3.6014
	2/3	3.5727	3.5828	3.6073	3.5928	3.6423
	$[Tr]$=240	3.9407	3.9418	3.9619	3.9734	4.0008
	$[Tr]$=360	3.5912	3.6005	3.6258	3.6091	3.6563

Table 17: Forecasting EWNY with Earning–Price Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	5.0249	5.0571	5.0842	5.4223	9.6600
	1/6	4.9193	4.9434	4.9656	5.1813	9.5667
1963.5	1/4	4.8053	4.8283	4.8547	5.0752	5.0157
To	1/3	4.4650	4.4554	4.4409	4.5015	4.5669
2007.12	1/2	4.1776	4.1958	4.1870	4.2122	4.2424
	5/8	3.6096	3.6661	3.6356	3.6683	3.6906
	2/3	3.6633	3.6993	3.6776	3.6975	3.7365
	$[Tr]$=240	4.1620	4.1775	4.1642	4.1899	4.2296
	$[Tr]$=360	3.6842	3.7162	3.6965	3.7152	3.7513

Table 18: Forecasting EVWNY with Earning–Price Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	4.2854	4.3289	4.3847	4.8713	8.6528
	1/6	4.2309	4.2691	4.3236	4.7164	8.5361
1963.5	1/4	4.2583	4.3017	4.3638	4.7242	4.4774
To	1/3	4.1021	4.1061	4.1139	4.1630	4.3074
2007.12	1/2	3.9898	4.0015	4.0206	4.0248	4.0580
	5/8	3.5298	3.5551	3.5716	3.5756	3.6114
	2/3	3.5696	3.5875	3.6087	3.5941	3.6512
	$[Tr]$=240	3.9377	3.9466	3.9621	3.9754	4.0074
	$[Tr]$=360	3.5880	3.6050	3.6268	3.6101	3.6654

Table 19: Forecasting EEWNY with Earning–Price Ratio: RMSFE×100

Sample	r	Model 1	Model 2	Model 3		
				c_1=1	c_2=2	c_3=3
	1/8	5.0460	5.0939	5.1227	5.4579	9.6246
	1/6	4.9368	4.9783	5.0021	5.2156	9.5253
1963.5	1/4	4.8187	4.8628	4.8923	5.1052	5.0483
To	1/3	4.4780	4.4879	4.4749	4.5321	4.6044
2007.12	1/2	4.1817	4.2061	4.1952	4.2229	4.2572
	5/8	3.6172	3.6767	3.6476	3.6795	3.7072
	2/3	3.6687	3.7097	3.6868	3.7081	3.7533
	$[Tr]$=240	4.1676	4.1895	4.1745	4.2010	4.2453
	$[Tr]$=360	3.6891	3.7267	3.7050	3.7258	3.7685

From the above tables, we can see that the predictive power of these two ratios are much weaker than that of dividend yield, regardless of whether linear or nonlinear models are used. The forecasting ability of these two ratios are restricted to only some special cases (with colour different from black) for each return series. Thus, the predictability of these two series, if exists, should be very weak.

6 Conclusion and Discussion

In this paper, we have established the uniform consistency and asymptotic normality of the Hermite series estimator of the proposed integrable cointegration model. The application of the estimator to stock return predictive models suggests existence of nonlinear predictive relationship between dividend yield and stock return. However, the book–to–market ratio and earning–price ratio may have neither linear nor nonlinear predictive power.

The theory and application of this paper can be extended. The choice of the truncation parameter should be discussed in more detail and a data driven choice of the truncation parameter should be discussed. The theory can be extended to an additive multivariate model with both stationary and nonstationary regressors or a partially linear cointegration model. The application can be extended to data sharing similar characteristics to return and dividend yield. For example, we may study the relationship of return of exchange rate and forward rate premium, where the exchange rate return is far less persistent than the forward rate premium.

7 Acknowledgements

The original version of this paper was presented at a lunch time seminar in the Department of Econometrics and Business Statistics, Monash University, Australia. The first author acknowledges some useful comments by the audience. The authors also acknowledge financial support by the Australian Research Council Discovery Grants Program under Grant Number: DP1096374.

8 Appendix

8.1 Appendix A. Proofs of Section 2.

Lemma A.1 Let $f(x)$ satisfy Assumption 3 and $F_j(x)$ be Hermite functions and θ_j be the projection coefficients of $f(x)$ on $F_j(x)$. Then, we have $\sup_x |f(x) - \sum\limits_{j=0}^{p-1} \theta_j F_j(x)| = O\left(p^{-\frac{r}{2}+1}\right)$.

Proof. Under Assumption 3, Lemma 3 of Schwartz (1967), the coefficients involved in the expansion $f(x) = \sum_{j=0}^{\infty} \theta_j F_j(x)$ satisfies

$$|\theta_j| < c_3(r)/(2j)^{r/2}, \tag{4.1}$$

where $c_3(r)$ is L_2 norm of $e^{x^2/2}(d^r/dx^r)[e^{-\frac{x^2}{2}} f(x)]$, which is bounded by Assumption 3.

Because $\{F_j(x)\}_{j=0,1,2...}$ are bounded uniformly in j and x, we have

$$\sup_x \left| \sum_{j=p}^{\infty} \theta_j F_j(x) \right| \le \sup_x \sum_{j=p}^{\infty} |\theta_j||F_j(x)| \le C \sum_{j=p}^{\infty} |\theta_j| \le C_2 \sum_{j=p}^{\infty} (2j)^{-r/2} = Cp^{-r/2+1}, \tag{4.2}$$

which completes the proof.

Proof of Theorem 1.

It's easy to show that

$$\begin{aligned} \left|\left|(F^{\tau}F)^{-1}F^{^{\tau}}\varepsilon\right|\right| &= \left\{\varepsilon^{\tau}F(F^{\tau}F)^{-2}F^{\tau}\varepsilon\right\}^{1/2} & (4.3) \\ &\leq \frac{1}{\sqrt{\sqrt{T}}}\left(1+O_P(1)\right)\{\varepsilon^{^{\tau}}F\left(F^{\tau}F\right)^{-1}F^{^{\tau}}\varepsilon\}^{1/2} \\ &= \frac{1}{\sqrt{\sqrt{T}}}\left(1+O_P(1)\right)\cdot\sqrt{p} = O_P\left(\sqrt{\sqrt{T}p}\right) & (4.4) \end{aligned}$$

using the result that $E\left[\varepsilon^{^{\tau}}FF^{^{\tau}}\varepsilon\right] = O(\sqrt{T}p)$ as proved in Theorem 3.1 of Dong and Gao (2012).

Meanwhile, by Theorem 4, we have

$$\begin{aligned} &\left|\left|[\frac{1}{\sqrt{T}}F^{\tau}F]^{-1}F^{\tau}(A-F\theta)/\sqrt{T}\right|\right| \\ &= \left\{(A-F\theta)^{\tau}/\sqrt{T}F[\frac{1}{\sqrt{\tau}}F^{\tau}F]^{-1}[\frac{1}{\sqrt{T}}F^{\tau}F]^{-1}F^{\tau}(A-F\theta)^{\tau}/\sqrt{T}\right\}^{1/2} \\ &\leq O_P(1)\left\{(A-F\theta)^{\tau}F[F^{T}F]^{-1}F^{T}(A-F\theta)/\sqrt{T}\right\}^{1/2} \\ &\leq O_P(1)\left\{(A-F\theta)^{\tau}(A-F\theta)/\sqrt{T}\right\}^{1/2} & (4.5) \end{aligned}$$

by using bounds of Rayleigh quotient, e.g., Exercise 7.53 of Abadir and Magnus (2005).

And (4.5) can be bounded by

$$\begin{aligned} &O_P(1)\cdot\left\{\sum_{i=1}^{T}(f(x_i)-\sum_{j=0}^{p-1}F_j(x_i)\theta_j)^2/\sqrt{T}\right\}^{1/2} \\ &\leq O_P(1)\cdot\{Tp^{-r+2}/\sqrt{T}\}^{1/2} = O_P(T^{1/4}p^{-r/2+1}). \end{aligned}$$

Therefore, by

$$\widehat{\theta}-\theta = [\frac{1}{\sqrt{T}}F^{\tau}F]^{-1}F^{\tau}(Y-A)/\sqrt{T} + [\frac{1}{\sqrt{T}}F^{\tau}F]^{-1}F^{\tau}(A-F\theta)/\sqrt{T},$$

we have

$$\begin{aligned} ||\widehat{\theta}-\theta|| &\leq \left|\left|[\frac{1}{\sqrt{T}}F^{\tau}F]^{-1}F^{\tau}\varepsilon/\sqrt{T}\right|\right| + \left|\left|[\frac{1}{\sqrt{T}}F^{\tau}F]^{-1}F^{\tau}(A-F\theta)/\sqrt{T}\right|\right| \\ &\leq O_P(p^{1/2}/T^{1/4}+T^{1/4}p^{-r/2+1}) \to 0 \end{aligned}$$

by Assumption 4.

Proof of Corollary 1. Note that

$$\sup_x|\widehat{f}(x)-f(x)| \leq \sup_x|Fp^{\tau}(x)(\widehat{\theta}-\theta)| + \sup_x|Fp^{\tau}(x)\theta-f(x)|, \quad (4.6)$$

By the uniformly boundedness of Hermite functions, we then have

$$\sup_x \|Fp(x)\| \leq Cp^{1/2}.$$

Thus, applying Cauchy Schwarz inequality and Theorem 1 for the first term and Lemma A.1 for the second term, we have

$$\begin{aligned} & \sup_x |Fp^\tau(x)(\widehat{\theta}-\theta)| + \sup_x |Fp^\tau(x)\theta - f(x)| \\ \leq\ & \sup_x \|Fp^\tau(x)\| \, ||\widehat{\theta}-\theta|| + \sup_x |Fp^\tau(x)\theta - f(x)| \\ \leq\ & O_P(p^{1/2}p^{1/2}/T^{1/4} + p^{1/2}T^{1/4}p^{-r/2+1} + p^{-r/2+1}) \to 0 \end{aligned}$$

by assumption A4.

Proof of Theorem 2. Observe that

$$\begin{aligned} & \sqrt[4]{T}\left(\sigma^2 \sum_{i=0}^{p-1} F_i^2(x)/L(1,0)\right)^{-\frac{1}{2}} \left[\widehat{f}(x) - f(x)\right] \\ =\ & \sqrt[4]{T}\left(\sigma^2 \sum_{i=0}^{p-1} F_i^2(x)/L(1,0)\right)^{-\frac{1}{2}} (A_1 + A_2). \end{aligned}$$

To complete the proof, it suffices to show that

$$\begin{aligned} & \sqrt[4]{T}(\sigma^2 \sum_{i=0}^{p-1} F_i^2(x)/L(1,0))^{-\frac{1}{2}} A_1 \to_D N(0,1), \\ & \sqrt[4]{T}(\sigma^2 \sum_{i=0}^{p-1} F_i^2(x)/L(1,0))^{-\frac{1}{2}} A_2 = o_P(1), \end{aligned} \tag{4.7}$$

where $A_1 = Fp^\tau(x)[\widehat{\theta}-\theta]$ and $A_2 = \sum_{j=p}^{\infty} \theta_j F_j(x)$.

We first show that $\sqrt[4]{T}\left(\sigma^2 \sum_{i=0}^{p-1} F_i^2(x)\right) A_2 = o_P(1)$. Note that

$$\begin{aligned} & \sqrt[4]{T}\left(\sigma^2 \sum_{i=0}^{p-1} F_i^2(x)\right)^{-\frac{1}{2}} A_2 = \sqrt[4]{T}\left(\sigma^2 \sum_{i=0}^{p-1} F_i^2(x)\right)^{\frac{-1}{2}} . \sum_{j=p}^{\infty} \theta_j F_j(x) \\ \leq\ & O_P\left(\sqrt[4]{T}\ p^{-1/2}p^{-r/2+1}\right) = O_P(\sqrt[4]{T}\ p^{-\frac{r-1}{2}}\right) = o_P(1) \end{aligned}$$

by Assumption 4.

We then show that

$$\sqrt[4]{T}(\sigma^2 \sum_{i=0}^{p-1} F_i^2(x)/L(1,0))^{-\frac{1}{2}} A_1 \to_D N(0,1).$$

Observe that

$$\sqrt[4]{T}\left(\sigma^2\sum_{i=0}^{p-1}F_i^2(x)/L(1,0)\right)^{-\frac{1}{2}}A_1=\sqrt[4]{T}\left(\sigma^2\sum_{i=0}^{p-1}F_i^2(x)/L(1,0)\right)^{\frac{-1}{2}}(A_{11}+A_{12}),$$

where

$$A_{11}=\left\{Fp^\tau(x)\sum_{t=1}^{T}[Fp(x_t)Fp(x_t)^\tau]^{-1}\sum_{t=1}^{T}Fp(x_t)\varepsilon_t\right\},$$
$$A_{12}=\left\{Fp^\tau(x)\left[\sum_{t=1}^{T}Fp(x_t)Fp(x_t)^\tau\right]^{-1}\sum_{t=1}^{T}Fp(x_t)\left[\sum_{j=0}^{\infty}\theta_jF_j(x_t)\right]-\theta\right\}.$$

We start to prove that

$$\sqrt[4]{T}\left(\sigma^2\sum_{i=0}^{p-1}F_i^2(x)/L(1,0)\right)^{-\frac{1}{2}}A_{11}\to_D N(0,1),$$
$$\sqrt[4]{T}\left(\sigma^2\sum_{i=0}^{p-1}F_i^2(x)\right)^{-\frac{1}{2}}A_{12}\to_P 0.$$

Meanwhile, by Assumption 4 and the proof of Theorem 1, we have

$$\begin{aligned}&\sqrt[4]{T}\left(\sigma^2\sum_{i=0}^{p-1}F_i^2(x)\right)^{-\frac{1}{2}}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Fp(x_t)Fp(x_t)^\tau\right)^{-1}\\&\times\frac{1}{\sqrt{T}}\sum_{t=1}^{T}Fp(x_t)\sum_{j=p+1}^{\infty}\theta_jF_j(x_t)\\&\le CT^{1/4}p^{-1/2}O_P(T^{1/4}p^{-r/2+1})=o_P\left(T^{1/2}p^{-r/2+1/2}\right)\to 0.\end{aligned}$$

We are then left to show that

$$\sqrt[4]{T}\left(\sigma^2\sum_{i=0}^{p-1}F_i^2(x)/L(1,0)\right)^{-\frac{1}{2}}A_{11}\to_D N(0,1).$$

Observe that

$$\begin{aligned}&\sqrt[4]{T}\left(\sigma^2\sum_{i=0}^{p-1}F_i^2(x)/L(1,0)\right)^{-\frac{1}{2}}\left\{Fp^\tau(x)\left[\sum_{t=1}^{T}Fp(x_t)Fp(x_t)^\tau\right]^{-1}\sum_{t=1}^{T}Fp(x_t)\varepsilon_t\right\}\\&=\left\{\left(\sigma^2L(1,0)\sum_{i=0}^{p-1}F_i^2(x)\right)^{-1/2}Fp^\tau(x)\frac{1}{\sqrt[4]{T}}\sum_{t=1}^{T}Fp(x_t)\varepsilon_t\right\}(1+o_P(1)).\end{aligned}$$

Thus, we need only to show that

$$\left(\sigma^2 \sum_{i=0}^{p-1} F_i^2(x) L(1,0)\right)^{-\frac{1}{2}} Fp^\tau(x) \frac{1}{\sqrt[4]{T}} \sum_{t=1}^{T} Fp(x_t)\varepsilon_t$$

$$\equiv L(1,0)^{-1/2}\sigma^{-1} \frac{1}{\sqrt[4]{T}} \sum_{t=1}^{T} m(x_t)\varepsilon_t \to_D N(0,1),$$

where $m(y) = \left[\sum\limits_{i=0}^{p-1} F_i^2(x)\right]^{-1/2} \sum\limits_{i=0}^{p} F_i(x)F_i(y)$.

To show this, we adopt similar approach as in Park and Phillips (2001). In the expanded space as discussed in the remark for Assumptions 1 and 2, we have $(W_{1Tr}, W_{2Tr}) \to_{a.s.} (W_{1r}, W_{2r})$.

Define

$$M_T(r) = \sqrt[4]{T} \sum_{t=1}^{k-1} m\left(\sqrt{T} W_{1T}\left(\frac{t-1}{T}\right)\right)\left(W_{2T}\left(\frac{\tau_{T,t}}{T}\right) - W\left(\frac{\tau_{T,t-1}}{T}\right)\right)$$
$$+ \sqrt[4]{T} m\left(\sqrt{T} W_{1T}\left(\frac{k-1}{T}\right)\right)\left(W_{2T}(r) - W\left(\frac{\tau_{T,k-1}}{T}\right)\right),$$

where $\tau_{T,t}$, $t = 1, \ldots T$ are stopping times such that $\tau_{T,k-1}/T < r \leq \tau_{T,k}/T$, and it can be seen that $M_T(r)$ is a continuous martingale such that

$$\frac{1}{\sqrt[4]{T}} \sum_{t=1}^{T} m(x_t)\varepsilon_t = M_T\left(\frac{\tau_{T,T}}{T}\right),$$
$$\sup\left|\left(\frac{\tau_{T,t}}{T} - \frac{\tau_{T,t-1}}{T}\right) - \frac{1}{T}\right| = o_{a.s.}(1).$$

The quadratic variation process $[M_T]$ of M_T is given by

$$[M_T]_r = \sqrt{T} \sum_{t=1}^{k}\left[\left(m\left(\sqrt{T} W_{1T}\left(\frac{t-1}{T}\right)\right)\right)^2\left(\frac{\tau_{T,t}}{T} - \frac{\tau_{T,t-1}}{T}\right)\right]$$
$$+ \sqrt{T} \sum_{t=1}^{k}\left(mF_i\left(\sqrt{T} W_{1T}\left(\frac{k-1}{T}\right)\right)\right)^2\left(r - \frac{\tau_{T,k-1}}{T}\right).$$

Note that $m^2(y) = (\sum\limits_{i=0}^{p} \lambda_i F_i(y))^2$, where $\lambda_i = [\sum\limits_{i=0}^{p-1} F_i^2(x)]^{-1/2} F_i(x)$. By Cauchy Schwarz inequality, we have

$$m^2(y) \leq \sum_{i=0}^{p} \lambda_i^2 \sum_{i=0}^{p} F_i^2(y) = \sum_{i=0}^{p} F_i^2(y).$$

Meanwhile, similarly to the proof of Theorem 3.1 of Dong and Gao (2012)[7], we have as $T \to \infty$

$$\frac{\Psi}{\sqrt{T}p} \sum_{t=1}^{T} \sum_{i=0}^{p} F_i^2(x_t) \to_P L(1,0). \tag{4.8}$$

[7]In their proof, they notice that (normalization of) the summation of infinite series of quadratic of Hermite functions converges to the so-called Ullman density, and thus a two–step procedure is used to prove the convergence result.

By the dominated convergence theorem, we have

$$[M_T]_r \to_P \sigma^2 \int \left(\sum_{i=0}^{p-1} \lambda_i F_i(s) \right)^2 ds L(r,0)(1 + o_{a.s.}(1))$$

$$= \quad \sigma^2 \sum_{i=0}^{p-1} \lambda_i^2 \int F_i^2(s) ds \; L(r,0) + o_P(1),$$

uniformly in $r \in [0,1]$. Thus, we have

$$[M_T]_1 \to_P \sum_{i=0}^{p-1} \lambda_i^2 \int F_i^2(s) ds L(1,0) = \sigma^2 \sum_{i=0}^{p-1} \lambda_i^2 F_i^2(x) L(1,0) = \sigma^2 L(1,0).$$

Moreover, if we denote $[M_n, W_{1T}]$ as the covariation process of M_n and W_{1T}, then we have

$$[M_n, W_{1T}]_r = \sqrt[4]{T} \sum_{t=1}^{k} m \left(\sqrt{T} W_{1T} \left(\frac{t-1}{T} \right) \right) \left(\frac{\tau_{T,t}}{T} - \frac{\tau_{T,t-1}}{T} \right) \sigma_{12}$$

$$+ \sqrt[4]{T} \sum_{t=1}^{k} m \left(\sqrt{T} W_{1T} \left(\frac{t-1}{T} \right) \right) \left(r - \frac{\tau_{T,k-1}}{T} \right) \sigma_{12},$$

$$= \sigma_{12} \sqrt[4]{T} \int_0^r m \left(\sqrt{T} W_1(s) \right) ds \cdot (1 + o_{a.s.}(1)),$$

where σ_{12} is the covariance of W_1 and W_2.

In addition, $p^{-11/12} m(y)$ is integrable and square integrable by Lemma B.7 listed in Appendix B below. Then, for all $r \in [0,1]$,

$$\left| \sqrt[4]{T} \int_0^r m(\sqrt{T} W_1(s)) ds \right| \leq \frac{1}{\sqrt[4]{T}} \left(\sqrt{T} \int \left| m \left(\sqrt{T} W_1(s) \right) \right| ds \right) = O_P(p^{11/12} / \sqrt[4]{T}).$$

According to Assumptions 1 and 2, we have

$$[M_n, W_{1T}]_r \leq O_P(p^{11/12} / \sqrt[4]{T}) = o_P(1).$$

Therefore, we have shown

$$\frac{1}{\sqrt[4]{T}} \sum_{t=1}^{T} m(x_t) \varepsilon_t \to_D MN(0, \sigma L(1,0)^{1/2}), \tag{4.9}$$

which implies

$$L(1,0)^{-1/2} \sigma^{-1} \frac{1}{\sqrt[4]{T}} \sum_{t=1}^{T} m(x_t) \varepsilon_t \to_D N(0,1). \tag{4.10}$$

Therefore, the proof of Theorem 2 is now completed.

Proof of Theorem 3. We have

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_0^2(x_t) \to_P L(1,0) \tag{4.11}$$

by Wang and Phillips (2009a).

Simple derivations imply as $T \to \infty$

$$\frac{1}{T}\sum_{t=1}^{T} y_t^2 = \frac{1}{T}\sum_{t=1}^{T}(f^2(x_t) + \varepsilon_t^2 + 2f(x_t)\varepsilon_t) \to_P \sigma^2,$$

because $\frac{1}{T}\sum\limits_{t=1}^{T} f^2(x_t) = O_P(T^{-1/2})$, $\frac{1}{T}\sum\limits_{t=1}^{T} 2f(x_t)\varepsilon_t = O_P(T^{-3/4})$ and $\frac{1}{T}\sum\limits_{t=1}^{T} \varepsilon_t^2 \to_P \sigma^2$. This completes the proof of Theorem 3.

8.2 Appendix B: Proof of Theorem 4.

We start some lemmas required to complete the proof of Theorem 4. The proofs of the lemmas are given before the proof of Theorem 4.

Lemma B.1. Let $\zeta_t = \sum\limits_{j=0}^{\infty} \pi_j e_{t-j}$ and $x_t = \sum\limits_{j=1}^{t} \zeta_t$, where $\{e_t\}$ is i.i.d. sequence with mean zero and variance 1 and $E(|e|_1^q) < \infty$ for some $q > 2$,, in which $\pi_0 = 1$, $\sum\limits_{j=0}^{\infty} |\pi_j| < \infty$ and $\sum\limits_{j=1}^{\infty} \pi_j = \Psi \neq 0$. Decompose $x_t = \Psi S_t + R_t^{'} + R_t^{''}$. Then, in an extended probability space, with $T^{1/q} \le h_T \le C_1(T\log T)^{1/2}$, we have

(1) $P(\max_{0\le k\le T} |S_k - W(k)| \ge h_T) \le C_2 T h_T^{-q}$ and $\max_{0\le k\le T} |S_k - W(k)| = o(T^{1/q})$ a.s.;

(2) $\forall\, T$, $P(\max_{1\le k\le T} |R_k^{'}| \ge h_T) \le C T h_T^{-q}$, $P(\max_{1\le k\le T} |R_k^{''}| \ge h_T) \le C T h_T^{-q}$ and $E(|R_k^{'}|^q) < \infty$ and $E(|R_k^{''}|^q) < \infty$.

Proof: The argument (1) follows from Komlos et al. (1976) or Einmahl (1989). The proof of argument (2) can be found in proof of Lemma 2 of Akonom (1993).

Lemma B.2 (Rothental Inequality) Let $q \ge 2$, and $\{w_i\}_{i=1,\dots,T}$ be a sequence of independent random variables with zero mean and $E|w_i^q| < \infty$. Then we have

$$E\left\{\left|\sum_{i=1}^{T} w_i\right|^q\right\} \le c_q \max\left\{\sum_{i=1}^{T} E\left[|w_i|^q\right], \left(\sum_{i=1}^{T} E\left(w_i^2\right)\right)^{q/2}\right\},$$

where c_q is constant depending on q.

Proof: See Rothental (1970).

Lemma B.3. Let $\zeta_t = \sum\limits_{j=0}^{\infty} \pi_j e_{t-j}$ and $x_t = \sum\limits_{j=1}^{t} \zeta_t$, where $\{e_t\}$ is i.i.d. sequence with mean zero and variance 1 and $E(|e_1|^q) < \infty$ for some $q > 2$, in which $\pi_0 = 1$, $\sum\limits_{j=0}^{\infty} |\pi_j| < \infty$ and $\sum\limits_{j=1}^{\infty} \pi_j = \Psi \neq 0$.

Then $\sup_r \, |\frac{1}{\Psi\sqrt{T}}x_{[Tr]} - W_1(r)| \to_P 0$ in a suitably extended space for some $W_1(r)$, and $\sup_r |\frac{1}{\sqrt{T}}\sum_{t=1}^{[Tr]} e_t - W_1(r)| \to_P 0$. That is, the random walk generated by the innovation process converges in probability to the same limiting process.

Proof: We have following decompositions:

$$x_t = \sum_{k=1}^{t} \zeta_k = \sum_{k=1}^{t}(\sum_{j=0}^{\infty} \pi_j e_{k-j}) = \pi_0 e_t + (\pi_0 + \pi_1)e_{t-1} + ... + (\pi_0 + \pi_1 + ... + \pi_{t-1})e_1$$
$$+ \sum_{k=0}^{\infty}(\pi_{k+1} + ... + \pi_{k+t})e_{-k} = \Psi S_t + R_t^{'} + R_t^{''},$$

where $R_t' = \sum_{k=1}^{t}(\Psi_{t-k} - \Psi)e_k$ and $R_t'' = \sum_{k=0}^{\infty}(\Psi_{t+k} - \Psi_k)e_{-k}$, in which $\Psi_k = \sum_{j=0}^{k}\pi_j$.

Thus, to show the result, it suffices to show that:

$$\max_{0\le t\le T}\left|\frac{1}{\sqrt{T}}R_t^{'}\right| \to_P 0 \;\text{ and }\; \max_{0\le t\le T}\left|\frac{1}{\sqrt{T}}R_t^{''}\right| \to_P 0.$$

According to Lemma 2 of Akonom (1993), we have

$$\max_{0\le t\le T}\left|R_t^{'}/\sqrt{T}\right| = o_P(T^{1/q-1/2}) \;\text{ and }\; \max_{0\le t\le T}\left|R_t^{''}/\sqrt{T}\right| = o_P(T^{1/q-1/2}).$$

This completes the proof of Lemma B.3.

Lemma B.4. Let $e_1, ..., e_T$ be a sequence of i.i.d. random variables with mean zero and variance 1, and $E(|e_1|^{6+\delta}) < \infty$ for some $\delta > 0$. One can construct a probability space (Ω, P, F) and random variable S_T and U_T such that $S_T = \sum_{t=1}^{T} e_t$ and $U_T = \sum_{t=1}^{T} u_t$ with u_t being i.i.d $N(0, 1)$ such that $E[|S_T - U_T|^2] \le CT^{2/3}$.

Proof: Let $E[|e_i|^q] < \infty$ for some $q \ge 1$. According to Akonomn (1993 p. 77), for h_T satisfying $T^{1/q} \le h_T \le C_1(T\log T)^{1/2}$, we can construct a probability space (Ω, P, F) such that

$$P(\max_{0\le k\le T}|e_1 + ... + e_k - U_k| \ge h_T) \le C_2 T h_T^{-q}.$$

Because $\{|S_T - U_T| \ge h_T\} \subseteq \{\max_{0\le k\le T}|e_1 + ... + e_k - U_k| \ge h_T\}$, we have

$$P(|S_T - U_T| \ge h_T) \le CTh_T^{-q},$$
$$E(|S_T - U_T|^2) \le h_T^2 + E[|S_T - U_T|^2 1_{\{|S_T - U_T| > h_T\}}]$$
$$\le\; h_T^2 + [E|S_T - U_T|^3]^{2/3}(P\{|S_T - U_T| > h_T\})^{1/3}.$$

Letting $h_T = T^\alpha$ with $\frac{2}{q} \le \alpha < \frac{1}{3}$, we have

$$(P\{|S_T - U_T| > \lambda\})^{1/3} \le [CTh_T^{-q}]^{1/3} = C^{1/3}T^{\frac{1}{3}(1-\alpha q)}.$$

By Minkowski inequality, we then have

$$E[|S_T - U_T|^3]^{2/3} \le 2\{E[|S_T|^3]^{2/3} + [E|U_T|^3]^{2/3}\}.$$

By Rothental inequality in Lemma B.2, we obtain

$$E|S_T|^3 \le C_1 T^{3/2},$$

and due to the property of Normal distribution,

$$[E|U_T|^3] \le CT^{3/2}.$$

Thus, $[E|S_T - U_T|^3]^{2/3} \le CT$. This, together with

$$[E|S_T - U_T|^3]^{2/3}(P\{|S_T - U_T| > h_T\})^{1/3} \le CT^{\frac{4}{3}-\frac{\alpha q}{3}},$$

implies

$$E(|S_T - U_T|^2) \le T^{2\alpha} + CT^{\frac{4}{3}-\frac{\alpha q}{3}} \le C\,T^{\frac{2}{3}}$$

by the choice of α. This completes the proof of Lemma B.4.

Lemma B.5. Write $\Psi_i = \sum_{j=0}^{i} \pi_j$, $\widetilde{S}_n = \sum_{i=0}^{n} \Psi_i e_i \equiv \sum_{i=0}^{n} \widetilde{e}_i$, and $\Lambda_n^2 = \sum_{i=0}^{n} \Psi_i^2$. Define $f_n(t) = Ee^{it\widetilde{S}_n/\Lambda_n}$ *and* let $h_n(x)$ be the density function of $\widetilde{S}_n/\Lambda_n$. Then

(1) $h_n(x)$ is uniformly bounded, and $E(\widetilde{S}_n)^2 = \Lambda_n^2$ implies $\widetilde{S}_n/\Lambda_n \to^d N(0,1)$ as $n \to \infty$.

(2) $\sup_x |h_n(x) - n(x)| = O(\frac{1}{\sqrt{n}})$, where $n(x)$ is the density of $U \sim N(0,1)$.

Proof: The first argument is due to the proof of Corollary 2.2 of Wang and Phillips (2009a). Now, we prove the second argument. By Fourier inverse transformation, we have

$$\sup_x |h_n(x) - n(x)| \le \frac{1}{2\pi} \int_{-\infty}^{\infty} |f_n(t) - e^{-t^2/2}| dt,$$

$$\frac{1}{2\pi} \int_{-\infty}^{\infty} |f_n(t) - e^{-t^2/2}| dt = \frac{1}{2\pi} \int_{|t|\le A} |f_n(t) - e^{-t^2/2}| dt + \frac{1}{2\pi} \int_{|t|>A} |f_n(t) - e^{-t^2/2}| dt$$

$$\equiv I_{1n} + I_{2n}.$$

It follows similarly from proof of Corollary 2.2 of Wang and Phillips (2009a), we have

$$I_{2n} \le 2 \int_{|t|>A} e^{-t^2/8} dt + C\eta^{n/2-1},$$

with $0 < \eta < 1$.

Notice that

$$
\begin{aligned}
e^{-t^2/2} &= e^{-\frac{t^2}{2}\cdot\frac{\sum_{j=0}^{n}(\Psi_j)^2}{\Lambda_n^2}} = \Pi_{j=0}^{n} e^{-\frac{t^2}{2}\cdot\frac{\Psi_j^2}{\Lambda_n^2}}, \\
E\left[e^{\frac{it\widetilde{S}_n}{\Lambda_n}}\right] &= \Pi_{j=0}^{n} E\left[e^{\frac{it\widetilde{e}_j}{\Lambda_n}}\right].
\end{aligned}
$$

We know that[8] for any complex value $|a_j| \le 1$ and $|b_j| \le 1$ for $j = 1, ..., n$,

$$
|a_1 ... a_n - b_1 ... b_n| \le \sum_{j=1}^{n} |a_j - b_j| .
$$

Thus, we have

$$
\begin{aligned}
&\left|Ee^{it\widetilde{S}_n/\Lambda_n} - e^{-t^2/2}\right| = e^{-t^2/2}\left|E\left[e^{it\widetilde{S}_n/\Lambda_n}\right]/e^{-t^2/2} - 1\right| \\
&= e^{-t^2/2}\left|\frac{\prod_{j=0}^{n} E\left[e^{it\widetilde{e}_j/\Lambda_n}\right]}{\prod_{j=0}^{n} e^{-\frac{t^2}{2}\cdot\frac{\Psi_j^2}{\Lambda_n^2}}} - 1\right| \le e^{-\frac{t^2}{2}} \sum_{j=0}^{n} \left|\frac{E\left[e^{\frac{it\widetilde{e}_j}{\Lambda_n}}\right]}{e^{-\frac{t^2}{2}\cdot\frac{\Psi_j^2}{\Lambda_n^2}}} - 1\right|.
\end{aligned}
$$

When $t < \min_j \sqrt{2\frac{\Lambda_n^2}{(\Psi_j)^2}} = C\sqrt{n}$ (due to the fact that $\Lambda_n^2 \sim d_n^2 \sim \Psi^2 n$, see Wang and Phillips (2009a, p. 731.), we have

$$
\begin{aligned}
&E\left[e^{it\widetilde{e}_j/\Lambda_n + \frac{t^2}{2}\cdot\frac{\Psi_j^2}{\Lambda_n^2}}\right] = 1 - \frac{(\Psi_j)^2 t^2}{2\Lambda_n^2} - i\frac{t^3 E(|\widetilde{e}_j^3|)}{6\Lambda_n^3} + \frac{(\Psi_j)^2 t^2}{2\Lambda_n^2} \\
&+ \frac{t^4}{4}\cdot\frac{\Psi_j^4}{2\Lambda_n^4} + o\left(\frac{t^3}{\Lambda_n^3}\right) + o\left(\frac{t^4}{\Lambda_n^4}\right).
\end{aligned}
$$

Hence, we have

$$
e^{-\frac{t^2}{2}} \sum_{j=0}^{n} \left|\frac{E\left[e^{\frac{it\widetilde{e}_j}{\Lambda_n}}\right]}{e^{-\frac{t^2}{2}\cdot\frac{\Psi_j^2}{\Lambda_n^2}}} - 1\right| \le Ce^{-t^2/2}(|t|^3/\Lambda_n + t^4/\Lambda_n^2),
$$

which suggests choosing $A = \min_j \sqrt{2\frac{\Lambda_n^2}{(\Psi_j)^2}} = C\sqrt{n}$ to imply

$$
\begin{aligned}
I_{1n} &\le C\frac{1}{\Lambda_n}\int_{|t|<A} e^{-t^2/2}|t|^3 dt + C\frac{1}{\Lambda_n^2}\int_{|t|<A} e^{-t^2/2}|t|^4 dt \\
&= O\left(\frac{1}{\Lambda_n}\right) = O\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

Obviously, I_{2n} is of order smaller than $\frac{1}{\sqrt{n}}$. Thus

$$
\sup_x |h_n(x) - n(x)| = O\left(\frac{1}{\sqrt{n}}\right). \tag{4.1}
$$

[8] See Feller (1971) p. 519.

Lemma B.6. Conditional on $F_{k,T} = \sigma(..., e_{k-1}, e_k)$, $\sup_{(l,k)\in\Omega_T(\eta)} \sup_{|u|<\delta} |h_{l-k}(u) - h_{l-k}(0)| \le C_1 \frac{1}{\sqrt{\eta T}} + C_2\delta$, where $\Omega_T(\eta) = \{(l,k) : \eta T \le k \le (1-\eta)T, k + \eta T \le l \le T\}$.

Proof: Observe that

$$\begin{aligned} x_l = \sum_{j=1}^{l} \sum_{i=-\infty}^{j} \pi_{j-i} e_i &= \sum_{j=1}^{k} \sum_{i=-\infty}^{j} \pi_{j-i} e_i + \sum_{j=k+1}^{l} \sum_{i=-\infty}^{j} \pi_{j-i} e_i \\ &= x_k + \sum_{j=k+1}^{l} \sum_{i=-\infty}^{k} \pi_{j-i} e_i + \sum_{j=k+1}^{l} \sum_{i=k+1}^{j} \pi_{j-i} e_i \\ &\equiv x_k + x_{1l} + x_{2l}. \end{aligned}$$

Note that $x_{2l} =_D \widetilde{S}_{l-k}$, where "$=_D$" means the same in distribution, and $\widetilde{S}_T = \sum_{i=1}^{T} \Psi_i e_i$ and $\Psi_i = \sum_{j=0}^{i} \pi_j$.

Conditional on $F_{k,T}$, $(x_l - x_k)/\Lambda_{l-k} = (x_{1l} + x_{2l})/\Lambda_{l-k}$ has a density[9] of the form $h_{l-k}(x - x_{1l}/\Lambda_{l-k})$, where $\Lambda_T^2 = \sum_{i=0}^{T} \Psi_i^2$. We then have

$$\begin{aligned} &\sup_{l-k\in\Omega_T} \sup_{|u|<\delta} |h_{l-k}(u - x_{1l}/\Lambda_{l-k}) - h_{l-k}(-x_{1l}/\Lambda_{l-k})| \\ \le &\sup_{l-k\in\Omega_T} \sup_{|u|<\delta} \Bigg| h_{l-k}(u - x_{1l}/\Lambda_{l-k}) - \frac{1}{\sqrt{2\pi}} e^{-(u-x_{1l}/\Lambda_{l-k})^2/2} + \frac{1}{\sqrt{2\pi}} e^{-(u-x_{1l}/\Lambda_{l-k})^2/2} \\ - &\frac{1}{\sqrt{2\pi}} e^{\frac{(-x_{1l}/\Lambda_{l-k})^2}{2}} + \frac{1}{\sqrt{2\pi}} e^{\frac{(-x_{1l}/\Lambda_{l-k})^2}{2}} - h_{l-k}(-x_{1l}/\Lambda_{l-k}) \Bigg| \\ \le &\ 2 \sup_{l-k\in\Omega_T} \sup_x \left| h_{l-k}(x) - \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \right| + \frac{1}{\sqrt{2\pi}} \sup_{|u|<\delta} \sup_x \left| e^{-(x+u)^2/2} - e^{-x^2/2} \right|. \end{aligned}$$

It is obvious that

$$\frac{1}{\sqrt{2\pi}} \sup_{|u|<\delta} \sup_x \left| e^{-(x+u)^2/2} - e^{-x^2/2} \right| \le C_2\delta.$$

Because of $l - k \ge \eta T$, according to Lemma B.5, we have

$$\sup_{l-k\in\Omega_T} \sup_x \left| h_{l-k}(x) - \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \right| \le C_1 \frac{1}{\sqrt{\eta T}},$$

which completes the proof of Lemma B.6.

Lemma B.7. Three Properties of Hermite Functions.

(a) $|F_i(x)| \le 0.816$, uniformly in $i = 0, 1, 2....$ and $x \in R$.

(b) $\int x^2 F_i^2(x) = \frac{i}{2} + \frac{i+1}{2}$.

[9] this equals to $(x_{l,T} - x_{k,T})/d_{l,k,T}$ as defined in Wang and Phillips 2009a, p. 731, line 13, and $d_{l,k,T}$ satisfies Assumption 2.3 of Wang and Phillips (2009a).

(c) $\int |F_i(x)|dx \leq C \times i^{5/12}$.

Proof: (1) Property (a) is Cramer's inequality, which can be found in Abramowit and Stegun (1965) or Schwartz (1967).

(2) To prove property (b), first notice the recursive relationship

$$
\begin{aligned}
xF_i(x) &= \sqrt{\frac{i}{2}}F_{i-1}(x) + \sqrt{\frac{i+1}{2}}F_{i+1}(x), i = 1, 2, ...,\\
x^2F_i^2(x) &= \frac{i}{2}F_{i-1}^2(x) + \frac{i+1}{2}F_{i+1}^2(x) + 2\sqrt{\frac{i}{2}}\sqrt{\frac{i+1}{2}}F_{i-1}(x)F_{i+1}(x).
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&\int x^2F_i^2(x) = \int \frac{i}{2}F_{i-1}^2(x) + \frac{i+1}{2}F_{i+1}^2(x) + 2\sqrt{\frac{i}{2}}\sqrt{\frac{i+1}{2}}F_{i-1}(x)F_{i+1}(x)\\
&= \frac{i}{2} + \frac{i+1}{2}.
\end{aligned}
$$

(3) Due to Muckenhoupt (1970a, 1970b), there exist positive constants C and D such that

$$
\begin{aligned}
|F_i(x)| &\leq C[|N - x^2| + N^{1/3}]^{-1/4}, \text{for } x^2 < N,\\
|F_i(x)| &\leq C \exp^{-Dx^2}, \text{for } x^2 \geq N,
\end{aligned}
$$

where $N = 2i + 1$.

Thus, we have

$$
\begin{aligned}
&\int |F_i(x)|\, dx = \int_{x^2<N} |F_i(x)|\, dx + \int_{x^2 \geq N} |F_i(x)|\, dx\\
&\leq C \int_{x^2<N} [|N - x^2| + N^{1/3}]^{-1/4} dx + C \int_{x^2 \geq N} \exp^{-Dx^2} dx\\
&\leq C \times i^{5/12} + o\left(i^{5/12}\right).
\end{aligned}
$$

Lemma B.8. Let $g(x)$ be a zero energy function, i.e., $\int g(x)dx = 0$, $g^2(x)$ is integrable and square integrable with $\int |g(t)|dt < \infty$ and $|\widehat{g}(t)| < C \min\{|t|, 1\}$, where $\widehat{g}(x) = \frac{1}{\sqrt{2\pi}} \int e^{itx} g(t)dt$ and C is positive constant. Then, as $T \to \infty$

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} E\left[g^2(x_t)\right] &= CT^{-1/2},\\
\frac{2}{T}\sum_{s=2}^{T}\sum_{t=1}^{s-1} E\left[|g(x_t)g(x_s)|\right] &= C\log(T) \cdot T^{-1/2}.
\end{aligned}
$$

Proof: Observe that

$$
\frac{1}{T}\sum_{t=1}^{T} E\left[g^2(x_t)\right] = \frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} E\left[g^2(x_t)\right] = \frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} E\left[g^2\left(\Psi\sqrt{T}\frac{x_t}{\Psi\sqrt{T}}\right)\right].
$$

Since $\frac{x_t}{\Psi\sqrt{t}}$ has a uniformly bounded density h_t (due to the proof of Corollary 2.2 of Wang and Phillips 2009a), we have

$$\frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}E\left[g^2\left(\Psi\sqrt{t}\frac{x_t}{\Psi\sqrt{t}}\right)\right] = \frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\int g^2(\sqrt{t}x)h_t(x)dx$$
$$= \frac{1}{T}\sum_{t=1}^{T}\int g^2(u)h_t(u)\frac{1}{\sqrt{t}}du \le \frac{C}{\sqrt{T}}$$

for some $C > 0$.

Using equation (3.8) of Lemma 3.2 of Wang and Phillips (2011), we have as $T \to \infty$

$$\frac{2}{T}\sum_{s=2}^{T}\sum_{t=1}^{s-1}E\left[|g(x_t)g(x_s)|\right] \le C\cdot\frac{1}{\sqrt{T}}\frac{1}{\sqrt{T}}\left(C+\sum_{k=1}^{T}\frac{1}{\sqrt{t}}\right)\left(C+\sum_{k=1}^{T}\frac{1}{t}\right) \le C\log(T)T^{-1/2},$$

which completes the proof.

Lemma B.9. Let $f_{ij}(x) = \frac{1}{i^{5/12}j^{5/12}}F_i(x)F_j(x)$ for $i = 1, ..., p-2$ and $j = i+1, ..., p-1$. Then, $f_{ij}(x)$ satisfy the conditions of Lemma B.8 uniformly over i and j.

Proof: (1) They are zero energy functions due to the orthogonality of $F_i(x)$ and $F_j(x)$ when $i \neq j$.

(2) The function $\left[\frac{1}{i^{5/12}j^{5/12}}F_i(x)F_j(x)\right]^2$ is integrable and square integrable due to the fact that $\frac{1}{i^{5/12}j^{5/12}}F_i(x)F_j(x) \in L_\infty$ and L_1 uniformly over i and j. Thus $\frac{1}{i^{5/12}j^{5/12}}F_i(x)F_j(x) \in L_k$, where $k \ge 1$ is an integer.

(3) They are integrable uniformly in i and j. By Cauchy Schwartz inequality, we have

$$\int |f_{ij}(x)|dx \le \frac{1}{i^{5/6}}\left[\int F_i^2(x)dx\int F_j^2(x)dx\right]^{1/2} = \frac{1}{i^{5/6}}.$$

(4) The property $\int |xf_{ij}(x)|dx < \infty$ holds uniformly in (i, j),

$$\int |xf_{ij}(x)|dx \le \int\left|x\frac{1}{\sqrt{j}}F_i(x)F_j(x)\right|dx$$
$$\le \frac{1}{i^{5/12}}\frac{1}{j^{5/12}}\left[\int F_j^2(x)dx\int x^2F_i^2(x)dx\right]^{1/2} = \frac{\sqrt{\frac{2i+1}{2}}}{i^{5/12}j^{5/12}},$$

which is bounded uniformly in (i, j) because $j > i$.

(5) $\int |\widehat{f}_{ij}(\varpi)|d\varpi < \infty$ uniformly in (i, j), where $\widehat{f}_{ij}(\varpi)$ is the fourier transformation of $f_{ij}(x)$. Note that the Hermite functions are eigenfunctions of itself such that

$$\widehat{F}_i(t) = (i)^jF_j(t) = \frac{1}{\sqrt{2\pi}}\int e^{itx}F_j(x)dx$$

And by convolution theorem, the Fourier transformation of $h(x) = \int \overline{f}(x)g(x+y)dy$ is $\widehat{h}(t) = \overline{\widehat{f}(t)}\widehat{g}(t)$. And by the property of double Fourier transformation, we have $\widehat{\widehat{h}}(t) = h(-t)$. Thus, we have

$$\widehat{f}_{ij}(t) = \frac{1}{\sqrt{2\pi}} \int e^{itx} \frac{1}{i^{5/12}} F_i(x) \frac{1}{j^{5/12}} F_j(x) dx$$
$$= \frac{1}{\sqrt{2\pi}} \int e^{itx} \frac{1}{i^{5/12}} (i)^{-i} \widehat{F}_i(x) \frac{1}{j^{5/12}} (i)^{-j} \widehat{F}_j(x) dx,$$

which implies

$$\left|\widehat{f}_{ij}(t)\right| = \left| \int \frac{1}{i^{5/12}} F_i(-t) \frac{1}{j^{5/12}} F_j(-t+y) dy \right|.$$

Therefore, we have

$$\int \left|\widehat{f}_{ij}(t)\right| dt = \frac{1}{i^{5/12}} \frac{1}{j^{5/12}} \int \left| \int F_i(-t) F_j(-t+y) dy \right| dt$$
$$\le \frac{1}{i^{5/12}} \frac{1}{j^{5/12}} \int |F_i(t)| dt \int |F_j(y)| dy.$$

By Lemma B.7 (c), we then have $\int |F_i(x)| dx = O\left(i^{\frac{5}{12}}\right)$. Thus, we have $\int |\widehat{f}_{ij}(t)| d\varpi \le C$. This completes the proof of part (5).

Lemmas B.1–B.9 are needed in the proofs of Lemmas B.10–B.13 below, which are needed for the proof of Theorem 4.

Proof of Theorem 4.

To prove the result, it suffices to show that as $T \to \infty$

$$E \left\| \frac{1}{\sqrt{T}} F^T F - L(1,0) I \right\|^2 = E \left(\sum_{i=0}^{p-1} \sum_{j=0,\neq i}^{p-1} \left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_i(x_t) F_j(x_t) \right]^2 \right)$$
$$+ \sum_{i=0}^{p-1} E \left(\left[\frac{1}{\sqrt{T}} \sum_{t=1}^{T} F_i^2(x_t) - L(1,0) \right]^2 \right) \to 0.$$

Due to Lemmas B.8 and B.9, we have

$$\sum_{i=0}^{p-1} \sum_{j=0,\neq i}^{p-1} E \left[\frac{1}{\sqrt{T}} \sum F_i(x_t) F_j(x_t) \right]^2$$
$$= \frac{1}{T} \sum_{i=0}^{p-1} \sum_{j=0,\neq i}^{p-1} \sum_{t=1}^{T} E\left[F_i^2(x_t) F_j^2(x_t)\right] + \frac{2}{T} \sum_{i=j+1}^{p-1} \sum_{j=0,\neq i}^{p-1} \sum_{t=s+1}^{T} \sum_{s=1}^{T-1} E\left[F_i(x_t) F_j(x_t) F_i(x_s) F_j(x_s)\right]$$
$$\le \frac{1}{T} \sum_{i=0}^{p-1} \sum_{j=0,\neq i}^{p-1} \sum_{t=1}^{T} E\left[F_i^2(x_t) F_j^2(x_t)\right] + \frac{4}{T} \sum_{i=j+1}^{p-1} \sum_{j=0}^{p-2} \sum_{t=s+1}^{T} \sum_{s=1}^{T-1} E\left[\left|F_i(x_t) F_j(x_t) F_i(x_s) F_j(x_s)\right|\right]$$
$$= \frac{1}{T} \sum_{i=0}^{p-1} \sum_{j=0,\neq i}^{p-1} \sum_{t=1}^{T} E\left[F_i^2(x_t) F_j^2(x_t)\right]$$

$$+ \frac{4}{T}\sum_{i=j+1}^{p-1}\sum_{j=0}^{p-2} i^{5/6} j^{5/6} \sum_{t=s+1}^{T}\sum_{s=1}^{T-1} E\left(\left|\frac{1}{i^{5/6}}\frac{1}{j^{5/6}}F_i(x_t)F_j(x_t)F_i(x_s)F_j(x_s)\right|\right)$$
$$\le Cp^2T^{-1/2} + Cp^2p^{5/3}(T^{-1/2} + \log(T)\cdot T^{-1/2}) \to 0$$

by Assumption 4.

Define

$$L_T = L_{Ti} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_i^2(x_t),$$
$$L_{T,\epsilon} = L_{Ti,\epsilon} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\int_{-\infty}^{\infty} F_i^2\left(\Psi\sqrt{T}\left(\frac{1}{\Psi\sqrt{T}}x_t + z\epsilon\right)\right)\phi(z)dz,$$

where $\phi(x) = \phi_1(x)$ and $\phi_\epsilon(x) = \frac{1}{\epsilon\sqrt{2\pi}}\exp(-\frac{x^2}{2\epsilon^2})$ for some $\epsilon > 0$.

Then, we have for each given i

$$\left(\frac{1}{\sqrt{T}}\sum F_i^2(x_t) - L(1,0)\right)^2 = \left(L_T - L_{T,\epsilon} + L_{T,\epsilon} - \frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right)\right.$$
$$\left. + \frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right) - \int_0^1 \phi_\epsilon\left(W_1(t)\right)dt + \int_0^1 \phi_\epsilon(W_1(t))dt - L(1,0)\right)^2.$$

By Cauchy Schwarz inequality, we have

$$E\left(\frac{1}{\sqrt{T}}\sum F_i^2(x_t) - L(1,0)\right)^2 \le 4E\left[L_T - L_{T,\epsilon}\right]^2 + 4E\left[L_{T,\epsilon} - \frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right)\right]^2$$
$$+ 4E\left[\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right) - \int_0^1 \phi_\epsilon(W_1(t))dt\right]^2 + 4E\left[\int_0^1 \phi_\epsilon\left(W_1(t)\right)dt - L(1,0)\right]^2$$
$$\equiv 4E(I_1^2) + 4E(I_2^2) + 4E(I_3^2) + 4E(I_4^2).$$

By Lemmas B.10–B.13 below, letting $\epsilon = T^{-1/2}$, we have

$$E\left[\frac{1}{\sqrt{T}}\sum F_i^2(x_t) - L(1,0)\right]^2 \le CT^{-1/4}.$$

Thus, we have

$$\sum_{i=1}^{p} E\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T} F_i^2(x_t) - L(1,0)\right]^2 \le \frac{C\ p}{\sqrt{\sqrt{T}}} \to 0$$

by Assumption 4.

Thus by Markov inequality

$$\left\|\frac{1}{\sqrt{T}}F^\tau F - L(1,0)I\right\| \to_P 0,$$

which completes the proof of Theorem 4.

Lemma B.10. $E(I_1^2) \le C\epsilon + C\epsilon^{1/2}$ as $T \to \infty$.

Proof: By Cauchy-Schwartz inequality and the fact that $\int \phi(z)dz = 1$, we have

$$E(I_1^2) \le \frac{1}{T} E\left[\int_{-\infty}^{\infty} \{\sum_{t=1}^{T} Y_{t,T}(z)\}^2 \phi(z)dz\right] \int_{-\infty}^{\infty} \phi(z)dz \tag{4.2}$$

$$= \frac{1}{T} E\left[\int_{-\infty}^{\infty} \{\sum_{t=1}^{T} Y_{t,T}(z)\}^2 \phi(z)dz\right] \le \frac{1}{T}\int_{-\infty}^{\infty} E\left\{\sum_{t=1}^{T} Y_{t,T}(z)\right\}^2 \phi(z)dz,$$

where $Y_{t,T} \equiv F_i^2(x_t) - F_i^2\left[\Psi\sqrt{T}\left(\frac{1}{\Psi\sqrt{T}}x_t + z\epsilon\right)\right]$.

Using the fact that the density of $\frac{x_t}{\Psi\sqrt{t}}$ is uniformly bound, we have

$$\frac{1}{T}\sum_{t=1}^{T} E\left[Y_{t,T}^2(z)\right]^2 = \frac{1}{T}\sum_{t=1}^{T}\int_{-\infty}^{\infty} \left|F_i^2\left(\sqrt{t}x + \Psi\sqrt{T}z\epsilon\right) - F_i^2\left(\sqrt{t}x\right)\right|^2 h_t(x)dx$$

$$= \frac{1}{T}\sum_{t=1}^{\infty}\frac{1}{\sqrt{t}}\int_{-\infty}^{\infty} \left|F_i^2\left(u + \Psi z\epsilon\right) - F_i^2(u)\right|^2 g_t\left(u/\sqrt{t}\right)du$$

$$\le \frac{2C}{T}\sum_{t=1}^{T}\frac{1}{\sqrt{t}}\int F_i^4(u)du = O(T^{-1/2}).$$

Let $\Omega_T = \{(l,k) : \eta T \le k \le (1-\eta)T, k + \eta T \le l \le T\}$ and let $\eta = \epsilon^{1/2}$. Then

$$\frac{2}{T}\sum_{k=1}^{T}\sum_{l=k+1}^{T} |E\{Y_{kT}(z)Y_{lT}(z)| = \frac{2}{T}\left(\sum_{l>k,(l,k)\notin\Omega_T} + \sum_{(l,k)\in\Omega_T}\right)|E\{Y_{kT}(z)Y_{lT}(z)|$$

$$\le \frac{1}{T^2}\sum_{l-k\le\eta T} A + \frac{A}{T^2}\sum_{k=1}^{T}\frac{\sqrt{T}}{\sqrt{k}}\sum_{l=k+1}^{T}\frac{\sqrt{T}}{\sqrt{l-k}}$$

$$\times\left[\int_{|y|\ge\sqrt{T}} |F_i^2(x)|dx + \sup_{(l,k)\in\Omega_T}\sup_{|u|<Cz\epsilon^{1/2}}\left|h_{l|k}(u) - h_{l|k}(0)\right|\right],$$

where $h_{l|k}(u)$ is the conditional density of $\frac{x_l - x_k}{\Lambda_{l-k}}$ with Λ_{l-k} being introduced in the proof of Lemma B.5, and $A = C\sqrt{n}$ as chosen in the proof of Lemma B.5.

The first term is of order $C\epsilon^{1/2}$ (see, for example, Page 1494 of Phillips 2009). Due to the property that there exist positive constants C and D, such that

$$|F_i(x)| \le C[|N - x^2| + N^{1/3}]^{-1/4}, \text{for } x^2 < N,$$
$$|F_i(x)| \le C\exp^{-Dx^2} \text{ for } x^2 \ge N,$$

for $N = 2i + 1$, and using the property of the exponential function and for $|x| > \sqrt{T}$, Assumption A2 implies $x^2 \ge N$ and then

$$\int_{|x|>\sqrt{T}} |F_i^2(x)|dx = C\left(\sqrt{T}\right)^{-c},$$

where c is a arbitrarily large constant.

According to Lemma B.6, we have

$$\sup_{(l,k)\in\Omega_T}\sup_{|u|<Cz\epsilon^{1/2}}|h_{l|k}(u)-h_{l|k}(0)|\leq Cz\epsilon^{1/2},$$

and when $\frac{1}{(\epsilon^{1/2}T)^{1/2}}<\epsilon^{1/2}$,

$$\frac{A}{T^2}\sum_{k=1}^{T}\frac{\sqrt{T}}{\sqrt{k}}\sum_{l=k+1}^{T}\frac{\sqrt{T}}{\sqrt{l-k}}\times\sup_{(l,k)\in\Omega_T}\sup_{|u|<Cz\epsilon}|h_{l|k}(u)-h_{l|k}(0)|\leq Cz\epsilon^{1/2}.$$

Therefore, by the dominated convergence theorem and the fact that $\int|x|^p\phi(x)dx<\infty$ for all p, we have

$$E(I_1^2)\leq C\epsilon^{1/2},$$

because $\int\limits_{|y|\geq\sqrt{T}}|F_i^2(x)|dx$ is of smaller magnitude than $\epsilon^{1/2}$.

Lemma B.11. For some $\delta>0$, we have $E(I_2^2)\leq C(T^{-1/2-\delta})$.

Proof: Define $G_T(y)=\int\limits_{-\infty}^{y}\sqrt{T}F_i^2(\sqrt{T}u)du$, and $G(y)=\int F_i^2(x)dx=1$ if $y\geq0$ and $G(y)=0$ if $y<0$. By construction, $G_T(y)\to G(y)$ for all continuity points of $G(y)$ and $G(b)-G(a)=0$ if $0\notin(a,b]$.

Now, $L_{T,\epsilon}$ has the following form: $\int\limits_{-\infty}^{\infty}(\frac{1}{T}\sum\limits_{t=1}^{T}\phi_\epsilon(y-\frac{1}{\Psi\sqrt{T}}x_t-x))dG_T(y)$. The difference between this and $\int_{|y<\nu|}(\frac{1}{T}\sum\limits_{t=1}^{T}\phi_\epsilon(y-\frac{1}{\Psi\sqrt{T}}x_t-x))dG_T(y)$ is bounded in absolute value by

$$C\left|\int_{|y>\nu|}dG_T(y)\right|=C\int_{|u>\nu\sqrt{T}|}F_i^2(u)du$$

because $\phi_\epsilon(y)$ is uniformly bounded over ϵ and y.

Because $F_i(u)$ are Hermite series, similarly to the arguments used in the proof of Lemma B.10, choosing $\nu=T^\alpha$ with $0>=\alpha>-1/2$, we have for any $q>0$

$$C\int_{|u>\nu\sqrt{T}|}F_i^2(u)du\leq C(T^{1/2-\alpha})^{-q}.$$

Define $y_{m,i},i=-m,...,0,..,m$ such that $y_{m,-m}=-[\nu]<y_{m,-m+1}<...<y_{m,m-1}<y_{m,m}=[\nu]$ and $\sup_i|y_{m,i}-y_{m,i-1}|=2*\frac{\nu}{m}$. Then, the difference between $\int_{|y<\nu|}(\frac{1}{T}\sum\limits_{t=1}^{T}\phi_\epsilon(y-\frac{1}{\Psi\sqrt{T}}x_t-x))dG_T(y)$ and $\sum\limits_{i=-m}^{m}(\frac{1}{T}\sum\limits_{t=1}^{T}\phi_\epsilon(y-\frac{1}{\Psi\sqrt{T}}x_t-x))\int\limits_{y_{m,i}}^{y_{m,i+1}}dG_T(y)$ is bounded in absolute value by

$$C\frac{\nu}{m}\int_{\{|y|\leq v\}}d|F_n|(y)\leq C\frac{\nu}{m}.$$

Further, the difference between $\sum_{i=-m}^{m}(\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(y-\frac{1}{\Psi\sqrt{T}}x_t-x))\int_{y_{m,i}}^{y_{m,i+1}}dG_T(y)$ and $\sum_{i=-m}^{m}(\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(y-\frac{1}{\Psi\sqrt{T}}x_t-x))\int_{y_{m,i}}^{y_{m,i+1}}dG(y)$ is bounded in absolute value by

$$C\sum_{i=-m}^{m}\left|\int_{y_{m,i}}^{y_{m,i+1}}d(G_T(y)-G(y))\right|.$$

When 0 is in $(y_{m,i},y_{m,i+1}]$, we have $\left|\int_{y_{m,i}}^{y_{m,i+1}}d(G_T(y)-G(y))\right|\leq C\frac{\nu}{m}$.

Let $\sqrt{T}\frac{\nu}{m}\to\infty$, as $T\to\infty$. When 0 is not in $(y_{m,i},y_{m,i+1}]$, and $y_{m,i}>0$, we have

$$\begin{aligned}\left|\int_{y_{m,i}}^{y_{m,i+1}}d(G_T(y)-G(y))\right| &= \left|\int_{y_{m,i}}^{y_{m,i+1}}dG_T(y)\right| = G_T(y)-G_T(y_{m,i})\\ &= \int_{y_{m,i}}^{y_{m,i+1}}\sqrt{T}F_i^2(\sqrt{T}u)du = \int_{\sqrt{T}y_{m,i}}^{\sqrt{T}y_{m,i+1}}F_i^2(\sqrt{T}u)d\sqrt{T}u\\ &\leq \int_{\sqrt{T}y_{m,i}}^{\infty}F_i^2(y)dy. \leq C(T^{1/2}\frac{\nu}{m})^{-l}\end{aligned}$$

by a similar argument to the proof in Lemma B.10.

When 0 is not in $(y_{m,i},y_{m,i+1}]$, and $y_{m,i+1}<0$, we have

$$\begin{aligned}\left|\int_{y_{m,i}}^{y_{m,i+1}}d(G_T(y)-G(y))\right| &= \left|\int_{y_{m,i}}^{y_{m,i+1}}dG_T(y)\right| = G_T(y)-G_T(y_{m,i})\\ &= \int_{y_{m,i}}^{y_{m,i+1}}\sqrt{T}F_i^2(\sqrt{T}u)du = \int_{\sqrt{T}y_{m,i}}^{\sqrt{T}y_{m,i+1}}F_i^2(\sqrt{T}u)d\sqrt{T}u\\ &\leq \int_{-\infty}^{\sqrt{T}y_{m,i+1}}F_i^2(y)dy. \leq C(T^{1/2}\frac{\nu}{m})^{-l}.\end{aligned}$$

Thus, the trem $C\sum_{i=-m}^{m}|\int_{y_{m,i}}^{y_{m,i+1}}d(G_T(y)-G(y))|$ is bounded by

$$C\frac{v}{m}+C(T^{1/2}\frac{\nu}{m})^{-l}.$$

As a consequence, $|I_2|$ is bounded surely b $C\cdot(T^{1/2-\alpha})^{-l}+(\frac{\nu}{m})+(\frac{\nu}{m}+(T^{1/2}\frac{\nu}{m})^{-l})$, where l can be sufficiently large.

Because the above approximation procedure is a deterministic procedure, we have

$$|I_2|^2\leq C\left[(T^{1/2-\alpha})^{-2l}+(\frac{\nu^2}{m^2})+C(T^{1/2}\frac{\nu}{m})^{-2l}\right].$$

Because ν and m can be chosen, and l is sufficiently large positive constant, we can choose $\frac{1}{T} < \frac{\nu^2}{m^2} < T^{-1/2-\delta}$ such that

$$|I_2|^2 \leq CT^{-1/2-\delta},$$

which can be improved for suitable choice of ν and m. However, for our purpose, the above choice is enough to guarantee convergence of the diagonal terms.

Lemma B.12. $E\left[I_3^2\right] \leq C(T^{-1/3})$ as $T \to \infty$.

Proof: Observe that

$$E\left[I_3^2\right] = E\left(\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right) - \int_0^1 \phi_\epsilon(W_1(t))dt\right)^2.$$

Then, we have

$$\begin{aligned}
& E\left|\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right) - \int_0^1 \phi_\epsilon(W_1(r))dr\right|^2 \\
= \ & E\left|\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right) - \frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(W_1\left(\frac{t}{T}\right)\right) - \int_0^1 \phi_\epsilon(W_1(r))dr + \frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(W_1\left(\frac{t}{T}\right)\right)\right|^2 \\
\leq \ & 2E\left[\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right) - \frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(W_1\left(\frac{t}{T}\right)\right)\right]^2 \\
& + 2E\left[-\int_0^1 \phi_\epsilon(W_1(r))dr + \frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(W_1\left(\frac{t}{T}\right)\right)\right]^2 \\
\leq \ & \frac{2}{T}\sum E\left[\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right) - \phi_\epsilon\left(W_1\left(t/T\right)\right)\right]^2 \\
& + E\left|-\int_0^1 \phi_\epsilon(W_1(r))dr + \frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(W_1\left(\frac{t}{T}\right)\right)\right|^2 \equiv J_1 + J_2.
\end{aligned}$$

For the J_1 term, we have

$$\left|\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right) - \phi_\epsilon(W_{1T}(t))\right| = \left|\phi_\epsilon^{'}(\xi)\left(\frac{1}{\Psi\sqrt{T}}x_t - W_1(t/T)\right)\right|,$$

where ξ lies between $\frac{1}{\sqrt{T}}x_t$ and $W_1(t/T)$. And we have

$$\left|\phi_\epsilon^{'}(x)\right| = \left|\frac{1}{\epsilon\sqrt{2\pi}}\exp(-\frac{x^2}{2\epsilon^2})(\frac{-x}{\epsilon^2})\right| \leq C \text{ uniformly over } x, \epsilon.$$

Thus, we have

$$\frac{1}{T}\sum E\left[\left|\phi_\epsilon\left(\frac{1}{\Psi\sqrt{T}}x_t\right) - \phi_\epsilon(W_1(t/T)\right)\right|^2 \leq \frac{C}{T}\sum E\left(\frac{1}{\Psi\sqrt{T}}x_t - W_1\left(\frac{t}{T}\right)\right)^2.$$

Recalling the following definitions and decompositions:

$$x_t = \sum_{k=1}^{t}(\sum_{j=0}^{\infty}\pi_k e_{t-j}) = \pi_0 e_t + (\pi_0+\pi_1)e_{t-1} + ... + \sum_{k=0}^{\infty}(\pi_{k+1}+...+\pi_{k+t})e_{-k},$$

$$\Psi_k = \sum_{j=0}^{k}\pi_j \quad \text{and} \quad \Psi = \sum_{j=0}^{\infty}\pi_j,$$

we have

$$x_t = \Psi S_t + \sum_{k=0}^{t-1}\Psi_k e_{t-k} + \sum_{k=0}^{\infty}(\Psi_{k+1}-\Psi_k)e_{-k},$$

or equivalently,

$$x_t = \Psi S_t + R_t^{'} + R_t^{''}$$

where $S_t = \sum_{k=1}^{t} e_t$, $R_t^{'} = \sum_{k=1}^{t}((\Psi_{k+1}-\Psi)e_k$ and $R_t^{''} = \sum_{k=1}^{t}((\Psi_{k+1}-\Psi_k)e_{-k}$. This means that x_t can be decomposed into a random walk plus two additional terms.

We then have

$$E\left(\frac{1}{\sqrt{T}}S_t + \frac{1}{\Psi\sqrt{T}}R_t^{'} + \frac{1}{\Psi\sqrt{T}}R_t^{''} - W_1(\frac{t}{T})\right)^2 = E\left(\chi_t + \frac{1}{\Psi\sqrt{T}}R_t^{'} + \frac{1}{\Psi\sqrt{T}}R_t^{''}\right)^2$$
$$\leq\ 2E\left[\chi_t^2\right] + 2E\left(\frac{1}{\Psi\sqrt{T}}R_t^{'}\right)^2 + 2E\left(\frac{1}{\Psi\sqrt{T}}R_t^{''}\right)^2,$$

where $\chi_t = \frac{1}{\sqrt{T}}S_t - W_1(\frac{t}{T})$.

Due to Lemma B.3, we have

$$E(\chi_T^2) \leq C(T^{-1/3}).$$

Using the property that χ_t has independent increments (because S_t and $W_1(t)$ have independent increments), we have

$$E\left(\chi_t^2\right) \leq C(T^{-1/3}\frac{t}{T}),$$

Therefore, we obtain

$$\frac{1}{T}\sum_{t=1}^{T}E\left[\chi_t^2\right] \leq C(T^{-1/3}).$$

By Akonom (1993 p. 76), we know that for all $t = 1, .., T$,

$$E\left(\frac{1}{\sqrt{T}}R_t^{'}\right)^2 \leq C_1/T,$$
$$E\left(\frac{1}{\sqrt{T}}R_t^{''}\right) = \frac{C_1}{T} \quad E\left(\frac{1}{\sqrt{T}}R_t^{''}\right)^2 \leq C_2/T,$$

both of which lead to

$$\frac{1}{T}\sum_{t=1}^{T} E\left(\frac{1}{\sqrt{T}}\Psi S_t + \frac{1}{\sqrt{T}}R_t^{'} + \frac{1}{\sqrt{T}}R_t^{''} - W_1\left(\frac{t}{\sqrt{T}}\right)\right)^2 \leq C\left(\frac{1}{T} + T^{-1/3}\right).$$

For the J_2 term, using the following equations:

$$\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon\left(W_1\left(\frac{t}{T}\right)\right) = \int_0^1 \phi_\epsilon\left(W_1\left(\frac{[rT]}{T}\right)\right) + \frac{1}{T}\phi_\epsilon\left(W_1\left(\frac{T}{T}\right)\right) - \frac{1}{T}\phi_\epsilon\left(W_1(0)\right)$$

$$= \int_0^1 \phi_\epsilon\left(W_1\left(\frac{[rT]}{T}\right)\right) + \frac{C}{T},$$

$$E\left(\int\left|\phi_\epsilon\left(W_1\left(\frac{[rT]}{T}\right)\right) - \phi_\epsilon(W_1(r))\right| dr\right)^2 \leq E\left(\int\left|\phi_\epsilon\left(W_1\left(\frac{[rT]}{T}\right)\right) - \phi_\epsilon\left(W_1(r)\right)\right|^2 dr\right)$$

$$= \left(\int E\left|\phi_\epsilon\left(W_1\left(\frac{[rT]}{T}\right)\right) - \phi_\epsilon(W_1(r))\right|^2 dr\right) \leq \int_0^1 \sup\left|\phi_\epsilon^{'}(x)\right|^2 E\left[\left|W_1\left(\frac{[rT]}{T}\right) - W_1(r)\right|^2\right] dr$$

$$\leq \frac{C}{T},$$

we have $J_2 \leq \frac{C}{T}$.

Combining the results for J_1 and J_2, we have derived that

$$E\left[I_3^2\right] \leq C\left(T^{-1/3}\right).$$

Lemma B.13. $E\left[I_4^2\right] \leq C\epsilon$ as $\epsilon \to 0$.

Proof: Observe that

$$E\left[I_4^2\right] = E\left[\int_{-\infty}^{\infty}\phi(x)L(1,\epsilon x)dx - \int_{-\infty}^{\infty}\phi(x)dx\ L(1,0)\right]^2.$$

Let $F(x) = \Phi(x) - I(x \geq 0)$, where $\Phi(x)$ is the CDF of normal variable $U \sim N(0,1)$ and $I(.)$ denote the conventional indicator function. We have $F(\infty) = F(-\infty) = 0$.

Note that $L(1,0) = \int_{-\infty}^{\infty} L(1,\epsilon x)I(x \geq 0)\ dx$ for any $\epsilon > 0$. We thus have

$$\epsilon^{-1/2}\left\{\int_{-\infty}^{\infty}\phi(x)L(1,\epsilon x)dx - L(1,0)\right\} = \epsilon^{-1/2}\int_{-\infty}^{\infty} L(1,\epsilon x)dF(x).$$

By the definition of $L(1,y)$ just above equation (1.2) and then equation (2.15) of Borodin (1986), we therefore obtain

$$E\left(\epsilon^{-1/2}\int_{-\infty}^{\infty} L(1,\epsilon x)dF(x)\right)^2 \leq C,$$

which implies

$$E\left[I_4^2\right] \leq C\epsilon,$$

which completes the proof of Lemma B.13.

References

Abadir, K.M. and J.R., Magnus, 2005. Matrix Algebra. Cambridge University Press.

Abramowit, M. and I., Stegun, 1965. Handbook of Mathematical Functions. Dover, New York.

Akonom, J., 1993. Comportement asymptotiques du temps d'occupation du processus des sommes partielles. Annales de l'I. H. P., 57-81.

Andrews, D.W.K., 1991. Asymptotic normality of series estimator for nonparametric and semiparametric models. Econometrica, 59, 307-345.

Alexits, G., 1961. Convergence Problems of Orthogonal Series. Pergamon Press, New York.

Blake, N.S. and T.B., Fomby, 1997. Threshold cointegration. International Economic Review, 38, 627-645.

Borodin, A.N., 1986. On the character of convergence to Brownian local time I. Probability Theory and Related Fields, 72, 231–250.

Campbell, J.Y., 1991. A variance decomposition of stock returns. Economic Journal, 101, 157-179.

Campbell, J.Y., 2000. Asset pricing at the millennium. Journal of Finance, 55, 1515-1567.

Campbell, J.Y., 2008. Viewpoint: estimating the equity premium. Canadian Journal of Economics, 41, 1-21.

Campbell, J.Y. and R.J., Shiller, 1988. The dividend price ratio and expectations of future dividends and discount factors. Review of Financial Studies, 1, 195-228.

Campbell, J.Y. and S.B., Thompson, 2008. Predicting excess stock returns out of sample: can anything beat the historical average? Review of Financial Studies, 21, 1509-1531.

Campbell, J.Y., and M., Yogo, 2006. Efficient tests of stock return predictability. Journal of Financial Economics, 81, 27-60.

Chen, J., J., Gao and D., Li 2012. Estimation in semi-parametric regression with non-stationary regressors. Bernoulli, 18, 678-702.

Chen, Q.Q. and Y.M., Hong, 2010. Predictability of equity returns over different time horizons: a nonparametric approach. Working paper at http://old.ccer.edu.cn/download/11861-1.pdf

Cohcrane, J.H., 1992. Explaining the variance of price–dividend ratios. Review of Financial Studies, 5, 243-280.

Dong, C. and J., Gao, 2011. Expansion of Lévy process functionals and its application in statistical estimation. Working paper at http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2013/wp03-13.pdf.

Dong, C. and J., Gao, 2012. Specification testing driven by orthogonal series in nonstationary time series models. Working paper at http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2012/wp20-12.pdf.

Einmahl, U., 1989. Extensions of results of Komlos, Major and Tusnady. Journal of Multivariate Analysis, 28, 85-96.

Engle, R.F. and C.W.J., Granger, 1987. Cointegration and error correction: representation, estimation, and testing. Econometrica, 55, 251-276.

Engstrom, E., 2003. The conditional relationship between equity risk and dividend price ratio. Working paper at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=355340.

Fama, E.F. and K.R., French, 1988. Dividend yields and expected stock returns. Journal of Financial Economics, 22, 3-25.

Feller, W., 1971. An Introduction to Probability Theory and Its Applications, Vol 2, 2nd ed. Wiely.

Gao, J., 2007. Nonlinear Time Series: Semiparametric and Nonparametric Methods. Chapman & Hall/CRC, London.

Gao, J. and P.C.B., Phillips, 2013. Semiparametric estimation in triangular system equations with nonstationarity. Journal of Econometrics, 176, 59–79.

Granger, C.W.J., 1995. Modelling nonlinear relationships between extended–memory variables. Econometrica, 63, 265-279.

Gonzalo, J. and J.Y., Pitarakis, 2012. Regime–specific predictability in predictive regression. Journal of Business and Economic Statistics, 20, 229-241.

Goyal, A. and I., Welch, 2003. Predicting the equity premium with dividend ratios. Management Science, 49, 639-654.

Goyal, A. and I., Welch, 2008. A comprehensive look at the empirical performance of equity premium prediction. Review of Financial Studies, 21, 1455-1508.

Hansen, P.R. and A., Timmermann, 2012. Choice of sample split in out–of–sample forecast evaluation. Manuscript, European University Institute and University of California at San Diego.

Karlsen, H.A. and D., Tjøstheim, 2001. Nonparametric estimation in null recurrent time series. Annal of Statistics, 29, 372-416.

Karlsen, H.A., T., Myklebust, and D., Tjøstheim, 2007. Nonparametric estimation in a nonlinear cointegration type model. Annal of Statistics, 35, 252-299.

Komlos, J., P., Major, and G., Tusnady, 1976. An approximation of partial sums of independent RV's, and the sample DF. II. Z. Wahrscheinlichkeitstheorie, 34, 33-58.

Kothari, S.P. and J., Shanken, 1997. Book–to–market, dividend yield, and expected market returns: a time series analysis. Journal of Financial Economics, 44, 169-203.

Lamont, O., 1998. Earnings and expected returns. Journal of Finance, 5, 1563-1587.

Lettau, M. and S.V., Niewerburgh, 2008. Reconciling the return predictability evidence. Review of Financial Studies, 21, 1607-1652.

Lewellen, L., 2004. Predicting returns with financial ratios. Journal of Financial Economics, 74, 209-235.

Mankiw, N.G. and M., Shapiro, 1986. Do we reject too often? small sample properties of tests of rational expectation models. Economic Letters, 20, 139-145.

Muckenhoupt, B., 1970a. Mean Convergence of Hermite and Laguerre series I. Translation of the American Mathematical Society, 147, 419-431.

Muckenhoupt, B., 1970b. Mean Convergence of Hermite and Laguerre series II. Translation of the American Mathematical Society, 147, 433-460.

Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. Journal of Econometrics, 79, 147-168.

Nelson, C. and M., Kim, 1993. Predictable stock returns: the role of small sample bias. Journal of Finance, 48, 641-661.

Park, J.Y. and P.C.B., Phillips, 1999. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory, 15, 269-298.

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

Pesaran, M.H. and A., Timmermann, 2007. Selection of estimation window in the presence of breaks. Journal of Econometrics, 137, 134-161.

Phillips, P.C.B., 2001. Descriptive econometrics for non-stationary time series with empirical applications. Journal of Applied Econometrics, 16, 389-413.

Phillips, P.C.B., 2009. Local limit theory and spurious nonparametric regression. Econometric Theory, 25, 1466-1497.

Pontiff, J. and L.D., Schall, 1998. Book–to–market ratios as predictors of market returns. Journal of Financial Economics, 49, 141-160.

Rapach, D.E. and G.F., Zhou, 2012. Forecasting stock returns. In Handbook of Economic Forecasting, Vol 2, edited by Elliott, G. and A. Timmermann.

Rothental, H.P., 1970. On the subspace L^p $(p \geq 2)$ spanned by a sequence of independent random variables. Israel Journal of Mathematics, 8, 273-303.

Schwartz, S.C., 1967. Estimation of probability density by an orthogonal series. Annals of Mathematical Statistics, 38, 1261-65.

Scz$\ddot{g}$o, G., 1959. Orthogonal Polynomials. American Mathematical Society Publication, 32.

Stambaugh, R., 1999. Predictive regression. Journal of Financial Economics, 54, 375-421.

Stone, M.H., 1927. Developments in Hermite polynomials. Annal of Mathematics, 29, 1-13.

Teräsvirta, T., D., Tjøstheim, and C.W.J., Granger, 2010. Modelling Nonlinear Economic Time Series. Oxford University Press, Oxford.

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

Wang, Q. and P.C.B., Phillips, 2009b. Structural nonparametric cointegrating regression. Econometrica, 77, 1901-1948.

Wang, Q. and P.C.B., Phillips, 2011. Asymptotic theory for zero energy functionals with nonparametric regression applications. Econometric Theory, 27, 235-259.