ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

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

Orthogonal Series Estimation in Nonlinear Cointegrating Models with Endogeneity

Biqing Cai, Chaohua Dong, Jiti Gao

September 2015

Working Paper 18/15

Orthogonal Series Estimation in Nonlinear Cointegrating Models with Endogeneity

Biqing Cai[†‡] and Chaohua Dong[†♭] and Jiti Gao[†] *

University of Bergen, Norway[‡] and Monash University, Australia[†]

and Southwestern University of Finance and Economics, China[♭]

September 30, 2015

Abstract

This paper considers a nonlinear time series model associated with both nonstationarity and endogeneity. The proposed model is then estimated by a nonparametric series method. An asymptotic theory is established in both point–wise and the space metric sense for the estimator. The Monte Carlo simulation results show that the performance of the proposed estimate is numerically satisfactory.

Key words: Cointegration, endogeneity, Hermite functions, 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 Balke 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 et al. 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,b). Meanwhile, Karlsen and Tjøstheim (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 et al. (2012) consider estimation issues in a partially linear model with non-stationary 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 another commonly used estimation 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), Chen and Shen (1998) and Gao (2007) for example. However, as far as we know, when the data set is assumed to be unit root nonstationary, there are only a couple of studies based on series estimation. Dong and Gao (2013, 2014) were among the first considering series expansion for nonstationary data. Dong and Gao (2013) discuss series expansion for Lévy processes which can be considered as an orthogonal series expansion based on time varying probability densities. By contrast, we propose using a 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 is easy to impose some types of restrictions, such as additive separability. It is also computationally convenient.

In this paper, we consider a class of integrable regression models and propose using a Hermite series estimation method for such a class of cointegration models where the time series regressor is nonstationary and endogenous with the error process. Without necessarily using an instrumental variable approach, we show that the proposed nonparametric series estimator is still asymptotically consistent and normally distributed under such a type of endogeneity. The nonparametric series based approach under endogeneity complements an existing kernel based method by Wang and Phillips (2014). It should be pointed out that while similar asymptotic results, such as Theorem 3.1 and Corollary 3.1 listed in Section 3.2 below, may be obtained by either the kernel or the series based method, both the establishments and the proofs of the asymptotic results are quite different. It should also be pointed out that while the class of integrable models may be restrictive, such integrable models have their own empirical applications for appropriately balancing the relationship between a stationary time series on the left–hand side and a highly nonstationary regressor on the right–hand side (see, for example, Marmer (2008)). Meanwhile, we establish an asymptotic distributional theory for a matrix of partial sums of nonlinear nonstationary time series in Theorem 3.2 listed in Section 3.3 below. Such an asymptotic result is generally applicable to deal with the inverses of matrices of unit root nonstationary time series. As a consequence, we are able to establish some uniform consistency results and an asymptotic normality for the series based estimator with a rate of $T^{-1/4}p^{1/2}$, where p is the truncation parameter involved in the series approximation and T is the sample size.

The organisation of this paper is as follows. In Section 2, we propose the model and discuss its estimation and assumptions. In Section 3, we derive the uniform consistency and asymptotic normality of the series estimator. In Section 4, we conduct Monte Carlo simulation to evaluate the finite sample performance of the nonparametric series estimator. Section 5 discusses potential extension, followed by Section 6 that concludes the paper. Several lemmas are present in Appendix A, which are crucial for the proof of our main results in Appendix B. The proofs of the lemmas listed in Appendix A, as well as a detailed proof of one theorem, are given in Appendix C of the paper.

Throughout this paper, $\to_D$, $\to_P$ and $\to_{a.s.}$ denote convergence in distribution, in probability and almost surely, respectively. For a vector $\|\cdot\|$ stands for the Euclidean norm and for a matrix $A = (a_{ij})_{n\times m}$, $\|A\|^2 = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2$. $\int g(x)dx$ stands for an integral over $(-\infty, +\infty)$.

2 Model estimation and assumptions

2.1 Preliminaries of the Hermite functions

In this paper, we use the Hermite functions to estimate square integrable functionals of a unit root process. Let $\{H_i(x)\}_{i=0}^{\infty}$ be the Hermite polynomial system orthogonal with respect to the weight function $\exp(-x^2)$ given by

$$H_i(x) = (-1)^i \exp(x^2)\frac{d^i}{dx^i}\exp(-x^2), \qquad i \geq 0. \tag{2.1}$$

It is known that $\{H_i(x)\}_{i=0}^{\infty}$ is a complete orthogonal system in the Hilbert space

$$L^2(\mathbb{R}, \exp(-x^2)) = \{g(x) : \int g^2(x)e^{-x^2}dx < \infty\}$$

satisfying the orthogonality $\int H_i(x)H_j(x)e^{-x^2}dx = \sqrt{\pi}2^i i!\delta_{ij}$, where $\mathbb{R} = (-\infty, \infty)$ and δ_{ij} is the Kronecker delta function. Define

$$F_i(x) = \frac{1}{\sqrt[4]{\pi}\sqrt{2^i i!}} H_i(x)\exp(-x^2/2), \qquad i \geq 0. \tag{2.2}$$

Then, $\{F_i(x)\}_{i=0}^{\infty}$ is the so-called Hermite series or Hermite functions in the literature, complete orthonormal in $L^2(\mathbb{R}) = \{g(x) : \int g^2(x)dx < \infty\}$ satisfying $\int F_i(x)F_j(x)dx = \delta_{ij}$. Consequently, any continuous function $f(x) \in L^2(\mathbb{R})$ has an infinite orthogonal series expansion

$$f(x) = \sum_{i=0}^{\infty} \theta_i F_i(x), \quad \text{where } \theta_i = \int f(x)F_i(x)dx. \tag{2.3}$$

Moreover, $F_i(x)$ is bounded uniformly in both i and $x \in \mathbb{R}$ (see Szego, 1975, p. 242).

2.2 Model estimation and assumptions

Consider a nonparametric regression model of the form

$$\begin{aligned} y_t &= f(x_t) + e_t, \\ x_t &= x_{t-1} + v_t, \; t = 1, 2, \cdots, T, \end{aligned} \tag{2.4}$$

where v_t is a stationary linear process, $x_0 = O_P(1)$, e_t is also a stationary linear process, and $f(\cdot) \in L^2(\mathbb{R})$. In view of (2.3), for each t we have $y_t = Z_p^{\tau}(x_t)\theta + \gamma_p(x_t) + e_t$, where p is some positive integer, $Z_p^{\tau}(\cdot) = (F_0(\cdot), ..., F_{p-1}(\cdot))$, $\theta^{\tau} = (\theta_0, \cdots, \theta_{p-1})$ and $\gamma_p(\cdot) = \sum_{j=p}^{\infty} \theta_j F_j(\cdot)$ a residue after truncation, or in a matrix form,

$$Y = Z\theta + \gamma + e, \tag{2.5}$$

where $Y^{\tau} = (y_1, \cdots, y_T)$, $Z = (Z_p(x_1), \cdots, Z_p(x_T))^{\tau}$ a $T \times p$ matrix, $\gamma = (\gamma_p(x_1), \cdots, \gamma_p(x_T))^{\tau}$ and $e = (e_1, \cdots, e_T)^{\tau}$. Hence, by the ordinary least–square (OLS) method θ is estimated by

$$\widehat{\theta} = (Z^{\tau}Z)^{-1} Z^{\tau}Y. \tag{2.6}$$

Then, naturally the series estimator of function $f(x)$ for any $x \in \mathbb{R}$ is $\widehat{f}(x) = Z_p^{\tau}(x)\widehat{\theta}$. To proceed further, we introduce the following technical assumptions.

Assumption 1. (a) Let $\{\epsilon_j, j \in \mathbb{Z}\}$ be a sequence of independent and identically distributed (iid) continuous random variables satisfying $E\epsilon_0 = 0$, $E\epsilon_0^2 = 1$ and $E\epsilon_0^4 < \infty$. Let $\varphi(u)$ be the characteristic function of ϵ_0 satisfy $\int |u\,\varphi(u)|du < \infty$.

(b) Let $\{v_t\}$ be a linear process defined by $v_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$, where $\psi_0 = 1$, $\psi := \sum_{j=0}^{\infty} \psi_j \neq 0$ and $\sum_{j=0}^{\infty} j|\psi_j| < \infty$.

(c) Let $x_t = x_{t-1} + v_t$ with $x_0 = O_P(1)$. Let $e_t = \sum_{j=0}^{\infty} \phi_j \epsilon_{t-j}$ with $\phi_0 = 1$, $\sum_{j=0}^{\infty} \phi_j \neq 0$ and $\sum_{j=0}^{\infty} j|\phi_j| < \infty$.

(d) For any given $u \in \mathbb{R}$, define $h(u) = \frac{\varphi'(u)}{\varphi(u)}$. Suppose that there is a nonnegative function $k(\lambda)$ such that $\max_{j\geq 0} |h(\lambda\phi_j)| \leq k(\lambda)$ and $\int_{-\infty}^{\infty} k(\lambda)\, |\Gamma(\lambda)|\, d\lambda < \infty$, where $\Gamma(\lambda) = \prod_{i=0}^{\infty} \varphi(\lambda\, \phi_i)$ is the characteristic function of e_t.

Condition (a) shows the requirement for the underlying process $\{\epsilon_j, j \in \mathbb{Z}\}$ that determines the properties of the regressor and the error term. The moment conditions are commonly used in the literature. The integrability of $|\lambda\varphi(\lambda)|$ in (1) is about to derive some properties for the density functions related to x_t, and the condition for $h(u)$ is satisfied in many cases, such as symmetric stable variables with $\alpha \in [1, 2]$, in which $h(u) = C_1\, u^{\alpha-1}$ and $k(u) = C_2 u^{\alpha-1}$ for some finite C_1 and C_2. Meanwhile, Assumption 1(d) is also satisfied with the case where $\phi(u) = 2/(e^u + e^{-u})$ and then $h(u) = \frac{e^{-u}-e^u}{e^u+e^{-u}}$ and $k(u) = 1$ (Lukacs, 1970, p.88).

The regressor x_t is integrated by the linear process v_t, while the linear processes v_t and e_t have the same i.i.d. sequence $\{\epsilon_j, j \in \mathbb{Z}\}$ as building blocks. The endogeneity of the structural cointegration model is incurred accordingly. While the same type of endogeneity is used in Wang and Phillips (2014) for the kernel estimation method, the estimation method as well as the establishment and the proof of the main results in this paper are quite different from the kernel method case.

By the Beveridge-Nelson decomposition (Phillips and Solo, 1992, p. 972), $v_t = \psi\epsilon_t + \tilde{v}_{t-1} - \tilde{v}_t$ where $\tilde{v}_t = \sum_{j=0}^{\infty} \tilde{\psi}_j \epsilon_{t-j}$ with $\tilde{\psi}_j = \sum_{k=j+1}^{\infty} \psi_k$. Note that $\tilde{v}_t$ is a stationary process since $\sum_{j=0}^{\infty} |\tilde{\psi}_j|^2 < \infty$ due to (b) of Assumption 1. A similar condition is used in Phillips and Solo (1992). It follows that $x_t = \psi \sum_{j=1}^{t} \epsilon_j + \tilde{v}_0 - \tilde{v}_t$ and hence $d_t := \sqrt{Ex_t^2} = |\psi|\sqrt{t}(1 + o(1))$. Define, for $0 \leq u \leq 1$,

$$W_T(u) = \frac{1}{d_T} x_{[Tu]}. \tag{2.7}$$

It is known that $W_T(u) \to_D B(u)$, a standard Brownian motion. Straightforwardly, $E[x_t e_t] = \psi \sum_{j=1}^{t} \phi_{t-j} + E[\tilde{v}_0 e_t] - E[\tilde{v}_t e_t]$ where $|E[\tilde{v}_0 e_t]| < t^{-1}$ and $E[\tilde{v}_t e_t] = \sum_{j=0}^{\infty} \tilde{\psi}_j \phi_j$ is a constant. Generally, $E[x_t e_t] \neq 0$ but $E[d_t^{-1} x_t e_t] \to 0$ as $t \to \infty$. This implies $d_t^{-1} x_t$ and e_t are asymptotically uncorrelated for large t. More importantly, in Lemma A.5 below we claim that $d_t^{-1} x_t$ and e_s are asymptotically independent for all large t and s. Our asymptotic theory is built upon the asymptotic independence.

Meanwhile, our asymptotic theory relies on the local time process $L_B(t, s)$ of $B(u)$ defined by

$$L_B(t, s) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \int_0^t I\{|B(u) - s| < \varepsilon\} du, \tag{2.8}$$

where $I(A)$ denotes the conventional indicator function. Roughly speaking, the local time can be interpreted as a spatial occupation density in s for Brownian motion $B(u)$. 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".

Assumption 2. Let $f(x) \in L^2(\mathbb{R})$ be differentiable. Moreover, there exists a positive integer r such that $x^i f^{(r-i)}(x) \in L^2(\mathbb{R})$ for all $i = 0, \cdots, r$.

Assumption 2 requires that $f(x)$ is sufficiently smooth with the thin tail such that the orthogonal expansion converges with a fast rate. See Lemma A.3 in Appendix A. The same assumption in a different form is used by Lemma 3 of Schwartz (1967). The classes of f includes Gaussian functions, Laplace functions and functions with compact support.

Assumption 3. Let the truncation parameter p of the Hermite series expansion satisfy $p = [c \cdot T^{\alpha}]$ where $c > 0$ is a constant and $\frac{1}{2(r-1)} < \alpha < \frac{1}{5}$.

Assumption 3 restricts the truncation parameter p to guarantee the convergence of the regression matrix $Z^{\tau}Z$ and the smoothness order r to ensure the truncation residue $\gamma_p(\cdot)$ does not affect the limit distribution studied below. The condition for r and α also implies $r > \frac{7}{2}$, which can be satisfied by $r \geq 4$ in Assumption 2.

3 Asymptotic theory

3.1 Consistency of series estimator

In this subsection, we discuss the asymptotic consistency of the series estimator.

Lemma 3.1. Under Assumptions 1–3, we have as $T \to \infty$, $\|\widehat{\theta} - \theta\| = o_P(1)$, and $\sup_x |\widehat{f}(x) - f(x)| = o_P(1)$.

Lemma 3.1 shows that the estimated coefficients converge to the true coefficients and the series estimator $\widehat{f}(x)$ for $f(x)$ has a uniform convergence.

When data are stationary time series, 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. By contrast, in nonparametric and nonstationary context it is very difficult, if not impossible, to obtain a uniform convergence on the entire real line using kernel method. Gao et al. (2009), Chan and Wang (2014) and Wang and Chan (2014) study the uniform convergence that, however, happen in a compact domain of the real line. In our study, due to the uniform boundedness of Hermite series, the uniform consistency requires the same conditions as those for the point–wise consistency. This is one of advantages that series estimation has in comparison with kernel estimation.

3.2 Asymptotic distribution

In this subsection, we shall establish asymptotical distribution for the series estimator. There are two kinds of approximation of $\widehat{f}(x)$ to $f(x)$: one is pointwise, $\widehat{f}(x)-f(x)=Z_p^\tau(x)(\widehat{\theta}-\theta)-\gamma_p(x)$ for any $x\in\mathbb{R}$; another one is in the L^2-sense, $\|\widehat{f}(x)-f(x)\|^2_{L^2(\mathbb{R})}=\|\widehat{\theta}-\theta\|^2+\|\gamma_p(x)\|^2_{L^2(\mathbb{R})}$, where by definition $\|g(x)\|_{L^2(\mathbb{R})}=\left(\int g^2(x)dx\right)^{1/2}$ the norm of $g(x)\in L^2(\mathbb{R})$. The following theorem gives the asymptotic distribution of $\widehat{f}(x)$ in both the pointwise and L^2-norm sense.

Theorem 3.1. Under Assumptions 1–3, we have as $T\to\infty$

$$\sqrt{\frac{L_B(1,0)}{\sigma_e^2\|Z_p(x)\|^2}\frac{T}{d_T}}\left(\widehat{f}(x)-f(x)\right)\to_D N(0,1), \tag{3.1}$$

and moreover

$$\sqrt{\frac{1}{\sigma_e^2\cdot p}\frac{T}{d_T}}\|\widehat{f}(x)-f(x)\|_{L^2(R)}\to_D L_B^{-1/2}(1,0), \tag{3.2}$$

where $L_B(1,0)$ is a local–time random variable with its cumulative distribution function being given by

$$F_L(x)=P\left(L_B(1,0)\le x\right)=\begin{cases}2\Phi(x)-1, & x\ge 0,\\ 0, & x<0,\end{cases} \tag{3.3}$$

in which $\Phi(x)$ is the cdf of $N(0,1)$.

Since $\|Z_p(x)\|^2=O(p)$ uniformly in x and $d_T=O(T^{1/2})$, the rates of convergence of the series estimator in both pointwise and L^2-norm sense are $(T^{1/4}p^{-1/2})^{-1}$. Meanwhile, the rate of convergence of the kernel estimator is $(T^{1/4}h^{1/2})^{-1}$ (see, for example, Wang and Phillips 2014), where h is the bandwidth parameter. Thus, they are equivalent when we replace h by p^{-1}.

Note also that there are three nuisance parameters involved in the large sample theory of (3.1), namely, ψ in $d_T=|\psi|\sqrt{T}(1+o(1))$, σ_e^2 and the local time $L_B(1,0)$, which should be replaced by their consistent estimates. However, noting the structure of $L_B(1,0)/d_T$ in (3.1) and the limit $\frac{d_T}{T}\sum_{t=1}^T\phi(x_t)\to_P L_B(1,0)$ in a rich probability space where $\phi(x)=\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$, we may estimate the ratio of $L_B(1,0)/|\psi|$ by $\frac{1}{\sqrt{T}}\sum_{t=1}^T\phi(x_t)$. Moreover, we estimate σ_e^2 by

$$\widehat{\sigma}_e^2:=\frac{1}{T}\sum_{t=1}^T\widehat{e}_t^2,\quad \text{where } \widehat{e}_t:=y_t-\widehat{f}(x_t). \tag{3.4}$$

It is also possible to estimate ψ individually if we stipulate a parametric structure for the linear process v_t in Assumption 1. See Dong and Gao (2014) for the details. Thus, in practice the limit in (3.2) can also be used for inference by noting that $L_B(1,0)$ follows the same distribution as $|N|$ where N is a standard normal variable. Nonetheless, we focus only on (3.1) since the limit is normal and it does not need an estimate of ψ.

Corollary 3.1. Under Assumptions 1–3, we have as $T \to \infty$, $\widehat{\sigma}_e^2 \to_P \sigma_e^2$, $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \phi(x_t) \to_P L_B(1,0)/|\psi|$; consequently,

$$\frac{1}{\widehat{\sigma}_e \|Z_p(x)\|} \sqrt{\sum_{t=1}^{T} \phi(x_t)} \left(\widehat{f}(x) - f(x) \right) \to_D N(0,1). \tag{3.5}$$

The proofs of Lemma 3.1, Theorem 3.1 and Corollary 3.1, which are given in Appendix B, employ an asymptotic approximation of the regression matrix $Z^\tau Z$ by a diagonal matrix listed in Theorem 3.2 in the next subsection.

3.3 Asymptotic property of $Z^\tau Z$

As mentioned in the introductory section and seen in the above discussion, the least squares estimator of θ involves an inverse matrix of $Z^\tau Z$, 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 $\widetilde{\theta} = Z^\tau Z \cdot \widehat{\theta}$ (see, for example, Dong and Gao 2014). 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 $\frac{d_T}{T} Z^\tau Z$ directly.

Theorem 3.2. Let $p = [c \cdot T^\alpha]$ for $c > 0$ and $0 < \alpha < \frac{1}{5}$. Suppose that Assumption 1 holds. Then, in an expanded probability space, we have as $T \to \infty$

$$\left\| \frac{d_T}{T} Z^\tau Z - L_B(1,0) I_p \right\| \to_P 0, \tag{3.6}$$

where I_p is an identity matrix of dimension $p \times p$.

It follows from the definition of Z that

$$\left\| \frac{d_T}{T} Z^\tau Z - L_B(1,0)\ I_p \right\|^2 = \sum_{i=0}^{p-1} \left(\frac{d_T}{T} \sum_{t=1}^{T} F_i^2(x_t) - L_B(1,0) \right)^2 + \sum_{i \neq j = 0}^{p-1} \left(\frac{d_T}{T} \sum_{t=1}^{T} F_i(x_t) F_j(x_t) \right)^2 .$$

Since $p \to \infty$, existing results (Wang and Phillips 2009a, 2011, for example) regarding all terms in the bracket are not applicable. Thus, the proof of Theorem 3.2 is not trivial because the key steps used in deriving the rates of convergence for the terms in the bracket use new ideas and various properties about the orthogonal series.

As frequently encountered in the nonparametric nonstationary series estimation context, Theorem 3.2 is of independent interest. The implication is that the regression matrix $Z^\tau Z$ for the parameterized model after normalization is asymptotically a diagonal matrix with $L_B(1,0)$ at its diagonal, and hence the eigenvalues satisfy $\lambda_{\min}(\frac{d_T}{T} Z^\tau Z) = L_B(1,0) + o_P(1)$ and $\lambda_{\max}(\frac{d_T}{T} Z^\tau Z) = L_B(1,0) + o_P(1)$. Our experience suggests that such convergence itself may be applicable to significantly simplify the construction of existing estimation and specification procedures, such as those discussed in Dong and Gao (2013, 2014).

The proof of Theorem 3.2 is given in Appendix C of the supplementary material. In Section 4 below, we examine the finite–sample performance of the series estimation.

4 Simulation study

In this section, we conduct Monte Carlo experiments to assess the finite sample performance of the proposed nonparametric series estimator. The data generation procedure is as follows. Let $\{\epsilon_t, e_t\}$ be an independent and identically distributed sequence, $\{\epsilon_t, e_t\} \sim N(0, \Sigma)$ with $\Sigma = 0.1^2 \left(\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right)$. The regressor x_t is integrated by an AR(1) process v_t, i.e.

$$x_t = x_{t-1} + v_t \quad \text{and} \quad v_t = 0.2\ v_{t-1} + \epsilon_t,$$

where $x_0 = O_P(1)$. The following models are used to investigate the performance:

$$\text{Model 1}: \ y_t = \frac{1}{1+x_t^4} + e_t, \quad t = 1, ..., T; \tag{4.1}$$

$$\text{Model 2}: \ y_t = (1+\sin(x_t)) \exp(-x_t^2/2) + e_t, \quad t = 1, ..., T. \tag{4.2}$$

We shall consider two cases for ρ: $\rho = 0$, implying the case of exogeneity, and $\rho = 0.9$, implying the existence of endogeneity.

4.1 Bias and standard deviation

Let $T = 400$, 800, 1200 and 1800 be the sample sizes. The number of replications is 2000. Using a generalised cross–validation method proposed in Gao et al. (2002), the truncation parameter is chosen as $p = [2 \cdot T^{1/8}]$ such that it varies along with the sample size and satisfies the theoretical requirement in Assumption 3.

The sample bias, standard deviation (Std) and root mean square error (RMSE) are defined by

$$\text{Bias} = \frac{1}{N}\frac{1}{T}\sum_{n=1}^{N}\sum_{t=1}^{T}\left| f(x_{n,t}) - \overline{\widehat{f}}(x_{n,t})\right|,$$

$$\text{Std} = \left(\frac{1}{N}\frac{1}{T}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(\widehat{f}(x_{n,t}) - \overline{\widehat{f}}(x_{n,t})\right)^2\right)^{1/2},$$

$$\text{RMSE} = \left(\frac{1}{N}\frac{1}{T}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(\widehat{f}(x_{n,t}) - f(x_{n,t})\right)^2\right)^{1/2},$$

respectively, where $(x_{n,1}, \cdots, x_{n,T})$ denotes the simulated data in $n-th$ replication, and by which $\widehat{f}(\cdot)$ is the series estimator of the regression function, and $\overline{\widehat{f}}(\cdot) = Z_p(\cdot)^\tau \overline{\widehat{\theta}}$ with $\overline{\widehat{\theta}}$ being the average of $\widehat{\theta}_n$ over Monte Carlo replications, and N is the number of replications. The results of the simulation are summarised in Table 1.

It should be pointed out that the sample size of simulation for nonstationary integrable regression models usually has to be much larger than that for stationary regression models. The reason is the slower rate of convergence in the former case.

It can be seen from Table 1 that both the bias and the standard deviation decrease with the increase of the sample size. These verify the approximation of the proposed estimator

Table 1: Simulation Results for Bias, Std and RMSE

		$\rho = 0$		$\rho = 0.9$	
	T	Model 1	Model 2	Model 1	Model 2
Bias	400	0.0819	0.0146	0.0899	0.0465
	800	0.0714	0.0132	0.0774	0.0333
	1200	0.0551	0.0121	0.0589	0.0221
	1800	0.0256	0.0110	0.0250	0.0067
Std	400	0.0904	0.0218	0.1037	0.0538
	800	0.0803	0.0194	0.0901	0.0385
	1200	0.0620	0.0185	0.0680	0.0293
	1800	0.0265	0.0171	0.0289	0.0104
RMSE	400	0.0583	0.0471	0.0586	0.0473
	800	0.0297	0.0080	0.0293	0.0075
	1200	0.0288	0.0071	0.0283	0.0066
	1800	0.0252	0.0061	0.0247	0.0054

to the true regression function. Comparing the results of the two models, however, Model 2 outperforms Model 1 in all sample sizes. According to our experience this may be mainly due to the difference in the tails of two regression functions, namely, the tail of $1/(1 + x^4)$ is much heavier than that of $(1+\sin(x)) \exp(-x^2/2)$. It is known that the heavier tail results in a slower convergence of the orthogonal series expansion. Consequently, Model 2 has better results than Model 1.

Additionally, the results for the case of $\rho = 0$ and the case $\rho = 0.9$ have no evidence to show how they are different. Based on these results, the endogeneity does not affect the nonparametric estimate in the proposed models, and more importantly, this coincides with our theoretical findings in the preceding section.

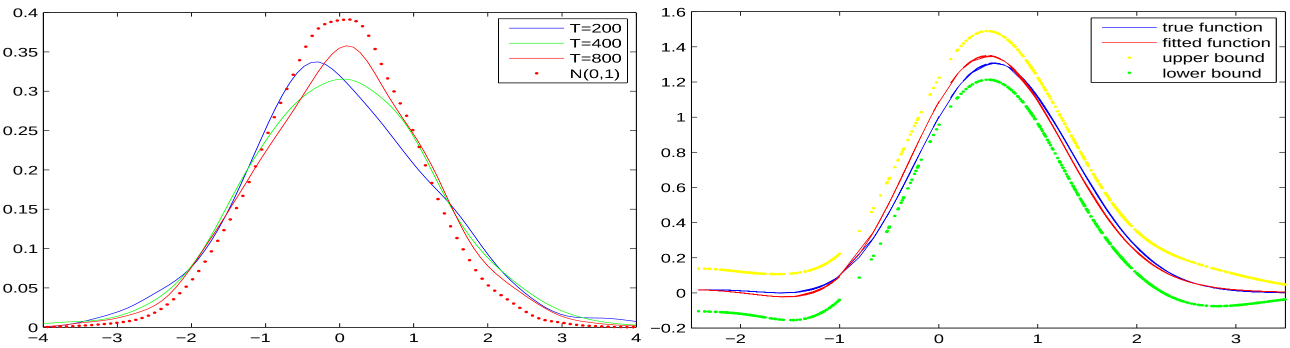
4.2 Normal approximation and confidence interval curves

Corollary 3.1 gives the normality of our estimator $\widehat{f}(x)$ with all nuisance parameters estimated by the observation $\{(x_t, y_t),\, t = 1, \cdots, T\}$. Accordingly, we are able to construct the confidence interval at a significance level and any point. This section devotes to the visualization of the normality.

To do so, using *ksdensity* function in MatLab we first estimate the density of a set of $\widehat{f}(x^*) - f(x^*)$ with normalization according to Corollary 3.1 for a particular point $x^* = 0$ for

Model 2 with $T = 200, 400$ and 800 for $\rho = 0$ and $N = 1000$, where the truncation parameter is taken using the same formula as before, viz., $p = [2 \times T^{1/8}]$.

Figure 1: Normal density approximation and confidence interval curves



(a) Normal density approximation (b) Confidence interval curves

Technically, we only use the replications that both the numbers of observations less than and larger than zero are greater than 0.2 T. The reason is that, due to the divergence of the integrated data, it is possible that the generated data $(x_{n1}, \cdots, x_{nT})$, where n corresponds to the n–th replication of the total number of replications, in one replication may be located mostly in one side of zero, which definitely gives a poor estimation of the density, particularly for the kernel method of *ksdensity* function in Matlab. Similar discussion is available in Section 5 of Karlsen et al. (2007).

Figure 1a shows three estimated density curves corresponding to the different sample size T. It can be seen that the densities gradually approach to the standard normal density with the increase of the sample size. We may conclude that the theoretical result of the normality in Corollary 3.1 is verified in this experiment.

Second, for significance level 95%, we draw for Model 2 the lower bound and upper bound confidence curves based on the result of (3.5), namely, $\widehat{f}(x) \pm 1.96\, \widehat{\sigma}_e \|Z_p(x)\| (\sum_{t=1}^{T} \phi(x_t))^{-1/2}$, where $\phi(\cdot)$ is the density function of a standard normal variable. Here, $T = 800$ and p is the same as before. Figure 1b displays the true regression function, the estimated function averaging over replications and the confidence interval curves. As can be seen, the estimated curve $\widehat{f}(x)$ is located exactly between the lower bound and the upper bound, implying the reliability of inference based on our estimator.

5 Discussion

It is worthy to discuss potential extensions of our method to deal with models where regression functions are not in $L^2(\mathbb{R})$. The following is a brief discussion on this issue. Consider $y_t = f(x_t) + e_t$ where x_t and e_t still satisfy Assumption 1 but $f(x) \in L^2(\mathbb{R}, e^{-x^2})$. It follows that $\tilde{f}(x) := f(x)\varphi(x) \in L^2(\mathbb{R})$ where $\varphi(x) = e^{-x^2/2}$. This motivates multiplying the both sides of the model by $\varphi(x_t)$, giving

$$\tilde{y}_t = \tilde{f}(x_t) + \tilde{e}_t, \quad t = 1, \cdots, T, \tag{5.1}$$

where $\tilde{y}_t = y_t\varphi(x_t)$, and $\tilde{e}_t = e_t\varphi(x_t)$. Now, model (5.1) is completely the same as model (2.4). Expand $\tilde{f}(x)$ into orthogonal series in terms of $\{F_i(x)\}$:

$$\tilde{f}(x) = \sum_{i=0}^{\infty} \tilde{\theta}_i F_i(x) = Z_p^{\tau}(x)\tilde{\theta} + \tilde{\gamma}(x), \quad \text{with } \tilde{\theta}_i = \int \tilde{f}(x)F_i(x)dx \tag{5.2}$$

where for any $p \geq 1$, $\tilde{\theta} = (\tilde{\theta}_0, \cdots, \tilde{\theta}_{p-1})^{\tau}$, $Z_p(x)$ is the same as before and $\tilde{\gamma}_p(x) = \sum_{i=p}^{\infty} \tilde{\theta}_i F_i(x)$.

Suppose further that $\tilde{f}(x)$ and truncation parameter p satisfy Assumptions 2 and 3. We are able to have an estimator of $\tilde{f}(x)$ following exactly the same procedure as in Section 2.2,

$$\widehat{\tilde{f}}(x) = Z_p^{\tau}(x)\widehat{\tilde{\theta}}, \quad \text{where } \widehat{\tilde{\theta}} = (Z^{\tau}Z)^{-1}Z^{\tau}\tilde{Y}, \tag{5.3}$$

in which $\widehat{\tilde{\theta}}$ is an estimate of $\tilde{\theta}$, Z is the same as before and $\tilde{Y} = (\tilde{y}_1, \cdots, \tilde{y}_T)^{\tau}$. Denote for later use that $\tilde{e} = (\tilde{e}_1, \cdots, \tilde{e}_T)^{\tau}$ and $\tilde{\gamma} = (\tilde{\gamma}_p(x_1), \cdots, \tilde{\gamma}_p(x_T))^{\tau}$.

To derive the asymptotic distribution of $\widehat{\tilde{f}}(x)$, notice that, for any $x \in \mathbb{R}$, $\widehat{\tilde{f}}(x) - \tilde{f}(x) = Z_p^{\tau}(x)(\widehat{\tilde{\theta}} - \tilde{\theta}) - \tilde{\gamma}(x)$ and

$$\begin{aligned} \|\widehat{\tilde{f}}(x) - \tilde{f}(x)\|^2_{L^2(r)} &= \int [\widehat{\tilde{f}}(x) - \tilde{f}(x)]^2 dx = \|\widehat{\tilde{\theta}} - \tilde{\theta}\|^2 + \int \tilde{\gamma}_p^2(x)dx - 2\int Z_p^{\tau}(x)(\widehat{\tilde{\theta}} - \tilde{\theta})\tilde{\gamma}_p(x)dx \\ &= \|\widehat{\tilde{\theta}} - \tilde{\theta}\|^2 + \|\tilde{\gamma}(x)\|^2_{L^2(\mathbb{R})}. \end{aligned}$$

Hence, following a similar fashion we may be able to establish the asymptotic distribution of $\widehat{\tilde{f}}(x)$ in both the point–wise and the L^2 sense.

Meanwhile, it is possible to extend the approach in Sections 2 and 3 to a partially linear single–index model of the form: $y_t = x_t^{\tau}\beta_0 + f(x_t^{\tau}\theta_0) + e_t$, where x_t is a vector of integrated time series, (β_0, θ_0) is a vector of unknown parameters and $f(\cdot)$ is an unknown integrable function. In empirical applications, a vector of macro-economic time variables, such as the income and real interest rate variable, may be chosen as x_t and y_t can be the expenditure variable when are interested in establishing the relationship between y_t and x_t. In order to establish similar results to Theorem 3.1 and Corollary 3.1, some new techniques may be needed. We therefore wish to leave such extensions to future research.

6 Conclusions

In this paper, we have established the uniform consistency and asymptotic distribution in both the point–wise and L^2 sense for the Hermite series estimator of the proposed integrable cointegration model accommodating endogeneity. The endogeneity is of a general form. Possible extensions from integrable models to non-integrable models have been discussed. The finite sample experiments show that the proposed series estimator performs well for models satisfying our assumptions.

Nonetheless, there are some problems that may be studied in our future research. The choice of the truncation parameter should be discussed in more detail and a data driven choice of the truncation parameter should be investigated. The theory may be extended to an additive multivariate model with both stationary and nonstationary regressors or a partially linear cointegration model.

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 comments by Farshid Vahid and the financial support by the Australian Research Council Discovery Grants Program under Grant Numbers: DP1096374 and DP1314229.

A Lemmas

Five useful lemmas are given in this section. All their proofs can be found in Appendix C of this paper. Throughout the rest of this paper, we use $0 < C < \infty$ to denote a generic constant which may have different values at different places. Meanwhile, we use $||\cdot||_{L^2}$ to simplify $||\cdot||_{L^2(\mathbb{R})}$ in the proofs.

We shall consider several versions of decomposition for x_t. Without loss of generality, in what follows let $x_0 = 0$ almost surely. It follows that

$$x_t = \sum_{\ell=1}^{t} v_\ell = \sum_{\ell=1}^{t} \sum_{i=-\infty}^{\ell} \psi_{\ell-i}\epsilon_i = \sum_{i=-\infty}^{t} \left(\sum_{\ell=\max(1,i)}^{t} \psi_{\ell-i} \right) \epsilon_i =: \sum_{i=-\infty}^{t} b_{t,i}\epsilon_i.$$

Let $j \le t$ be fixed. Thus we have

$$x_t = b_{t,j}\epsilon_j + x_{t/j}, \quad \text{with } x_{t/j} := \sum_{i=-\infty,\neq j}^{t} b_{t,i}\epsilon_i, \tag{A.1}$$

where $x_{t/j}$ is the variable deducting the term containing ϵ_j in x_t. Obviously, $x_{t/j}$ and ϵ_j are mutually independent.

Additionally, letting $1 \le s < j \le t$, x_t also has the following decomposition:

$$x_t = x_s^* + x_{ts} = x_s^* + b_{t,j}\epsilon_j + x_{ts/j}, \tag{A.2}$$

where $x_s^* = x_s + \bar{x}_s$ with $\bar{x}_s = \sum_{i=s+1}^{t} \sum_{a=-\infty}^{s} \psi_{i-a}\epsilon_a$ containing all the information available up to s and $x_{ts} = \sum_{i=s+1}^{t} b_{t,i}\epsilon_i$, while obviously $x_{ts/j} = \sum_{i=s+1,\neq j}^{t} b_{t,i}\epsilon_i$. Evidently, x_{ts} captures all the information contained in x_t on the time periods $(s, t]$, while $x_{ts/j}$ captures all information contained in x_t on the time periods $(s, j) \cup (j, t]$. Let $d_{ts} := (Ex_{ts}^2)^{1/2}$ throughout the rest of this paper. Moreover, $\bar{x}_s = O_P(1)$ by virtue of Assumption 1.

Lemma A.1. Suppose that Assumption 1 holds. For t or $t - s$ is large,

(1) $d_t^{-1}x_t$ have uniformly bounded densities $f_t(x)$ over all t and x satisfying a uniform Lipschitz condition $\sup_x |f_t(x+y) - f_t(x)| \le C|y|$ for any y and some constant $C > 0$. In addition, $\sup_x |f_t(x) - \phi(x)| \to 0$ as $t \to \infty$ where $\phi(x)$ is the standard normal density function.

Let $1 \le s < t$. $d_{ts}^{-1}x_{ts}$, where x_{ts} is given by (A.2), have uniformly bounded densities $f_{ts}(x)$ over all t, s and x satisfying the above uniform Lipschitz condition as well.

(2) Let $j \le t$. $d_t^{-1}x_{t/j}$, where $x_{t/j}$ is given by (A.1), have uniformly bounded densities $f_{t/j}(x)$ over all t, j and x satisfying uniform Lipschitz condition in the above form.

Let $1 \le s < j \le t$. $d_{ts}^{-1}x_{ts/j}$, where $x_{ts/j}$ is given by (A.2), have uniformly bounded densities $f_{ts/j}(x)$ over all t, j and s, x satisfying the above uniform Lipschitz condition as well.

Lemma A.2. Suppose that Assumption 1 holds. Let j be a fixed integer and $j \le t$. For any functions U and $g : \mathbb{R} \mapsto \mathbb{R}$ such that $\int |U(w)|dw < \infty$ and $E|\epsilon_j g(\epsilon_j)| < \infty$, and for large t or $t - s$, we have

(1) $E[U(x_t)g(\epsilon_j)] = E[U(x_{t/j})]E[g(\epsilon_j)] + c_U \frac{1}{t}$ where $x_{t/j}$ defined by (A.1) is independent of ϵ_j, c_U is a quantity such that $|c_U| \le C\ E|\epsilon_j g(\epsilon_j)| \int |U(w)|dw$. In particular, if $Eg(\epsilon_j) = 0$, then $E[U(x_t)g(\epsilon_j)] = c_U \frac{1}{t}$;

(2) $E|U(x_t)g(\epsilon_j)| \le C\ \frac{1}{\sqrt{t}} E|g(\epsilon_j)| \int |U(w)|dw$;

(3) For any ℓ: $j \ne \ell \le t$, $E[U(x_t)g(\epsilon_j)|\epsilon_\ell] = E[U(x_{t/j})|\epsilon_\ell]E[g(\epsilon_j)] + \frac{1}{t}\eta_\ell$ where $x_{t/j}$ is defined by (A.1), η_ℓ is a random variable depending on ϵ_ℓ such that $|\eta_\ell| \le C\ \int |U(w)|dw$ almost surely. If $E[g(\epsilon_j)] = 0$, then $E[U(x_t)g(\epsilon_j)|\epsilon_\ell] = \frac{1}{t}\eta_\ell$.

Meanwhile, $E[|U(x_t)g(\epsilon_j)||\epsilon_\ell] \le C\ \frac{1}{\sqrt{t}} E|g(\epsilon_j)| \int |U(w)|dw$ almost surely.

(4) For $1 \le s < j \le t$, $E[U(x_t)g(\epsilon_j)|\mathcal{F}_s] = E[g(\epsilon_j)]E[U(x_{t/j})|\mathcal{F}_s] + \frac{1}{t-s}\xi_s$, where $|\xi_s| \le O(1)E|\epsilon_j g(\epsilon_j)| \int |U(x)|dx$ almost surely; meanwhile, $E[|U(x_t)g(\epsilon_j)||\mathcal{F}_s] \le O(1)\frac{1}{\sqrt{t-s}}E[|g(\epsilon_j)|] \int |U(w)|dw$ a.s..

Lemma A.3. (1) (i) $\sup_x \|Z_p(x)\|^2 = O(1)p$; (ii) $\int \|Z_p(x)\|^2 dx = p$; (iii) $\int \|Z_p(x)\|dx = O(1)p^{11/12}$; (iv) $\int x^2 F_i^2(x)dx = (2i+1)/2$.

(2) Let Assumption 2 hold. Then, (i) $\sup_x |\gamma_p(x)| = o\left(p^{-(r-1)/2-1/12}\right)$; (ii) $\int \gamma_p^2(x)dx = o(1)p^{-r}$.

Lemma A.4. $Z_p^\tau(x)Z_p(y) \to \delta(x-y)$ as $p \to \infty$, where $\delta(u)$ is the Dirac delta function.

The Dirac delta function $\delta(u)$ is a generalized function satisfying $\delta(u) = 0$ for any $u \ne 0$ and $\int \delta(u)du = 1$. See Kanwal (1983, p. 5).

Lemma A.5. Let m_T be a sequence such that $m_T \to \infty$ and $m_T/T \to 0$ as $T \to \infty$. Let $\{a_j\}$ be any sequence of nonnegative real numbers satisfying $\sum_{i=m_T}^T a_i = 1$.

(1) For any $s \ge 1$ and $t \ge m_T$, e_s and $d_t^{-1}x_t$ are asymptotically independent. Consequently, for any given k and $t \ge m_T$, ϵ_k and $d_t^{-1}x_t$ are asymptotically independent.

(2) For any $s \ge 1$, e_s and $\sum_{j=m_T}^T a_j d_j^{-1}x_j$ are asymptotically independent.

B Proofs of the main results

This appendix gives the proofs of Lemma 3.1, Theorem 3.1 and Corollary 3.1.

Proof of Lemma 3.1. Notice by (2.6) and Theorem 3.2 that

$$\begin{aligned}\widehat{\theta}-\theta =&(Z^\tau Z)^{-1}Z^\tau Y-\theta=(Z^\tau Z)^{-1}Z^\tau(e+\gamma)\\ =&(1+o_P(1))L_B^{-1}(1,0)\frac{d_T}{T}Z^\tau(e+\gamma):=(1+o_P(1))L_B^{-1}(1,0)(A_{1T}+A_{2T}),\end{aligned}$$

where we will ignore the small order "$o_P(1)$" for notational simplicity in the rest of the derivations. We will also use "$O(1)$" to denote any finite positive constant. Observe that

$$\begin{aligned}\|A_{1T}\|^2 =&\frac{d_T^2}{T^2}\left\|\sum_{t=1}^T Z_p(x_t)e_t\right\|^2=\frac{d_T^2}{T^2}\left(\sum_{t=1}^T Z_p(x_t)e_t\right)^\tau\sum_{t=1}^T Z_p(x_t)e_t\\ =&\frac{d_T^2}{T^2}\sum_{t=1}^T\|Z_p(x_t)\|^2e_t^2+2\frac{d_T^2}{T^2}\sum_{t=2}^T\sum_{s=1}^{t-1}Z_p^\tau(x_t)Z_p(x_s)e_te_s\\ =&\sigma_e^2\frac{d_T^2}{T^2}\sum_{t=1}^T\|Z_p(x_t)\|^2+\frac{d_T^2}{T^2}\sum_{t=1}^T\|Z_p(x_t)\|^2(e_t^2-\sigma_e^2)+2\frac{d_T^2}{T^2}\sum_{t=2}^T\sum_{s=1}^{t-1}Z_p^\tau(x_t)Z_p(x_s)e_te_s\\ :=&A_{11T}+A_{12T}+2A_{13T},\quad \text{say.}\end{aligned}$$

Using the density $f_t(x)$ of $d_t^{-1}x_t$ and its uniform boundedness, we have

$$\begin{aligned}\frac{d_T^2}{T^2}\sum_{t=1}^T E\|Z_p(x_t)\|^2 =&\frac{d_T^2}{T^2}\sum_{t=1}^T d_t^{-1}\int\|Z_p(x)\|^2f_t(d_t^{-1}x)dx\\ \le& O(1)\frac{1}{T}\sum_{t=1}^T d_t^{-1}\int\|Z_p(x)\|^2dx=O(1)T^{-1/2}p,\end{aligned}$$

implying that $A_{11T}=O_P(T^{-1/2}p)$.

For the second term A_{12T}, notice that $e_t^2-\sigma_e^2=\sum_{j=0}^\infty\phi_j^2(\epsilon_{t-j}^2-1)+\sum_{j_1=0,\neq j_2}^\infty\phi_{j_1}\phi_{j_2}\epsilon_{t-j_1}\epsilon_{t-j_2}$. Moreover, by (1) and (3) of Lemma A.2 and conditional argument, we have

$$\begin{aligned}&|E[A_{12T}]|\le\frac{d_T^2}{T^2}\sum_{t=1}^T\sum_{j=-\infty}^t\phi_{t-j}^2|E[\|Z_p(x_t)\|^2(\epsilon_j^2-1)]|\\ &+\frac{d_T^2}{T^2}\sum_{t=1}^T\sum_{j_1-\infty}^t\sum_{j_2=-\infty}^{j_1-1}|\phi_{t-j_1}\phi_{t-j_2}E[\|Z_p(x_t)\|^2\epsilon_{j_1}\epsilon_{j_2}]|\\ &\le O(1)\frac{1}{T}\sum_{t=1}^T d_t^{-2}\sum_{j=-\infty}^t\phi_{t-j}^2\int\|Z_p(x)\|^2dx\\ &+O(1)\frac{1}{T}\sum_{t=1}^T d_t^{-2}\sum_{j_1-\infty}^t\sum_{j_2=-\infty}^{j_1-1}|\phi_{t-j_1}\phi_{t-j_2}|\int\|Z_p(x)\|^2dx=O(1)\frac{1}{T}p\ln(T).\end{aligned}$$

For A_{13T}, notice that, for $t>s$,

$$e_te_s=e_s\sum_{j=s+1}^t\phi_{t-j}\epsilon_j+e_s\sum_{j=-\infty}^s\phi_{t-j}\epsilon_j$$

$$=e_s\sum_{j=s+1}^{t}\phi_{t-j}\epsilon_j+\sum_{j=-\infty}^{s}\phi_{s-j}\phi_{t-j}\epsilon_j^2+\sum_{j_1=-\infty}^{s}\sum_{j_2=-\infty}^{j_1-1}\phi_{s-j_2}\phi_{t-j_1}\epsilon_{j_1}\epsilon_{j_2}.$$

Meanwhile, we only need to tackle the terms in A_{13T} where $t-s>m_T$ with m_T satisfying $m_T^4/T\to 0$ and $m_T\to\infty$ as $T\to\infty$. Because for the rest we are able to control them as small as we wish. Moreover, since the probability of $x_t=x_s$ is zero, we exclude these regions in the following calculation of expectations. Normally, the exclusion does not make any difference but due to Lemma A.4 this time it really matters.

Thus, it follows from (3) and (4) of Lemma A.2 that

$$\begin{aligned}|E[A_{13T}]|\leq&\frac{d_T^2}{T^2}\sum_{t=m_T+1}^{T}\sum_{s=1}^{t-m_T}\sum_{j=s+1}^{t}|\phi_{t-j}||E[Z_p^\tau(x_t)Z_p(x_s)e_s\epsilon_j]|\\&+\frac{d_T^2}{T^2}\sum_{t=m_T+1}^{T}\sum_{s=1}^{t-m_T}\sum_{j=-\infty}^{s}|\phi_{s-j}\phi_{t-j}||E[Z_p^\tau(x_t)Z_p(x_s)\epsilon_j^2]|\\&+\frac{d_T^2}{T^2}\sum_{t=m_T+1}^{T}\sum_{s=1}^{t-m_T}\sum_{j_1=-\infty}^{s}\sum_{j_2=-\infty}^{j_1-1}|\phi_{s-j_2}\phi_{t-j_1}||E[Z_p^\tau(x_t)Z_p(x_s)\epsilon_{j_1}\epsilon_{j_2}]|\\\leq&O(1)\frac{1}{T}\sum_{t=m_T+1}^{T}\sum_{s=1}^{t-m_T}d_{ts}^{-2}d_s^{-1}\sum_{j=s+1}^{t}|\phi_{t-j}|\iint_{x\neq y}|Z_p^\tau(x)Z_p(y)|dxdy\\&+O(1)\frac{1}{T}\sum_{t=m_T+1}^{T}\sum_{s=1}^{t-m_T}d_{ts}^{-1}d_s^{-1}\sum_{j=-\infty}^{s}|\phi_{s-j}\phi_{t-j}|\iint_{x\neq y}|Z_p^\tau(x)Z_p(y)|dxdy\\&+O(1)\frac{1}{T}\sum_{t=m_T+1}^{T}\sum_{s=1}^{t-m_T}d_{ts}^{-1}d_s^{-1}\sum_{j_1=-\infty}^{s}\sum_{j_2=-\infty}^{j_1-1}|\phi_{s-j_2}\phi_{t-j_1}|\iint_{x\neq y}|Z_p^\tau(x)Z_p(y)|dxdy\\\leq&o(1)\frac{1}{T}\sqrt{T}+o(1)\frac{1}{T}\sum_{t=m_T+1}^{T}\sum_{s=1}^{t-m_T}d_{ts}^{-1}d_s^{-1}(t-s)^{-1}=o\left(T^{-1/2}\right),\end{aligned}$$

where we have used Assumption 1 that $\sum_j j|\phi_j|<\infty$ to derive $\sum_{j\leq s}|\phi_{t-j}|=O(1)(t-s)^{-1}$, and Lemma A.4 has used to have the double integral being $o(1)$ by virtue of the Dirac function. These calculations give that $A_{1T}=O_P(1)T^{-1/4}p^{1/2}$.

Next, consider the term A_{2T}. Note that

$$\begin{aligned}E\|A_{2T}\|\leq&\frac{d_T}{T}\sum_{t=1}^{T}E\|Z_p(x_t)\gamma_p(x_t)\|\leq O(1)\frac{d_T}{T}\sum_{t=1}^{T}d_t^{-1}\int\|Z_p(x)\gamma_p(x)\|dx\\\leq&O(1)\left(\int\|Z_p(x)\|^2dx\int|\gamma_p(x)|^2dx\right)^{1/2}=O(1)p^{1/2}\|\gamma_p(x)\|_{L^2}\end{aligned}$$

The assertion for $\widehat{\theta}$ follows, in view of Lemma A.3. Precisely, $\|\widehat{\theta}-\theta\|=O_P(1)p^{1/2}\max(T^{-1/4},\|\gamma_p(x)\|_{L^2})$.

For the second part, by the result for $\widehat{\theta}-\theta$,

$$\begin{aligned}&\sup_x|\widehat{f}(x)-f(x)|\leq\sup_x|Z_p^\tau(x)(\widehat{\theta}-\theta)|+\sup_x|\gamma_p(x)|\\\leq&\sup_x\|Z_p(x)\|\|\widehat{\theta}-\theta\|+\sup_x|\gamma_p(x)|=O_P(1)p\max(T^{-1/4},\|\gamma_p(x)\|_{L^2})+\sup_x|\gamma_p(x)|=o_P(1).\end{aligned}$$

□

Proof of Theorem 3.1. This proof includes two parts, Part One and Part Two, to show (3.1) and (3.2), respectively.

Part One. Notice by Theorem 3.2 that

$$\begin{aligned}\widehat{f}(x)-f(x)=&Z_p^\tau(x)(\widehat{\theta}-\theta)-\gamma_p(x)=Z_p^\tau(x)(Z^\tau Z)^{-1}Z^\tau(\gamma+e)-\gamma_p(x)\\=&\frac{d_T}{T}L_B^{-1}(1,0)Z_p^\tau(x)Z^\tau(\gamma+e)(1+o_P(1))-\gamma_p(x)\\=&\frac{d_T}{T}L_B^{-1}(1,0)Z_p^\tau(x)Z'e+\frac{d_T}{T}L_B^{-1}(1,0)Z_p^\tau(x)Z^\tau\gamma-\gamma_p(x)\\=&\frac{d_T}{T}L_B^{-1}(1,0)\sum_{t=1}^{n}Z_p^\tau(x)Z_p(x_t)e_t+\frac{d_T}{T}L_B^{-1}(1,0)Z_p^\tau(x)Z^\tau\gamma-\gamma_p(x),\end{aligned}$$

where $Z_p^\tau(x)Z_p(x_t)=\sum_{i=0}^{p-1}F_i(x)F_i(x_t)$, and we have replaced "$(1+o_P(1))$" by 1.

It follows that

$$\begin{aligned}\frac{T}{d_T}L_B(1,0)[\widehat{f}(x)-f(x)]=&\sum_{t=1}^{T}Z_p^\tau(x)Z_p(x_t)e_t+Z_p^\tau(x)Z^\tau\gamma-\frac{T}{d_T}L_B(1,0)\gamma_p(x)\\:=&A_{1T}+A_{2T}-A_{3T},\quad\text{say.}\end{aligned}\tag{B.1}$$

Moreover, choose $m_T\to\infty$ and $\frac{m_T}{p\sqrt{T}}\to 0$. Similar to the evaluation of A''_{1T} below, we may show that $A_{1T}=\sum_{t=1}^{T}Z_p^\tau(x)Z_p(x_t)e_t=\sum_{t=1}^{m_T-1}Z_p^\tau(x)Z_p(x_t)e_t+\sum_{t=m_T}^{T}Z_p^\tau(x)Z_p(x_t)e_t=(1+o_P(1))\cdot\sum_{t=m_T}^{T}Z_p^\tau(x)Z_p(x_t)e_t$. Denote $A_{1T}^*=\sum_{t=m_T}^{T}Z_p^\tau(x)Z_p(x_t)e_t$ for convenience.

Let $B_T=B_T(x;x_{m_T},\cdots,x_T)=\left[\sum_{t=m_T}^{T}\left(Z_p^\tau(x)Z_p(x_t)\right)^2\right]^{1/2}$. We then show that

$$B_T^{-1}A_{1T}\to_D N\left(0,\sigma_e^2\right)\quad\text{and}\quad B_T^{-1}A_{iT}\to_P 0\tag{B.2}$$

for $i=2,3$, where $\sigma_e^2=E[e_1^2]$. We start to prove the first part of (B.2). Towards this end, it suffices to show that $B_T^{-1}A_{1T}^*\to_D N\left(0,\sigma_e^2\right)$ as $T\to\infty$. As shown in Lemma A.5, for any $s\geq 1$ and $t\geq m_T$, e_s and $d_t^{-1}x_t$ are asymptotically independent. Therefore, the dominated convergence theorem implies that as $T\to\infty$

$$\left|P(B_T^{-1}A_{1T}^*<u)-\Phi(u)\right|\leq E\left|E\left[I(B_T^{-1}A_{1T}^*<u)|d_{m_T}^{-1}x_{m_T},\cdots,d_T^{-1}x_T\right]-\Phi(u)\right|\to 0,\tag{B.3}$$

if, for any $u\in\mathbb{R}$,

$$\left|P(B_T^{-1}A_{1T}^*<u|d_{m_T}^{-1}x_{m_T},\cdots,d_T^{-1}x_T)-\Phi(u)\right|\to_P 0,\tag{B.4}$$

as $T\to\infty$, where $\Phi(u)$ is the distributional function of a standard normal random variable.

Recall that $e_t=\sum_{j=0}^{\infty}\phi_j\epsilon_{t-j}$. Notice that

$$\begin{aligned}A_{1T}^*=&\sum_{t=m_T}^{T}Z_p^\tau(x)Z_p(x_t)e_t=\sum_{t=m_T}^{T}Z_p^\tau(x)Z_p(x_t)\sum_{j=0}^{\infty}\phi_j\epsilon_{t-j}\\=&\sum_{t=m_T}^{T}Z_p^\tau(x)Z_p(x_t)\sum_{j=-\infty}^{t}\phi_{t-j}\epsilon_j=\sum_{j=-\infty}^{T}\left(\sum_{t=\max(m_T,j)}^{T}Z_p^\tau(x)Z_p(x_t)\phi_{t-j}\right)\epsilon_j\\=&\sum_{j=m_T}^{T}\left(\sum_{t=j}^{T}Z_p^\tau(x)Z_p(x_t)\phi_{t-j}\right)\epsilon_j+\sum_{j=-\infty}^{m_T-1}\left(\sum_{t=m_T}^{T}Z_p^\tau(x)Z_p(x_t)\phi_{t-j}\right)\epsilon_j\\=&A'_{1T}+A''_{1T},\quad\text{say.}\end{aligned}$$

Let $D_T := D_T(x; x_{m_T}, \cdots, x_T) = \left[\sum_{j=m_T}^T \left(\sum_{t=j}^T Z_p^\tau(x) Z_p(x_t)\phi_{t-j}\right)^2\right]^{1/2}$. We then have

$$L_B(1,0)\frac{T}{d_T} D_T^{-1}[\widehat{f}(x) - f(x)] = D_T^{-1} A_{1T}^*(1 + o_P(1)) + D_T^{-1} A_{2T} - D_T^{-1} A_{3T}$$
$$= D_T^{-1} A_{1T}' + D_T^{-1} A_{1T}'' + D_T^{-1} A_{2T} - D_T^{-1} A_{3T}.$$

Denote $D_T^{-1} A_{1T}' = \sum_{j=m_T}^T v_{jT}\epsilon_j$, in which $v_{jT} = D_T^{-1}\sum_{t=j}^T Z_p^\tau(x) Z_p(x_t)\phi_{t-j}$. As shown in the derivation of D_T below, $D_T^2 = B_T^2\, \sigma_e^2\, (1 + o_P(1))$. In order to prove (B.4), in view of Lemma A.5 that ϵ_j and $d_t^{-1} x_t$ are asymptotically independent for any $j \geq m_T$ and $t \geq m_T$, it suffices to show that for $u \in \mathbb{R}$ and as $T \to \infty$,

$$\left|P(D_T^{-1} A_{1T}' < u | d_{m_T}^{-1} x_{m_T}, \cdots, d_T^{-1} x_T) - \Phi(u)\right| \to_P 0. \tag{B.5}$$

We now employ Lemma 1 of Robinson (1997) to prove (B.5). Observe that the condition (2.2) of the lemma is satisfied automatically due to $\sum_{j=1}^T v_{jT}^2 = 1$, and hence what we need to show is that the condition (2.3) is fulfilled, i.e., $\lim_{T\to\infty} \max_{1\leq j\leq T} |v_{jT}| = 0$ in probability. To begin, note that

$$D_T^2 = \sum_{j=m_T}^T \left(\sum_{t=j}^T Z_p^\tau(x) Z_p(x_t)\phi_{t-j}\right)^2 = \sum_{j=m_T}^T \sum_{t=j}^T \left(Z_p^\tau(x) Z_p(x_t)\phi_{t-j}\right)^2$$
$$+ 2\sum_{j=m_T}^T \sum_{t=j+1}^T \sum_{s=j}^{t-1} Z_p^\tau(x) Z_p(x_t)\phi_{t-j} Z_p^\tau(x) Z_p(x_s)\phi_{s-j} =: D_{1T} + D_{2T}.$$

The leading term of D_T^2 is the first term D_{1T}. Indeed,

$$D_{1T} = \sum_{j=m_T}^T \sum_{t=j}^T \left(Z_p^\tau(x) Z_p(x_t)\phi_{t-j}\right)^2 = \sum_{t=m_T}^T [Z_p^\tau(x) Z_p(x_t)]^2 \sum_{j=m_T}^t \phi_{t-j}^2$$
$$= \sum_{t=m_T}^T [Z_p^\tau(x) Z_p(x_t)]^2 \sum_{j=0}^{t-m_T} \phi_j^2 = \sigma_e^2(1 + o(1)) \sum_{t=m_T}^T [Z_p^\tau(x) Z_p(x_t)]^2$$
$$= \sigma_e^2(1 + o(1)) B_T^2$$

where $\sum_{j=0}^{t-m_T} \phi_j^2 = \sigma_e^2(1 + o(1))$ uniformly in $t \geq m_T + m_T^q$ with any $0 < q < 1$.

Notice further that using the densities of $d_t^{-1} x_t$ in Lemma A.1,

$$E[B_T^2] = \sum_{t=m_T}^T E[Z_p^\tau(x) Z_p(x_t)]^2 = \sum_{t=m_T}^T \int [Z_p^\tau(x) Z_p(d_t y)]^2 f_t(y) dy$$
$$= \sum_{t=m_T}^T d_t^{-1} \int [Z_p^\tau(x) Z_p(y)]^2 f_t(d_t^{-1} y) dy = \phi(0) \sum_{t=m_T}^T d_t^{-1} \int [Z_p^\tau(x) Z_p(y)]^2 dy$$
$$+ \sum_{t=m_T}^T d_t^{-1} \int [Z_p^\tau(x) Z_p(y)]^2 [f_t(d_t^{-1} y) - f_t(0) + f_t(0) - \phi(0)] dy$$
$$= C\sqrt{T} Z_p^\tau(x) \int Z_p(y) Z_p^\tau(y) dy Z_p(x)(1 + o(1)) = O(1)\sqrt{T}\|Z_p(x)\|^2$$
$$= O(1)\sqrt{T} p,$$

where we have used the orthogonality of the basis to derive $\int Z_p(y) Z_p^\tau(y) dy = I_p$, and Lipschitz condition for $f_t(\cdot)$ and the uniform approximation of $\sup_x |f_t(x) - \phi(x)| \to 0$ in Lemma A.1.

For D_{2T}, using the densities of $d_{ts}^{-1}x_{ts}$ and $d_s^{-1}x_s$ in Lemma A.1, we have

$$
\begin{aligned}
E|D_{2T}| &= E\left|\sum_{t=m_T+1}^{T}\sum_{s=m_T}^{t-1} Z_p^\tau(x)Z_p(x_t)Z_p^\tau(x)Z_p(x_s)\sum_{j=0}^{s-1}\phi_{t-s+j}\phi_j\right| \\
\le &\sum_{t=m_T+1}^{T}\sum_{s=m_T}^{t-1} E|Z_p^\tau(x)Z_p(x_t)Z_p^\tau(x)Z_p(x_s)|\sum_{j=0}^{s-1}|\phi_{t-s+j}\phi_j| \\
\le &O(1)\sum_{t=m_T+1}^{T}\sum_{s=m_T}^{t-1} d_{ts}^{-1}d_s^{-1}\iint |Z_p^\tau(x)Z_p(y)Z_p^\tau(x)Z_p(z)|dydz\sum_{j=0}^{s-1}|\phi_{t-s+j}\phi_j| \\
\le &O(1)\sum_{t=m_T+1}^{T}\sum_{s=m_T}^{t-1}(t-s)^{-3/2}s^{-1/2}\left(\int |Z_p^\tau(x)Z_p(y)|dy\right)^2 \\
= &O(1)\sqrt{T}\left(\int |Z_p^\tau(x)Z_p(y)|dy\right)^2 = O(1)\sqrt{T},
\end{aligned}
$$

where we have used $\int |Z_p^\tau(x)Z_p(y)|dy \to \int \delta(x-y)dy = 1$ as $p\to\infty$ by Lemma A.4 and the convergence of $\sum_k k|\phi_k| < \infty$.

Thus, $D_T^2 = O_P\left(p\,T^{1/2}\right)$. In order to prove $\lim_{T\to\infty}\max_{m_T\le j\le T}|v_{jT}| = 0$ in probability, it therefore suffices to show that $p^{-1/2}T^{-1/4}\max_{m_T\le j\le T}\left|\sum_{t=j}^T Z_p^\tau(x)Z_p(x_t)\phi_{t-j}\right| = o_P(1)$. In fact,

$$
\begin{aligned}
&p^{-1/2}T^{-1/4}\max_{m_T\le j\le T}\left|\sum_{t=j}^{T} Z_p^\tau(x)Z_p(x_t)\phi_{t-j}\right| \le p^{-1/2}T^{-1/4}\max_{m_T\le j\le T}\sum_{t=j}^{T}\|Z_p(x)\|\|Z_p(x_t)\||\phi_{t-j}| \\
=&O(1)p^{1/2}T^{-1/4}\max_{m_T\le j\le T}\sum_{t=j}^{T}|\phi_{t-j}| \le O(1)p^{1/2}T^{-1/4}\sum_{j=0}^{\infty}|\phi_j| = o(1)
\end{aligned}
$$

almost surely by Assumption 3. Therefore, (B.5) holds.

In what follows, we will prove that $p^{-1/2}T^{-1/4}A_{1T}'' = o_P(1)$, $p^{-1/2}T^{-1/4}A_{2T} = o_P(1)$ and $p^{-1/2}T^{-1/4}A_{3T} = o_P(1)$, in view of $D_T = O_P\left(p^{1/2}T^{1/4}\right)$. Indeed, by Lemma A.2,

$$
\begin{aligned}
&p^{-1}T^{-1/2}E[A_{1T}'']^2 = T^{-1/2}p^{-1}E\left[\sum_{j=-\infty}^{m_T-1}\left(\sum_{t=m_T}^{T} Z_p^\tau(x)Z_p(x_t)\phi_{t-j}\right)\epsilon_j\right]^2 \\
=&T^{-1/2}p^{-1}E\sum_{j=-\infty}^{m_T-1}\left[\left(\sum_{t=m_T}^{T} Z_p^\tau(x)Z_p(x_t)\phi_{t-j}\right)\epsilon_j\right]^2 \\
&+2T^{-1/2}p^{-1}E\sum_{j=-\infty}^{m_T-1}\sum_{i=-\infty}^{j-1}\left(\sum_{t=m_T}^{T} Z_p^\tau(x)Z_p(x_t)\phi_{t-j}\right)\epsilon_j\left(\sum_{t=m_T}^{T} Z_p^\tau(x)Z_p(x_t)\phi_{t-i}\right)\epsilon_i \\
=&T^{-1/2}p^{-1}E\sum_{j=-\infty}^{m_T-1}\sum_{t=m_T}^{T}[Z_p^\tau(x)Z_p(x_t)]^2\phi_{t-j}^2\epsilon_j^2 \\
&+4T^{-1/2}p^{-1}E\sum_{j=-\infty}^{m_T-1}\sum_{t=m_T+1}^{T}\sum_{s=m_T}^{t-1} Z_p^\tau(x)Z_p(x_t)\phi_{t-j}Z_p^\tau(x)Z_p(x_s)\phi_{s-j}\epsilon_j^2 \\
&+2T^{-1/2}p^{-1}E\sum_{j=-\infty}^{m_T-1}\sum_{i=-\infty}^{j-1}\sum_{t=m_T}^{T}[Z_p^\tau(x)Z_p(x_t)]^2\phi_{t-j}\phi_{t-i}\epsilon_j\epsilon_i \\
&+4T^{-1/2}p^{-1}E\sum_{j=-\infty}^{m_T-1}\sum_{i=-\infty}^{j-1}\sum_{t=m_T+1}^{T}\sum_{s=m_T}^{t-1} Z_p^\tau(x)Z_p(x_t)\phi_{t-j}Z_p^\tau(x)Z_p(x_s)\phi_{s-i}\epsilon_j\epsilon_i
\end{aligned}
$$

$$
\begin{aligned}
\leq & T^{-1/2}p^{-1}\sum_{t=m_T}^{T}\frac{1}{\sqrt{t}}\int[Z_p^\tau(x)Z_p(y)]^2dy\sum_{j=-\infty}^{m_T-1}\phi_{t-j}^2\\
&+4T^{-1/2}p^{-1}\sum_{t=m_T+1}^{T}\sum_{s=m_T}^{t-1}\frac{1}{\sqrt{t-s}}\frac{1}{\sqrt{s}}\int|Z_p^\tau(x)Z_p(y)|dy\int|Z_p^\tau(x)Z_p(z)|dz\sum_{j=-\infty}^{m_T-1}|\phi_{s-j}\phi_{t-j}|\\
&+2T^{-1/2}p^{-1}\sum_{t=m_T}^{T}\frac{1}{t}\int[Z_p^\tau(x)Z_p(y)]^2dy\sum_{j=-\infty}^{m_T-1}\sum_{i=-\infty}^{j-1}|\phi_{t-j}\phi_{t-i}|\\
&+4T^{-1/2}p^{-1}\sum_{t=m_T+1}^{T}\sum_{s=m_T}^{t-1}\frac{1}{t-s}\frac{1}{\sqrt{s}}\int|Z_p^\tau(x)Z_p(y)|dy\int|Z_p^\tau(x)Z_p(z)|dz\sum_{j=-\infty}^{m_T-1}\sum_{i=-\infty}^{j-1}|\phi_{t-j}\phi_{s-i}|\\
\leq & O(1)T^{-1/2}p\sum_{t=m_T}^{T}\frac{1}{\sqrt{t}}\sum_{j=t}^{\infty}\phi_j^2+O(1)T^{-1/2}p^{11/6}\sum_{t=m_T+1}^{T}\sum_{s=m_T}^{t-1}\frac{1}{\sqrt{t-s}}\frac{1}{\sqrt{s}}\sum_{j=s}^{\infty}|\phi_j\phi_{t-s+j}|\\
&+O(1)T^{-1/2}p\sum_{t=1}^{T}\frac{1}{t}\sum_{j=-\infty}^{m_T}\sum_{i=-\infty}^{j-1}|\phi_{t-j}\phi_{t-i}|\\
&+O(1)T^{-1/2}p^{11/6}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\frac{1}{t-s}\frac{1}{\sqrt{s}}\sum_{j=m_T}^{\infty}\sum_{i=j+1}^{\infty}|\phi_{t+j}\phi_{s+i}|\\
&\leq O(1)T^{-1/2}p+O(1)T^{-1/2}p^{11/6}=o(1),
\end{aligned}
$$

where we have used $\sum_j j|\phi_j|<\infty$ and the assumption for p in Assumption 3.

Moreover, we obtain

$$
\begin{aligned}
&T^{-1/4}p^{-1/2}E|A_{2T}|=T^{-1/4}p^{-1/2}E\left|\sum_{t=m_T}^{T}Z_p^\tau(x)Z_p(x_t)\gamma_p(x_t)\right|\\
\leq & O(1)T^{-1/4}\sum_{t=m_T}^{T}E\|Z_p(x_t)\||\gamma_p(x_t)|\leq O(1)T^{-1/4}\sum_{t=m_T}^{T}\frac{1}{\sqrt{t}}\int\|Z_p(y)\||\gamma_p(y)|dy\\
\leq & O(1)T^{1/4}\left(\int\|Z_p(y)\|^2dy\int|\gamma_p(y)|^2dy\right)^{1/2}=o(1)T^{1/4}(pp^{-r})^{1/2}\\
= & o(1)T^{1/4-(r-1)\alpha/2}=o(1)
\end{aligned}
$$

again by Assumption 3.

Finally, for A_{3T}, note that $T^{1/4}p^{-1/2}|\gamma_p(x)|=o(1)T^{1/4}p^{-1/2}p^{-(r-1)/2-1/12}=o(1)T^{1/4-r\alpha/2-1/12}=o(1)$ due to Lemma A.3. The proof of (3.1) is completed.

Part Two. To show the limit in (3.2), it suffices to show the limit for $\|\widehat{f}(x)-f(x)\|_{L^2}^2$ due to the continuous mapping theorem. Notice that

$$
\begin{aligned}
\|\widehat{f}(x)-f(x)\|_{L^2}^2=&\int(\widehat{f}(x)-f(x))^2dx=\int[Z_p^\tau(x)(\widehat{\theta}-\theta)-\gamma_p(x)]^2dx\\
=&\int[Z_p^\tau(x)(\widehat{\theta}-\theta)]^2dx+\int[\gamma_p(x)]^2dx-2\left(\widehat{\theta}-\theta\right)^\tau\int Z_p(x)\gamma_p(x)dx\\
=&\|\widehat{\theta}-\theta\|^2+\|\gamma_p(x)\|_{L^2}^2,
\end{aligned}
$$

because of the orthogonality of the Hermite function sequence. It follows from Taylor expansion and Theorem 3.2 that

$$
\|\widehat{\theta}-\theta\|^2=\|(Z^\tau Z)^{-1}Z^\tau(e+\gamma)\|^2=(e+\gamma)^\tau Z(Z^\tau Z)^{-2}Z^\tau(e+\gamma)
$$

$$
\begin{aligned}
=&\frac{d_T^2}{T^2}(e+\gamma)^\tau Z\left(L_B(1,0)I_p+\frac{d_T}{T}Z^\tau Z-L_B(1,0)I_p\right)^{-2}Z^\tau(e+\gamma)\\
=&\frac{d_T^2}{T^2}(e+\gamma)^\tau Z\left(L_B^{-2}(1,0)I_p+O_p(1)\|\frac{d_T}{T}Z^\tau Z-L_B(1,0)I_p\|\right)Z^\tau(e+\gamma)\\
=&L_B^{-2}(1,0)\frac{d_T^2}{T^2}(e+\gamma)^\tau ZZ^\tau(e+\gamma)(1+o_P(1))\\
=&L_B^{-2}(1,0)\frac{d_T^2}{T^2}[e^\tau ZZ^\tau e+\gamma^\tau ZZ^\tau\gamma+2e^\tau ZZ^\tau\gamma].
\end{aligned}
$$

Rescaling gives

$$
\frac{T}{d_T\,p}\|\widehat{f}(x)-f(x)\|_{L^2}^2=L_B^{-2}(1,0)\frac{d_T}{T\,p}[e^\tau ZZ^\tau e+\gamma^\tau ZZ^\tau\gamma+2e^\tau ZZ^\tau\gamma]+\frac{T}{d_T\,p}\|\gamma_p(x)\|_{L^2}^2.
$$

Noticing $e^\tau ZZ^\tau e=\sum_{t=1}^T\|Z_p(x_t)\|^2e_t^2+2\sum_{t=2}^T\sum_{s=1}^{t-1}Z_p^\tau(x_t)Z_p(x_s)e_te_s$, using exactly the same way as L_{1n} proved in Theorem 3.1 of Dong and Gao (2014), we have

$$
\frac{d_T}{T\,p}\sum_{t=1}^T\|Z_p(x_t)\|^2e_t^2\to_D\sigma_e^2L_B(1,0)\quad\text{and}\quad\frac{d_T}{T\,p}\sum_{t=2}^T\sum_{s=1}^{t-1}Z_p^\tau(x_t)Z_p(x_s)e_te_s=o_P(1)
$$

as $T\to\infty$. To fulfill the proof of (3.2), by virtue of Cauchy-Schwarz inequality we only need to show that

$$
\frac{d_T}{T\,p}\gamma^\tau ZZ^\tau\gamma=o_P(1)\quad\text{and}\quad\frac{T}{d_T\,p}\|\gamma_p(x)\|_{L^2}^2=o(1).
$$

In fact, by (2) of Lemma A.3, $\frac{T}{d_T\,p}\|\gamma_p(x)\|^2=o(1)T^{1/2}p^{-1}p^{-r}=o(1)T^{1/2-(1+r)\alpha}=o(1)$ due to Assumption 3. Moreover, using the densities in Lemma A.1, we have

$$
\begin{aligned}
\frac{d_T}{T\,p}E[\gamma^\tau ZZ^\tau\gamma]=&\frac{d_T}{T\,p}\sum_{t=1}^TE[\|Z_p(x_t)\|^2\gamma_p^2(x_t)]+2\frac{d_T}{T\,p}\sum_{t=2}^T\sum_{s=1}^{t-1}E[Z_p^\tau(x_t)Z_p(x_s)\gamma_p(x_t)\gamma_p(x_s)]\\
\le&O(1)\frac{d_T}{T\,p}\sum_{t=1}^T\frac{1}{\sqrt{t}}\int\|Z_p(x)\|^2\gamma_p^2(x)dx\\
&+O(1)\frac{d_T}{T\,p}\sum_{t=2}^T\sum_{s=1}^{t-1}\frac{1}{\sqrt{t-s}}\frac{1}{\sqrt{s}}\iint|Z_p^\tau(x)Z_p(y)\gamma_p(x)\gamma_p(y)|dxdy\\
\le&O(1)\int\gamma_p^2(x)dx+O(1)T^{1/2}p^{-1}\left(\int\|Z_p(x)\||\gamma_p(x)|dx\right)^2\\
\le&o(1)+O(1)T^{1/2}p^{-1}\int\|Z_p(x)\|^2dx\int|\gamma_p(x)|^2dx\\
=&O(1)T^{1/2}p^{-r}=O(1)T^{1/2-r\alpha}=o(1)
\end{aligned}
$$

in view of Lemma A.3 and again Cauchy-Schwarz inequality. □

Proof of Corollary 3.1. We first show that $\widehat{\sigma}_e^2\to_P\sigma_e^2$ as $T\to\infty$. Note that

$$
\widehat{\sigma}_e^2=\frac{1}{T}\sum_{t=1}^T(e_t+f(x_t)-\widehat{f}(x_t))^2=\frac{1}{n}\sum_{t=1}^Te_t^2+\frac{1}{T}\sum_{t=1}^T(f(x_t)-\widehat{f}(x_t))^2+2\frac{1}{T}\sum_{t=1}^Te_t(f(x_t)-\widehat{f}(x_t)).
$$

To begin with, we shall show that $\frac{1}{T}\sum_{t=1}^Te_t^2\to_P\sigma_e^2$. Recall that $\sigma_e^2=Ee_t^2=\sum_{j=0}^\infty\phi_j^2$. By the independence of $\{\epsilon_i\}$, we obtain

$$
E\left(\frac{1}{T}\sum_{t=1}^Te_t^2-\sigma_e^2\right)^2=\frac{1}{T^2}\sum_{t=1}^TE(e_t^2-\sigma_e^2)^2+2\frac{1}{T^2}E\sum_{t=2}^T\sum_{s=1}^{t-1}(e_t^2-\sigma_e^2)(e_s^2-\sigma_e^2)
$$

$$
\begin{aligned}
=&\frac{1}{T}Var(e_1^2)+2\frac{1}{T^2}E\sum_{t=2}^{T}\sum_{s=1}^{t-1}(e_t^2-\sigma_e^2)(e_s^2-\sigma_e^2)\\
=&2\frac{1}{T^2}E\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=-\infty}^{s}\phi_{t-j}^2(\epsilon_j^2-1)\sum_{\ell=-\infty}^{s}\phi_{s-\ell}^2(\epsilon_\ell^2-1)\\
&+2\frac{1}{T^2}E\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=-\infty}^{s}\sum_{j_1=-\infty,\neq j}^{s}\phi_{t-j}\phi_{t-j_1}\epsilon_j\epsilon_{j_1}\sum_{\ell=-\infty}^{s}\sum_{\ell_1=-\infty,\neq\ell}^{s}\phi_{s-\ell}\phi_{s-\ell_1}\epsilon_\ell\epsilon_{\ell_1}\\
=&2\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=-\infty}^{s}\phi_{t-j}^2\phi_{s-j}^2E(\epsilon_j^2-1)^2+4\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=-\infty}^{s-1}\sum_{j_1=j+1}^{s}\phi_{t-j}\phi_{t-j_1}\phi_{s-j}\phi_{s-j_1}E\epsilon_j^2\epsilon_{j_1}^2\\
\le&O(1)\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=-\infty}^{s}\phi_{t-j}^2+O(1)\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\sum_{j=-\infty}^{s-1}\sum_{j_1=j+1}^{s}|\phi_{t-j}\phi_{t-j_1}|\\
=&O(1)\frac{1}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}(t-s)^{-2}=O(1)\frac{1}{T^2}T=o(1),
\end{aligned}
$$

by virtue of $\sum_j j|\phi_j|<\infty$.

Moreover, by Lemma 3.1 $\sup_x|f(x)-\widehat{f}(x)|=o_P(1)$, implying $|f(x_t)-\widehat{f}(x_t)|=o_P(1)$ uniformly in t. Thus, the second term is $o_P(1)$ and so is the third one.

The other two assertions are trivially valid in view of Theorem 3.1.

□

Appendix C: Proofs of Lemmas A.1–A.5

The related notation is rephrased. Without loss of generality, in what follows let $x_0=0$ almost surely. It follows that

$$x_t=\sum_{\ell=1}^{t}v_\ell=\sum_{\ell=1}^{t}\sum_{i=-\infty}^{\ell}\psi_{\ell-i}\epsilon_i=\sum_{i=-\infty}^{t}\left(\sum_{\ell=\max(1,i)}^{t}\psi_{\ell-i}\right)\epsilon_i=:\sum_{i=-\infty}^{t}b_{t,i}\epsilon_i.$$

Let $j\le t$ be fixed. Thus we have

$$x_t=b_{t,j}\epsilon_j+x_{t/j},\quad \text{with } x_{t/j}:=\sum_{i=-\infty,\neq j}^{t}b_{t,i}\epsilon_i,$$

where $x_{t/j}$ is the variable deducting the term containing ϵ_j in x_t. Obviously, $x_{t/j}$ and ϵ_j are mutually independent.

Additionally, letting $1\le s<j\le t$, x_t also has the following decomposition:

$$x_t=x_s^*+x_{ts}=x_s^*+b_{t,j}\epsilon_j+x_{ts/j},$$

where $x_s^*=x_s+\bar{x}_s$ with $\bar{x}_s=\sum_{i=s+1}^{t}\sum_{a=-\infty}^{s}\psi_{i-a}\epsilon_a$ containing all information available up to s and $x_{ts}=\sum_{i=s+1}^{t}b_{t,i}\epsilon_i$, while obviously $x_{ts/j}=\sum_{i=s+1,\neq j}^{t}b_{t,i}\epsilon_i$. Evidently, x_{ts} captures all information containing in x_t on the time periods $(s,t]$, while $x_{ts/j}$ captures all information containing in x_t on the time periods $(s,j)\cup(j,t]$. Let $d_{ts}:=(Ex_{ts}^2)^{1/2}$ for late use. Moreover, $\bar{x}_s=O_P(1)$ by virtue of Assumption 1.

Lemma A.1. Suppose that Assumption 1 holds. For t or $t-s$ is large,

(1) $d_t^{-1}x_t$ have uniformly bounded densities $f_t(x)$ over all t and x satisfying a uniform Lipschitz condition $\sup_x |f_t(x+y) - f_t(x)| \leq C|y|$ for any y and some constant $C > 0$. In addition, $\sup_x |f_t(x) - \phi(x)| \to 0$ as $t \to \infty$ where $\phi(x)$ is the standard normal density function.

Let $1 \leq s < t$. $d_{ts}^{-1}x_{ts}$ have uniformly bounded densities $f_{ts}(x)$ over all (t,s) and x satisfying the above uniform Lipschitz condition as well.

(2) Let $j \leq t$. $d_t^{-1}x_{t/j}$ have uniformly bounded densities $f_{t/j}(x)$ over all (t,j) and x satisfying uniform Lipschitz condition in the above form.

Let $1 \leq s < j \leq t$. $d_{ts}^{-1}x_{ts/j}$ have uniformly bounded densities $f_{ts/j}(x)$ over all (t,j) and (s,x) satisfying the above uniform Lipschitz condition as well.

Proof of Lemma A.1: We shall prove the assertion about $d_t^{-1}x_t$ only. All the other claims follow in the same fashion.

Denote by $\varphi(\lambda)$ the characteristic function of ϵ_0. Under Assumption 1, $\int |\lambda\varphi(\lambda)|d\lambda < \infty$. Let $\Phi_t(\alpha)$ be the characteristic function of $d_t^{-1}x_t$ for $\alpha \in \mathbb{R}$. Denote $x_t = x_t^+ + x_t^-$, where x_t^+ includes all ϵ_j with $j > 0$ in x_t, while x_t^- includes all ϵ_j with $j \leq 0$ in x_t. It follows that

$$\int |\alpha||\Phi_t(\alpha)|d\alpha = \int |\alpha||E\exp(i\alpha d_t^{-1}x_t)|d\alpha \leq \int |\alpha||E\exp(i\alpha d_t^{-1}x_t^+)|d\alpha$$
$$= \int |\alpha| \left| E\exp\left[i\left(\alpha d_t^{-1}\sum_{j=1}^{t} b_{t,j}\epsilon_j \right) \right] \right| d\alpha = \int |\alpha| \left| \prod_{j=1}^{t} E\exp\left[i\alpha d_t^{-1}b_{t,j}\epsilon_j \right] \right| d\alpha$$
$$= \int |\alpha| \prod_{j=1}^{t} \left| \varphi(\alpha d_t^{-1}b_{t,j}) \right| d\alpha.$$

It is clear that there exists a $\delta_0 > 0$ such that $|\varphi(\lambda)| < e^{-|\lambda|^2/4}$ whenever $|\lambda| \leq \delta_0$ and $|\varphi(\lambda)| < \eta$ if $|\lambda| > \delta_0$ for some $0 < \eta < 1$ (Wang and Phillips, 2009a, p. 730).

Note also that $b_{t,j} = \psi_0 + \cdots + \psi_{t-j}$. If $t-j$ is large, $b_{t,j} = \psi(1 + o(1))$ where $\psi = \sum_j \psi_j \neq 0$. Let $\nu = \nu_t$ be a function of t such that $\nu \to \infty$ and $\nu/t \to 0$ as $t \to \infty$. Thus, for $1 \leq j \leq t - \nu$, there exist constants c_1, c_2 such that $0 < c_1 < c_2 < \infty$ and $c_1 < |b_{t,j}| < c_2$. Indeed, we may take $c_1 = |\psi|/2$ and $c_2 = 3|\psi|/2$. Therefore, letting $\delta = \delta_0/c_2$,

$$\int |\alpha| \prod_{j=1}^{t} \left| \varphi(\alpha d_t^{-1}b_{t,j}) \right| d\alpha \leq \int |\alpha| \prod_{j=1}^{t-\nu} \left| \varphi(\alpha d_t^{-1}b_{t,j}) \right| d\alpha$$
$$= \left(\int_{|\alpha| \leq d_t\delta} + \int_{|\alpha| > d_t\delta} \right) |\alpha| \prod_{j=1}^{t-\nu} \left| \varphi(\alpha d_t^{-1}b_{t,j}) \right| d\alpha$$
$$\leq \int_{|\alpha| \leq d_t\delta} |\alpha| e^{-\alpha^2 d_t^{-2}\sum_{j=1}^{t-\nu} b_{t,j}^2/4} d\alpha + \eta^{t-\nu-1} \int_{|\alpha| > d_t\delta} |\alpha| \left| \varphi\left(\alpha d_t^{-1}b_{t,1} \right) \right| d\alpha$$
$$\leq \int_{|\alpha| \leq d_t\delta} |\alpha| e^{-\alpha^2 c_1(1-\nu/t)/4} d\alpha + b_{t,1}^{-2}d_t^2\eta^{t-\nu-1} \int_{|\alpha| > \delta} |\alpha| \left| \varphi(\alpha) \right| d\alpha$$
$$\leq \int |\alpha| e^{-\alpha^2 c_1/4} d\alpha + b_{t,1}^{-2}d_t^2\eta^{t-\nu-1} \int |\alpha| \left| \varphi(\alpha) \right| d\alpha < \infty,$$

where we have used the fact that $d_t^2\eta^{t-\nu-1} \to 0$ and $b_{t,1} \to \psi \neq 0$ as $t \to \infty$.

The integrability of $|\Phi_t(\alpha)|$ implies the uniform boundedness of the densities $f_t(x)$ due to the inverse formula. Similarly, the integrability of $|\alpha||\Phi_t(\alpha)|$ gives the uniform boundedness of the derivative of

$f_t(x)$. As a matter of fact, we have

$$\left|\frac{d}{dx}f_t(x)\right| = \frac{1}{2\pi}\left|\frac{d}{dx}\int e^{-i\alpha x}\Phi_t(\alpha)d\alpha\right| = \frac{1}{2\pi}\left|\int(-i\alpha)e^{-i\alpha x}\Phi_t(\alpha)d\alpha\right|$$
$$\leq \frac{1}{2\pi}\int|\alpha||\Phi_t(\alpha)|d\alpha \leq C.$$

It follows immediately from the mean value theorem that $\sup_x|f_t(x+y)-f_t(x)|\leq C|y|$.

The normality approximation can be found in literature, for example, equation (5.11) of Wang and Phillips (2009a, p. 729).

□

Lemma A.2. Suppose that Assumption 1 holds. Let j be a fixed integer and $j\leq t$. For any functions U and $g:\mathbb{R}\mapsto\mathbb{R}$ such that $\int|U(w)|dw<\infty$ and $E|\epsilon_j g(\epsilon_j)|<\infty$, and for large t or $t-s$, we have

(1) $E[U(x_t)g(\epsilon_j)] = E[U(x_{t/j})]E[g(\epsilon_j)] + c_U d_t^{-2}$ where c_U is such that $|c_U|\leq O(1)E|\epsilon_j g(\epsilon_j)|$ $\times\int|U(w)|dw$. In particular, if $Eg(\epsilon_j)=0$, then $E[U(x_t)g(\epsilon_j)] = c_U d_t^{-2}$;

(2) $E|U(x_t)g(\epsilon_j)|\leq O(1)d_t^{-1}E|g(\epsilon_j)|\int|U(w)|dw$;

(3) For any ℓ: $j\neq\ell\leq t$, $E[U(x_t)g(\epsilon_j)|\epsilon_\ell] = E[U(x_{t/j})|\epsilon_\ell]E[g(\epsilon_j)] + d_t^{-2}\eta_\ell$ where η_ℓ is a random variable depending on ϵ_ℓ such that $|\eta_\ell|\leq O(1)\int|U(w)|dw$ almost surely. If $E[g(\epsilon_j)]=0$, then $E[U(x_t)g(\epsilon_j)|\epsilon_\ell] = d_t^{-2}\eta_\ell$.

Meanwhile, $E[|U(x_t)g(\epsilon_j)||\epsilon_\ell]\leq O(1)d_t^{-1}E|g(\epsilon_j)|\int|U(w)|dw$ almost surely.

(4) For $1\leq s<j\leq t$, $E[U(x_t)g(\epsilon_j)|\mathcal{F}_s] = E[g(\epsilon_j)]E[U(x_{t/j})|\mathcal{F}_s] + d_{ts}^{-2}\xi_s$ where $|\xi_s|\leq O(1)E|\epsilon_j g(\epsilon_j)|$ $\int|U(x)|dx$ almost surely; meanwhile, $E[|U(x_t)g(\epsilon_j)||\mathcal{F}_s]\leq O(1)d_{ts}^{-1}E[|g(\epsilon_j)|]\int|U(w)|dw$ a.s.

Proof of Lemma A.2:

(1) Let $f_\epsilon(\cdot)$ be the density of ϵ_0. Recalling that $x_t = b_{t,j}\epsilon_j + x_{t/j}$ with $b_{t,j}=\sum_{i=1\vee j}^t\psi_{i-j}=O(1)$, $d_t^{-1}x_{t/j}$ has a uniformly bounded density $f_{t/j}(x)$ satisfying Lipschitz condition, we have

$$\begin{aligned}
&E[U(x_t)g(\epsilon_j)] = E[U(b_{t,j}\epsilon_j + x_{t/j})g(\epsilon_j)]\\
=&\iint U(b_{t,j}v + d_t w)g(v)f_\epsilon(v)f_{t/j}(w)dvdw\\
=&d_t^{-1}\iint U(w)g(v)f_\epsilon(v)f_{t/j}\left(\frac{w-b_{t,j}v}{d_t}\right)dvdw\\
=&d_t^{-1}\iint U(w)g(v)f_\epsilon(v)f_{t/j}\left(\frac{w}{d_t}\right)dvdw\\
&+d_t^{-1}\iint U(w)g(v)f_\epsilon(v)\left[f_{t/j}\left(\frac{w-b_{t,j}v}{d_t}\right)-f_{t/j}\left(\frac{w}{d_t}\right)\right]dvdw\\
=&d_t^{-1}\int g(v)f_\epsilon(v)dv\int U(w)f_{t/j}\left(\frac{w}{d_t}\right)dw + d_t^{-2}c_U\\
=&E[g(\epsilon_j)]\int U(d_t w)f_{t/j}(w)dw + d_t^{-2}c_U = E[U(x_{t/j})]E[g(\epsilon_j)] + d_t^{-2}c_U,
\end{aligned}$$

where $c_U := d_t\iint U(w)g(v)h_\epsilon(v)\left[f_{t/j}\left(\frac{w-b_{t,j}v}{d_t}\right)-f_{t/j}\left(\frac{w}{d_t}\right)\right]dvdw$ satisfies

$$|c_U|\leq d_t\int|g(v)|f_\epsilon(v)\int|U(w)|\left|f_{t/j}\left(\frac{w-b_{t,j}v}{d_t}\right)-f_{t/j}\left(\frac{w}{d_t}\right)\right|dwdv$$

$$\leq O(1)\int |g(v)|f_\epsilon(v)\int |U(w)||b_{t,j}v|dwdv = O(1)E|\epsilon_j g(\epsilon_j)|\int |U(w)|dw,$$

using Lipschitz condition for $f_{t/j}$ in Lemma A.1. Clearly, $E[U(x_t)f(\epsilon_j)] = d_t^{-2}c_U$, if $Ef(\epsilon_j) = 0$.

(2) It follows that

$$\begin{aligned}
E|U(x_t)g(\epsilon_j)| =& E|U(b_{t,j}\epsilon_j + x_{t/j})g(\epsilon_j)| \\
=& \iint |U(b_{t,j}v + d_t w)g(v)|f_\epsilon(v)f_{t/j}(w)dvdw \\
=& d_t^{-1}\iint |U(w)g(v)|f_\epsilon(v)f_{t/j}\left(\frac{w - b_{t,j}v}{d_t}\right)dvdw \\
\leq& O(1)d_t^{-1}\iint |U(w)g(v)|f_\epsilon(v)dvdw \\
=& O(1)d_t^{-1}\int |g(v)|f_\epsilon(v)dv\int |U(w)|dw \\
=& O(1)d_t^{-1}E|g(\epsilon_j)|\int |U(w)|dw.
\end{aligned}$$

(3) Because of similarity we only consider here $\ell > j > 0$. In this case, we have the following decomposition,

$$x_t = b_{t,j}\epsilon_j + b_{t,\ell}\epsilon_\ell + x_{t/j\ell},$$

where $x_{t/j\ell}$ includes all terms in x_t except those terms involving ϵ_ℓ and ϵ_j.

Moreover, $Ex_{t/j\ell}^2 = Ex_t^2 - b_{t,j}^2 - b_{t,\ell}^2 = O(1)t$ and, similar to Lemma A.1, we may show that $d_t^{-1}x_{t/j\ell}$ has density $f_{t/j\ell}(x)$ and $f_{t/j\ell}(x)$ satisfies Lipschtiz condition uniformly on $\mathbb{R}$. Recalling that ϵ_j has density $f_\epsilon(v)$,

$$\begin{aligned}
&E[U(x_t)g(\epsilon_j)|\epsilon_\ell] = E[U(b_{t,j}\epsilon_j + b_{t,\ell}\epsilon_\ell + x_{t/j\ell})g(\epsilon_j)|\epsilon_\ell] \\
=& \iint U(b_{t,j}v + b_{t,\ell}\epsilon_\ell + d_t w)g(v)f_\epsilon(v)f_{t/j\ell}(w)dvdw \\
=& d_t^{-1}\iint U(w)g(v)f_\epsilon(v)f_{t/j\ell}\left(\frac{w - b_{t,j}v - b_{t,\ell}\epsilon_\ell}{d_t}\right)dvdw \\
=& d_t^{-1}\iint U(w)g(v)f_\epsilon(v)f_{t/j\ell}\left(\frac{w - b_{t,\ell}\epsilon_\ell}{d_t}\right)dvdw \\
&+ d_t^{-1}\iint U(w)g(v)f_\epsilon(v)\left[f_{t/j\ell}\left(\frac{w - b_{t,j}v - b_{t,\ell}\epsilon_\ell}{d_t}\right) - f_{t/j\ell}\left(\frac{w - b_{t,\ell}\epsilon_\ell}{d_t}\right)\right]dvdw \\
=& E[U(x_{t/j\ell} + b_{t,\ell}\epsilon_\ell)|\epsilon_\ell]E[g(\epsilon_j)] + d_t^{-2}\eta_\ell = E[U(x_{t/j})|\epsilon_\ell]E[g(\epsilon_j)] + d_t^{-2}\eta_\ell
\end{aligned}$$

and using Lipschitz condition

$$|\eta_\ell| \leq O(1)\int |g(v)|f_\epsilon(v)\int |U(w)||b_{t,j}v|dwdv = O(1)\int |vg(v)|f_\epsilon(v)dv\int |U(w)|dw.$$

When $E[g(\epsilon_j)] = 0$ we shall have $E[U(x_t)g(\epsilon_j)|\epsilon_\ell] = d_t^{-2}\eta_\ell$. Additionally,

$$\begin{aligned}
&E[|U(x_t)g(\epsilon_j)||\epsilon_\ell] = \iint |U(b_{t,j}v + b_{t,\ell}\epsilon_\ell + d_t w)g(v)|f_\epsilon(v)f_{t/j\ell}(w)dvdw \\
=& d_t^{-1}\iint |U(w)g(v)f_\epsilon(v)|f_{t/j\ell}\left(\frac{w - b_{t,j}v - b_{t,\ell}\epsilon_\ell}{d_t}\right)dvdw \\
\leq& O(1)d_t^{-1}\iint |U(w)g(v)f_\epsilon(v)|dvdw = O(1)d_t^{-1}E|g(\epsilon_j)|\int |U(w)|dw,
\end{aligned}$$

almost surely.

(4) Recalling that $x_t = x_s^* + b_{t,j}\epsilon_j + x_{ts/j}$ and $d_{ts}^{-1}x_{ts/j}$ has a uniformly bounded density $f_{ts/j}(x)$ satisfying uniform Lipschitz condition,

$$\begin{aligned}
&E[U(x_t)g(\epsilon_j)|\mathcal{F}_s] = E[U(x_s^* + b_{t,j}\epsilon_j + x_{ts/j})g(\epsilon_j)|\mathcal{F}_s] \\
=&\iint U(x_s^* + b_{t,j}v + d_{ts}x)g(v)f_\epsilon(v)f_{ts/j}(x)dxdv \\
=&d_{ts}^{-1}\iint U(x)g(v)f_\epsilon(v)f_{ts/j}\left(\frac{x - b_{t,j}v - x_s^*}{d_{ts}}\right)dvdx \\
=&d_{ts}^{-1}\iint U(x)g(v)f_\epsilon(v)f_{ts/j}\left(\frac{x - x_s^*}{d_{ts}}\right)dvdx \\
&+ d_{ts}^{-1}\iint U(x)g(v)f_\epsilon(v)\left[f_{ts/j}\left(\frac{x - b_{t,j}v - x_s^*}{d_{ts}}\right) - f_{ts/j}\left(\frac{x - x_s^*}{d_{ts}}\right)\right]dvdx \\
=&d_{ts}^{-1}\int g(v)f_\epsilon(v)dv\int U(x)f_{ts/j}\left(\frac{x - x_s^*}{d_{ts}}\right)dvdx + d_{ts}^{-2}\xi_s \\
=&E[g(\epsilon_j)]E[U(x_{ts/j} + x_s^*)|\mathcal{F}_s] + d_{ts}^{-2}\xi_s,
\end{aligned}$$

where $\xi_s = d_{ts}\iint U(x)g(v)f_\epsilon(v)\left[f_{ts/j}\left(\frac{x-b_{t,j}v-x_s^*}{d_{ts}}\right) - f_{ts/j}\left(\frac{x-x_s^*}{d_{ts}}\right)\right]dvdx$, and using Lipschitz condition,

$$|\xi_s| \le C\iint |U(x)g(v)|f_\epsilon(v)|b_{t,j}v|dvdx = O(1)E|\epsilon_j g(\epsilon_j)|\int |U(x)|dx \quad a.s.$$

Consequently, when $E[g(\epsilon_j)] = 0$, $E[U(x_t)g(\epsilon_j)|\mathcal{F}_s] = d_{ts}^{-2}\xi_s$. Meanwhile,

$$\begin{aligned}
&E[|U(x_t)g(\epsilon_j)||\mathcal{F}_s] = \iint |U(x_s^* + b_{t,j}v + d_{ts}x)g(v)|f_\epsilon(v)f_{ts/j}(x)dxdv \\
=&d_{ts}^{-1}\iint |U(x)g(v)|f_\epsilon(v)f_{ts/j}\left(\frac{x - b_{t,j}v - x_s^*}{d_{ts}}\right)dvdx \\
\le&O(1)d_{ts}^{-1}\iint |U(x)g(v)|g_\epsilon(v)dxdv = O(1)d_{ts}^{-1}E|g(\epsilon_j)|\int |U(x)|dx.
\end{aligned}$$

□

Lemma A.3. (1) (i) $\|Z_p(x)\|^2 = O(1)p$; (ii) $\int \|Z_p(x)\|^2dx = p$; (iii) $\int \|Z_p(x)\|dx = O(1)p^{11/12}$; (iv) $\int x^2F_i^2(x)dx = (2i+1)/2$.

(2) Let Assumption 2 hold. Then, (i) $\sup_x |\gamma_p(x)| = o\left(p^{-(r-1)/2-1/12}\right)$; (ii) $\int \gamma_p^2(x)dx = o(1)p^{-r}$.

Proof of Lemma A.3:

(1) The assertions of (i) and (ii) follows trivially, since $F_i(x)$ are uniformly bounded and $\int F_i^2(x)dx = 1$. To prove (iii), from Christoffel-Darboux formula, $\|Z_p(x)\|^2 = pF_{p-1}^2(x) - \sqrt{(p-1)p}F_{p-2}(x)F_p(x)$, which implies

$$\|Z_p(x)\| \le \sqrt{p}|F_{p-1}(x)| + \sqrt[4]{(p-1)p}\sqrt{|F_{p-2}(x)F_p(x)|}.$$

Meanwhile, by Askey and Wainger (1965, p. 700) there exist two positive constants c_1 and c_2 such that $|F_i(x)| \le c_1(|N - x^2| + N^{1/3})^{-1/4}$ whenever $x^2 < N = 2i+1$, otherwise $|F_i(x)| < c_1\exp(-c_2x^2)$. Straightforward calculation yields $\int |F_i(x)|dx = O(1)i^{5/12}$ that, along with the above inequality, implies the assertion.

The assertion of (iv) holds because of the recursion relation for Hermite functions:

$$xF_0(x) = \frac{1}{\sqrt{2}}F_1(x), \quad xF_i(x) = \frac{1}{\sqrt{2}}(\sqrt{i}F_{i-1}(x) + \sqrt{i+1}F_{i+1}(x)),$$

and the orthogonality of the Hermite functions.

(2) We calculate the coefficient θ_i in the orthogonal expansion (2.3). Let $\phi(x) = \exp(-x^2)$ and $b_i^2 = \sqrt{\pi}2^i i!$. For i large, integration by parts gives

$$\begin{aligned}
\theta_i(g) = &\int g(x)F_i(x)dx = \frac{1}{b_i}\int g(x)H_i(x)e^{-x^2/2}dx \\
=&(-1)^i\frac{1}{b_i}\int g(x)\phi^{(i)}(x)e^{x^2/2}dx = (-1)^i\frac{1}{b_i}\int g(x)e^{x^2/2}d\phi^{(i-1)}(x) \\
=&(-1)^i\frac{1}{b_i}g(x)e^{x^2/2}\phi^{(i-1)}(x)|_{-\infty}^{\infty} - (-1)^i\frac{1}{b_i}\int \phi^{(i-1)}(x)[g(x)e^{x^2/2}]'dx \\
=&(-1)^{i-1}\frac{1}{b_i}\int \phi^{(i-1)}(x)[g(x)e^{x^2/2}]'dx \\
=&\frac{1}{b_i}\int [g(x)e^{x^2/2}]'H_{i-1}(x)\phi(x)dx = \frac{b_{i-1}}{b_i}\int [g(x)e^{x^2/2}]'e^{-x^2/2}F_{i-1}(x)dx \\
=&\frac{1}{\sqrt{2i}}\theta_{i-1}(\widetilde{g}_1).
\end{aligned}$$

where we define $\widetilde{g}_m = [g\phi^{-1/2}]^{(m)}\phi^{1/2}$ for positive integer m for notational convenience.

We obtain by repeatedly use of the above derivation that

$$\begin{aligned}
|\gamma_p(x)| = &\left|\sum_{i=p}^{\infty}\theta_i F_i(x)\right| = O(1)\left|\sum_{i=p}^{\infty} i^{-r/2}\theta_{i-r}(\widetilde{g}_r)F_i(x)\right| \leq o(1)\left(\sum_{i=p}^{\infty} i^{-r}F_i^2(x)\right)^{1/2} \\
=&o(1)\left(\sum_{i=p}^{\infty} i^{-r-1/6}\right)^{1/2} = o(1)p^{-(r-1)/2-1/12},
\end{aligned}$$

hence, $|\gamma_p(x)| = o(1)p^{-(r-1)/2-1/12}$ uniformly in x.

In addition, by the orthogonality,

$$\int \gamma_p^2(x)dx = \sum_{i=p}^{\infty}\theta_i^2 = O(1)\sum_{i=p}^{\infty} i^{-r}\theta_{i-r}^2(\widetilde{g}_r) \leq O(1)p^{-r}\sum_{i=p}^{\infty}\theta_{i-r}^2(\widetilde{g}_r) = o(1)p^{-r}.$$

□

Lemma A.4. $Z_p^\tau(x)Z_p(y) \to \delta(x-y)$ as $p \to \infty$ where $\delta(u)$ is the Dirac delta function.

Proof of Lemma A.4: For any smooth $f(x) \in L^2(\mathbb{R})$, we have

$$\begin{aligned}
&\lim_{p\to\infty}\int f(x)[Z_p^\tau(x)Z_p(y)]dx = \lim_{p\to\infty}\int f(x)\sum_{i=0}^{p-1}F_i(x)F_i(y)dx \\
=&\lim_{p\to\infty}\sum_{i=0}^{p-1}\int f(x)F_i(x)dx\ F_i(y) = \lim_{p\to\infty}\sum_{i=0}^{p-1}\theta_i F_i(y) = f(y)
\end{aligned}$$

due to smoothness of $f(x)$. Hence, $Z_p^\tau(x)Z_p(y)$ is a delta-convergent sequence defined in Kanwal (1983, p. 14). Then, the assertion holds by the definition of the delta-convergent sequence. □

Lemma A.5. Let m_T be a sequence such that $m_T \to \infty$ and $m_T/T \to 0$ as $T \to \infty$. Let $\{a_j\}$ be any sequence of nonnegative real numbers satisfying $\sum_{i=m_T}^{T} a_i = 1$.

(1) For any $s \geq 1$ and $t \geq m_T$, e_s and $d_t^{-1}x_t$ are asymptotically independent. Consequently, for any given k and $t \geq m_T$, ϵ_k and $d_t^{-1}x_t$ are asymptotically independent.

(2) For any $s \geq 1$, e_s and $\sum_{j=m_T}^{T} a_j d_j^{-1} x_j$ are asymptotically independent.

Proof of Lemma A.5:

(1) Since e_s is a stationary process, its density and characteristic function are independent of s.

Denote by $\rho(u)$ the density of e_s and by $\kappa_t(x,u)$ the joint density of $(d_t^{-1}x_t, e_s)$. Let $\Psi_t(\alpha,\lambda)$, $\Phi_t(\alpha)$ and $\Gamma(\lambda)$ be the characteristic functions of $(d_t^{-1}x_t, e_s)$, $d_t^{-1}x_t$ and e_s, respectively.

Recall that $x_t = \sum_{j=-\infty}^{t} b_{t,j}\epsilon_j$. Here, $b_{t,j} = \psi_0 + \cdots + \psi_{t-j}$ for $j \geq 1$ and $b_{t,j} = \psi_{1-j} + \cdots + \psi_{t-j}$ for $j \leq 0$. Thus, the sequence $\{|b_{t,j}|\}_j$ is uniformly bounded by the summability of $b_0 := \sum_{j=0}^{\infty} |\psi_j| < \infty$.

Let ν_t be a positive sequence chosen such that $\nu_t \to \infty$ and $\frac{\nu_t}{\sqrt{t}} \to 0$ as $t \to \infty$. Note that $e_s = \sum_{j=-\infty}^{s} \phi_{s-j}\epsilon_j$ and $\sum_{j=0}^{\infty} |\phi_j| < \infty$. As a result, $\phi_{\nu_t} = o(1)$ for large t and without loss of generality assume that $|\phi_j| \leq |\phi_{\nu_t}|$ for all $j \geq \nu_t$ since $\phi_j \to 0$ as $j \to \infty$.

For any given $\epsilon > 0$, denote $R_\epsilon \equiv \{(\alpha,\lambda) : \; d_t^{-1} b_0 |\alpha| + |\phi_{\nu_t}\lambda| < \epsilon\}$ a region in $\mathbb{R}^2$ and R'_ϵ its complement.

From the inverse formula we have

$$\begin{aligned}
&\kappa_t(x,u) - f_t(x)\rho(u) \\
=&\frac{1}{(2\pi)^2}\iint e^{-i(\alpha x + \lambda u)}[\Psi_t(\alpha,\lambda) - \Phi_t(\alpha)\Gamma(\lambda)]d\alpha d\lambda \\
=&\frac{1}{(2\pi)^2}\iint_{R_\epsilon} e^{-i(\alpha x + \lambda u)}[\Psi_t(\alpha,\lambda) - \Phi_t(\alpha)\Gamma(\lambda)]d\alpha d\lambda \\
&+ \frac{1}{(2\pi)^2}\iint_{R'_\epsilon} e^{-i(\alpha x + \lambda u)}[\Psi_t(\alpha,\lambda) - \Phi_t(\alpha)\Gamma(\lambda)]d\alpha d\lambda := T_1 + T_2,
\end{aligned}$$

where $f_t(x)$ is the density of $d_t^{-1}x_t$.

It has been shown that $\int |\alpha||\Phi_t(\alpha)|d\alpha < \infty$ in Lemma A.1 and similarly we may show that $\int \|(\alpha,\lambda)\||\Psi_t(\alpha,\lambda)|d\alpha d\lambda < \infty$ and $\int |\lambda||\Gamma(\lambda)|d\lambda < \infty$. By virtue of these, it is easily seen that $T_2 = o(1)$. In fact,

$$\begin{aligned}
|T_2| \leq&\frac{1}{(2\pi)^2}\iint_{R'_\epsilon} |\Psi_t(\alpha,\lambda) - \Phi_t(\alpha)\Gamma(\lambda)|d\alpha d\lambda \\
\leq&\frac{1}{(2\pi)^2\epsilon}\iint_{R'_\epsilon} (d_t^{-1}|\alpha| + |\phi_{\nu_t}\lambda|)|\Psi_t(\alpha,\lambda) - \Phi_t(\alpha)\Gamma(\lambda)|d\alpha d\lambda \\
\leq&C_1 d_t^{-1} + C_2|\phi_{\nu_t}| = o(1).
\end{aligned}$$

We then deal with T_1 in the sequel. For the sake of exposition, suppose that $t \geq s$ and it can be seen that the case $t < s$ is similar and easier. To see this, note that for $t < s$, all ϵ_j, $j = t+1, \cdots, s$ are not included in x_t so that $\{\epsilon_j : j = t+1, \cdots, s\}$ are independent of x_t, while when $t \geq s$, all the information in e_s is contained in x_t.

Observe that

$$\Psi_t(\alpha,\lambda) = E\exp[i(\alpha d_t^{-1}x_t + \lambda e_s)]$$

$$
\begin{aligned}
=&E\exp\left[i\sum_{j=s+1}^{t}\alpha d_t^{-1}b_{t,j}\epsilon_j+i\sum_{j=-\infty}^{s}(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j})\epsilon_j\right]\\
=&\prod_{j=s+1}^{t}\varphi(\alpha d_t^{-1}b_{t,j})\prod_{j=-\infty}^{s}\varphi(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j}),
\end{aligned}
$$

where $\varphi(\cdot)$ is the characteristic function of ϵ_1. Meanwhile,

$$
\Phi_t(\alpha)=E\exp(i\alpha d_t^{-1}x_t)=E\exp\left[i\sum_{j=-\infty}^{t}\alpha d_t^{-1}b_{t,j}\epsilon_j\right]=\prod_{j=-\infty}^{t}\varphi(\alpha d_t^{-1}b_{t,j}),
$$

and

$$
\Gamma(\lambda)=E\exp(i\lambda e_s)=E\exp\left[i\sum_{j=-\infty}^{s}\lambda\phi_{s-j}\epsilon_j\right]=\prod_{j=-\infty}^{s}\varphi(\lambda\phi_{s-j}).
$$

Hence,

$$
\begin{aligned}
\Psi_t(\alpha,\lambda)-\Phi_t(\alpha)\Gamma(\lambda)=\Phi_t(\alpha)\Gamma(\lambda)\left[\frac{\Psi_t(\alpha,\lambda)}{\Phi_t(\alpha)\Gamma(\lambda)}-1\right]\\
=\Phi_t(\alpha)\Gamma(\lambda)\left[\prod_{j=-\infty}^{s}\frac{\varphi(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j})}{\varphi(\alpha d_t^{-1}b_{t,j})\varphi(\lambda\phi_{s-j})}-1\right].
\end{aligned}
$$

We now consider

$$
\begin{aligned}
&\prod_{j=-\infty}^{s}\frac{\varphi(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j})}{\varphi(\alpha d_t^{-1}b_{t,j})\varphi(\lambda\phi_{s-j})}-1\\
=&\prod_{j=-\infty}^{s-\nu_t}\frac{\varphi(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j})}{\varphi(\alpha d_t^{-1}b_{t,j})\varphi(\lambda\phi_{s-j})}\times\prod_{j=s-\nu_t+1}^{s}\frac{1}{\varphi(\alpha d_t^{-1}b_{t,j})}\times\prod_{j=s-\nu_t+1}^{s}\frac{\varphi(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j})}{\varphi(\lambda\phi_{s-j})}-1\\
:=&A_1\times A_2\times A_3-1,\quad \text{say.}
\end{aligned}
$$

We shall show $A_1-1=o(1)$, $A_2-1=o(1)$ and $A_3-1=o(1)$, which imply that $A_1A_2A_3-1=o(1)$.

By the definition of R_ϵ, for $j\le s-\nu_t$, we have $|\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j}|<\epsilon$, $|\alpha d_t^{-1}b_{t,j}|<\epsilon$ and $|\lambda\phi_{s-j}|<\epsilon$ simultaneously on R_ϵ. As a result, all characteristic functions in A_1 can be expanded at zero by Taylor expansion. That is, for $j\le s-\nu_t$,

$$
\begin{aligned}
\varphi(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j})=&1-\frac{1}{2}(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j})^2(1+o(1)),\\
\varphi(\alpha d_t^{-1}b_{t,j})=&1-\frac{1}{2}(\alpha d_t^{-1}b_{t,j})^2(1+o(1)),\\
\varphi(\lambda\phi_{s-j})=&1-\frac{1}{2}(\lambda\phi_{s-j})^2(1+o(1)).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
A_1-1=&\prod_{j=-\infty}^{s-\nu_t}\frac{\varphi(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j})}{\varphi(\alpha d_t^{-1}b_{t,j})\varphi(\lambda\phi_{s-j})}-1\\
=&\frac{\prod_{j=-\infty}^{s-\nu_t}[1-\frac{1}{2}(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j})^2]}{\prod_{j=-\infty}^{s-\nu_t}[1-\frac{1}{2}(\alpha d_t^{-1}b_{t,j})^2]\prod_{j=-\infty}^{s-\nu_t}[1-\frac{1}{2}(\lambda\phi_{s-j})^2]}-1\\
=&-\frac{1}{2}\sum_{j=-\infty}^{s-\nu_t}(\alpha d_t^{-1}b_{t,j}+\lambda\phi_{s-j})^2+\frac{1}{2}\sum_{j=-\infty}^{s-\nu_t}(\alpha d_t^{-1}b_{t,j})^2+\frac{1}{2}\sum_{j=-\infty}^{s-\nu_t}(\lambda\phi_{s-j})^2
\end{aligned}
$$

$$= -\alpha\lambda \sum_{j=-\infty}^{s-\nu_t} d_t^{-1} b_{t,j}\phi_{s-j},$$

implying that $|A_1 - 1| \le C|\alpha\lambda| d_t^{-1}$ due to $|b_{t,j}| \le b_0$ and $\sum_{j=0}^{\infty} |\phi_j| < \infty$.

Meanwhile, we have

$$\begin{aligned}
A_2 - 1 &= \prod_{j=s-\nu_t+1}^{s} \frac{1}{\varphi(\alpha d_t^{-1} b_{t,j})} - 1 = \sum_{j=s-\nu_t+1}^{s} \frac{1}{2}(\alpha d_t^{-1} b_{t,j})^2 = \frac{1}{2}\alpha^2 d_t^{-2} \sum_{j=s-\nu_t+1}^{s} b_{t,j}^2 \\
&\le \frac{1}{2} b_0^2 \alpha^2 d_t^{-2} \nu_t \le C\epsilon |\alpha| d_t^{-1} \nu_t,
\end{aligned}$$

due to $b_{t,j}^2 < b_0^2$ and $b_0|\alpha| d_t^{-1} < \epsilon$ on R_ϵ.

Moreover, since $|\alpha d_t^{-1} b_{t,j}| < \epsilon$ on R_ϵ, for $s - \nu_t + 1 \le j \le s$ we may expand $\varphi(\alpha d_t^{-1} b_{t,j} + \lambda\phi_{s-j})$ at the point $\lambda\phi_{s-j}$, giving that $\varphi(\alpha d_t^{-1} b_{t,j} + \lambda\phi_{s-j}) = \varphi(\lambda\phi_{s-j}) + \varphi'(\lambda\phi_{s-j})\alpha d_t^{-1} b_{t,j}(1 + o(1))$. It follows that

$$\begin{aligned}
A_3 - 1 &= \prod_{j=s-\nu_t+1}^{s} \frac{\varphi(\alpha d_t^{-1} b_{t,j} + \lambda\phi_{s-j})}{\varphi(\lambda\phi_{s-j})} - 1 \\
&= \prod_{j=s-\nu_t+1}^{s} \left[1 + \frac{\varphi'(\lambda\phi_{s-j})}{\varphi(\lambda\phi_{s-j})}\alpha d_t^{-1} b_{t,j}\right] - 1 \\
&= \sum_{j=s-\nu_t+1}^{s} \frac{\varphi'(\lambda\phi_{s-j})}{\varphi(\lambda\phi_{s-j})}\alpha d_t^{-1} b_{t,j} = \sum_{j=s-\nu_t+1}^{s} h(\lambda\phi_{s-j})\alpha d_t^{-1} b_{t,j},
\end{aligned}$$

where we omit the higher order terms and $h(u) = \varphi'(u)/\varphi(u)$ defined in Assumption 1(d).

Invoking the condition on $h(u)$ in Assumption 1 gives

$$|A_3 - 1| \le \sum_{j=s-\nu_t+1}^{s} |h(\lambda\phi_{s-j})\alpha d_t^{-1} b_{t,j}| \le |\alpha| d_t^{-1} b_0 \sum_{j=0}^{\nu_t} |h(\lambda\phi_j)| \le C\, d_t^{-1}\, \nu_t\, |\alpha|\, k(\lambda),$$

where we have used Assumption 1(d) to deduce $\max_{j\ge 0} |h(\lambda\phi_j)| \le k(\lambda)$, and $0 < C < \infty$ is a constant.

Finally, it follows that

$$\begin{aligned}
|T_1| \le& \frac{1}{(2\pi)^2} \iint_{R_\epsilon} |\Psi_t(\alpha,\lambda) - \Phi_t(\alpha)\Gamma(\lambda)| d\alpha d\lambda \\
=& \frac{1}{(2\pi)^2} \iint_{R_\epsilon} |\Phi_t(\alpha)\Gamma(\lambda)| \left|\frac{\Psi_t(\alpha,\lambda)}{\Phi_t(\alpha)\Gamma(\lambda)} - 1\right| d\alpha d\lambda \\
=& \frac{1}{(2\pi)^2} \iint_{R_\epsilon} |\Phi_t(\alpha)\Gamma(\lambda)|\, |A_1A_2A_3 - 1|\, d\alpha d\lambda \\
=& C \iint_{R_\epsilon} |\Phi_t(\alpha)\Gamma(\lambda)|(|A_1 - 1| + |A_2 - 1| + |A_3 - 1|) d\alpha d\lambda\, (1 + o(1)) \\
\le& C_1 d_t^{-1} \iint_{R_\epsilon} |\alpha\lambda\, \Phi_t(\alpha)\Gamma(\lambda)| d\alpha d\lambda + C_2 \epsilon d_t^{-1} \nu_t \iint_{R_\epsilon} |\alpha||\Phi_t(\alpha)\Gamma(\lambda)| d\alpha d\lambda \\
&+ C_3 d_t^{-1} \iint_{R_\epsilon} |\alpha v(\lambda)\Phi_t(\alpha)\Gamma(\lambda)| d\alpha d\lambda \\
\le& C_1 d_t^{-1} \iint |\alpha\lambda\Phi_t(\alpha)\Gamma(\lambda)| d\alpha d\lambda + C_2 d_t^{-1} \nu_t \epsilon \iint |\alpha||\Phi_t(\alpha)\Gamma(\lambda)| d\alpha d\lambda \\
&+ C_3 d_t^{-1} \nu_t \iint |\alpha\, k(\lambda)\Phi_t(\alpha)\Gamma(\lambda)| d\alpha d\lambda = o(1),
\end{aligned}$$

where we have used Assumption 1(d) again. This shows that $|\kappa_t(x,u) - f_t(x)\rho(u)| \to 0$ as $t \to \infty$.

(2) Recalling that $x_i = \sum_{j=-\infty}^{i} b_{i,j}\epsilon_j$, for any nonnegative real numbers $\{a_i\}$ satisfying $\sum_{i=m_T}^{T} a_i = 1$, we define

$$
\begin{aligned}
\xi_T = \sum_{i=m_T}^{T} a_i d_i^{-1} x_i &= \sum_{i=m_T}^{T} a_i d_i^{-1} \sum_{j=-\infty}^{i} b_{i,j}\epsilon_j \\
&= \sum_{j=-\infty}^{T} \left(\sum_{i=\max(j,m_T)}^{T} a_i d_i^{-1} b_{i,j} \right) \epsilon_j.
\end{aligned}
$$

For a better exposition, let $B_{T,j} := \sum_{i=\max(j,m_T)}^{T} a_i d_i^{-1} b_{i,j}$ in what follows.

In order to prove part (2) of Lemma A.5, it suffices to show that e_s and ξ_T are asymptotically independent. Notice that $B_{T,j}$ perform similarly as $d_t^{-1} b_{t,j}$ in x_t. First, $b_{i,j}$ are uniformly bounded across both i and j, i.e. $|b_{i,j}| < b_0 = \sum_k |\psi_k|$ hold for all i and j. Second, for given a_i, $|B_{T,j}| \le c_0 d_{m_T}^{-1}$ for some constant c_0. Therefore, ξ_T can be treated as $d_T^{-1} x_T$ and hence following the same schedule as in (1), we can show the asymptotic independence of ξ_T and e_s for any given $s \ge 1$.

□

Proof of Theorem 3.2

Note that $Z^\tau Z$ has element $\sum_{t=1}^{T} F_{i-1}(x_t) F_{j-1}(x_t)$ at the place of (i,j). For the sake of convenience, denote that $F_{ij}(\cdot) = F_i(\cdot) F_j(\cdot)$ for $i \ne j$. It follows that

$$
\begin{aligned}
\left\| \frac{d_T}{T} Z^\tau Z - L_B(1,0) I_p \right\|^2 =& \sum_{i=0}^{p-1} \sum_{j=0,\ne i}^{p-1} \left(\frac{d_T}{T} \sum_{t=1}^{T} F_{ij}(x_t) \right)^2 + \sum_{i=0}^{p-1} \left(\frac{d_T}{T} \sum_{t=1}^{T} F_i^2(x_t) - L_B(1,0) \right)^2 \\
:=& A_{1T} + A_{2T}, \quad \text{say.}
\end{aligned}
$$

Firstly, we shall show that $A_{1T} = o_P(1)$ by proving $E[A_{1T}] \to 0$ as $T \to \infty$. Noting that

$$
\begin{aligned}
A_{1T} =& \sum_{i=0}^{p-1} \sum_{j=0,\ne i}^{p-1} \frac{d_T^2}{T^2} \sum_{t=1}^{T} F_{ij}^2(x_t) + 2 \sum_{i=0}^{p} \sum_{j=0,\ne i}^{p-1} \frac{d_T^2}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{t-1} [F_{ij}(x_t) F_{ij}(x_s)] \\
:=& A_{a1T} + A_{b1T},
\end{aligned}
$$

it suffices to show $E[A_{a1T}] \to 0$ and $E[A_{b1T}] \to 0$ as $T \to \infty$.

Using the uniformly boundedness of the density $f_t(x)$ for $d_t^{-1} x_t$ from Lemma A.1,

$$
\begin{aligned}
E[A_{a1T}] =& \sum_{i=0}^{p-1} \sum_{j=0,\ne i}^{p-1} \frac{d_T^2}{T^2} \sum_{t=1}^{n} E[F_{ij}^2(x_t)] = \sum_{i=0}^{p-1} \sum_{j=0,\ne i}^{p-1} \frac{d_T^2}{T^2} \sum_{t=1}^{T} d_t^{-1} \int F_{ij}^2(x) f_t(d_t^{-1} x) dx \\
\le& C \frac{d_T^2}{T^2} \sum_{t=1}^{T} t^{-1/2} \sum_{i=0}^{p-1} \sum_{j=0,\ne i}^{p-1} \int F_i^2(x) F_j^2(x) dx = C(p^2 - p) \frac{1}{\sqrt{T}} = O(1) p^2 T^{-1/2} = o(1),
\end{aligned}
$$

by Assumption 3, where we have used the fact that $\int F_i^2(x) = 1$ and the uniform boundedness of the sequence $F_i^2(x)$ as well as $f_t(x)$.

Note that, for $t > s$, $x_t = x_{ts} + x_s^*$. Recall Lemma A.1 that $d_{ts}^{-1} x_{ts}$ have densities $f_{ts}(x)$ that are uniformly bounded over all t, s and $x \in \mathbb{R}$, and that satisfy Lipschitz condition. Thus, by the orthogonality $\int F_{ij}(x) dx = 0$, we have

$$
E[A_{b1T}] = \sum_{i=0}^{p-1} \sum_{j=0,\ne i}^{p-1} \frac{d_T^2}{T^2} \sum_{t=2}^{T} \sum_{s=1}^{t-1} E[F_{ij}(x_t) F_{ij}(x_s)]
$$

$$=\sum_{i=0}^{p-1}\sum_{j=0,\neq i}^{p-1}\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}E[F_{ij}(x_{ts}+x_s^*)F_{ij}(x_s)]$$
$$=\sum_{i=0}^{p-1}\sum_{j=0,\neq i}^{p-1}\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}d_{ts}^{-1}E\left[\int F_{ij}(x)f_{ts}\left(\frac{x-x_s^*}{d_{ts}}\right)dxF_{ij}(x_s)\right]$$
$$=\sum_{i=0}^{p-1}\sum_{j=0,\neq i}^{p-1}\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}d_{ts}^{-1}E\left\{\int F_{ij}(x)\left[f_{ts}\left(\frac{x-x_s^*}{d_{ts}}\right)-f_{ts}\left(\frac{-x_s^*}{d_{ts}}\right)\right]dxF_{ij}(x_s)\right\}$$

which gives

$$\begin{aligned}|E[A_{b1T}]| \leq& C\sum_{i=0}^{p-1}\sum_{j=0,\neq i}^{p-1}\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}d_{ts}^{-2}\int|xF_{ij}(x)|dxE|F_{ij}(x_s)|\\ \leq& C\sum_{i=0}^{p-1}\sum_{j=0,\neq i}^{p-1}\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}d_{ts}^{-2}d_s^{-1}\int|xF_{ij}(x)|dx\int|F_{ij}(x)|dx\\ \leq& C\sum_{i=2}^{p-1}\sum_{j=0}^{i-1}\sqrt{2j+1}\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}(t-s)^{-1}s^{-1/2}\\ =& O(1)p^{5/2}\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}(t-s)^{-1}s^{-1/2}=O(1)p^{5/2}T^{-1/2}\ln(T)=o(1),\end{aligned}$$

where we have used the facts that $\int|xF_{ij}(x)|dx=\int|xF_i(x)F_j(x)|dx\leq\left(\int x^2F_j^2(x)dx\right)^{1/2}=(j+1/2)^{1/2}$ by a recursive relation for Hermite functions and $\int|F_{ij}(x)|dx\leq 1$ due to Cauchy-Schwarz inequality.

Secondly, to tackle A_{2T}, define for any $i: 0\leq i\leq p-1$ and $\epsilon>0$, $\phi_\epsilon(x)=\frac{1}{\sqrt{2\pi\epsilon}}e^{-x^2/(2\epsilon^2)}$, $\phi(z)=\phi_1(z)$, and $U_T(i,\epsilon)=\frac{d_T}{T}\sum_{t=1}^T\int F_i^2(x_t+d_Tx\epsilon)\phi(x)dx$.

Observe that

$$\begin{aligned}A_{2T}=&\sum_{i=0}^{p-1}\left(\frac{d_T}{T}\sum_{t=1}^{T}F_i^2(x_t)-L_B(1,0)\right)^2\\ \leq& 4\sum_{i=0}^{p-1}\left(\frac{d_T}{T}\sum_{t=1}^{T}F_i^2(x_t)-U_T(i,\epsilon)\right)^2+4\sum_{i=0}^{p-1}\left(U_T(i,\epsilon)-\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(d_T^{-1}x_t)\right)^2\\ &+4p\left(\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(d_T^{-1}x_t)-\int_0^1\phi_\epsilon(B(r))dr\right)^2+4p\left(\int_0^1\phi_\epsilon(B(r))dr-L_B(1,0)\right)^2\\ :=& 4A_{a2T}+4A_{b2T}+4A_{c2T}+4A_{d2T}.\end{aligned}$$

Hence, it is sufficient to show that $E[A_{\ell 2T}]=o(1)$ for $\ell=a,b,c$ and d.

For the first term A_{a2T}, notice that

$$\left(\frac{d_T}{T}\sum_{t=1}^{T}F_i^2(x_t)-U_T(i,\epsilon)\right)^2=\left(\frac{d_T}{T}\sum_{t=1}^{T}\int[F_i^2(x_t)-F_i^2(x_t+d_Tx\epsilon)]\phi(x)dx\right)^2$$
$$=\frac{d_T^2}{T^2}\left(\int\sum_{t=1}^{T}[F_i^2(x_t)-F_i^2(x_t+d_Tx\epsilon)]\phi(x)dx\right)^2\leq\frac{d_T^2}{T^2}\int\left(\sum_{t=1}^{T}[F_i^2(x_t)-F_i^2(x_t+d_Tx\epsilon)]\right)^2\phi(x)dx,$$

by Cauchy-Schwarz inequality.

Recalling $x_t = x_{ts} + x_s^*$ and $d_{ts}^{-1} x_{ts}$ has density $f_{ts}(\cdot)$ which satisfies Lipschitz condition, we then have

$$\begin{aligned}
&\frac{d_T^2}{T^2} E\left(\sum_{t=1}^{T}[F_i^2(x_t) - F_i^2(x_t + d_T x\epsilon)]\right)^2 \\
\leq & \frac{d_T^2}{T^2}\sum_{t=1}^{T} E[F_i^2(x_t) - F_i^2(x_t + d_T x\epsilon)]^2 \\
&+ 2\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\left|E\{[F_i^2(x_t) - F_i^2(x_t + d_T x\epsilon)]\cdot[F_i^2(x_s) - F_i^2(x_s + d_T x\epsilon)]\}\right| \\
\leq & C\frac{d_T^2}{T^2}\sum_{t=1}^{T} d_t^{-1}\int[F_i^2(z) - F_i^2(z + d_T x\epsilon)]^2 f_t\left(\frac{z}{d_t}\right)dz + 2\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} d_{ts}^{-1} \\
&\times\left|E\int\left[F_i^2(u) - F_i^2(u + d_T x\epsilon)\right] f_{ts}\left(\frac{u - x_s^*}{d_{ts}}\right)du\cdot\left[F_i^2(x_s) - F_i^2(x_s + d_T x\epsilon)\right]\right| \\
\leq & C_1\frac{d_T^2}{T^2}\sum_{t=1}^{T} d_t^{-1} + C_2\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} d_{ts}^{-1} \\
&\times\left|E\int F_i^2(u)\left[f_{ts}\left(\frac{u - x_s^*}{d_{ts}}\right) - f_{ts}\left(\frac{u - d_T x\epsilon - x_s^*}{d_{ts}}\right)\right]du\cdot\left[F_i^2(x_s) - F_i^2(x_s + d_T x\epsilon)\right]\right| \\
\leq & C_1\cdot T^{-1/2} + C_2|x|\epsilon d_T\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} d_{ts}^{-2} E|F_i^2(x_s) - F_i^2(x_s + d_T x\epsilon)| \\
\leq & C_1\cdot T^{-1/2} + C_2|x|\epsilon d_T\frac{d_T^2}{T^2}\sum_{t=2}^{T}\sum_{s=1}^{t-1} d_{ts}^{-2} d_s^{-1}\int|F_i^2(u) - F_i^2(u + d_T x\epsilon)|du \\
\leq & C_1\cdot T^{-1/2} + C_2|x|\epsilon\ln(T)
\end{aligned}$$

where $0 < C_1, C_2 < \infty$ are some constants which may be different at each appearance, and we have used variable change in the integrals, i.e.

$$\int\left[F_i^2(u) - F_i^2(u + d_T x\epsilon)\right] f_{ts}\left(\frac{u - x_s^*}{d_{ts}}\right)du = \int F_i^2(u)\left[f_{ts}\left(\frac{u - x_s^*}{d_{ts}}\right) - f_{ts}\left(\frac{u - d_T x\epsilon - x_s^*}{d_{ts}}\right)\right]du$$

and then the Lipschitz condition is applied; meanwhile, $\int|F_i^2(u) - F_i^2(u + d_T x\epsilon)|du \leq 2\int F_i^2(u)du = 2$ is also derived from a variable change.

Then, noting that there is $\phi(x)$ in the above equation, $E[A_{a2T}]$ is bounded by

$$\sum_{i=0}^{p-1}\left(C_1\ T^{-1/2} + C_2\ \epsilon\ T^{-1/2}\ln(T)\int|x|\phi(x)dx\right) = C_1\ T^{-1/2}\ p + C_2\ \epsilon\ \ln(T)\ p = o(1)$$

by a proper choice of ϵ.

To deal with the second term A_{b2T}, denote $G_{iT}(x) := d_T\int^x F_i^2(d_T u)du = \int^{d_T x} F_i^2(u)du$, so that $dG_{iT}(x) = d_T F_i^2(d_T x)dx$. Also define, $G(x) = 1$ if $x > 0$, and $G(x) = 0$ if $x < 0$. We have for any fixed i, $G_{iT}(x) \to G(x)$ as $T \to \infty$ at all continuous points of $G(x)$ since $\int F_i^2(x)dx = 1$. Notice that

$$\begin{aligned}
U_T(i,\epsilon) =& \frac{d_T}{T}\sum_{t=1}^{T}\int F_i^2(x_t + d_T x\epsilon)\phi(x)dx = \frac{d_T}{T}\sum_{t=1}^{T}\int F_i^2(d_T x)\phi_\epsilon(x - d_T^{-1}x_t)dx \\
=& \int\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(x - d_T^{-1}x_t))dG_{iT}(x).
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
A_{b2T} =&\sum_{i=0}^{p-1}\left(U_T(i,\epsilon)-\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(d_T^{-1}x_t)\right)^2\\
=&\sum_{i=0}^{p-1}\left(\int\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(x-d_T^{-1}x_t))dG_{iT}(x)-\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(d_T^{-1}x_t)\right)^2\\
\le&2\sum_{i=0}^{p-1}\left(\int_{|x|\le v}\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(x-d_T^{-1}x_t)dG_{iT}(x)-\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(d_T^{-1}x_t)\right)^2\\
&+2\sum_{i=0}^{p-1}\left(\int_{|x|>v}\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(x-d_T^{-1}x_t))dG_{iT}(x)\right)^2,
\end{aligned}$$

where $v>0$ is sufficiently large and fixed. For the second term above, observe that

$$\begin{aligned}
&\sum_{i=0}^{p-1}\left(\int_{|x|>v}\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(x-d_T^{-1}x_t))dG_{iT}(x)\right)^2\le O_P(1)\sum_{i=0}^{p-1}\left(\int_{|x|>v}dG_{iT}(x)\right)^2\\
=&O_P(1)\sum_{i=0}^{p-1}\left(\int_{|x|>d_Tv}F_i^2(x)dx\right)^2\le O_P(1)\frac{1}{(d_Tv)^2}\sum_{i=0}^{p-1}\left(\int|x|F_i^2(x)dx\right)^2\\
\le&O_P(1)\frac{1}{(d_Tv)^2}\sum_{i=0}^{p-1}\int x^2F_i^2(x)dx=O_P(1)\frac{1}{(d_Tv)^2}\sum_{i=0}^{p-1}i=O_P(1)p^2d_T^{-2},
\end{aligned}$$

by Cauchy-Swarchitz inequality and (iv) of Lemma A.3, $\int x^2F_i^2(x)dx=(2i+1)/2$, where we have used $\phi_\epsilon(x-d_T^{-1}x_t))=O_P(1)$ on the region $|x|>v$ for all $\epsilon>0$ since $d_T^{-1}x_t=O_P(1)$ and v is large enough so that $x-d_T^{-1}x_t\neq 0$ in probability on the region $|x|>v$.

Now, divide the interval $[-v,v]$ into $2m+1$ subintervals with equal length by a grid $\{s_{m,\ell}:\ell=-m,\cdots,m\}$ where $s_{m,-m}=-v<s_{m,-m+1}<\cdots<s_{m,m}<s_{m,m+1}=v$. Note also that $0\in(s_{m,0},s_{m,1})$. We have

$$\begin{aligned}
&\left|\int_{|x|\le v}\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(x-d_T^{-1}x_t))dG_{iT}(x)-\sum_{\ell=-m}^{m-1}\frac{1}{T}\sum_{t=1}^{T}\int_{s_{m,\ell}}^{s_{m,\ell+1}}\phi_\epsilon(s_{m,\ell}-d_T^{-1}x_t)dG_{iT}(x)\right|\\
=&\left|\sum_{\ell=-m}^{m}\frac{1}{T}\sum_{t=1}^{T}\left[\int_{s_{m,\ell}}^{s_{m,\ell+1}}(\phi_\epsilon(x-d_T^{-1}x_t)-\phi_\epsilon(s_{m,\ell}-d_T^{-1}x_t))dG_{iT}(x)\right]\right|\\
\le&C\frac{2v}{2m+1}\int_{|x|\le v}dG_{iT}(x)=C\frac{2v}{2m+1}\int_{|z|\le d_Tv}F_i^2(x)dx\le C\frac{2v}{2m+1}
\end{aligned}$$

due to the boundedness of the derivative of $\phi_\epsilon(\cdot)$.

Moreover, we have

$$\begin{aligned}
&\left|\sum_{\ell=-m}^{m}\frac{1}{T}\sum_{t=1}^{T}\int_{s_{m,\ell}}^{s_{m,\ell+1}}\phi_\epsilon(s_{m,\ell}-d_T^{-1}x_t)dG_{iT}(x)-\sum_{\ell=-m}^{m-1}\frac{1}{T}\sum_{t=1}^{T}\int_{s_{m,\ell}}^{s_{m,\ell+1}}\phi_\epsilon(s_{m,\ell}-d_T^{-1}x_t)dG(x)\right|\\
=&\left|\sum_{\ell=-m}^{m-1}\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(s_{m,\ell}-d_T^{-1}x_t)\int_{s_{m,\ell}}^{s_{m,\ell+1}}d(G_{iT}(x)-G(x))\right|\\
\le&\sum_{\ell=-m}^{m-1}\left|\int_{s_{m,\ell}}^{s_{m,\ell+1}}d(G_{iT}(x)-G(x))\right|=\left|\int_{s_{m,0}}^{s_{m,1}}d(G_{iT}(x)-G(x))\right|+2\sum_{\ell=1}^{m}\left|\int_{s_{m,\ell}}^{s_{m,\ell+1}}dG_{iT}(x)\right|\\
\le&|G_{iT}(s_{m,1})-1-G_{iT}(s_{m,0})|+2\sum_{\ell=1}^{m}\int_{d_Ts_{m,\ell}}^{d_Ts_{m,\ell+1}}F_i^2(x)dx
\end{aligned}$$

$$
\begin{aligned}
=&\left|\int^{d_T s_{m,1}} F_i^2(x)dx - 1 - \int^{d_T s_{m,0}} F_i^2(x)dx\right| + 2\int_{d_T s_{m,1}}^{\infty} F_i^2(x)dx \\
=&\left|\int_{d_T s_{m,1}} F_i^2(x)dx + \int^{d_T s_{m,0}} F_i^2(x)dx\right| + 2\int_{d_T s_{m,1}}^{\infty} F_i^2(x)dx \\
=&4\int_{d_T s_{m,1}}^{\infty} F_i^2(x)dx \le 4(d_T s_{m,1})^{-1}\left(\int x^2 F_i^2(x)dx\right)^{1/2} = O(1)(d_T v/m)^{-1}\sqrt{i},
\end{aligned}
$$

where we have used the facts that $F_i^2(x)$ is an even function as $F_i(x)$ is either even or odd function, $\int F_i^2(x)dz = 1$, $\int x^2 F_i^2(x)dx = O(1)i$ and Cauchy-Schwarz inequality.

In addition, by $\int_{s_{m,\ell}}^{s_{m,\ell+1}} dG(x) = 0$ if $0 \notin (s_{m,\ell}, s_{m,\ell+1})$ and 1 otherwise, we have

$$
\begin{aligned}
&\left|\sum_{\ell=-m}^{m-1} \frac{1}{T}\sum_{t=1}^{T} \phi_\epsilon(s_{m,\ell} - d_T^{-1}x_t)\int_{s_{m,\ell}}^{s_{m,\ell+1}} dG(x) - \frac{1}{T}\sum_{t=1}^{n}\phi_\epsilon(d_T^{-1}x_t)\right| \\
=&\left|\frac{1}{T}\sum_{t=1}^{T}[\phi_\epsilon(s_{m,0} - d_T^{-1}x_t) - \phi_\epsilon(-d_T^{-1}x_t)]\right| \le C|s_{m,0}| = C\frac{v}{m},
\end{aligned}
$$

by the symmetry of $\phi_\epsilon(\cdot)$ and the boundedness of its derivative.

It follows that the second term T_{b2T} is surely bounded by

$$
\sum_{i=0}^{p-1}\left(U_T(i,\epsilon) - \frac{d_T}{T}\sum_{t=1}^{T}\phi_\epsilon(d_T^{-1}x_t)\right)^2 \le C(p^2 d_T^{-2} + p(v/m)^2 + p^2(d_T v/m)^{-2}).
$$

Noting that v is fixed, $pd_T^{-1} = o(1)$ as $T \to \infty$, we may choose a $m = m_T = T^\tau$ such that not only $p/m^2 \to 0$ but also $pm^2 d_T^{-2} \to 0$. This is fulfilled if we let τ satisfy $\alpha < 2\tau < 1 - \alpha$. Such τ does exist due to Assumption 3.

For the third term A_{c2T}, notice that

$$
\begin{aligned}
&\left(\frac{1}{T}\sum_{t=1}^{T}\phi_\epsilon(d_T^{-1}x_t) - \int_0^1 \phi_\epsilon(B(r))dr\right)^2 = \left(\int_0^1 \phi_\epsilon(W_T(r))dr - \int_0^1 \phi_\epsilon(B(r))dr + \frac{1}{T}\phi_\epsilon(d_T^{-1}x_T)\right)^2 \\
\le& 2\left(\int_0^1 [\phi_\epsilon(W_T(r)) - \phi_\epsilon(B(r))]dr\right)^2 + O_P\left(\frac{1}{T^2}\right) \le 2\int_0^1 [\phi_\epsilon(W_T(r)) - \phi_\epsilon(B(r))]^2 dr + O_P\left(\frac{1}{T^2}\right),
\end{aligned}
$$

where $W_T(r) := d_T^{-1}x_{[Tr]}$.

By the strong approximation, in a richer probability space we have $\sup_{0\le r\le 1}|W_T(r) - B(r)| = o(T^{-1/4}\log(T))$ a.s. (Phillips, 2001, p. 391) in an expanded probability space. Thus, the third term A_{c2T} is almost surely bounded by $o_P\left((pT^{-1/2}\log^2(T))\right) + O_P\left(pT^{-2}\right)$, by noting that the derivative of $\phi_\epsilon(\cdot)$ is bounded and neglecting some constants.

For the last term A_{d2T}, note that the derivative of the Heaviside function $I(x \ge 0)$ is the Deric delta function $\delta(x)$, i.e., $\frac{d}{dx}I(x \ge 0) = \delta(x)$, or equivalently, Heaviside function $I(x \ge 0)$ is the distribution function of the delta function. Thus, using the property of the delta function (i.e. $\int f(x)\delta(x)dx = f(0)$ for any continuous function $f(x)$), for any $\epsilon > 0$ we have $\int L_B(1,\epsilon x)\delta(x)dx = \int L_B(1,x)\delta(x)dx = L_B(1,0)$ because $\delta(\epsilon x) = \delta(x)/\epsilon$. See Gel'fand and Shilov (1964) for detailed facts on the generalized functions. It follows from the occupation time formula that

$$
\int_0^1 \phi_\epsilon(B(r))dr - L_B(1,0) = \int \phi_\epsilon(x)L_B(1,x)dx - L_B(1,0)
$$

$$= \int \phi(x) L_B(1, \epsilon x) dx - \int L_B(1, \epsilon x)\delta(x) dx = \int L_B(1, \epsilon x) d(\Phi(x) - I(x \geq 0))$$
$$:= \int L_B(1, \epsilon x) d\Psi(x),$$

where $\Phi(x)$ is the distribution function of a standard normal variable and $\Psi(x) := \Phi(x) - I(x \geq 0)$, so that $\Psi(\infty) = \Psi(-\infty) = 0$.

Lemma 2.1 of Borodin (1986, p. 239) shows that $E\left|\int L_B(1, \epsilon x) d\Psi(x)\right|^\lambda \leq C\epsilon^{\lambda/2}$ for any $\lambda = 1, 2, \cdots$. Therefore, taking $\lambda = 1$, $E[A_{d2T}]^2 \leq p\epsilon$. If we choose $\epsilon = o(p^{-1})$, we will conclude the assertion. □

References

Andrews, D. W. K. (1991). Asymptotic normality of series estimator for nonparametric and semiparametric models. *Econometrica*, 59:307–345.

Askey, R. and Wainger, S. (1965). Mean convergence of expansions in Laguerre and Hermite series. *American Journal of Mathematics*, 87:695–708.

Balke, N. S. and Fomby, T. B. (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.

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

Chen, J., Gao, J., and Li, D. (2012). Estimation in semiparametric regression with nonstationary regressors. *Bernoulli*, 18:678–702.

Chen, X. and Shen, X. (1998). Sieve extremum estimates for weakly dependent data. *Econometrica*, 66:289–314.

Dong, C. and Gao, J. (2013). Orthogonal expansion of functionals of lévy processes: theory and practice. http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2013/03-13.php.

Dong, C. and Gao, J. (2014). Specification testing in structural nonparametric cointegration: theory and practice. http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2014/wp02-14.php.

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

Gao, J. (2007). *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman & Hall/CRC, New York.

Gao, J., Li, D., and Tjøstheim, D. (2009). Uniform consistency for nonparametric estimators in null recurrent time series. http://economics.adelaide.edu.au/research/papers/doc/wp2009-26.pdf.

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

Gel'fand, I. M. and Shilov, G. E. (1964). *Generalized Functions*. Academic Press, New York.

Granger, C. W. J. (1995). Modelling nonlinear relationship between extended-memory variables. *Econometrica*, 63(2):265–279.

Kanwal, R. P. (1983). *Generalized Functions: Theory and Technique*. Academic Press, New York.

Karlsen, H. A., Mykelbust, T., and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics*, 35:252–299.

Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Annals of Statistics*, 29:372–416.

Lukacs, E. (1970). *Characteristic Functions*. Griffin, London, second edition edition.

Marmer, V. (2008). Nonlinearity, nonstationarity and spurious forecasts. *Journal of Econometrics*, 142:1–27.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79:147–168.

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

Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear regression with integreted time series. *Econometrica*, 69(1):117–161.

Phillips, P. C. B. (2001). Descriptive econometrics for nonstationary time series with empirical applications. *Journal of Applied Econometrics*, 16:389–413.

Phillips, P. C. B. and Solo, V. (1992). Asymptotics for linear processes. *Annals of Statistics*, 20(2):971–1001.

Robinson, P. M. (1997). Large-sample inference for nonparametric regression with dependent errors. *Annals of Statistics*, 25:2054–2083.

Schwartz, S. C. (1967). Estimation of probability density by an orthogonal series. *Annals of Mathematical Statistics*, 38:1261–1265.

Szego, G. (1975). *Orthogonal Polynomials*. Colloquium publications XXIII. American Mathematical Association, Providence, Rhode Island.

Teräsvirta, T., Tjøstheim, D., and Granger, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. Advanced Texts in Econometrics. Oxford University Press, New York.

Wang, Q. and Chan, N. (2014). Uniform convergence rates for a class of martingales with application in nonlinear cointegrating regression. *Bernoulli*, 20:207–230.

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

Wang, Q. and Phillips, P. C. B. (2009b). Structure nonparametric cointegrating regression. *Econometrica*, 77:1901–1948.

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

Wang, Q. and Phillips, P. C. B. (2014). Nonparametric cointegrating regression with endogeneity and long memory. Working paper available at http://www.maths.usyd.edu.au/u/pubs/publist/preprints/2014/wang-12.pdf.