Journal of Econometrics 207 (2018) 212–236

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Additive nonparametric models with time variable and both stationary and nonstationary regressors

Chaohua Dong [a,b,*], Oliver Linton [c]

[a] *Zhongnan University of Economics and Law, China*
[b] *Southwestern University of Finance and Economics, China*
[c] *University of Cambridge, UK*

ARTICLE INFO

Article history:
Received 5 October 2016
Received in revised form 8 January 2018
Accepted 14 May 2018
Available online 17 August 2018

JEL classification:
C13
G14
C22

Keywords:
Additive nonparametric models
Deterministic trend
Pairs trading
Series estimator
Stationary and locally stationary processes
Unit root process

ABSTRACT

This paper considers nonparametric additive models that have a deterministic time trend and both stationary and integrated variables as components. The diverse nature of the regressors caters for applications in a variety of settings. In addition, we extend the analysis to allow the stationary regressor to be instead locally stationary, and we allow the models to include a linear form of the integrated variable. Heteroscedasticity is allowed for in all models. We propose an estimation strategy based on orthogonal series expansion that takes account of the different type of stationarity/nonstationarity possessed by each covariate. We establish pointwise asymptotic distribution theory jointly for all estimators of unknown functions and also show the conventional optimal convergence rates jointly in the L_2 sense. In spite of the entanglement of different kinds of regressors, we can separate out the distribution theory for each estimator. We provide Monte Carlo simulations that establish the favorable properties of our procedures in moderate sized samples. Finally, we apply our techniques to the study of a pairs trading strategy.

1. Introduction

This paper is devoted to the investigation of additively separable nonparametric regressions with deterministic time trend, stationary and nonstationary variables. In practice all these types of variables are important in applications in economics, finance and related fields. For example, aggregate consumption, disposable income and share prices are widely accepted as being (globally) nonstationary variables, while interest rates and the volume of share trading are often taken as stationary variables or locally stationary variables with mild trends. Some variables may also contain a deterministic time trend. Therefore, from a practical point of view, it is necessary to study regression with different kinds of regressors. The choice of functional form is also important, and we should not like to restrict the shape of the regression functions, though quite hard to address in the presence of nonstationarity, which is the purpose of our study.

Grenander and Rosenblatt (1957) is a classic treatment of parametric deterministic trend models, while Phillips (2007, 2010) provide an update and discussion. There are a number of papers that develop theory for nonparametric regression with nonstationary variables alone. Karlsen et al. (2007) investigate the nonparametric regression situation where the single covariate is a recurrent Markov chain. Schienle (2008) investigates additive nonparametric regressions with Harris recurrent covariates and obtained a limit theory for kernel smooth backfitting estimators. Wang and Phillips (2009) consider an

* Correspondence to: School of Statistics and Mathematics, Zhongnan University of Economics and Law, Wuhan, Hubei 430073, China.
E-mail address: dchaohua@hotmail.com (C. Dong).

https://doi.org/10.1016/j.jeconom.2018.05.007
0304-4076/

alternative treatment by making use of local time limit theory and, instead of recurrent Markov chains, worked with partial sum representations. Phillips et al. (2017) consider a functional coefficient model where the covariates are unit root processes and the functional coefficient is driven by rescaled time. Wang (2015) gives an excellent overview of the tools needed for distribution theory in a variety of these settings.

To the best of our knowledge, there are no theoretical studies that accommodate these three kinds of regressors in a nonparametric setting. The closest study is Chang et al. (2001) where, though all the three regressors are contained, a nonlinear parametric model is studied, that is, all functions are supposed to be known. In addition, there are a number of studies that contain regressors with two of these features and most of them are linear regression with perhaps functional coefficients. Park and Hahn (1999) study linear regression with I(1) regressor and time varying coefficients depending on fixed design; Xiao (2009) studies functional-coefficient cointegration regression where the coefficients depend on a stationary variable and the regressor is an I(1) vector; Cai et al. (2009) study a similar model with more flexibility; more interestingly, Li et al. (2016) recently investigate the convergence of sample covariances which have I(1) process and a variable that can be a fixed design or a random design but not both.

In this paper, we mainly consider the model

$$y_t = \beta(t/n) + g(z_t) + m(x_t) + e_t, \ \ t = 1, \dots, n \tag{1.1}$$

where β, g and m are unknown smooth functions, z_t and x_t are stationary and integrated processes, respectively, but may be correlated, e_t is an error term. Here, $\beta(\cdot)$ is defined on [0,1], $g(\cdot)$ is defined on V_z, the support of z_1, and $m(\cdot)$ is supposed to be integrable and defined on $\mathbb{R}$. Notice that V_z could be a finite interval like $[a, b]$ or an infinite interval like $(-\infty, \infty)$ or $(0, \infty)$.

All unknown functions will be estimated by the series method, which is particularly convenient in additive models (Andrews and Whang, 1990), compared with the kernel method that requires an iterative "backfitting technique" (Mammen et al., 1999). Indeed, the series method gives an explicit solution for the estimators obtained by the ordinary least squares, which facilitates the asymptotic analyses. In contrast, the smooth backfitting technique needs two steps, in order to derive the estimators. See, for example, Vogt (2012, p. 2612).

Moreover, the setting of model (1.1) is quite different from existing papers such as Dong et al. (2016) and Phillips et al. (2017). Note that Dong et al. (2016) mainly investigates a single-index model with an integrated regressor that does not contain either deterministic trend or stationary variable, while Phillips et al. (2017) deals with a functional-coefficient model. In particular, the approach of deriving asymptotic distribution makes much improvement in this paper as simultaneously three types of variables are involved in nonparametric models.

The most important feature of model (1.1) is the diverse nature of the regressors, which permits a wide variety of applications. This, however, gives rise to a challenge for the asymptotic analyses. Our findings include that: (1) the interactions between $m(x_t)$, properly normalized, and any one of the other components eventually vanishes; (2) although different kinds of variables are entangled inside the estimators, each has its own separable convergence rate; (3) conventional optimal convergence rates are attainable.

We further extend the model (1.1) in two respects. We shall relax the stationary process z_t to be a locally stationary process. That is, we consider also

$$y_t = \beta(t/n) + g(z_{nt}) + m(x_t) + e_t, \tag{1.2}$$

where $t = 1, \dots, n$, all ingredients are the same as in model (1.1) except that z_{nt} is a locally stationary process defined below. This class of processes has received a lot of attention recently, (see, Vogt (2012)), and it captures an important notion that there is slowly evolving change. In addition, since the integrability of the function $m(\cdot)$ excludes the polynomial form in x_t, we extend the model below to contain a linear form of the integrated process. It is clear that this linear form may be substituted by any polynomial without constant and similar theoretical results remain true.

We work with scalar covariates although it is easy to extend the theory to allow the stationary or locally stationary regressor z_{nt} to be a vector $(z_{nt;j}, j = 1, \dots, d)$ and $g(z_{nt}) = \sum_{j=1}^{d} g_j(z_{nt;j})$, but we have eschewed this further complication due to its notational cost.

Our procedure is easy to implement and we verify in simulation experiments that the distribution theory we obtain well captures the finite sample behavior of our estimators. We apply our methodology to the study of pairs trading (Gatev et al., 2006). We consider the stock prices of Coke and Pepsi and build a model that links these prices and allows for globally nonstationary components, slowly moving deterministic trends, and a stationary or locally stationary covariate, in our case the relative trading volume of the two common stocks. We find that our model captures important nonlinearity and evolutionary behavior in the relationship between the two stock prices that the usual linear cointegrating relationship ignores. The value of our approach is quantified through out of sample forecast and trading profits relative to the linear alternative.

The organization of the rest is as follows. Section 2 describes the procedure of estimation; Section 3 gives the entire asymptotic theory that covers the normality of estimators for model (1.1) in Section 3.1, that for model (1.2) in Section 3.2 and that for the extended model which contains an extra linear form of x_t in Section 3.3; Monte Carlo experiment is conducted in Section 4, followed by an empirical study in Section 5, and Section 6 concludes. Appendix A contains all technical lemmas whose proofs are relegated to the supplementary material of the paper; Appendix B gives the proofs of theorems in Sections 3.1 and 3.2 while that of all other theorems, proposition and corollaries are shown in the supplement.

Throughout the paper, I_k is the identity matrix of dimension k; $\|v\|$ is Euclidean norm for any vector v and $\|A\|$ is entry-wise norm for any matrix; $\int f(x)dx$ is an integral on the entire $\mathbb{R}$; $C, C_1, \ldots$, can be any positive constants and may be different at each appearance.

2. Assumptions and estimation procedure

This section gives assumptions on the regressors and the error term as well as the procedure by which the unknown functions are estimated.

2.1. Assumptions

We first give the structure of the integrated regressor x_t that we shall assume.

Assumption A.

A.1 Let $\{\epsilon_j, -\infty < j < \infty\}$ be a scalar sequence of independent and identically distributed (i.i.d.) random variables having an absolutely continuous distribution with respect to the Lebesgue measure and satisfying $\mathbb{E}[\epsilon_1] = 0$, $\mathbb{E}[\epsilon_1^2] = 1$, $\mathbb{E}|\epsilon_1|^{q_1} < \infty$ for some $q_1 \geq 4$. The characteristic function of ϵ_1 satisfies that $\int |\lambda||\mathbb{E}\exp(i\lambda\epsilon_1)|d\lambda < \infty$.

A.2 Let $w_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}$ where $\sum_{j=0}^{\infty} j|\psi_j| < \infty$ and $\psi := \sum_{j=0}^{\infty} \psi_j \neq 0$.

A.3 For $t \geq 1$, $x_t = x_{t-1} + w_t$, and $x_0 = O_P(1)$.

The conditions in Assumption A are commonly used in the literature concerning unit root time series (see, e.g. Park and Phillips, 1999, 2001 and Dong et al., 2016). The innovation variables $\{\epsilon_j\}$ are building blocks for the linear process w_t from which the regressor is obtained by integration. All crucial properties of x_t for our theoretical development given in Lemma A.1 are derived from the $I(1)$ structure.

Meanwhile, from the structure of x_t, we may have $d_n^2 := \mathbb{E}(x_n^2) = \psi^2 n(1 + o(1))$ simply by virtue of the BN decomposition for w_t (Phillips and Solo, 1992 p. 972). It follows that for $r \in [0, 1]$, $d_n^{-1} x_{[nr]} \to_D W(r)$ in the space $D[0, 1]$ as $n \to \infty$, where $[\cdot]$ is the biggest integer not exceeding the argument. Here, $D[0, 1]$ is the Skorokhod space on $[0, 1]$, that is, the collection of functions defined on $[0, 1]$ that are everywhere right-continuous and have left limits everywhere; $W(r)$ is a standard Brownian motion and our theory developed below depends on its local time process defined by $L_W(r, a) = \lim_{\epsilon\to 0} \epsilon^{-1} \int_0^r I(|W(u) - a| < \epsilon)du$, where $I(\cdot)$ is the indicator function. Note that $L_W(r, a)$ stands for the sojourn time of the process $W(\cdot)$ at the spatial point a over the time period $[0, r]$, and Revuz and Yor (2005) is a standard book introducing the local time of Brownian motion.

Assumption B.

B.1 Suppose that either (a) z_t is a strictly stationary and α-mixing process with mixing coefficients $\alpha(i)$ such that $\sum_{i=1}^{\infty} \alpha^{\delta/(2+\delta)}(i) < \infty$ for some $\delta > 0$, and z_t are independent of $\{\epsilon_j, -\infty < j < \infty\}$ defined in Assumption A; or (b) $z_t = \rho(\epsilon_t, \ldots, \epsilon_{t-d+1}; \eta_t, \ldots, \eta_{t-d+1})$ with fixed $d \geq 1$ and measurable function $\rho : \mathbb{R}^{2d} \mapsto \mathbb{R}$, and z_t have finite second moment, where i.i.d.(0,1) sequence $\{\eta_j\}$ is independent of $\{\epsilon_j\}$.

B.2 There exists an orthogonal function sequence $\{p_i(z), i \geq 0\}$ on the support V_z of z_1 and the orthogonality is with respect to $dF(z)$ where $F(z)$ is a distribution function on V_z. In addition, for $\delta > 0$ given by Assumption B.1, we have either (a) $\mathbb{E}|p_j(z_1)|^{2(2+\delta)} = O(j)$ for large j or (b) $\sup_{j\geq 0} \mathbb{E}|p_j(z_1)|^{2(2+\delta)} < \infty$.

B.3 There is a filtration sequence $\mathcal{F}_{n,t}$ such that $(e_t, \mathcal{F}_{n,t})$ form a martingale difference sequence and (z_t, x_t) is adapted to $\mathcal{F}_{n,t-1}$. Moreover, almost surely $\mathbb{E}(e_t^2|\mathcal{F}_{n,t-1}) = \sigma^2(t/n)$, where $\sigma^2(\cdot)$ is a positive continuous function on $[0, 1]$ and $\max_{1\leq t\leq n} \mathbb{E}(|e_t|^{q_2}|\mathcal{F}_{n,t-1}) < \infty$ for some $q_2 \geq 4$.

Condition B.1 takes into account two cases for z_t. In (a), z_t is an α-mixing stationary process (a common assumption that we only refer the readers to Gao (2007) and independent of x_t, while in (b), z_t is correlated with x_t by sharing the same $\epsilon_t, \ldots, \epsilon_{t-d+1}$. These two conditions are different but overlap, because z_t in (b) is d-dependent, a subclass of α-mixing process, while in terms of the relationship with x_t they are mutually exclusive. Definitely, the presence of the correlation between x_t and z_t would give rise to a challenge for our theoretical derivation. To tackle the issue, we show the probability properties of x_t in Lemmas A.1 and A.2 and the correlation for functions of x_t and z_t in Lemma A.3 below. Particularly, the results of Lemma A.3 imply that, comparing with the independence, the correlation gives an extra term which is infinitesimal and therefore negligible in our derivation. Hence, it is due to these lemmas that we are able to deal with the correlation in model (1.1) and that our model is applicable broadly.

Condition B.2 stipulates an orthogonal sequence $\{p_i(z), i \geq 0\}$ on the support $V(\equiv V_z$, the subscript is suppressed here and below) that is used to approximate the unknown function $g(\cdot)$ in the regression model.

Given a support $V \subset \mathbb{R}$, the choice of the density $dF(z)$ determines what function space we shall work with. It is well known that an orthogonal polynomial sequence can be constructed on a support with respect to a density by the Gram–Schmidt orthonormalization theorem. See, for example, Dudley (2003, p. 168). If z_1 is normal, $V = \mathbb{R}$, the sequence is consisting of Hermite polynomials given $dF(z) = (2\pi)^{-1/2} e^{-z^2/2} dz$; if z_1 has support $V = [0, \infty)$, the sequence is consisting of Laguerre polynomials given $dF(z) = e^{-z}dz$; if $V = [0, 1]$, orthogonal trigonometric polynomials could be used; if $V = [-1, 1]$, the orthogonal polynomials are Chebyshev or Legendre polynomials.

Notice also that Conditions (a) and (b) in B.2 are about how to control the high order moments of the basis $p_j(x)$ and are used to measure the divergence of certain partial sum below. Because we do not specify the interval V of the variable z_1, there are two cases considered herein. B.2(a) is tackling the case that V is an infinite interval where the high order moment of $p_j(x)$ diverges with j, while B.2(b) is mainly for the case where V is a compact set (e.g. $[0, 1]$, $[-1, 1]$ and so on) such that the high order moment is uniformly bounded with j. The moment condition is mild and commonly used. In the literature, B.2(a) is used in Assumption 3 of Dong et al. (2015) and B.2(b) is used in Assumption 3 of Su and Jin (2012). It is worth to point out that the similar assumption for bases used to estimate $\beta(\cdot)$ and $m(\cdot)$ (i.e. $\varphi_j(\cdot)$ and $\mathscr{H}_j(\cdot)$ below) is not necessary since these are specified bounded functions.

The martingale difference structure for the error term is extensively used in the literature such as Park and Phillips (1999, 2001) and Gao and Phillips (2013) among others. However, Condition B.3 here allows heteroscedasticity that is a function depending on the normalized time t/n. This makes our theoretical results more applicable, but the function $\sigma^2(\cdot)$ might be multivariate to contain additionally either z_t or x_t even both. This possibility would affect a bit the conditional variance matrices studied below while the main results still hold. To preserve space, we do not consider all possibilities in this regard.

In order to be more applicable, we may allow z_t in model (1.1) to be a locally stationary process, which is defined as follows.

Definition 2.1 (*Locally Stationary Process*)**.** Process $\{z_{nt}\}$ is locally stationary if for each rescaled time point $v \in [0, 1]$ there exists an associated process $\{z_t(v)\}$ satisfying:

(i) $\{z_t(v)\}$ is strictly stationary;

(ii) it holds that

$$|z_{nt} - z_t(v)| \leq \left(\left|\frac{t}{n} - v\right| + \frac{1}{n}\right) U_{nt}(v) \quad \text{a.s.},$$

where $U_{nt}(v)$ is a process of positive variables such that $\mathbb{E}[(U_{nt}(v))^{q_3}] < C$ for some $q_3 \geq 1$ and $C > 0$ independent of v, t and n.

This definition of locally stationarity accommodates a variety of financial datasets. Koo and Linton (2012) give sufficient conditions under which a time-inhomogeneous diffusion process is locally stationary and meanwhile, Vogt (2012) studies nonparametric regression for locally stationary time series. Certainly, each stationary process is locally stationary.

Assumption B*. Suppose that

B*.1 $\{z_{nt}\}$ is locally stationary with associated process $\{z_t(v)\}$, and all z_{nt} ($1 \leq t \leq n$) have the same compact support $V_z = [a_{\min}, a_{\max}]$. Moreover, the density $f(v, z)$ of $z_t(v)$ is smooth in v.

B*.2 For all t and any $v \in [0, 1]$, either (a) $z_t(v)$ satisfies Assumption B.1.a, or (b) $z_t(v)$ satisfies Assumption B.1.b.

B*.3 There exists an orthogonal function sequence $\{p_i(z), i \geq 0\}$ on the support $[a_{\min}, a_{\max}]$ with respect to $dF(z)$ such that $\sup_{v\in[0,1]}\sup_{j\geq 0}\mathbb{E}|p_j(z_1(v))| < \infty$.

B*.4 Suppose that there is a filtration sequence $\mathcal{F}_{nt}$ such that $(e_t, \mathcal{F}_{n,t})$ form a martingale difference sequence and $(z_t(t/n), x_t)$ is adapted with $\mathcal{F}_{n,t-1}$. Meanwhile, $\mathbb{E}(e_t^2|\mathcal{F}_{n,t-1}) = \sigma^2(t/n)$ almost surely with continuous and nonzero function $\sigma(\cdot)$ and for some $q_3 \geq 4$, $\max_{1\leq t\leq n}\mathbb{E}(|e_t|^{q_3}|\mathcal{F}_{n,t-1}) < \infty$.

This assumption allows us to approximate the locally stationary variable z_{nt} by stationary variable $z_t(v)$ when t/n is in a small neighborhood of v. Thus, the theoretical results below may be applicable. As studied in Koo and Linton (2012, p. 212), $\{z_{nt}\}$ may have a common domain of closed interval. Thus, we simply require the support of the locally stationary process to be compact in this paper. Moreover, $\{z_t(v)\}$ possibly is α-mixing and β-mixing, as studied in Koo and Linton (2012) and Chen et al. (2010). Moreover, Theorem 3.3 of Vogt (2012) shows, under certain conditions, the density $f(v, z)$ of $z_t(v)$ is smooth in v. Here again, by B*.2 we allow the associated stationary process to be either independent of or correlated with the nonstationary process x_t, for which we have the same comment as that for Assumption B.1.

Basically, Assumption B*.1 is particularly for the local stationary process, while Assumptions B*.2–B*.4 are a generalized version of Assumption B.1–B.3, that take into account the dependence of the locally stationarity on the normalized time $v \in [0, 1]$. As z_{nt} is approximated asymptotically by the stationary process $z_t(t/n)$, the condition of e_t in B*.4 is assumed to be a martingale difference sequence with respect to a filtration that satisfies conditions similar to B.3 of Assumption B.

2.2. Estimation procedure

The least squares series estimation method is used to estimate all unknown functions in models (1.1) and (1.2). By nature these functions belong to different function spaces, and therefore we introduce these function spaces and their orthonormal bases.

First, suppose that $\beta(\cdot) \in L^2[0,1] = \{u(r) : \int_0^1 u^2(r)dr < \infty\}$, in which the inner product is given by $\langle u_1, u_2\rangle = \int_0^1 u_1(r)u_2(r)dr$ and the induced norm $\|u\|^2 = \langle u, u\rangle$. Let $\varphi_0(r) \equiv 1$, and for $j \geq 1$, $\varphi_j(r) = \sqrt{2}\cos(\pi jr)$. Then, $\{\varphi_j(r)\}$ is an orthonormal basis in the Hilbert space $L^2[0,1]$, $\langle\varphi_i(r), \varphi_j(r)\rangle = \delta_{ij}$ the Kronecker delta. The basis $\{\varphi_j(r)\}$ is used to expand the unknown continuous function $\beta(r) \in L^2[0,1]$ into orthogonal series, that is,

$$\beta(r) = \sum_{j=0}^{\infty} c_{1,j}\varphi_j(r), \quad \text{where } c_{1,j} = \langle\beta(r), \varphi_j(r)\rangle. \tag{2.1}$$

It is noteworthy that $\{\varphi_j(r)\}$ can be replaced by any other orthonormal basis in $L^2[0,1]$, as shown in Chen and Shen (1998), Gao et al. (2001) and Phillips (2005) among others. However, with this specific basis other than a general one we do not need any assumption on it, and all quantities related to the basis are easily and directly calculated. See Lemma A.2 below.

Second, in order to expand $g(z_t)$, suppose that the function $g(\cdot)$ is in Hilbert space $L^2(V, dF(x)) = \{q(x) : \int_V q^2(x)dF(x) < \infty\}$ where $F(x)$ is a distribution on the support V that may not be compact. The sequence $\{p_j(x), j \geq 0\}$ in Assumption B.2 is an orthonormal basis in $L^2(V, dF(x))$ where an inner product is given by $\langle q_1, q_2\rangle = \int_V q_1(x)q_2(x)dF(x)$ and the induced norm $\|q\|^2 = \langle q, q\rangle$. Hence, the unknown function $g(x)$ has an orthogonal series expansion in terms of the basis of $\{p_j(x), j \geq 0\}$, viz.,

$$g(x) = \sum_{j=0}^{\infty} c_{2,j}p_j(x), \quad \text{where } c_{2,j} = \langle g(x), p_j(x)\rangle. \tag{2.2}$$

Third, because of $x_t = O_P(\sqrt{t})$, the support of $m(\cdot)$ has to be $\mathbb{R}$. We thus suppose $m(\cdot) \in L^2(\mathbb{R}) = \{f(x) : \int f^2(x)dx < \infty\}$ in which an inner product is given by $\langle f_1, f_2\rangle = \int f_1(x)f_2(x)dx$ and the induced norm $\|f\|^2 = \langle f, f\rangle$. To expand $m(x)$, recall the Hermite polynomials $\{H_j(x)\}$ and the Hermite functions $\{\mathscr{H}_j(x)\}$. By definition

$$H_j(x) = (-1)^j \exp(x^2)\frac{d^j}{dx^j}\exp(-x^2), \qquad j \geq 0, \tag{2.3}$$

are Hermite polynomials such that $\int H_i(x)H_j(x)\exp(-x^2)dx = \sqrt{\pi}2^j j!\delta_{ij}$, meaning that they are orthogonal with respect to the density $\exp(-x^2)$. It is known that

$$\mathscr{H}_j(x) = (\sqrt{\pi}2^j j!)^{-1/2}H_j(x)\exp\left(-\frac{x^2}{2}\right), \qquad j \geq 0, \tag{2.4}$$

are called Hermite functions in the relevant literature.

The orthogonality of the Hermite polynomials implies that $\langle\mathscr{H}_i(x), \mathscr{H}_j(x)\rangle = \delta_{ij}$. In addition, $\{\mathscr{H}_j(x)\}$ is bounded uniformly in both j and $x \in \mathbb{R}$. See Szego (1975, p. 242). Moreover, $\{\mathscr{H}_j(x)\}$ is an orthonormal basis in Hilbert space $L^2(\mathbb{R})$. The unknown function $m(x)$ thence has an orthogonal series expansion in terms of $\{\mathscr{H}_j(x)\}$, viz.,

$$m(x) = \sum_{j=0}^{\infty} c_{3,j}\mathscr{H}_j(x), \text{ where } c_{3,j} = \langle m(x), \mathscr{H}_j(x)\rangle. \tag{2.5}$$

2.2.1. Estimation procedure for model (1.1)

Let k_i, $i = 1, 2, 3$, be positive integers. Define truncation series with truncation parameter k_1 for $\beta(r)$ as $\beta_{k_1}(r) = \sum_{j=1}^{k_1} c_{1,j}\varphi_j(r)$ (noting by Assumption C.2 below that $c_{1,0} = 0$) and residue after truncation $\gamma_{1k_1}(r) = \sum_{j=k_1+1}^{\infty} c_{1,j}\varphi_j(r)$. It is known that $\beta_{k_1}(r) \to \beta(r)$ as $k_1 \to \infty$ in pointwise sense for smooth $\beta(r)$. Similarly, define the truncation series for $g(x)$ as $g_{k_2}(x) = \sum_{j=0}^{k_2-1} c_{2,j}p_j(x)$ and residue after truncation as $\gamma_{2k_2}(x) = \sum_{j=k_2}^{\infty} c_{2,j}p_j(x)$; for $m(x)$ as $m_{k_3}(x) = \sum_{j=0}^{k_3-1} c_{3,j}\mathscr{H}_j(x)$ and residue after truncation as $\gamma_{3k_3}(x) = \sum_{j=k_3}^{\infty} c_{3,j}\mathscr{H}_j(x)$. It follows that $g_{k_2}(x) \to g(x)$ and $m_{k_3}(x) \to m(x)$ as $k_2, k_3 \to \infty$ in some sense under certain condition. We omit mathematical details in order not to deviate from our main course.

Denote $\phi_{k_1}(r) = (\varphi_1(r), \ldots, \varphi_{k_1}(r))^\intercal$ and $c_1 = (c_{1,1}, \ldots, c_{1,k_1})^\intercal$. We then have $\beta_{k_1}(r) = \phi_{k_1}(r)^\intercal c_1$. Denote also $a_{k_2}(x) = (p_0(x), \ldots, p_{k_2-1}(x))^\intercal$, $b_{k_3}(x) = (\mathscr{H}_0(x), \ldots, \mathscr{H}_{k_3-1}(x))^\intercal$, and $c_i = (c_{i,0}, \ldots, c_{i,k_i-1})^\intercal$, $i = 2,3$. Accordingly, $g_{k_2}(x) = a_{k_2}(x)^\intercal c_2$ and $m_{k_3}(x) = b_{k_3}(x)^\intercal c_3$. Thus, model (1.1) is written as

$$\begin{aligned} y_t =&\phi_{k_1}(t/n)^\intercal c_1 + a_{k_2}(z_t)^\intercal c_2 + b_{k_3}(x_t)^\intercal c_3 \\ &+ \gamma_{1k_1}(t/n) + \gamma_{2k_2}(z_t) + \gamma_{3k_3}(x_t) + e_t, \end{aligned} \tag{2.6}$$

where $t = 1, \ldots, n$.

To write all equations in (2.6) into a matrix form, let $y = (y_1, \ldots, y_n)^\intercal$, $c = (c_1^\intercal, c_2^\intercal, c_3^\intercal)^\intercal$, $e = (e_1, \ldots, e_n)^\intercal$, $\gamma = (\gamma(1), \ldots, \gamma(n))^\intercal$ where $\gamma(t) = \gamma_{1k_1}(t/n) + \gamma_{2k_2}(z_t) + \gamma_{3k_3}(x_t)$, $t = 1, \ldots, n$, and

$$B_{nk} = \begin{pmatrix} \phi_{k_1}(1/n)^\intercal & a_{k_2}(z_1)^\intercal & b_{k_3}(x_1)^\intercal \\ \vdots & \vdots & \vdots \\ \phi_{k_1}(1)^\intercal & a_{k_2}(z_n)^\intercal & b_{k_3}(x_n)^\intercal \end{pmatrix}$$

a $n \times k$ matrix with $k = k_1 + k_2 + k_3$ for convenience. Consequently, we have

$$y = B_{nk}c + \gamma + e \tag{2.7}$$

which by the ordinary least squares (OLS) gives $\widehat{c} = (\widehat{c}_1^\intercal, \widehat{c}_2^\intercal, \widehat{c}_3^\intercal)^\intercal = (B_{nk}^\intercal B_{nk})^{-1}B_{nk}^\intercal y$ provided that the matrix $B_{nk}^\intercal B_{nk}$ is non-singular (which will be so under our conditions with high probability).

Therefore, for any $r \in [0, 1]$, $z \in V$ and $x \in \mathbb{R}$ define naturally $\widehat{\beta}_n(r) = \phi_{k_1}(r)^\intercal \widehat{c}_1$, $\widehat{g}_n(z) = a_{k_2}(z)^\intercal \widehat{c}_2$ and $\widehat{m}_n(x) = b_{k_3}(x)^\intercal \widehat{c}_3$ as estimators of the unknown functions β, g and m, which can be wrapped up in a vector

$$(\widehat{\beta}_n(r), \widehat{g}_n(z), \widehat{m}_n(x))^\intercal = \Psi(r, z, x)^\intercal \widehat{c}, \tag{2.8}$$

where $\Psi(r, z, x)$ is a block matrix given by

$$\Psi(r, z, x) = \begin{pmatrix} \phi_{k_1}(r) & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & a_{k_2}(z) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & b_{k_3}(x) \end{pmatrix} \tag{2.9}$$

in which $\mathbf{0}$'s are zero column vectors that have different dimensions over each row. We study the asymptotics of the estimators in the next section.

2.2.2. Estimation procedure for model (1.2)

In model (1.2) where the regressor z_t is replaced by a locally stationary process z_{nt}, the procedure of estimation remains the same, but notice that, $a_{k_2}(z_t)$ in B_{nk} in this case are replaced by $a_{k_2}(z_{nt})$, $t = 1, \dots, n$. Let $\widetilde{B}_{nk}$ be the counterpart of B_{nk} in the previous setting. Meanwhile, the estimator in (2.8) should be adjusted by using the coefficient vector $\widehat{c}$ calculated from $\widetilde{B}_{nk}$, as the model can be written as $y = \widetilde{B}_{nk}c + \widetilde{\gamma} + e$, where $\widetilde{\gamma} = (\widetilde{\gamma}(1), \dots, \widetilde{\gamma}(n))^\intercal$ with $\widetilde{\gamma}(t) = \gamma_{1k_1}(t/n) + \gamma_{2k_2}(z_{nt}) + \gamma_{3k_3}(x_t)$, $t = 1, \dots, n$. As a result, $\widehat{c} = (\widetilde{B}_{nk}^\intercal \widetilde{B}_{nk})^{-1}\widetilde{B}_{nk}^\intercal y$. The asymptotics of these estimators will be studied in the next section as well.

3. Asymptotic theory

We shall first study the asymptotics of the estimators defined in (2.8) for model (1.1). After this, the estimators for model (1.2) where z_t is replaced by a locally stationary process z_{nt} are investigated. Additionally, we also consider in the third subsection an extension of our model.

3.1. Estimators for model (1.1)

Note by Eq. (2.7) that $\widehat{c} - c = (B_{nk}^\intercal B_{nk})^{-1}B_{nk}^\intercal(\gamma + e)$. Thus, it is necessary to study first the asymptotics of $B_{nk}^\intercal B_{nk}$, which is done under the following assumptions and given by Lemma A.5.

Assumption C.

- C.1 The functions $\beta(\cdot)$, $g(\cdot)$ and $m(\cdot)$ are continuously differentiable up to s_1, s_2 and s_3, respectively. Moreover, $\beta^{(s_1)}(\cdot)$, $g^{(s_2)}(\cdot)$ and $m^{(s_3)}(\cdot)$ belong to the Hilbert spaces which contain the original functions, respectively.
- C.2 For $\beta(\cdot)$ function, let $\int_0^1 \beta(r)dr = 0$.

Since we need not only the convergence of all orthogonal expansions discussed before but also quicker rates for them, the smoothness of the unknown functions is necessary to guarantee a certain rate of the convergence. The concrete requirements on s_i will be shown below, combining with sample size as well as truncation parameters. Note that C.2 is an identification condition since in both the expansions of $\beta(\cdot)$ and $g(\cdot)$ there is constant term that could not be distinguished one from another in the following regression. Notice also that C.2 is sufficient as $m(\cdot)$ is integrable on $\mathbb{R}$.

Assumption D. All k_i, $i = 1, 2, 3$, diverge with n such that:

- D.1 If B.2(a) holds, (1) $k_2^{2+2/(2+\delta)} = o(n)$, $k_3^5 = o(n)$, (2) $k_1 k_2^{1+1/(2+\delta)} = o(n)$, $k_1^2 k_3^3 = o(n)$, $k_2^2 k_3^{3/2} = o(n)$; if B.2(b) holds, (3) $k_2^2 = o(n)$, $k_3^5 = o(n)$, (4) $k_1 k_2 = o(n)$, $k_1^2 k_3^3 = o(n)$, $k_2^3 k_3^3 = o(n)$.
- D.2 Suppose that as $n \to \infty$, (5) $n k_1^{-(2s_1-1)} = o(1)$, $n k_2^{-(s_2-1)} = o(1)$ and $n^{1/2} k_3^{-(s_3-1)} = o(1)$ and (6) $n k_2 k_1^{-2s_1} = o(1)$, $n k_3 k_1^{-2s_1} = o(1)$, $n k_1 k_2^{-s_2} = o(1)$, $n k_3 k_2^{-s_2} = o(1)$, $n^{1/2} k_1 k_3^{-s_3} = o(1)$, $n^{1/2} k_2 k_3^{-s_3} = o(1)$.

This assumption imposes the divergence rates for k_i, $i = 1, 2, 3$, which guarantee the convergence of the estimators. Because of the divergence of the moment of $p_j(z_1)$ with j in B.2(a), the requirement for k_2 in (1) and (2) is harsher than its counterpart in (3) and (4). Due to the nonstationarity of x_t, k_3 diverges very slowly, the rate of which is similar to the related study purely on integrated time series, see, for example, Dong et al. (2016). Anyway, if we simply take $k_i = \tilde{k}$ for $i = 1, 2, 3$, then $\tilde{k}^6 = o(n)$ is a concise condition.

Additionally, note that the conditions in (2) and (4) are for two of k_i's, while (1) and (3) are for each of k_2 and k_3. This is due to the structure of $B_{nk}^\intercal B_{nk} := (\Pi_{ij})_{3\times 3}$ a block symmetric matrix. Note also that the conditions in (2) are made for the blocks like $\Pi_{12} = \sum_{t=1}^{n}\phi_{k_1}(t/n)a_{k_2}(z_t)^{\intercal}$ under B.2(a), whereas that in (4) are made the same blocks but under B.2(b). More importantly, k_1 is not included in (1) and (3). This is because $\Pi_{11} := \sum_{t=1}^{n}\phi_{k_1}(t/n)\phi_{k_1}(t/n)^{\intercal}$ is convergent so fast that the condition derived from Π_{11} is substituted by the slower ones that are derived from Π_{12} and Π_{13}.

Given the smoothness of the unknown functions in Condition C.1, Condition D.2 demands that the smoothness orders be large enough such that the residues after truncation ($\gamma_{ik_i}, i = 1, 2, 3$) converging to zero rapidly enough and do not affect the convergence of the estimators. This can be understood as an undersmoothing condition (see Comment 4.3 in Belloni et al. (2015, p. 352)). The combination of the requirements in Assumption D for k_i implies that we have a minimum demand on the smoothness. We here emphasize that all requirements on k_i are compatible. For example, in an extreme case that $k_i = [n^{\tau}]$ for all $i = 1, 2, 3$ with $0 < \tau < 1/5$, along with $s_1 \ge 3, s_2 \ge 6$ and $s_3 \ge 4$, Assumption D is fulfilled.

Before showing the large sample theory for the estimators, we introduce some notation and preliminary results. Let $D_n = \mathrm{diag}(\sqrt{n}I_{k_1}, \sqrt{n}I_{k_2}, \sqrt{n/d_n}I_{k_3})$ a diagonal matrix of $k\times k\,(k = k_1+k_2+k_3)$. Then, as shown in Lemma A.5, $D_n^{-1}B_{nk}^{\intercal}B_{nk}D_n^{-1}$ is asymptotically approximated by a matrix U_k in probability, viz., $\|D_n^{-1}B_{nk}^{\intercal}B_{nk}D_n^{-1} - U_k\| = o_P(1)$ as $n \to \infty$ on a richer probability space. Here, $U_k = \mathrm{diag}(I_{k_1}, U_{k_2}, L_W(1, 0)I_{k_3})$ where $L_W(1, 0)$ is the local time of $W(r)$ given in Section 2, and $U_{k_2} = \mathbb{E}[a_{k_2}(z_1)a_{k_2}(z_1)^{\intercal}]$.

In addition, in order to tackle the heteroskedasticity we also need to consider the limit of the conditional covariance matrix $B_{nk}^{\intercal}\Sigma_n B_{nk}$ where $\Sigma_n = \mathrm{diag}(\sigma^2(1/n), \sigma^2(2/n), \ldots, \sigma^2(1))$. Note that $\|D_n^{-1}B_{nk}^{\intercal}\Sigma_n B_{nk}D_n^{-1} - V_k\| = o_P(1)$ where $V_k = \mathrm{diag}\left(V_*, \int_0^1 \sigma^2(r)dL_W(r, 0)I_{k_3}\right)$ in which $V_* = (V_{*ij})$ is a 2×2 symmetric block matrix with

$$V_{*11} = \int_0^1 \phi_{k_1}(r)\phi_{k_1}(r)^{\intercal}\sigma^2(r)dr,$$

$$V_{*12} = \int_0^1 \phi_{k_1}(r)\sigma^2(r)dr\,\mathbb{E}(a_{k_2}(z_1)^{\intercal}),$$

$$V_{*22} = \int_0^1 \sigma^2(r)dr\,\mathbb{E}(a_{k_2}(z_1)a_{k_2}(z_1)^{\intercal}).$$

This is given by Lemma A.7. In the homoskedastic case, $V_k = \sigma^2 U_k$, where $\sigma^2(\cdot) \equiv \sigma^2$. To show the following theorem, denote by $\overline{\Psi}(r, z, x)$ the normalized version of $\Psi(r, z, x)$ defined in Section 2, i.e. post-multiplying $\mathrm{diag}(\|\phi_{k_1}(r)\|, \|a_{k_2}(z)\|, \|b_{k_3}(x)\|)^{-1}$ to $\Psi(r, z, x)$ such that all block vectors in $\overline{\Psi}(r, z, x)$ are unit, $\bar{U}_k = \mathrm{diag}(I_{k_1}, U_{k_2}, I_{k_3})$ and $\bar{V}_k = \mathrm{diag}(V_*, I_{k_3})$.

Theorem 3.1. *Suppose that uniformly over all n, all eigenvalues of U_{k_2} and V_* are bounded below from zero and above from infinity, and that Assumptions A–D hold. Then, for any $r \in [0, 1]$, $z \in V$ and $x \in \mathbb{R}$,*

$$\Omega_n^{-1/2}\begin{pmatrix} \dfrac{\sqrt{n}}{\|\phi_{k_1}(r)\|}[\widehat{\beta}_n(r) - \beta(r)] \\ \dfrac{\sqrt{n}}{\|a_{k_2}(z)\|}[\widehat{g}_n(z) - g(z)] \\ \sqrt{\dfrac{n}{d_n}}\dfrac{1}{\|b_{k_3}(x)\|}[\widehat{m}_n(x) - m(x)] \end{pmatrix} \to_D N\left(\mathbf{0}, \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & a^2 \end{pmatrix}\right) \tag{3.1}$$

as $n \to \infty$ where $\mathbf{0}$ is a 3-dimensional zero column vector, $a^2 := L_W^{-2}(1, 0)\int_0^1 \sigma^2(r)dL_W(r, 0)$ and $\Omega_n := \overline{\Psi}(r, z, x)^{\intercal}\bar{U}_k^{-1}\bar{V}_k\bar{U}_k^{-1}\overline{\Psi}(r, z, x)$ is a 3×3 deterministic matrix.

The proof is relegated to Appendix B below. Here, the estimator has a mixed normal limiting distribution. As argued in Park and Phillips (2001, p. 122), the random variable a is independent of the underling normal distribution due to the integrability of $m(\cdot)$. This applies to the following theorems too.

The boundedness of all eigenvalues of the deterministic matrices U_{k_2} and V_* is a commonly used assumption in the literature. See, Condition A.2 in Belloni et al. (2015, p. 347) and Assumptions 1.3 and 1.4 in Hansen (2015) among others. Here, $U_{k_2} = \mathbb{E}[a_{k_2}(z_1)a_{k_2}(z_1)^{\intercal}]$ and V_* is formed in the same way but from one deterministic basis functions and another basis functions of variable z_t. This condition, along with the block diagonal structure of U_k and V_k containing the local time $L_W(1, 0)$, is sufficient in the derivation of the normality in the theorem. This is because $L_W(1, 0) = O_P(1)$ in the sense that, for any $\epsilon > 0$, there exists a constant $M > 0$ such that $P(M^{-1} \le L_W(1, 0) \le M) \ge 1 - \epsilon$ (so is $L_W^{-1}(1, 0) = O_P(1)$). This is easy to be verified by virtue of the distribution function of $L_W(1, 0)$, viz., $2\Phi(x) - 1$ with $\Phi(x)$ being the standard normal distribution.

In the homoskedastic case $V_k = \sigma^2 U_k$, two requirements on U_{k_2} and V_* are reduced to that about U_{k_2} and researchers often normalize U_{k_2} to be the identity matrix. See, for example, Eq. 11 of Chen and Christensen (2015, p. 450) and the normalization of Belloni et al. (2015, p. 347).

Note that the matrix Ω_n has a diagonal block form

$$\Omega_n = \mathrm{diag}\left(\overline{\Psi}_{12}(r, z)^{\intercal}U_*^{-1}V_*U_*^{-1}\overline{\Psi}_{12}(r, z), 1\right),$$

where we denote by $\overline{\Psi}_{12}(r,z)$ the left-top 2×2 sub-matrix of $\overline{\Psi}(r,z,x)$ defined right before Theorem 3.1 and $U_*:=\mathrm{diag}(I_{k_1},U_{k_2})$. This reveals some crucial asymptotic behaviors for the variables. Due to the divergence of the I(1) process x_t, all interactions between $m(x_t)$ and each one of $\beta(t/n)$ and $g(z_t)$ with proper normalization are asymptotically negligible and thence Ω_n has the above diagonal block form. The details can be found in Lemmas A.5 and A.7 below.

Therefore, we may separate the estimator $\widehat{m}_n(x)$ from the other estimators in (3.1). That is, as $n\to\infty$,

$$[\overline{\Psi}_{12}(r,z)^{\intercal}U_*^{-1}V_*U_*^{-1}\overline{\Psi}_1(r,z)]^{-1/2}\begin{pmatrix}\dfrac{\sqrt{n}}{\|\phi_{k_1}(r)\|}[\widehat{\beta}_n(r)-\beta(r)]\\ \dfrac{\sqrt{n}}{\|a_{k_2}(z)\|}[\widehat{g}_n(z)-g(z)]\end{pmatrix}\to_D N(0,I_2) \tag{3.2}$$

$$\sqrt{\frac{n}{d_n}}\frac{1}{\|b_{k_3}(x)\|}(\widehat{m}_n(x)-m(x))\to_D N(0,a^2). \tag{3.3}$$

They are all comparable with the literature in the corresponding context. To see this, observe that $\overline{\Psi}_{12}(r,z)^{\intercal}U_*^{-1}V_*U_*^{-1}\overline{\Psi}_1(r,z)$ has eigenvalues bounded below from zero and above from infinity due to the condition on U_{k_2} and V_*. Then, the rates of (3.2) are $[\sqrt{n}/\|\phi_{k_1}(r)\|]^{-1}$ and $[\sqrt{n}/\|a_{k_2}(z)\|]^{-1}$ for $\widehat{\beta}_n(r)-\beta(r)$ and $\widehat{g}_n(z)-g(z)$, respectively, the same as the estimators in Theorem 2 of Newey (1997) and Theorem 3.1 of Chen and Christensen (2015) in the case that the functional of the estimator in the papers is identical.

On the other hand, the rate in (3.3) is about $n^{-1/4}k_3$, very slow due to the divergence of x_t and the integrability of $m(x)$. This is the same as that in Theorem 3.3 of Dong et al. (2016). Overall, although the additive model has the mixture of deterministic trend, nonparametric function of stationary variable and nonparametric integrable function of the unit root variable, the estimators have their own separable rate of convergence.

Note that the matrices U_* and V_* could be further simplified in the special case that the function sequence $\{p_j(x)\}$ is orthogonal with respect to the density of z_1 (i.e., $dF(x)$ in the space $L^2(V,dF(x))$ is the density of z_1). Hence, $\mathbb{E}(a_{k_2}(z_1))=0$ and $\mathbb{E}(a_{k_2}(z_1)a_{k_2}(z_1)^{\intercal})=I_{k_2}$. Particularly, when $\sigma^2(\cdot)\equiv\sigma^2$, $V_*=\sigma^2 I_{k_1+k_2}$ and $U_*=I_{k_1+k_2}$. Therefore, the statement about the limits for $\widehat{\beta}_n(r)-\beta(r)$ and $\widehat{g}_n(z)-g(z)$ in (3.2) would be simplified too.

More importantly, the conventional optimal convergence rates for $\|\widehat{\beta}_n(r)-\beta(r)\|$ and $\|\widehat{g}_n(z)-g(z)\|$ can be jointly established where $\|\cdot\|$ stands for the norm of functions in different spaces defined in Section 2. Here, the conventional optimal rates are in the sense studied in Stone (1982, 1985).

Proposition 3.1. *Suppose that Assumptions A–D hold. In the model (1.1) we have jointly* $\|\widehat{\beta}_n(r)-\beta(r)\|=O_P(\sqrt{k_1/n}+k_1^{-s_1})$, $\|\widehat{g}_n(z)-g(z)\|=O_P(\sqrt{k_2/n}+k_2^{-s_2})$ *and* $\|\widehat{m}_n(x)-m(x)\|=O_P(\sqrt{k_3/\sqrt[4]{n}}+k_3^{-s_3/2})$ *as* $n\to\infty$, *where the norms are of* L_2 *sense in the function spaces, respectively.*

The proposition implies that the optimal rates of Stone (1982, 1985) are attainable jointly for the estimators $\widehat{\beta}_n(r)$ and $\widehat{g}_n(z)$. Indeed, if $k_i=O(n^{1/(2s_i+1)})$, the rates will be $O_P(n^{-s_i/(2s_i+1)})$, $i=1,2$, which are exactly the optimal rates in Stone (1982, 1985). Note also that in the literature as far as we know, there is no study dwelling on the optimal rates with respect to unit root regressor. While Newey (1997) and Chen and Christensen (2015, p.451) obtain optimal rates for sieve estimator in some situations, Proposition 3.1 establishes the optimal rates jointly for two nonparametric functions in an additive model.

In order to make statistical inference, there is a need to estimate the function $\sigma^2(\cdot)$. Though the estimation is possible by nonparametric method using the estimated residues, the main purpose of the paper would be deviated if we were about to do so. In what follows, we focus on the inference in a simpler case, the case of homoskedasticity. It can be seen from (3.2)–(3.3) that $V_*=\sigma^2 U_*$ and we need to estimate σ^2 and $nL_W(1,0)/d_n$ because of $\int_0^1\sigma^2(r)dL_W(r,0)=\sigma^2 L_W(1,0)$. Here, as an unknown parameter in d_n, viz., ψ, can be offset from the estimate of $L_W(1,0)$, we simply estimate the quantity $nL_W(1,0)/d_n$ directly. Define

$$\widehat{\sigma}=\left[\frac{1}{n}\sum_{t=1}^{n}(y_t-\widehat{\beta}_n(t/n)-\widehat{g}_n(z_t)-\widehat{m}_n(x_t))^2\right]^{1/2},$$

$$\Lambda_n=\sum_{t=1}^{n}f(x_t),\qquad \text{where } f(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}.$$

We then have the following corollary.

Corollary 3.1. *Suppose that Assumptions A–D hold. Then,* $\widehat{\sigma}\to_P\sigma$ *and* $\Lambda_n/(nL_W(1,0)/d_n)\to_P 1$ *as* $n\to\infty$. *As a result, with the replacement of* σ *by* $\widehat{\sigma}$ *and* $nL_W(1,0)/d_n$ *by* Λ_n, *we have,* $n\to\infty$,

$$[\widehat{\sigma}^2\overline{\Psi}_{12}(r,z)^{\intercal}U_*^{-1}\overline{\Psi}_1(r,z)]^{-1/2}\begin{pmatrix}\dfrac{\sqrt{n}}{\|\phi_{k_1}(r)\|}[\widehat{\beta}_n(r)-\beta(r)]\\ \dfrac{\sqrt{n}}{\|a_{k_2}(z)\|}[\widehat{g}_n(z)-g(z)]\end{pmatrix}\to_D N(0,I_2) \tag{3.4}$$

$$\sqrt{\Lambda_n}\frac{1}{\widehat{\sigma}\|b_{k_3}(x)\|}(\widehat{m}_n(x)-m(x))\to_D N(0,1). \tag{3.5}$$

3.2. Estimators for model (1.2)

In this case we have $\widehat{c}-c=(\widetilde{B}_{nk}^\intercal\widetilde{B}_{nk})^{-1}\widetilde{B}_{nk}^\intercal(\widetilde{\gamma}+e)$. The asymptotics of $\widetilde{B}_{nk}^\intercal\widetilde{B}_{nk}$ is given by Lemma A.6. Note that $\widetilde{B}_{nk}$ is the same as B_{nk} but the stationary process z_t is replaced by the locally stationary process z_{nt}. The replacement only affects Π_{12} (Π_{21}), Π_{23} (Π_{32}) and Π_{22}, denoted respectively by $\widetilde{\Pi}_{12}$, $\widetilde{\Pi}_{23}$ and $\widetilde{\Pi}_{22}$ the resulting counterparts. Precisely, $\widetilde{\Pi}_{12}=\sum_{t=1}^n\phi_{k_1}(t/n)a_{k_2}(z_{nt})^\intercal$, $\widetilde{\Pi}_{22}=\sum_{t=1}^n a_{k_2}(z_{nt})a_{k_2}(z_{nt})^\intercal$, and $\widetilde{\Pi}_{23}=\sum_{t=1}^n a_{k_2}(z_{nt})b_{k_3}(x_t)^\intercal$.

Define $\widetilde{U}_k=\mathrm{diag}(\widetilde{U}_*, L_W(1,0)I_{k_3})$, where $\widetilde{U}_*=(\widetilde{U}_{*ij})$ is a symmetric 2×2 block matrix of order $(k_1+k_2)\times(k_1+k_2)$ with $\widetilde{U}_{*11}=I_{k_1}$, $\widetilde{U}_{*12}=\int_0^1\phi_{k_1}(r)\mathbb{E}[a_{k_2}(z_1(r))^\intercal]dr$ with elements $\int_0^1\varphi_i(r)\mathbb{E}[p_j(z_1(r))]dr$ for $1\le i\le k_1, 0\le j\le k_2-1$, and $\widetilde{U}_{*22}=\int_0^1\mathbb{E}[a_{k_2}(z_1(r))a_{k_2}(z_1(r))^\intercal]dr$ with elements $\int_0^1\mathbb{E}[p_i(z_1(r))p_j(z_1(r))]dr$ for $i,j=0,\ldots,k_2-1$. As shown in Lemma A.6, under certain condition we have $\|D_n^{-1}\widetilde{B}_{nk}^\intercal\widetilde{B}_{nk}D_n^{-1}-\widetilde{U}_k\|=o_P(1)$ where D_n is the same as before.

Meanwhile, due to the heteroskedasticity, we also consider the limit of $\widetilde{B}_{nk}^\intercal\Sigma_n\widetilde{B}_{nk}$ where Σ_n is the same as in the preceding section. The result is given by Lemma A.8, that is, $\|D_n^{-1}\widetilde{B}_{nk}^\intercal\Sigma_n\widetilde{B}_{nk}D_n^{-1}-\widetilde{V}_k\|=o_P(1)$, where $\widetilde{V}_k=\mathrm{diag}\left(\widetilde{V}_*, \int_0^1\sigma^2(r)dL_W(r,0)I_{k_3}\right)$ in which $\widetilde{V}_*=(\widetilde{V}_{*ij})$ is a 2×2 symmetric block matrix with $\widetilde{V}_{*11}=V_{*11}$, $\widetilde{V}_{*12}=\int_0^1\phi_{k_1}(r)\sigma^2(r)\mathbb{E}(a_{k_2}(z_1(r))^\intercal)dr$ and $\widetilde{V}_{*22}=\int_0^1\sigma^2(r)\mathbb{E}(a_{k_2}(z_1(r))a_{k_2}(z_1(r))^\intercal)dr$.

Denote $\widetilde{\Omega}_n=\mathrm{diag}(\overline{\Psi}_{12}(r,z)^\intercal\widetilde{U}_*^{-1}\widetilde{V}_*\widetilde{U}_*^{-1}\overline{\Psi}_{12}(r,z), 1)$ a deterministic matrix of 3×3 with the same $\overline{\Psi}_{12}(r,z)$ as before. We then have the following theorem.

Theorem 3.2. *Suppose that uniformly over all n, all eigenvalues of $\widetilde{U}_*$ and $\widetilde{V}_*$ are bounded below from zero and above from infinity, and that Assumptions* A, B*, C *and* D *hold. Then, for any $r\in[0,1]$, $z\in V$ and $x\in\mathbb{R}$, the estimators of the unknown functions in model* (1.2) *obey*

$$\widetilde{\Omega}_n^{-1}\begin{pmatrix}\dfrac{\sqrt{n}}{\|\phi_{k_1}(r)\|}[\widehat{\beta}_n(r)-\beta(r)]\\ \dfrac{\sqrt{n}}{\|a_{k_2}(z)\|}[\widehat{g}_n(z)-g(z)]\\ \sqrt{\dfrac{n}{d_n}}\dfrac{1}{\|b_{k_3}(x)\|}[\widehat{m}_n(x)-m(x)]\end{pmatrix}\to_D N\left(\mathbf{0},\begin{pmatrix}1&0&0\\0&1&0\\0&0&a^2\end{pmatrix}\right) \tag{3.6}$$

as $n\to\infty$ where **0** *is a 3-dimensional zero column vector and a^2 is the same as in the previous theorem.*

The proof is relegated to Appendix B below. The main contribution of the theorem is the relaxation of the stationary process in model (1.1) to the locally stationary process in model (1.2). It is readily seen that if the distribution of the associated process $z_t(v)$ does not depend on v, implying that $\mathbb{E}[p_j(z_1(r))]=\mathbb{E}[p_j(z_1)]$, $\widetilde{U}_k$ would reduce to U_k and $\widetilde{V}_k$ would reduce to V_k. Consequently, in this degenerated case $\widetilde{\Omega}_n=\Omega_n$ and essentially model (1.2) would reduce to model (1.1).

We have similar comments for Theorem 3.2 as that for Theorem 3.1. In particular, the condition on the eigenvalues of the deterministic matrices $\widetilde{U}_*$ and $\widetilde{V}_*$ is often encountered in the sieve literature such as Condition A.2 in Belloni et al. (2015, p. 347). For the statistical inference purpose, under homoskedasticity the unknown parameter in (3.6) may be estimated similar to Corollary 3.1, which is omitted for brevity.

3.3. Extension of model (1.1)

Since the function $m(\cdot)$ is integrable on $\mathbb{R}$, model (1.1) is impossible to have any polynomial form of the regressor x_t. This possibly is a restriction in some situations. Thus, it is worth to extend model (1.1) to be

$$y_t=\beta(t/n)+g(z_t)+\theta_0x_t+m(x_t)+e_t, \tag{3.7}$$

where $t=1,\ldots,n$, β, g and m are unknown smooth functions and θ_0 is an unknown scalar, z_t, x_t and e_t are the same as before. It can be seen later that the linear form of x_t may be replaced by any polynomial form $\theta_{01}x_t+\cdots+\theta_{0d}x_t^d$ with d being known and a similar result remains true.

To estimate $\beta(\cdot)$, $g(\cdot)$ and $m(\cdot)$, the same bases are used for their orthogonal expansions. Notice that θ_0 can be estimated along with the estimate of the coefficients in the expansions and this can be viewed as an advantage of the series method because it parameterizes the nonparametric variables. Using previous notation model (3.7) is written as

$$\begin{aligned}y_t=&\phi_{k_1}(t/n)^\intercal c_1+a_{k_2}(z_t)^\intercal c_2+\theta_0x_t+b_{k_3}(x_t)^\intercal c_3\\&+\gamma_{1k_1}(t/n)+\gamma_{2k_2}(z_t)+\gamma_{3k_3}(x_t)+e_t,\end{aligned} \tag{3.8}$$

and we define

$$A_{nk}=\begin{pmatrix}\phi_{k_1}(1/n)^{\intercal} & x_1 & a_{k_2}(z_1)^{\intercal} & b_{k_3}(x_1)^{\intercal}\\ \vdots & \vdots & \vdots & \vdots\\ \phi_{k_1}(1)^{\intercal} & x_n & a_{k_2}(z_n)^{\intercal} & b_{k_3}(x_n)^{\intercal}\end{pmatrix}$$

a $n\times k$ matrix with $k=k_1+k_2+k_3+1$ for convenience. Consequently, we have

$$y=A_{nk}c+\gamma+e \tag{3.9}$$

which by the ordinary least squares (OLS) gives $\widehat{c}=(\widehat{c}_1^{\intercal},\widehat{\theta},\widehat{c}_2^{\intercal},\widehat{c}_3^{\intercal})^{\intercal}=(A_{nk}^{\intercal}A_{nk})^{-1}A_{nk}^{\intercal}y$ provided that $A_{nk}^{\intercal}A_{nk}$ is non-singular (that is true with high probability).

Similarly, for any $r\in[0,1]$, $z\in V$ and $x\in\mathbb{R}$ define $\widehat{\beta}_n(r)=\phi_{k_1}(r)^{\intercal}\widehat{c}_1$, $\widehat{g}_n(z)=a_{k_2}(z)^{\intercal}\widehat{c}_2$ and $\widehat{m}_n(x)=b_{k_3}(x)^{\intercal}\widehat{c}_3$ as estimators of the unknown functions, which together with the estimator of θ_0 can be wrapped up in a vector

$$(\widehat{\beta}_n(r),\widehat{\theta},\widehat{g}_n(z),\widehat{m}_n(x))^{\intercal}=\Phi(r,z,x)^{\intercal}\widehat{c}, \tag{3.10}$$

where $\Phi(r,z,x)$ is a block matrix given by

$$\Phi(r,z,x)=\begin{pmatrix}\phi_{k_1}(r) & \mathbf{0} & \mathbf{0} & \mathbf{0}\\ 0 & 1 & 0 & 0\\ \mathbf{0} & \mathbf{0} & a_{k_2}(z) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & b_{k_3}(x)\end{pmatrix}$$

in which **0**'s are zero column vectors that have different dimension over each row while 0's are scalar.

As before, we introduce first some notation and preliminary results. Let $M_n=\mathrm{diag}(\sqrt{n}I_{k_1},\sqrt{n}d_n,\sqrt{n}I_{k_2},\sqrt{n/d_n}I_{k_3})$ a diagonal matrix of $k\times k$. Then, $M_n^{-1}A_{nk}^{\intercal}A_{nk}M_n^{-1}$ is asymptotically approximated by a matrix in probability, viz., $\|M_n^{-1}A_{nk}^{\intercal}A_{nk}M_n^{-1}-Q_k\|=o_P(1)$ as $n\to\infty$ as shown in Lemma A.9. Here $Q_k=\mathrm{diag}(Q_*,L_W(1,0)I_{k_3})$ and Q_* has a 3×3 block form (Q_{*ij}): $Q_{*11}=I_{k_1}$, $Q_{*12}=\int_0^1\phi_{k_1}(r)W(r)dr$ of $k_1\times 1$, $Q_{*13}=\mathbf{0}$ of $k_1\times k_2$, $Q_{*22}=\int_0^1 W^2(r)dr$ a scalar, $Q_{*23}=\mathbb{E}[a_{k_2}(z_1)^{\intercal}]\int_0^1 W(r)dr$ of $k_2\times 1$ and $Q_{*33}=\mathbb{E}[a_{k_2}(z_1)a_{k_2}(z_1)^{\intercal}]$.

In addition, in order to tackle the heteroskedasticity we also need to consider the limit of the conditional covariance matrix $A_{nk}^{\intercal}\Sigma_nA_{nk}$ where $\Sigma_n=\mathrm{diag}(\sigma^2(1/n),\ldots,\sigma^2(1))$. By Lemma A.11, $\|M_n^{-1}A_{nk}^{\intercal}\Sigma_nA_{nk}M_n^{-1}-P_k\|=o_P(1)$ where $P_k=\mathrm{diag}\left(P_*,\int_0^1\sigma^2(r)dL_W(r,0)I_{k_3}\right)$ in which $P_*=(P_{*ij})$ is a 3×3 symmetric block matrix with

$$P_{*11}=\int_0^1\phi_{k_1}(r)\phi_{k_1}(r)^{\intercal}\sigma^2(r)dr,\qquad P_{*13}=\int_0^1\phi_{k_1}(r)\sigma^2(r)dr\mathbb{E}(a_{k_2}(z_1)^{\intercal}),$$

$$P_{*12}=\int_0^1\phi_{k_1}(r)\sigma^2(r)W(r)dr,\qquad P_{*22}=\int_0^1\sigma^2(r)W^2(r)dr,$$

$$P_{*23}=\int_0^1\sigma^2(r)W(r)dr\mathbb{E}(a_{k_2}(z_1)^{\intercal}),\qquad P_{*33}=\int_0^1\sigma^2(r)dr\mathbb{E}(a_{k_2}(z_1)a_{k_2}(z_1)^{\intercal}).$$

Once the model reduces to the case of homoskedasticity, $P_k=\sigma^2Q_k$ where $\sigma^2(\cdot)\equiv\sigma^2$, as expected.

Denote $\Xi_n=\overline{\Phi}(r,z,x)^{\intercal}\bar{Q}_k^{-1}\bar{P}_k\bar{Q}_k^{-1}\overline{\Phi}(r,z,x)$ a matrix of 4-by-4, where $\overline{\Phi}$ is the normalized version of Φ, i.e. the $\phi_{k_1}(r)$, $a_{k_2}(z)$ and $b_{k_3}(x)$ in Φ are replaced by the $\phi_{k_1}(r)/\|\phi_{k_1}(r)\|$, $a_{k_2}(z)/\|a_{k_2}(z)\|$ and $b_{k_3}(x)/\|b_{k_3}(x)\|$, respectively; $\bar{Q}_k=\mathrm{diag}(Q_*,I_{k_3})$ and $\bar{P}_k=\mathrm{diag}\left(P_*,I_{k_3}\right)$. Hence, $\Xi_n=\mathrm{diag}(\Xi_{1n},1)$ where Ξ_{1n} is of 3-by-3 and $\Xi_{1n}=\overline{\Phi}_{13}(r,z)^{\intercal}Q_*^{-1}P_*Q_*^{-1}\overline{\Phi}_{13}(r,z)$ where $\overline{\Phi}_{13}(r,z)$ is the left-top 3-by-3 block submatrix of $\overline{\Phi}(r,z,x)$.

Note that the Brownian motion $W(r)$ is contained in Q_* and P_*, we thus need to strengthen the conditions on e_t in Assumptions B and B*.

Assumption E. The limit Brownian motion $W(r)$ derived from x_t is independent of $\{e_t,t\geq 1\}$.

This assumption would facilitate the establishment of the following asymptotic normality for our estimators. The condition can be fulfilled if $\{\epsilon_j\}$ in Assumption A is independent of $\{e_t\}$ and x_t is substituted by $x_t'=x_t+f(e_{t-1})$ for some measurable function $f(\cdot)$. Notice that, x_t' still has limit $W(r)$, $d_n^{-1}x'_{[nr]}\to_D W(r)$, as long as $E|f(e_t)|<\infty$ and therefore Assumption E is satisfied. A stronger one to replace Assumption E is that x_t is independent of e_s for all t and s.

Theorem 3.3. *In addition to Assumptions A–E, suppose that uniformly over all n, all eigenvalues of Q_* and P_* are bounded below from zero and above from infinity almost surely, and $\Xi_{1n}\to_P\Xi_1$ when $n\to\infty$. Then, for any $r\in[0,1]$, $z\in V$ and $x\in\mathbb{R}$,*

$$\begin{pmatrix} \frac{\sqrt{n}}{\|\phi_{k_1}(r)\|}[\widehat{\beta}_n(r)-\beta(r)] \\ \sqrt{n}d_n[\widehat{\theta}-\theta_0] \\ \frac{\sqrt{n}}{\|a_{k_2}(z)\|}[\widehat{g}_n(z)-g(z)] \\ \frac{\sqrt{n}}{\|b_{k_3}(x)\|\sqrt{d_n}}[\widehat{m}_n(x)-m(x)] \end{pmatrix} \to_D N\left(\mathbf{0}, \begin{pmatrix} \Xi_1 & \\ & a^2 \end{pmatrix}\right) \tag{3.11}$$

as $n \to \infty$ where $\mathbf{0}$ is a 4-dimensional zero column vector, and a^2 is the same as in the previous theorem.

The proof is relegated to Appendix D in the supplementary material. We have similar comment as that for Theorem 3.1. Note that the covariance matrix in the limit has a diagonal block form diag(Ξ_1, a^2). This is similar to but more than the situation in Theorem 3.1. First, all interactions between $m(x_t)$ and one of $\beta(t/n)$, $g(z_t)$ and x_t with proper normalization are asymptotically negligible and thence the covariance has the above diagonal block form; second, interestingly, the interactions between x_t and each of $\beta(t/n)$ and $g(z_t)$ with the same normalization last ultimately, and thereby the block in Ξ_1 is a square matrix of order 3 that in general cannot be reduced further. The details can be found in Lemmas A.9 and A.10 below.

Therefore, we may isolate the estimator $\widehat{m}_n(x)$ from the other estimators in (3.11). That is, as $n \to \infty$,

$$\begin{pmatrix} \frac{\sqrt{n}}{\|\phi_{k_1}(r)\|}[\widehat{\beta}_n(r)-\beta(r)] \\ \sqrt{n}d_n[\widehat{\theta}-\theta_0] \\ \frac{\sqrt{n}}{\|a_{k_2}(z)\|}[\widehat{g}_n(z)-g(z)] \end{pmatrix} \to_D N(0, \Xi_1) \tag{3.12}$$

$$\frac{\sqrt{n}}{\|b_{k_3}(x)\|\sqrt{d_n}}(\widehat{m}_n(x)-m(x)) \to_D N(0, a^2). \tag{3.13}$$

Here, (3.13) is exactly the same as (3.3), meaning that the estimate of $m(\cdot)$ is not affected by the linear form of x_t at all, while since $W(r)$ is involved in Ξ_{1n}, the other estimators are affected more or less. All function estimators have the same order as before, whereas $\widehat{\theta}-\theta_0$ has a super rate $O_P(n^{-1})$ in view of $d_n \sim \sqrt{n}$. The normalizer $\sqrt{n}d_n$ in the front of $\widehat{\theta}-\theta_0$ has an extra d_n comparing with the usual stationary regression. That is due to the convergence of $d_n^{-1}x_{[nr]} \to_D W(r)$ and thus results in the super rate. Because the linear form $\theta_0 x_t$ is one particular kind of H-regular function defined in Park and Phillips (2001), the order of $\widehat{\theta}-\theta_0$ is comparable with its counterpart in Theorem 7 of Chang et al. (2001, p. 13). Overall, the estimators in this additive model, where each component is different dramatically in terms of regressors and functions, have their own separable rate of convergence.

Observe that both the matrices Q_* and P_* are almost surely positive definite by their structures. Note further that the matrices Q_* and P_* could be further simplified in the special case aforementioned, which gives $\mathbb{E}(a_{k_2}(z_1)) = 0$ and $\mathbb{E}(a_{k_2}(z_1)a_{k_2}(z_1)^\intercal) = I_{k_2}$. Particularly, when $\sigma^2(\cdot) \equiv \sigma^2$, $P_{*11} = \sigma^2 I_{k_1}$, $P_{*13} = 0$, $P_{*22} = \sigma^2 \int_0^1 W^2(r)dr$, $P_{*23} = 0$ and $P_{*33} = \sigma^2 I_{k_2}$, but normally $P_{*12} = \sigma^2 \int_0^1 \phi(r)W(r)dr \neq 0$. The same situation applies to Q_*. Therefore, Q_* and P_* are reduced to diagonal block matrices and thus the limit for $\widehat{g}_n(z)-g(z)$ in (3.12) can be isolated from the other two, that however cannot be broken up any more due to $P_{*12} \neq 0$.

The case that z_t in model (3.7) is replaced by z_{nt} is considered now, that is,

$$y_t = \beta(t/n) + g(z_{nt}) + \theta_0 x_t + m(x_t) + e_t, \tag{3.14}$$

where $t = 1, \ldots, n$.

With the same estimation procedure, in this case we have $\widehat{c} - c = (\widetilde{A}_{nk}^\intercal \widetilde{A}_{nk})^{-1}\widetilde{A}_{nk}^\intercal(\widetilde{\gamma} + e)$. Here, $\widetilde{A}_{nk}$ is the counterpart of A_{nk} with z_t substituted by z_{nt}. The asymptotics of $\widetilde{A}_{nk}^\intercal \widetilde{A}_{nk}$ is given by Lemma A.10.

Define $\widetilde{Q}_k = \mathrm{diag}(\widetilde{Q}_*, L_W(1,0)I_{k_3})$, where $\widetilde{Q}_* = (\widetilde{Q}_{*ij})$ is a symmetric 3×3 block matrix of order $(k_1+k_2+1)\times(k_1+k_2+1)$ with $\widetilde{Q}_{*11} = I_{k_1}$, $\widetilde{Q}_{*12} = \int_0^1 \phi_{k_1}(r)W(r)dr$, $\widetilde{Q}_{*13} = \int_0^1 \phi_{k_1}(r)\mathbb{E}[a_{k_2}(z_1(r))^\intercal]dr$ with elements $\int_0^1 \varphi_i(r)\mathbb{E}[p_j(z_1(r))]dr$ for $i = 1, \ldots, k_1$, $j = 0, \ldots, k_2-1$, $\widetilde{Q}_{*22} = \int_0^1 W^2(r)dr$ a scalar, $\widetilde{Q}_{*23} = \int_0^1 \mathbb{E}[a_{k_2}(z_1(r))^\intercal]W(r)dr$ and $\widetilde{Q}_{*33} = \int_0^1 \mathbb{E}[a_{k_2}(z_1(r))a_{k_2}(z_1(r))^\intercal]dr$ with elements $\int_0^1 \mathbb{E}[p_i(z_1(r))p_j(z_1(r))]dr$ for $i, j = 0, \ldots, k_2-1$. As shown in Lemma A.10, under certain condition we have $\|M_n^{-1}\widetilde{A}_{nk}^\intercal\widetilde{A}_{nk}M_n^{-1} - \widetilde{Q}_k\| = o_P(1)$ where M_n is the same as before.

Meanwhile, due to the heteroskedasticity, we also consider the limit of $\widetilde{A}_{nk}^\intercal \Sigma_n \widetilde{A}_{nk}$ where Σ_n is the same as in the preceding section. The result is given by Lemma A.12, that is, $\|M_n^{-1}\widetilde{A}_{nk}^\intercal \Sigma_n \widetilde{A}_{nk}M_n^{-1} - \widetilde{P}_k\| = o_P(1)$, where $\widetilde{P}_k = \mathrm{diag}\left(\widetilde{P}_*, \int_0^1 \sigma^2(r)dL_W(r,0)I_{k_3}\right)$ in which $\widetilde{P}_* = (\widetilde{P}_{*ij})$ is a 3×3 symmetric block matrix with $\widetilde{P}_{*11} = P_{*11}$, $\widetilde{P}_{*22} = P_{*22}$, $\widetilde{P}_{*12} = P_{*12}$, while $\widetilde{P}_{*13} = \int_0^1 \phi_{k_1}(r)\sigma^2(r)\mathbb{E}(a_{k_2}(z_1(r))^\intercal)dr$ and $\widetilde{P}_{*33} = \int_0^1 \sigma^2(r)\mathbb{E}(a_{k_2}(z_1(r))a_{k_2}(z_1(r))^\intercal)dr$ and $\widetilde{P}_{*23} = \int_0^1 \sigma^2(r)\mathbb{E}(a_{k_2}(z_1(r))^\intercal)W(r)dr$.

Define $\widetilde{\Xi}_{1n} = \overline{\Phi}_{13}(r,z)^{\intercal}\widetilde{Q}_*^{-1}\widetilde{P}_*\widetilde{Q}_*^{-1}\overline{\Phi}_{13}(r,z)$ an 3-by-3 matrix with $\overline{\Phi}_{13}(r,z)$ defined as before. We then have the following theorem.

Theorem 3.4. *In addition to Assumptions A, B*, C–E, suppose that uniformly over all n, all eigenvalues of $\widetilde{Q}_*$ and $\widetilde{P}_*$ are bounded below from zero and above from infinity, and $\widetilde{\Xi}_{1n}\to_P\widetilde{\Xi}_1$ as $n\to\infty$. Then, for any $r\in[0,1]$, $z\in V$ and $x\in\mathbb{R}$, the estimators for model (3.14) obey*

$$\begin{pmatrix} \dfrac{\sqrt{n}}{\|\phi_{k_1}(r)\|}[\widehat{\beta}_n(r)-\beta(r)] \\ \sqrt{n}d_n[\widehat{\theta}-\theta_0] \\ \dfrac{\sqrt{n}}{\|a_{k_2}(z)\|}[\widehat{g}_n(z)-g(z)] \\ \dfrac{\sqrt{n}}{\|b_{k_3}(x)\|\sqrt{d_n}}[\widehat{m}_n(x)-m(x)] \end{pmatrix} \to_D N\left(\mathbf{0},\begin{pmatrix}\widetilde{\Xi}_1 & \\ & a^2\end{pmatrix}\right) \tag{3.15}$$

as $n\to\infty$ where $\mathbf{0}$ is a 4-dimensional zero column vector and a^2 is the same as in the previous theorem.

The proof is relegated to Appendix D in the supplementary material. The main contribution of the theorem is the relaxation of the stationary process in model (3.7) to the locally stationary process in model (3.14). It is readily seen that if the distribution of the associated process $z_t(v)$ does not depend on v, implying that $\mathbb{E}[p_j(z_1(r))] = \mathbb{E}[p_j(z_1)]$, $\widetilde{Q}_k$ would reduce to Q_k and $\widetilde{P}_k$ would reduce to P_k. Consequently, in this degenerated case $\widetilde{\Xi}_n = \Xi_n$ and essentially model (3.14) would reduce to model (3.7).

We have similar comments for Theorem 3.4 as that for Theorem 3.3, which is omitted for brevity.

4. Simulation

In this section we conduct Monte Carlo simulation to investigate the performance of our estimators proposed in the last section in the finite sample situation. We mainly focus on model (1.1). Let $M = 1000$ be the number of replication and n the sample size.

Example 1. Let $z_t \sim i.i.U[-1,1]$ and $g(z) = z^2+\sin(z)$. The Chebyshev polynomials of the first kind, $p_j(x) = \cos(j\ \arccos(x))$, $j\geq 0$, are used to approximate the function $g(\cdot)$.

Suppose that $\epsilon_i \sim N(0,1)$, $w_t = \rho w_{t-1}+\epsilon_t$ with $\rho = 0.2$ and $w_0 \sim N(0, 1/(1-\rho^2))$. This is the theoretical distribution of w_0 in the AR(1) process. Let $x_0 = 0$, $x_t = x_{t-1}+w_t$, $t\geq 1$. Put $m(x) = 1/(1+x^4)$. The Hermite functions are used for the orthogonal expansion of $m(x)$.

Moreover, let $\beta(r) = r-1/2$ satisfying $\int_0^1\beta(r)dr = 0$. The cosine sequence given in Section 2 is utilized for the expansion of $\beta(r)$. Then, $y_t = \beta(t/n)+g(z_t)+m(x_t)+e_t$, $t = 1,\ldots,n$, where $e_t\sim N(0,1)$.

In the experiments below, the sample size is $n = 400{,}600$ and 1200, respectively, and the truncation parameters $k_1 = k_2 = 3, 4, 5$ for $\beta(\cdot)$ and $g(\cdot)$ and $k_3 = 3, 4, 6$ for $m(\cdot)$, corresponding to each sample size. This indicates the move of the truncation parameters with the sample size. It is noteworthy that, though in stationary case one may use the Generalized Cross Validation (GCV) (see, e.g., Gao, 2007) to determine the truncation parameter, similar approach is not available in nonstationary case.

After we obtain all estimators by the procedure described in Section 2, we shall calculate the bias (denoted by $B_\beta(n)$, $B_g(n)$ and $B_m(n)$), standard deviation (denoted by $\pi_\beta(n)$, $\pi_g(n)$ and $\pi_m(n)$) and RMSE (denoted by $\Pi_\beta(n)$, $\Pi_g(n)$ and $\Pi_m(n)$) of estimators, that is,

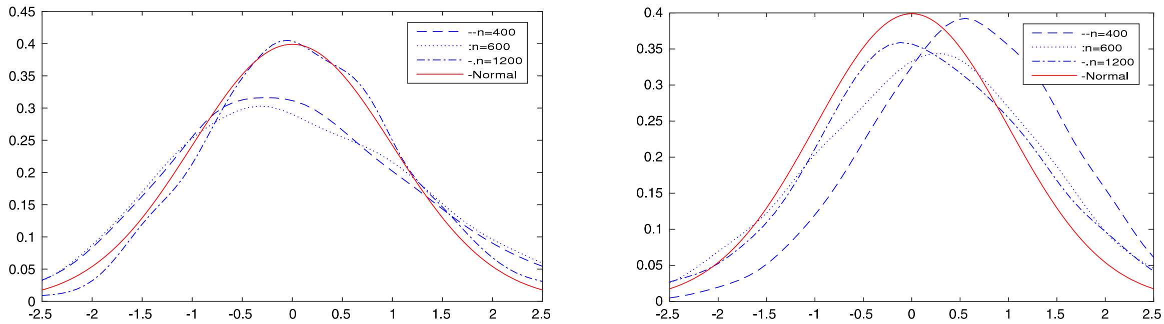
$$B_\beta(n) := \frac{1}{Mn}\sum_{t=1}^{n}\sum_{\ell=1}^{M}[\beta(t/n)-\overline{\widehat{\beta}}(t/n)], \qquad \pi_\beta(n) := \left(\frac{1}{Mn}\sum_{t=1}^{n}\sum_{\ell=1}^{M}[\widehat{\beta}^{\ell}(t/n)-\overline{\widehat{\beta}}(t/n)]^2\right)^{1/2},$$

$$B_g(n) := \frac{1}{Mn}\sum_{t=1}^{n}\sum_{\ell=1}^{M}[g^{\ell}(z_t)-\overline{\widehat{g}}(z_t)], \qquad \pi_g(n) := \left(\frac{1}{Mn}\sum_{t=1}^{n}\sum_{\ell=1}^{M}[\widehat{g}^{\ell}(z_t)-\overline{\widehat{g}}(z_t)]^2\right)^{1/2},$$

$$B_m(n) := \frac{1}{Mn}\sum_{t=1}^{n}\sum_{\ell=1}^{M}[m^{\ell}(x_t)-\overline{\widehat{m}}(x_t)], \qquad \pi_m(n) := \left(\frac{1}{Mn}\sum_{t=1}^{n}\sum_{\ell=1}^{M}[\widehat{m}^{\ell}(x_t)-\overline{\widehat{m}}(x_t)]^2\right)^{1/2},$$

where the superscript ℓ indicates the ℓth replication, $\overline{\widehat{\beta}}(\cdot) = \phi_{k_1}(\cdot)^{\intercal}\overline{\widehat{c}_1}$, $\overline{\widehat{g}}(\cdot) = a_{k_2}(\cdot)^{\intercal}\overline{\widehat{c}_2}$ and $\overline{\widehat{m}}(\cdot) = b_{k_3}(\cdot)^{\intercal}\overline{\widehat{c}_3}$ are the average of $\widehat{\beta}^{\ell}(\cdot)$, $\widehat{g}^{\ell}(\cdot)$ and $\widehat{m}^{\ell}(\cdot)$, respectively, over Monte Carlo replications $\ell = 1,\ldots,M$, $g^{\ell}(z_t)$ and $m^{\ell}(x_t)$ means the values of g and m are evaluated for the z_t and x_t, respectively, in the ℓth replication; and

$$\Pi_\beta(n) := \left(\frac{1}{Mn}\sum_{t=1}^{n}\sum_{\ell=1}^{M}[\beta(t/n)-\widehat{\beta}^{\ell}(t/n)]^2\right)^{1/2},$$



(a) Estimated density of $\widehat{\beta}_n(0.5) - \beta(0.5)$ (b) Estimated density of $\widehat{g}(-0.4) - g(-0.4)$

Fig. 1. The plot of estimated density functions.

Table 1
Bias and S.d. of the estimators.

	Bias			S.d.		
n	$B_\beta(n)$	$B_g(n)$	$B_m(n)$	$\pi_\beta(n)$	$\pi_g(n)$	$\pi_m(n)$
400	0.0012	−0.0605	0.0863	0.1040	0.1093	0.2117
600	0.0004	−0.0496	0.0804	0.0992	0.0990	0.1429
1200	0.0001	−0.0431	0.0497	0.0761	0.0725	0.1193

Table 2
RMSE of the estimators.

n	$\Pi_\beta(n)$	$\Pi_g(n)$	$\Pi_m(n)$
400	0.0917	0.0831	0.1063
600	0.0831	0.0775	0.0975
1200	0.0707	0.0624	0.0774

$$\Pi_g(n) := \left(\frac{1}{Mn} \sum_{t=1}^{n} \sum_{\ell=1}^{M} [g^\ell(z_t) - \widehat{g}^\ell(z_t)]^2 \right)^{1/2},$$

$$\Pi_m(n) := \left(\frac{1}{Mn} \sum_{t=1}^{n} \sum_{\ell=1}^{M} [m^\ell(x_t) - \widehat{m}^\ell(x_t)]^2 \right)^{1/2}.$$

It can be seen from Tables 1 and 2 that all the statistics perform very well as all quantities are decreasing reasonably with the increase of the sample size. Nevertheless, there might be a visible slower rate for the estimator of m function than the other two. This possibly is because the convergence rate of the estimator $\widehat{m}_n(x)$ to $m(x)$ is the slowest among all estimators, in view of Theorem 3.1.

In addition, with the same estimators, we also calculate their values at particular points, i.e., $\widehat{\beta}^\ell(0.5), \widehat{g}^\ell(-0.4)$ and $\widehat{m}^\ell(1.2)$ for all $\ell = 1, \ldots, M$. Then we may estimate the densities of $\widehat{\beta}^\ell(0.5) - \beta(0.5), \widehat{g}^\ell(-0.4) - g(-0.4)$ and $\widehat{m}^\ell(1.2) - m(1.2)$ with normalization in Corollary 3.1. These are done in Matlab by the `ksdensity` function and are plotted in the following figures.

From the three pictures in Figs. 1 and 2, the curves of the estimated densities for $\widehat{\beta}_n(0.5) - \beta(0.5), \widehat{g}(-0.4) - g(-0.4)$ and $\widehat{m}_n(1.2) - m(1.2)$ are gradually approaching the standard normal density. Particularly, the first two estimations seem visually to have a quicker convergence, which coincides again with our theoretical results in the preceding section.

Example 2. Let all settings be the same as in Example 1 except that $z_t = \Delta x_t = w_t$. Hereby, z_t and x_t share infinite many innovations ϵ_i. Although we cannot establish our theory on this situation, this example implies that the estimation procedure might be still workable. We report the results of the experiments in the following tables. In addition, in this correlated case we also calculate the proportion of $\widehat{\beta}_n(0.1), \widehat{g}_n(0.4)$ and $\widehat{m}_n(-0.5)$ dropping into the theoretical confidence intervals at 95% significant level according to Corollary 3.1.

It can be seen from Tables 3 and 4 that the three statistics and the probability of the estimators in the confidence intervals perform satisfactorily, and, comparing with the results in Example 1, it seems that in our settings the correlation between x_t and z_t does not affect the implementation of our estimating procedure. In particular, the probabilities are very high, and therefore sharing infinite many innovations for x_t and z_t might not affect statistical inference.

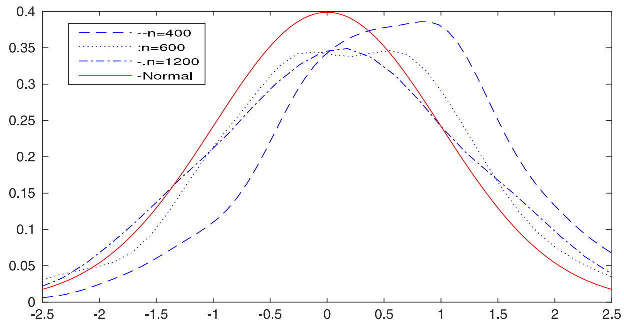
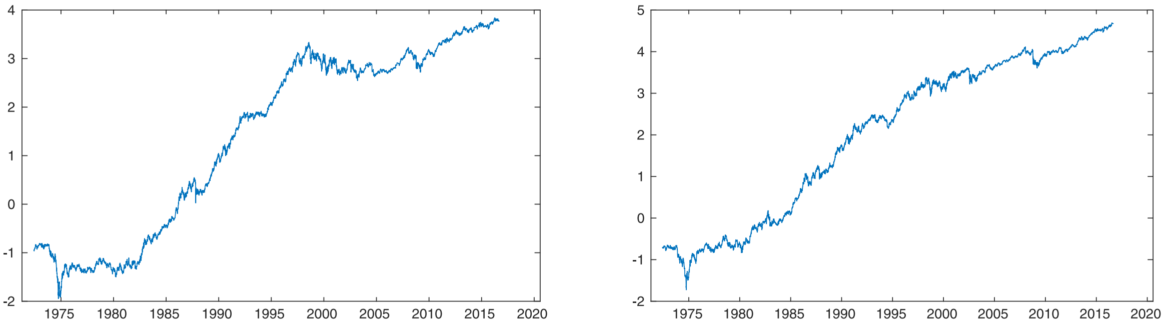


Fig. 2. Estimated density of $\widehat{m}_n(1.2) - m(1.2)$.



(a) Log price of Coke (b) Log price of Pepsi

Fig. 3. Plot data about Coke and Pepsi.

Table 3
Bias and S.d. of estimators in correlated case.

	Bias			S.d.		
n	$B_\beta(n)$	$B_g(n)$	$B_m(n)$	$\pi_\beta(n)$	$\pi_g(n)$	$\pi_m(n)$
400	0.0010	−0.0539	0.0662	0.1099	0.1104	0.1683
600	0.0006	−0.0472	0.0389	0.1008	0.0967	0.1438
1200	0.0001	−0.0224	−0.0226	0.0738	0.0717	0.1075

Table 4
RMSE and coverage probability of estimators in correlated case.

	RMSE			Coverage Probability		
n	$\Pi_\beta(n)$	$\Pi_g(n)$	$\Pi_m(n)$	$\widehat{\beta}_n(0.1)$	$\widehat{g}_n(0.4)$	$\widehat{m}_n(-0.5)$
400	0.1118	0.1250	0.1439	0.9938	0.9933	1
600	0.1034	0.1076	0.1255	0.9970	0.9950	1
1200	0.0745	0.0752	0.0887	1	1	1

5. Empirical study

This section provides an investigation of the relationship between the stock prices of Coke and Pepsi. Let Y_t be the log adjusted close price of Coke, X_t be the log adjusted close price of Pepsi and let z_t be the ratio of the trading volume for Coke and that for Coke plus Pepsi such that we always have $0 \leq z_t \leq 1$. The time span is from the first of June, 1972 to the 31st of August, 2016. Excluding all weekends and public holidays, we have $n = 11163$ observations. In Figs. 3 and 4 are the plots of Y_t and X_t as well as z_t, respectively.

To verify whether X_t is a unit root process, the ADF test is employed. The test fails to reject the null hypothesis that X_t is a unit root process with the p-Value 0.9901. The same test is implemented on Y_t and results in the p-Value 0.9627, a unit root process as well. We also plot the daily returns of Coke and Pepsi in Fig. 5, in order to visualize the unit root processes. The marginal price series appear to contain drifts and be non-recurrent, that is, we may suppose that $X_t = \mu_1 + X_{t-1} + \xi_t$ and $Y_t = \mu_2 + Y_{t-1} + \zeta_t$, with $\mu_1, \mu_2 \neq 0$. This implies that $X_t - \mu_1 t = X_0 + \sum_{j=1}^{t} \xi_j$ and $Y_t - \mu_2 t = Y_0 + \sum_{j=1}^{t} \zeta_j$ are recurrent processes that satisfy the theoretical requirement in the preceding sections. We work with $x_t = X_t - \hat{\mu}_1 t$ and

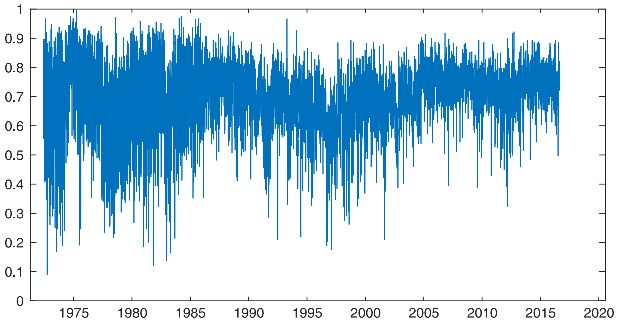
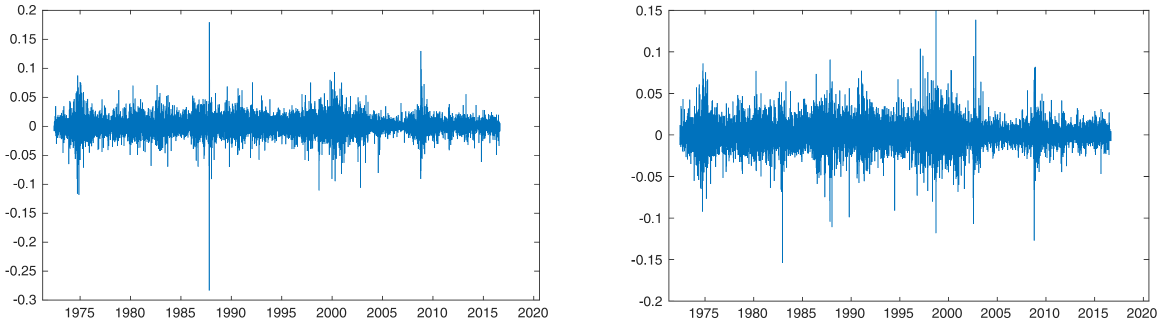


Fig. 4. Volume weight.



(a) Daily return of Coke (b) Daily return of Pepsi

Fig. 5. Daily returns of Coke and Pepsi.

$y_t = Y_t - \hat{\mu}_2 t$, where $\hat{\mu}_1 = (X_n - X_0)/n$ and $\hat{\mu}_2 = (Y_n - Y_0)/n$ are clearly consistent estimators of μ_1 and μ_2, respectively. More importantly, z_t and x_t might have certain correlation which our theory can deal with (see Assumption B.1.(b)).

We shall look into the relationship of the variables y_t, t/n, z_t and x_t through the model

$$y_t = \beta(t/n) + g(z_t) + m(x_t) + e_t, \tag{5.1}$$

for $t = 1, \ldots, n$, where all functions $\beta(\cdot)$, $g(\cdot)$ and $m(\cdot)$ are unknown and will be estimated.

Since both $\beta(\cdot)$ and $g(\cdot)$ are defined on $[0, 1]$, we use the cosine basis for their expansions, and for $m(\cdot)$ we use the Hermite sequence. All of these bases can be found in Section 2.

A key issue in using the series method in practice is the determination of the truncation parameters in the orthogonal expansions. The model can be estimated by the proposed procedure only if the truncation parameters are specified. However, there is no theoretical guide for the choice of such parameters, in particular in the case where both stationary and integrated processes are present. Since forecasting ability is one of the most important characteristics for a model, we shall choose the truncation parameters for our model through the best forecasting ability.

The forecasting ability for a model is measured by the so-called Out-of-Sample mean square errors (mse). That is, we use part of data, $1 \le t \le n_1$ ($n_1 < n$), say, to estimate the model for given k_i ($i = 1, 2, 3$), then using the estimated model we may forecast the dependent variable at $t = n_1 + 1$, obtaining $\widehat{y}_{n_1+1}$. The Out-of-Sample mse with the given truncation parameters is defined by $J^{-1}\sum_{j=1}^{J}(\widehat{y}_{n_j+1} - y_{n_j+1})^2$ where $n_j < n_{j+1} < n$ for $j = 1, \ldots, J-1$. The model that has the smaller Out-of-Sample mse has better forecasting ability.

In this example, let $J = 20$, $n_j = 9162 + 100j$, $1 \le j \le J$. In view of the nature of the dataset, we shall use the same truncation parameter for $\beta(\cdot)$ and $g(\cdot)$, $k_1 = k_2$, while the parameter for $m(\cdot)$ is still denoted by k_3. The Out-of-Sample mse's are calculated for all feasible k_i, that is, for all k_i that are not too large since from the complexity point of view this requirement is reasonable for a model. The results are reported in Table 5. From the table we can see that with $\widehat{k}_1 = \widehat{k}_2 = 2$ and $\widehat{k}_3 = 1$ the model has the smallest Out-of-Sample mse 0.0146, viz., the best forecasting ability. For the dataset we thus suggest the unknown functions in model (5.1) have the form $\widehat{\beta}(r) = \beta_2(r)$, $\widehat{g}(z) = g_2(z)$ and $\widehat{m}(x) = m_1(r)$. After the estimation procedure, we obtain

$$\begin{aligned}
\widehat{\beta}(r) &= -\,0.0223\varphi_1(r) - 0.0115\varphi_2(r), \quad r \in [0, 1], \\
\widehat{g}(z) &= -\,2.7906 + 0.1461\varphi_1(z), \quad z \in [0, 1], \\
\widehat{m}(x) &= 3.4201 e^{-x^2/2}, \quad x \in \mathbb{R},
\end{aligned} \tag{5.2}$$

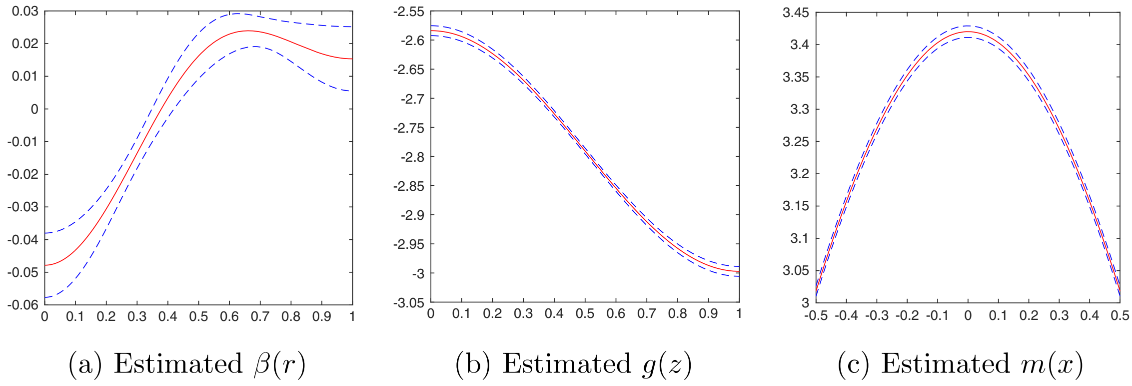


Fig. 6. Plot of estimated functions and confidence curves at 95% level.

Table 5
Out-of-sample mean square errors for model (5.1).

k_3	$k_1(=k_2)$						
	2	3	4	5	6	7	8
1	0.0146	0.0515	0.0241	0.0364	0.0358	0.0251	0.0227
2	0.0752	0.0392	0.0190	0.0251	0.0454	0.0378	0.0342
3	0.0529	0.0316	0.0150	0.0191	0.0380	0.0332	0.0314
4	0.0329	0.0293	0.0197	0.0225	0.0367	0.0330	0.0318
5	0.0315	0.0290	0.0196	0.0224	0.0407	0.0383	0.0368
6	0.0260	0.0299	0.0226	0.0248	0.0388	0.0356	0.0338

Table 6
Estimation and related statistics for model (5.3).

a_0	0.045	(0.0173, 0.0727)	a_1	0.1304	(0.1125,0.1482)
a_2	−0.3357	(−0.3735, −0.2980)	a_3	1.2609	(1.2506,1.2711)
		$R^2 = 0.8985$			$F = 32945$
		$f = 0$			$p = 0.0491$

where $\varphi_j(r) = \sqrt{2}\cos(\pi jr)$ for $j \geq 1$. We plot the pictures of $\widehat{\beta}(r)$, $\widehat{g}(z)$ and $\widehat{m}(x)$ and their confidence curves at 95% level in Fig. 6. The effect of relative trading volume is estimated as negative and close to linear, meaning that large amounts of trading in Coke relative to Pepsi is predictive of a decline in the price of Coke, ceteris paribus. The effect of Pepsi price on Coke is symmetrical around zero, implying that Pepsi price far away from its central range in either direction has a negative effect on the price of Coke, ceteris paribus. The estimated trend seems to be upward during the sample and bottoming out at the end, meaning that the price of Coke has increased over the sample period relative to the value predicted by a time invariant relationship based on the chosen covariates.

Comparison. In what follows the proposed model is compared with some potential competing models. One is a pure linear parametric model and another one is the model studied in Section 3.3,

$$y_t = a_0 + a_1\frac{t}{n} + a_2 z_t + a_3 x_t + \varepsilon_{1t} \tag{5.3}$$

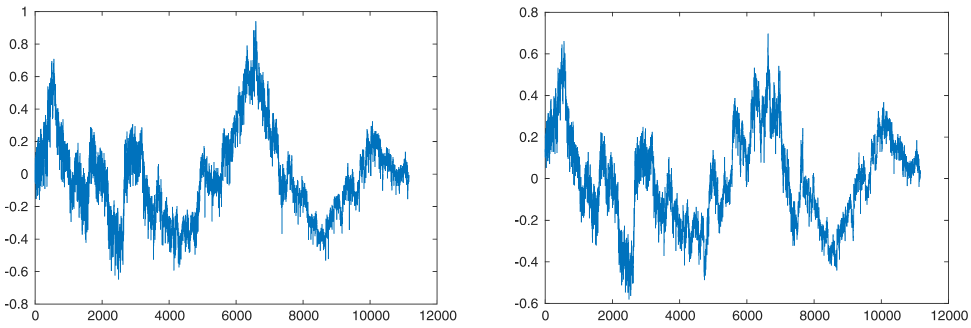
$$y_t = \beta_1\left(\frac{t}{n}\right) + g_1(z_t) + \theta_0 x_t + m_1(x_t) + \varepsilon_{2t}. \tag{5.4}$$

The models are still measured by their forecasting ability.

For model (5.3), using the full data we have the estimated coefficients, confidence intervals at 95% significance level and related statistics reported in Table 6. The linear model is fitted well as the R^2 is close to one, $F \gg f$ and $p < 0.05$. However, it is easily to calculate that the Out-of-Sample mse for model (5.3) is 0.0453, much larger than that of the proposed model with functions in (5.2). Nevertheless, the residual plot looks quite similar for the two models in Fig. 7.

For model (5.4) we compute the Out-of-Sample mse's with different combinations of feasible truncation parameters, showing in Table 7. It can be seen that the smallest Out-of-Sample mse is 0.0154 that corresponds to model (5.4) with $\widehat{k}_1 = \widehat{k}_2 = 4$ and $\widehat{k}_3 = 2$. Though we seek the model that has the best forecasting ability in a broad area for the truncation parameters, the resulting Out-of-Sample mse is larger than that calculated for model (5.1) with $\widehat{k}_1 = \widehat{k}_2 = 2$ and $\widehat{k}_3 = 1$.

Taking both models (5.3) and (5.4) into account, in terms of Out-of-Sample mse we still recommend model (5.1) with functions in (5.2) for the given dataset.



(a) Residuals of the proposed model (b) Residuals of the linear model

Fig. 7. Plot of residuals for the proposed and linear models.

Table 7
Out-of-sample mean square errors for model (5.4).

k_3	$k_1(=k_2)$						
	2	3	4	5	6	7	8
1	0.0462	0.0366	0.0167	0.0218	0.0319	0.0240	0.0227
2	0.0508	0.0313	0.0154	0.0195	0.0371	0.0322	0.0306
3	0.0457	0.0324	0.0181	0.0225	0.0356	0.0300	0.0286
4	0.0315	0.0289	0.0196	0.0223	0.0408	0.0386	0.0371
5	0.0313	0.0288	0.0195	0.0224	0.0408	0.0392	0.0376
6	0.0261	0.0293	0.0230	0.0259	0.0388	0.0370	0.0341

Trading strategy. We consider the performance of our proposed model in a pair trading strategy. The strategy has at least a 30-year history on Wall Street and is among the proprietary 'statistical arbitrage' tools currently used by hedge funds as well as investment banks. The strategy makes use of the idea of cointegration between two related stocks: it opens short/long positions when they diverge and closes the positions when they converge. See Gatev et al. (2006) for details. However, usually the cointegration is depicted by a linear form equation in the related literature. By contrast, we shall use nonparametric nonlinear cointegration in defining the pair strategy.

Let $n_0 \in (1, n)$ be an integer. With the proposed model (5.1) and (5.2), we have $\widehat{e}_t = y_t - \widehat{\beta}(t/n_0) - \widehat{g}(z_t) - \widehat{m}(x_t)$, $1 \le t \le n_0$. Let α be a significance level specified below. Find the empirical lower $(\alpha/2)$-quantile $\ell(\alpha/2)$ and upper $(\alpha/2)$-quantile $L(\alpha/2)$ from $\{\widehat{e}_t : 1 \le t \le n_0\}$.

The trading rule is as follows. From $t = n_0 + 1$ to $t = n$, calculate $\widehat{e}_t = y_t - \widehat{\beta}(1) - \widehat{g}(z_t) - \widehat{m}(x_t)$. If $\widehat{e}_t > L(\alpha/2)$, short one dollar in Coke and long one dollar in Pepsi; if $\widehat{e}_t < \ell(\alpha/2)$, long one dollar in Coke and short one dollar in Pepsi; otherwise, close all positions held if any, and put positive gain into a risk free bond account with rate r_0 and offset negative gain from the account. At the last trading day, all positions shall be closed ignoring the location of the residual.

Mathematically, at date $t \ge n_0 + 1$, if $\widehat{e}_t > L(\alpha/2)$, we owe $1/Y_t$ share of Coke and buy $1/X_t$ share of Pepsi; if $\widehat{e}_t < \ell(\alpha/2)$, we owe $1/X_t$ share of Pepsi and buy $1/Y_t$ share of Coke; otherwise, we clear all positions held since last date of closing positions, say, date k, that is, we obtain $\sum_{j=k}^{t-1} \Delta_j^t$, where

$$\Delta_j^t = \begin{cases} (X_t/X_j - Y_t/Y_j)(1+r_0)^{n-t}, & \text{if } \widehat{e}_j > L(\alpha/2) \text{ and } X_t/X_j - Y_t/Y_j \ge 0, \\ X_t/X_j - Y_t/Y_j, & \text{if } \widehat{e}_j > L(\alpha/2) \text{ and } X_t/X_j - Y_t/Y_j < 0, \\ (Y_t/Y_j - X_t/X_j)(1+r_0)^{n-t}, & \text{if } \widehat{e}_j < \ell(\alpha/2) \text{ and } Y_t/Y_j - X_t/X_j \ge 0, \\ Y_t/Y_j - X_t/X_j, & \text{if } \widehat{e}_j < \ell(\alpha/2) \text{ and } Y_t/Y_j - X_t/X_j < 0. \end{cases}$$

Then, the total profit of the trading period is $\sum_{t\in A} \sum_{j=k}^{t-1} \Delta_j^t$ where A is the collection of all clearing dates.

Let $\alpha = 0.01$ and 0.05, and put $r_0 = 0.02/250$ per day. Here, we do not consider any cost in the trading like transaction fee or price impact. We report the trading results in Table 8. In order to compare with the linear model, we also show the trading results in the same table using model (5.3). It can be seen that normally the results are sensitive to the length of the data history that determines the thresholds of taking action. In terms of profit, the proposed nonlinear cointegration model outperforms the linear model. Also, it seems no action token place for $t > 9000$ for both but with $\alpha = 0.01$ the linear model in the experiment always has nothing to gain. The results imply that nonlinear cointegration might be a better alternative relationship to the linear cointegration in the literature of pair trading strategy.

6. Conclusion and extension

This paper has studied additive models that have nonparametrically time trend, stationary and integrated variables as their components. Meanwhile, in order to accommodate more practical situations, the stationary variable has been relaxed

Table 8
Pair trading for Coke and Pepsi.

	α	Nonlinear cointegration			Linear cointegration		
		$L(\alpha/2)$	$\ell(\alpha/2)$	Profit	$L(\alpha/2)$	$\ell(\alpha/2)$	Profit
$n_0 = 7000$	0.01	0.3511	−1.2710	0.0227	0.5678	−0.4937	0
	0.05	0.1130	−1.2025	0.6525	0.4631	−0.4324	0.0767
$n_0 = 7500$	0.01	0.3450	−1.2669	0.0227	0.5680	−0.4874	0
	0.05	0.1012	−1.1963	0.8162	0.4614	−0.4236	0.1389
$n_0 = 8000$	0.01	0.3401	−1.2647	0.0227	0.5681	−0.4828	0
	0.05	0.0806	−1.1913	0.9117	0.4580	−0.4167	0.1931
$n_0 = 8500$	0.01	0.3318	−1.2561	0.0145	0.5646	−0.4780	0
	0.05	0.0704	−1.1963	0.7515	0.4562	−0.4122	0.5708
$n_0 = 9000$	0.01	0.3234	−1.2622	0	0.5635	−0.4734	0
	0.05	0.0580	−1.2059	0	0.4547	−0.4153	0

to be locally stationary; the correlation between regressors is allowed; the models have been extended to include an extra linear form of the integrated process that compensates a possible shortcoming in some particular cases. All these efforts provide with practitioners a variety of options, as illustrated by the empirical study.

As far as we know, it seems the first time in the literature that such models are investigated. All nonparametric functions are estimated by orthogonal series method; the central limit theorems for all proposed estimators have been established; the conventional optimal convergence rates are attainable; Monte Carlo experiment has conducted to verify the performance of the estimators with finite sample and an empirical study is provided.

The series estimators are convenient, but they are known in other contexts to be inefficient in the sense considered in Fan (1993). Following Linton (1997), Liu et al. (2013), and Linton and Wang (2016) we may consider efficiency improvement by one step kernel estimation. However, given the orthogonality between the estimated components, it is likely that the efficiency improvement is minimal, which is why we have not pursued this here. In addition, it is desirable to investigate the situation where z_t and x_t may be sharing infinite many innovations. Our next study would relax this condition to make the estimation procedure more applicable.

Acknowledgments

We would like to thank co-editor Prof Jianqing Fan, the Associate Editor and three anonymous referees for their constructive comments and suggestions that improve the paper considerably. In addition, Dong thanks the support from National Natural Science Foundation of China under Grant 71671143.

Appendix A. Lemmas

This section presents all technical lemmas while their proofs are relegated in Appendix C in the supplementary material of the paper.

We first study some properties about x_t. Without loss of generality, let $x_0 = 0$ almost surely. It follows that

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

Taking into account that in Assumption B.1.(b), z_t maybe contains $\epsilon_t, \ldots, \epsilon_{t-d+1}$, we decompose, for $t > d$,

$$x_t = \sum_{i=t-d+1}^{t} b_{t,i}\epsilon_i + \sum_{i=-\infty}^{t-d} b_{t,i}\epsilon_i := x_t^{(d)} + x_t^{(t-d)}. \tag{A.2}$$

Thus, $x_t^{(d)}$ and $x_t^{(t-d)}$ are mutually independent, and $x_t^{(d)}$ is stationary since it is a combination of $\epsilon_t, \ldots, \epsilon_{t-d+1}$ with fixed coefficients $\psi_0, \ldots, \sum_{\ell=0}^{d-1}\psi_\ell$ (i.e., a MA(d) process), while $x_t^{(t-d)}$ is still nonstationary as we only take out fixed number of ϵ's from x_t.

Letting $1 \le s < t$, x_t also has the following decomposition:

$$x_t = x_s^* + x_{ts},$$

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 $x_{ts} = \sum_{i=s+1}^{t} b_{t,i}\epsilon_i$ which captures all information containing in x_t on the time periods $(s, t]$. Let $d_{ts} := (Ex_{ts}^2)^{1/2} \sim \sqrt{t-s}$ for large $t - s$. Moreover, $\bar{x}_s = O_P(1)$ by virtue of Assumption A.

Additionally, taking into account of that z_t and z_s maybe have $\epsilon_t, \ldots, \epsilon_{t-d}$ and $\epsilon_s, \ldots, \epsilon_{s-d}$ for $t-s \geq d$, we decompose

$$x_t = x_t^{(d)} + x_{ts}^{(d)} + x_s^{(d*)} + x_s^{(s-d*)}, \tag{A.3}$$

where $x_t^{(d)} = \sum_{i=t-d+1}^{t} b_{t,i}\epsilon_i$, $x_{ts}^{(d)} = \sum_{i=s+1}^{t-d} b_{t,i}\epsilon_i$

$$x_s^{(d*)} = x_s^{(d)} + \bar{x}_s^{(d)}, \quad x_s^{(s-d*)} = x_s^{(s-d)} + \bar{x}_s^{(s-d)},$$

recalling that $x_s^{(d)}$ and $\bar{x}_s^{(d)}$ are the sums of the first d terms of x_s and $\bar{x}_s$, respectively, whereas $x_s^{(s-d)}$ and $\bar{x}_s^{(s-d)}$ are the rests of them in x_s and $\bar{x}_s$, respectively. Obviously, all four components in (A.3) are mutually independent.

Lemma A.1. *Suppose that Assumption A 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.*

(2) *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.*

Lemma A.2. *Suppose that Assumption A holds. For t or $t-s$ is large,*

(1) *Let $\tilde{d}_t^2 = E[(x_t^{(t-d)})^2]$. $\tilde{d}_t^{-1}x_t^{(t-d)}$ have uniformly bounded densities $f_{t/d}(x)$ over all t and x satisfying a uniform Lipschitz condition $\sup_x|f_{t/d}(x+y)-f_{t/d}(x)| \leq C|y|$ for any y and some constant $C>0$. In addition, $\sup_x|f_{t/d}(x)-\phi(x)| \to 0$ as $t\to\infty$ where $\phi(x)$ is the standard normal density function.*

(2) *For $1 \leq s < t$ and $t-s>d$, let $\tilde{d}_{ts}^2 = E[(x_{ts}^{(t-d)})^2]$. $\tilde{d}_{ts}^{-1}x_{ts}^{(t-d)}$ have uniformly bounded densities $f_{ts/d}(x)$ over all (t,s) and x satisfying the above uniform Lipschitz condition as well.*

It is noteworthy that $\tilde{d}_t \sim \sqrt{t}$, the same order as d_t for large t, and $\tilde{d}_{ts} \sim \sqrt{t-s}$, the same order as d_{ts}, for large $t-s$ noting by that d is fixed. This fact will be used frequently in the following derivation which, for simplicity, will not be mentioned repeatedly.

Lemma A.3. *Suppose that Assumptions A and B.1(b) hold.*

(1) *Let $p(\cdot)$ be a function such that $\mathbb{E}|p(z_t)| < \infty$, $h(\cdot)$ be an integrable function on $\mathbb{R}$, i.e. $\int|h(x)|dx < \infty$. Then, for large t, $|\mathbb{E}p(z_t)h(x_t)| < C\tilde{d}_t^{-1}\mathbb{E}|p(z_t)|\int|h(x)|dx(1+O(\tilde{d}_t^{-1}))$.*

(2) *Let $p_1(\cdot)$ and $p_2(\cdot)$ satisfy the above condition for $p(\cdot)$; and $h_1(\cdot)$ is integrable and $h_2(\cdot)$ is such that $\int|xh_2(x)|dx<\infty$. For $1 \leq s < t$ and $t-s>d$, $|\mathbb{E}[p_1(z_t)p_2(z_s)h_1(x_t)h_2(x_s)]| \leq C\tilde{d}_{ts}^{-1}\tilde{d}_s^{-1}\mathbb{E}|p_1(z_t)|\mathbb{E}|p_2(z_s)|\int|h_1(x)|dx\int|h_2(x)|dx(1+O(\tilde{d}_{ts}^{-1}))$.*

This lemma is sufficient to deal with the correlation between z_t and x_t stipulated in Assumptions B and B*.

All notation used below can be found in the text and thus is omitted for brevity.

Lemma A.4. *(1) $\left\|\frac{1}{n}\sum_{t=1}^n\phi_{k_1}(t/n)\phi_{k_1}^\intercal(t/n)-I_{k_1}\right\|^2 = O(n^{-2}k_1^2)$ as $k_1/n\to 0$;*
(2) $\sup_{0\leq r\leq 1}\|\phi_{k_1}(r)\|^2 = k_1+O(1)$ as $k_1\to\infty$.

Lemma A.5. *Let $D_n = \mathrm{diag}(\sqrt{n}I_{k_1}, \sqrt{n}I_{k_2}, \sqrt{n/d_n}I_{k_3})$. Then, under Assumptions A, B and D, $\|D_n^{-1}B_{nk}^\intercal B_{nk}D_n^{-1}-U_k\| = o_P(1)$ as $n\to\infty$ on a richer probability space. Particularly, $\|\frac{1}{n}\sum_{t=1}^n\phi_{k_1}(t/n)\phi_{k_1}(t/n)^\intercal - I_{k_1}\| = o(1)$, $\|\frac{1}{n}\sum_{t=1}^n a_{k_2}(z_t)a_{k_2}(z_t)^\intercal - U_{*22}\| = o_P(1)$ and $\|\frac{d_n}{n}\sum_{t=1}^n b_{k_2}(x_t)b_{k_2}(x_t)^\intercal - L_W(1,0)I_{k_3}\| = o_P(1)$.*

Lemma A.6. *Under Assumptions A, B* and D, $\|D_n^{-1}\widetilde{B}_{nk}^\intercal\widetilde{B}_{nk}D_n^{-1}-\widetilde{U}_k\| = o_P(1)$ as $n\to\infty$ on a richer probability space, where D_n is given in Lemma A.5.*

Lemma A.7. *Under Assumptions A, B and D, $\|D_n^{-1}B_{nk}^\intercal\Sigma_nB_{nk}D_n^{-1}-V_k\| = o_P(1)$ as $n\to\infty$ on a richer probability space, where $\Sigma_n = diag(\sigma^2(1/n),\ldots,\sigma^2(1))$ and D_n is given in Lemma A.5.*

Lemma A.8. *Under Assumptions A, B* and D, $\|D_n^{-1}\widetilde{B}_{nk}^\intercal\Sigma_n\widetilde{B}_{nk}D_n^{-1}-\widetilde{V}_k\| = o_P(1)$ as $n\to\infty$ on a richer probability space, where D_n is given in Lemma A.5.*

Lemma A.9. *Let $M_n = \mathrm{diag}(\sqrt{n}I_{k_1}, \sqrt{n}d_n, \sqrt{n}I_{k_2}, \sqrt{n/d_n}I_{k_3})$. Then, under Assumptions A, B and D, $\|M_n^{-1}A_{nk}^\intercal A_{nk}M_n^{-1}-Q_k\| = o_P(1)$ as $n\to\infty$ on a richer probability space.*

Lemma A.10. *Under Assumptions A, B* and D, $\|M_n^{-1}\widetilde{A}_{nk}^\intercal\widetilde{A}_{nk}M_n^{-1}-\widetilde{Q}_k\| = o_P(1)$ as $n\to\infty$ on a richer probability space, where M_n is the same as in Lemma A.9.*

Lemma A.11. *Under Assumptions A, B and D, $\|M_n^{-1}A_{nk}^{\top}\Sigma_n A_{nk}M_n^{-1}-P_k\|=o_P(1)$ as $n\to\infty$ on a richer probability space, where M_n is the same as in Lemma A.9.*

Lemma A.12. *Under Assumptions A, B* and D, $\|M_n^{-1}\widetilde{A}_{nk}^{\top}\Sigma_n\widetilde{A}_{nk}M_n^{-1}-\widetilde{P}_k\|=o_P(1)$ as $n\to\infty$ on a richer probability space, where M_n is the same as in Lemma A.9.*

Appendix B. Proof of the main result

In this appendix only the proofs of Theorems 3.1 and 3.2 are provided, while that for other theorems, proposition and corollaries are relegated to the supplement of the paper.

Proof of Theorem 3.1. The theorem will be shown via Cramér–Wold theorem. Notice that

$$\widehat{c}-c=(B_{nk}^{\top}B_{nk})^{-1}B_{nk}^{\top}(\gamma+e)=D_n^{-1}[D_n^{-1}B_{nk}^{\top}B_{nk}D_n^{-1}]^{-1}D_n^{-1}B_{nk}^{\top}(\gamma+e), \tag{B.1}$$

which implies

$$D_n(\widehat{c}-c)=[D_n^{-1}B_{nk}^{\top}B_{nk}D_n^{-1}]^{-1}D_n^{-1}B_{nk}^{\top}(\gamma+e).$$

Hence, for any $r\in[0,1]$, $z\in V$ and $x\in\mathbb{R}$,

$$\begin{pmatrix}\dfrac{\sqrt{n}}{\|\phi_{k_1}(r)\|}[\widehat{\beta}_n(r)-\beta(r)]\\ \dfrac{\sqrt{n}}{\|a_{k_2}(z)\|}[\widehat{g}_n(z)-g(z)]\\ \dfrac{\sqrt{n}}{\|b_{k_3}(x)\|\sqrt{d_n}}[\widehat{m}_n(x)-m(x)]\end{pmatrix}=\overline{\Psi}(r,z,x)^{\top}D_n(\widehat{c}-c)-\begin{pmatrix}\dfrac{\sqrt{n}}{\|\phi_{k_1}(r)\|}\gamma_{1k_1}(r)\\ \dfrac{\sqrt{n}}{\|a_{k_2}(z)\|}\gamma_{2k_2}(z)\\ \dfrac{\sqrt{n}}{\|b_{k_3}(x)\|\sqrt{d_n}}\gamma_{3k_3}(x)\end{pmatrix}$$

$$=\overline{\Psi}(r,z,x)^{\top}[D_n^{-1}B_{nk}^{\top}B_{nk}D_n^{-1}]^{-1}D_n^{-1}B_{nk}^{\top}e \tag{B.2}$$

$$+x s\overline{\Psi}(r,z,x)^{\top}[D_n^{-1}B_{nk}^{\top}B_{nk}D_n^{-1}]^{-1}D_n^{-1}B_{nk}^{\top}\gamma-\begin{pmatrix}\dfrac{\sqrt{n}}{\|\phi_{k_1}(r)\|}\gamma_{1k_1}(r)\\ \dfrac{\sqrt{n}}{\|a_{k_2}(z)\|}\gamma_{2k_2}(z)\\ \dfrac{\sqrt{n}}{\|b_{k_3}(x)\|\sqrt{d_n}}\gamma_{3k_3}(x)\end{pmatrix}, \tag{B.3}$$

where $\overline{\Psi}(r,z,x)$ is the $\Psi(r,z,x)$ defined in Section 2 postmultiplying by $\mathrm{diag}(\|\phi_{k_1}(r)\|^{-1},\|a_{k_2}(z)\|^{-1},\|b_{k_3}(z)\|^{-1})$, so that each block in $\overline{\Psi}(r,z,x)$ is a unit vector. Here, the leading term in the above is $\overline{\Psi}(r,z,x)^{\top}[D_n^{-1}B_{nk}^{\top}B_{nk}D_n^{-1}]^{-1}D_n^{-1}B_{nk}^{\top}e$ that will be dealt with firstly. To begin, by Lemma A.5, $\|D_n^{-1}B_{nk}^{\top}B_{nk}D_n^{-1}-U_k\|=o_P(1)$ as $n\to\infty$, and making use of the block diagonal structure of U_k, it follows that

$$\begin{aligned}\overline{\Psi}(r,z,x)^{\top}[D_n^{-1}B_{nk}^{\top}B_{nk}D_n^{-1}]^{-1}D_n^{-1}B_{nk}^{\top}e&=\overline{\Psi}(r,z,x)^{\top}U_k^{-1}D_n^{-1}B_{nk}^{\top}e(1+o_P(1))\\&=L_3^{-1}\overline{\Psi}(r,z,x)^{\top}\bar{U}_k^{-1}D_n^{-1}B_{nk}^{\top}e(1+o_P(1)),\end{aligned} \tag{B.4}$$

where $L_3=\mathrm{diag}(1,1,L_W(1,0))$ and $\bar{U}_k=\mathrm{diag}(I_{k_1},U_{k_2},I_{k_3})$. As L_3 is independent of the sample size, we now focus on $\overline{\Psi}(r,z,x)^{\top}\bar{U}_k^{-1}D_n^{-1}B_{nk}^{\top}e$.

Write

$$\overline{\Psi}(r,z,x)^{\top}\bar{U}_k^{-1}D_n^{-1}B_{nk}^{\top}e=\sum_{t=1}^{n}\xi_{nt}e_t$$

where we denote

$$\xi_{nt}:=\overline{\Psi}(r,z,x)^{\top}\bar{U}_k^{-1}D_n^{-1}\begin{pmatrix}\phi_{k_1}(t/n)\\ a_{k_2}(z_t)\\ b_{k_3}(x_t)\end{pmatrix}.$$

Since $(e_t,\mathcal{F}_{nt})$ is a martingale difference sequence stipulated in Assumption B, $\sum_{t=1}^{n}\xi_{nt}e_t$ is a martingale due to Assumption B.4. We calculate the conditional variance as follows:

$$\sum_{t=1}^{n}\mathbb{E}[\xi_{nt}\xi_{nt}^{\top}e_t^2|\mathcal{F}_{n,t-1}]=\sum_{t=1}^{n}\xi_{nt}\xi_{nt}^{\top}\sigma^2(t/n)$$

$$
\begin{aligned}
&=\overline{\Psi}(r,z,x)^{\intercal}\bar{U}_k^{-1}D_n^{-1}B_{nk}^{\intercal}\Sigma_n B_{nk}D_n^{-1}\bar{U}_k^{-1}\overline{\Psi}(r,z,x)\\
&=\overline{\Psi}(r,z,x)^{\intercal}\bar{U}_k^{-1}V_k\bar{U}_k^{-1}\overline{\Psi}(r,z,x)(1+o_P(1))\\
&=\overline{\Psi}(r,z,x)^{\intercal}\bar{U}_k^{-1}\bar{V}_k\bar{U}_k^{-1}\overline{\Psi}(r,z,x)L_\sigma(1+o_P(1))\\
&:=\Omega_n L_\sigma(1+o_P(1))
\end{aligned} \tag{B.5}
$$

by Lemma A.7 and the structure of V_k, where $L_\sigma=\mathrm{diag}(1,1,\int_0^1\sigma^2(r)dL(r,1))$, $\bar{V}_k=\mathrm{diag}(V_*,I_{k_3})$ a deterministic matrix, and $\Omega_n:=\overline{\Psi}(r,z,x)^{\intercal}\bar{U}_k^{-1}\bar{V}_k\bar{U}_k^{-1}\overline{\Psi}(r,z,x)$ is a 3×3 deterministic matrix as well. This means that the conditional variance of $\Omega_n^{-1/2}\sum_{t=1}^n\xi_{nt}e_t$ is approximated by L_σ in probability.

Here, we emphasize that $\Omega_n^{-1/2}$ is exchangeable with L_3, i.e. $\Omega_n^{-1/2}L_3=L_3\Omega_n^{-1/2}$. Indeed, notice that

$$
\begin{aligned}
\Omega_n^{-1/2}&=\left[\overline{\Psi}(r,z,x)^{\intercal}\begin{pmatrix}U_*^{-1}V_*U_*^{-1} & 0\\ 0 & I_{k_3}\end{pmatrix}\overline{\Psi}(r,z,x)\right]^{-1/2}\\
&=\begin{pmatrix}[\overline{\Psi}_{12}(r,z)^{\intercal}U_*^{-1}V_*U_*^{-1}\overline{\Psi}_{12}(r,z)]^{-1/2} & 0\\ 0 & 1\end{pmatrix},
\end{aligned}
$$

where $\overline{\Psi}_{12}(r,z):=\mathrm{diag}(\phi_{k_1}(r)/\|\phi_{k_1}(r)\|,a_{k_2}(z)/\|a_{k_2}(z)\|)$ the left-top 2-by-2 sub-block matrix of $\overline{\Psi}(r,z,x)$, while the right-bottom block of $\overline{\Psi}(r,z,x)$ is $b_{k_3}(x)/\|b_{k_3}(x)\|$, $U_*=\mathrm{diag}(I_{k_1},U_{k_2})$. Then, it is obvious that $\Omega_n^{-1/2}$ is exchangeable with L_3. This point allows us to normalize the left hand side of Eq. (B.2) and the martingale $\sum_{t=1}^n\xi_{nt}e_t$ by $\Omega_n^{-1/2}$ simultaneously.

Hence, we shall show that the martingale $\Omega_n^{-1/2}\sum_{t=1}^n\xi_{nt}e_t$ converges to $N(0,L_\sigma)$ by Cramér–Wold theorem and Corollary 3.1 of Hall and Heyde (1980, p. 58).

To this end, let $\lambda=(\lambda_1,\lambda_2,\lambda_3)\neq 0$ and we need to check for

$$
\xi_n:=\sum_{t=1}^n\lambda\Omega_n^{-1/2}\xi_{nt}e_t,
$$

whether (1) Lindeberg condition and (2) the convergence of the conditional variance are fulfilled.

(1). The Lindeberg condition is fulfilled if we show that $\sum_{t=1}^n\mathbb{E}[(\lambda\xi_{nt}e_t)^4|\mathcal{F}_{n,t-1}]\to_P 0$ as $n\to\infty$. Indeed, denoting $\mu_4:=\max_{1\le t\le n}\mathbb{E}[e_t^4|\mathcal{F}_{n,t-1}]$,

$$
\begin{aligned}
&\sum_{t=1}^n\mathbb{E}[(\lambda\xi_{nt}e_t)^4|\mathcal{F}_{n,t-1}]\le\mu_4\sum_{t=1}^n(\lambda\xi_{nt})^4\\
=&\mu_4\sum_{t=1}^n[\lambda\overline{\Psi}(r,z,x)^{\intercal}\bar{U}_k^{-1}D_n^{-1}(\phi_{k_1}(t/n)^{\intercal},a_{k_2}(z_t)^{\intercal},b_{k_3}(x_t)^{\intercal})^{\intercal}]^4\\
=&\mu_4\sum_{t=1}^n\Bigg(\lambda_1\frac{1}{\sqrt{n}}\frac{\phi_{k_1}(r)^{\intercal}}{\|\phi_{k_1}(r)\|}\phi_{k_1}(t/n)+\lambda_2\frac{1}{\sqrt{n}}\frac{a_{k_2}(z)^{\intercal}}{\|a_{k_2}(z)\|}U_{k_2}^{-1}a_{k_2}(z_t)\\
&\qquad+\sqrt{\frac{d_n}{n}}\lambda_3\|b_{k_3}(x)\|^{-1}b_{k_3}(x)^{\intercal}b_{k_3}(x_t)\Bigg)^4\\
\le&C_1\lambda_1^4\frac{1}{n^2}\sum_{t=1}^n\frac{1}{\|\phi_{k_1}(r)\|^4}[\phi_{k_1}(r)^{\intercal}\phi_{k_1}(t/n)]^4+C_2\lambda_2^4\frac{1}{n^2}\sum_{t=1}^n\frac{1}{\|a_{k_2}(z)\|}[a_{k_2}(z)^{\intercal}a_{k_2}(z_t)]^4\\
&+C_3\frac{d_n^2}{n^2}\sum_{t=1}^n[\lambda_4\|b_{k_3}(x)\|^{-1}b_{k_3}(x)^{\intercal}b_{k_3}(x_t)]^4,
\end{aligned}
$$

because U_{k_2} has eigenvalues greater than zero and bounded from above uniformly. Denote $u_1=\phi_{k_1}(r)/\|\phi_{k_1}(r)\|$ and $u_2=a_{k_2}(z)/\|a_{k_2}(z)\|$ two unit vectors with dimensions k_1 and k_2, respectively. It follows that

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{t=1}^n[u_1^{\intercal}\phi_{k_1}(t/n)]^4=\frac{1}{n}\int_0^1[u_1^{\intercal}\phi_{k_1}(s)]^4ds+O(n^{-2})\\
\le&\frac{1}{n}\int_0^1\|\phi_{k_1}(s)\|^4ds=O(n^{-1}k_1^2)\to 0,
\end{aligned}
$$

by Cauchy–Schwarz inequality and $\sup_{r\in[0,1]}\|\phi_{k_1}(s)\|^2=O(k_1)$. Also, in order to show that $\frac{1}{n^2}\sum_{t=1}^n[u_2^{\intercal}a_{k_2}(z_t)]^4\to_P 0$, note that

$$
\frac{1}{n^2}\mathbb{E}\sum_{t=1}^n(u_2^{\intercal}a_{k_2}(z_t))^4=\frac{1}{n^2}\mathbb{E}\sum_{t=1}^n\left(\sum_{i=0}^{k_2-1}u_{2i}p_i(z_t)\right)^4
$$

$$
\begin{aligned}
=&\frac{1}{n^2}\sum_{t=1}^{n}\sum_{i=0}^{k_2-1}u_{2i}^4\mathbb{E}p_i^4(z_t)+6\frac{1}{n^2}\sum_{t=1}^{n}\sum_{i=1}^{k_2-1}\sum_{j=1}^{i-1}u_{2i}^2u_{2j}^2\mathbb{E}[p_i^2(z_t)p_j^2(z_t)]\\
&+4\frac{1}{n^2}\sum_{t=1}^{n}\sum_{i=1}^{k_2-1}\sum_{j=1}^{i-1}u_{2i}u_{2j}^3\mathbb{E}[(p_i(z_t))p_j^3(z_t)]\\
&+8\frac{1}{n^2}\sum_{t=1}^{n}\sum_{i_1=3}^{k_2-1}\sum_{i_2=2}^{i_1-1}\sum_{i_3=1}^{i_2-1}\sum_{i_4=0}^{i_3-1}u_{2i_1}u_{2i_2}u_{2i_3}u_{2i_4}\mathbb{E}[p_{i_1}(z_t)p_{i_2}(z_t)p_{i_3}(z_t)p_{i_4}(z_t)]\\
\leq&\frac{1}{n}k_2\sum_{i=1}^{k_2}u_{2i}^4+6\frac{1}{n}k_2\sum_{i=1}^{k_2}\sum_{j=0}^{i-1}u_{2i}^2u_{2j}^2+4\frac{1}{n}k_2\sum_{i=1}^{k_2}\sum_{j=1}^{i-1}|u_{2i}||u_{2j}|^3\\
&+8\frac{1}{n}k_2\sum_{i_1=3}^{k_2}\sum_{i_2=2}^{i_1-1}\sum_{i_3=1}^{i_2-1}\sum_{i_4=0}^{i_3-1}|u_{2i_1}u_{2i_2}u_{2i_3}u_{2i_4}|\\
\leq&\frac{1}{n}k_2+4\frac{1}{n}k_2k_2^{1/2}+8\frac{1}{n}k_2k_2^2=o(1),
\end{aligned}
$$

where we denote $u_2=(u_{21},\ldots,u_{2k_2})^\intercal$, and Assumption B.2(a) is used for $\mathbb{E}p_i^4(z_t)=O(i)$ for i large, Cauchy–Schwarz inequality to derive $\mathbb{E}|(p_i(z_t))p_j^3(z_t)|\leq(\mathbb{E}|(p_i(z_t))|^4)^{1/4}(\mathbb{E}|p_j(z_t)|^4)^{4/3}$ as well as other similar terms; meanwhile, $\sum_{i=0}^{k_2-1}|u_{2i}|\leq k_2^{1/2}$. The third term is much easier to be dealt with. Let $u_3:=\|b_{k_3}(x)\|^{-1}b_{k_3}(x)$ a unit vector, and notice that $\|b_{k_3}(\cdot)\|^2\leq Ck_3$ uniformly by the uniform boundedness of Hermite functions. We have, by Lemma A.1,

$$
\begin{aligned}
&\frac{d_n^2}{n^2}\mathbb{E}\sum_{t=1}^{n}(u_3^\intercal b_{k_3}(x_t))^4\leq Ck_3\frac{d_n^2}{n^2}\sum_{t=1}^{n}\mathbb{E}(u_3^\intercal b_{k_3}(x_t))^2\\
=&Ck_3\frac{d_n^2}{n^2}\sum_{t=1}^{n}\int(u_3^\intercal b_{k_3}(d_tx))^2f_t(x)dx=Ck_3\frac{d_n^2}{n^2}\sum_{t=1}^{n}\frac{1}{d_t}\int(u_3^\intercal b_{k_3}(x))^2f_t(d_t^{-1}x)dx\\
\leq&Ck_3\frac{d_n^2}{n^2}\sum_{t=1}^{n}\frac{1}{d_t}\int(u_3^\intercal b_{k_3}(x))^2dx=Ck_3\frac{d_n^2}{n^2}\sum_{t=1}^{n}\frac{1}{d_t}\\
=&Ck_3n^{-1/2}=o(1),
\end{aligned}
$$

where $\int(u_3^\intercal b_{k_3}(x))^2dx=\|u_3\|^2=1$ by the orthogonality. This finishes the Lindeberg condition.

(2). For the conditional variance, it is clear by (B.5) that the martingale ξ_n has conditional variance approaching $\lambda L_\sigma\lambda^\intercal$ in probability. The normality therefore is shown.

To finish the proof, we next demonstrate that all reminder terms in (B.3) are negligible, that is, as $n\to\infty$,

$$
\begin{aligned}
&\sum_{t=1}^{n}\xi_{nt}\gamma(t)=o_P(1),\quad\sqrt{n}\|\phi_{k_1}(r)\|^{-1}\gamma_{1k_1}(r)=o(1),\\
&\sqrt{n}\|a_{k_2}(z)\|^{-1}\gamma_{2k_2}(z)=o(1),\quad\sqrt{n/d_n}\|b_{k_3}(x)\|^{-1}\gamma_{3k_3}(x)=o(1).
\end{aligned}
$$

Here, we omit the normalizer Ω_n since it is positive definite and has eigenvalues bounded below from zero and above from infinity due to the condition on U_{k_2} and V_*.

In view of the structures of ξ_{nt}, we need to show

$$(3)\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\|\phi_{k_1}(r)\|^{-1}\phi_{k_1}(r)^\intercal\phi_{k_1}(t/n)\gamma(t)=o(1),$$

$$(4)\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\|a_{k_2}(z)\|^{-1}a_{k_2}(z)^\intercal a_{k_2}(z_t)\gamma(t)=o_P(1),$$

$$(5)\ \sqrt{n}\|\phi_{k_1}(r)\|^{-1}\gamma_{1k_1}(r)=o(1),\quad\sqrt{n}\|a_{k_2}(z)\|^{-1}\gamma_{2k_2}(z)=o(1),$$

$$(6)\ \sqrt{\frac{d_n}{n}}\sum_{t=1}^{n}\frac{1}{\|b_{k_3}(x)\|}b_{k_3}(x)^\intercal b_{k_3}(x_t)\gamma(t)=o_P(1),$$

$$(7)\ \sqrt{\frac{n}{d_n}}\frac{1}{\|b_{k_3}(x)\|}\gamma_{3k_3}(x)=o(1).$$

To fulfill (3)–(5), it suffices to show

$$A_{1n} := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \|\phi_{k_1}(t/n)\| |\gamma(t)| = o_P(1), \qquad B_{1n} := \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \|a_{k_2}(z_t)\| |\gamma(t)| = o_P(1),$$

$$A_{2n} := \sqrt{n} \frac{1}{\|\phi_{k_1}(r)\|} |\gamma_{1k_1}(r)| = o(1), \qquad B_{2n} := \sqrt{n} \frac{1}{\|a_{k_2}(z)\|} |\gamma_{2k_2}(z)| = o(1).$$

Indeed, note that $\max_{r\in[0,1]} |\gamma_{1k_1}(r)| = O(k_1^{-s_1})$ and $\mathbb{E}|\gamma_{2k_2}(z_t)|^2 = O(k_2^{-s_2})$ by Newey (1997) and Chen and Christensen (2015) where s_1 and s_2 are respectively the smoothness order of $\beta(\cdot)$ and $g(\cdot)$, whereas using the density for $d_t^{-1}x_t$ in Lemma A.1 and the result of Lemma C.1 in Dong et al. (2016), we have $\mathbb{E}|\gamma_{3k_3}(x_t)|^2 \le Cd_t^{-1} \int |\gamma_{3k_3}(x)|^2 dx = d_t^{-1} O(k_3^{-s_3})$. Notice further that,

$$\begin{aligned}
\mathbb{E}|A_{1n}| \le & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \|\phi_{k_1}(t/n)\| \mathbb{E}|\gamma(t)| \\
\le & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \|\phi_{k_1}(t/n)\| |\gamma_{1k_1}(t/n)| \\
& + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \|\phi_{k_1}(t/n)\| \mathbb{E}|\gamma_{2k_2}(z_t)| \\
& + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \|\phi_{k_1}(t/n)\| \mathbb{E}|\gamma_{3k_3}(x_t)| \\
\le & \sqrt{nk_1} \max_{r\in[0,1]} |\gamma_{1k_1}(r)| + \sqrt{nk_1} O(k_2^{-s_2/2}) \\
& + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \|\phi_{k_1}(t/n)\| d_t^{-1/2} O(k_3^{-s_3/2}) \\
\le & \sqrt{nk_1} O(k_1^{-s_1}) + \sqrt{nk_1} O(k_2^{-s_2/2}) + n^{1/4} \sqrt{k_1} O(k_3^{-s_3/2}) \\
= & o(1)
\end{aligned}$$

by Assumption D, implying $A_{1n} = o_P(1)$. Similarly, it is readily seen that $A_{2n} = o(1)$ as well. For B_{1n}, denoting $u_2 = \|a_{k_2}(z)\|^{-1} a_{k_2}(z)$ temporarily,

$$\begin{aligned}
\mathbb{E}|B_{1n}| \le & \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbb{E}\|a_{k_2}(z_t)\gamma(t)\| \le \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\mathbb{E}\|a_{k_2}(z_t)\|^2 \mathbb{E}|\gamma(t)|^2\right]^{1/2} \\
\le & C \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left[\mathbb{E}\|a_{k_2}(z_t)\|^2\right]^{1/2} \left[|\gamma_{1k_1}(t/n)|^2 + \mathbb{E}|\gamma_{2k_2}(z_t)|^2 + \mathbb{E}|\gamma_{3k_3}(x_t)|^2\right]^{1/2} \\
= & C\sqrt{n} k_2^{1/2} \max_{r\in[0,1]} |\gamma_{1k_1}(r)| + C\sqrt{n} k_2^{1/2} O(k_2^{-s_2/2}) \\
& + Ck_2^{1/2} n^{1/4} O(k_3^{-s_3/2}) \\
= & C\sqrt{n} k_2^{1/2} O(k_1^{-s_1}) + C\sqrt{n} k_2^{1/2} O(k_2^{-s_2/2}) + Ck_2^{1/2} n^{1/4} O(k_3^{-s_3/2}),
\end{aligned}$$

due to Assumption D where $\mathbb{E}\|a_{k_2}(z_t)\|^2 \le Ck_2$ for some constant C since $\mathbb{E}[a_{k_2}(z_t)a_{k_2}(z_t)^\intercal]$ a block in Lemma A.5 has bounded eigenvalues. In addition,

$$\begin{aligned}
|B_{2n}| = & \frac{1}{\|a_{k_2}(z)\|} \sqrt{n} |\gamma_{2k_2}(z)| = \frac{1}{\|a_{k_2}(z) f_z(z)\|} \sqrt{n} |\gamma_{2k_2}(z) f_z(z)| \\
= & O(k_2^{-1/2}) \sqrt{n} k_2^{-s_2/2} = o(1),
\end{aligned}$$

where we have used $\|a_{k_2}(z) f_z(z)\|^2 = O(k_2)$ for fixed z and pointwise convergence $|\gamma_{2k_2}(z) f_z(z)| = o(k_2^{-s_2/2})$.

For (6), letting $u_3 = \|b_{k_3}(x)\|^{-1} b_{k_3}(x)$ as before and by Lemma A.1,

$$\begin{aligned}
& \sqrt{\frac{d_n}{n}} \sum_{t=1}^{n} \mathbb{E}|u_3^\intercal b_{k_3}(x_t)\gamma(t)| \\
\le & \sqrt{\frac{d_n}{n}} \sum_{t=1}^{n} \mathbb{E}|u_3^\intercal b_{k_3}(x_t)| |\gamma_{1k_1}(t/n)|
\end{aligned}$$

$$
\begin{aligned}
&+\sqrt{\frac{d_n}{n}}\sum_{t=1}^{n}\mathbb{E}|u_3^{\intercal}b_{k_3}(x_t)||\gamma_{2k_2}(z_t)| \\
&+\sqrt{\frac{d_n}{n}}\sum_{t=1}^{n}\mathbb{E}|u_3^{\intercal}b_{k_3}(x_t)\gamma_{3k_3}(x_t)|| \\
\leq&\sqrt{\frac{d_n}{n}}\max_{r\in[0,1]}|\gamma_{1k_1}(r)|\sum_{t=1}^{n}\left[\mathbb{E}\|b_{k_3}(x_t)\|^2\right]^{1/2} \\
&+\sqrt{\frac{d_n}{n}}k_2^{-s_2/2}\sum_{t=1}^{n}[\mathbb{E}\|b_{k_3}(x_t)\|^2]^{1/2} \\
&+\sqrt{\frac{d_n}{n}}\sum_{t=1}^{n}[\mathbb{E}\|b_{k_3}(x_t)\|^2\mathbb{E}|\gamma_{3k_3}(x_t)|^2]^{1/2} \\
\leq& C_1 n^{-1/4}k_1^{-s_1}k_3^{1/2}n^{3/4}+C_2 n^{-1/4}k_2^{-s_2/2}k_3^{1/2}n^{3/4} \\
&+C_3\sqrt{\frac{d_n}{n}}\sum_{t=1}^{n}d_t^{-1}\left[\int\|b_{k_3}(x)\|^2dx\int|\gamma_{3k_3}(x)|^2dx\right]^{1/2} \\
=&C_1 n^{1/2}k_1^{-s_1}k_3^{1/2}+C_2 n^{1/2}k_2^{-s_2/2}k_3^{1/2}+C_3 n^{1/4}k_3^{-s_3/2}k_3^{1/2} \\
=&o(1)
\end{aligned}
$$

due to Assumption D where we have used the boundedness of the density $f_t(x)$ for x_t/d_t by Lemma A.1. In the mean time, for (7),

$$
\begin{aligned}
\frac{1}{\|b_{k_3}(x)\|}\sqrt{n/d_n}|\gamma_{3k_3}(x)| =&O(k_3^{-1/2})O(n^{1/4})o(k_3^{-(s_3-1)/2-1/12}) \\
=&o(n^{1/4}k_3^{-s_3/2-1/12})=o(1),
\end{aligned}
$$

where $\sup_x|\gamma_{3k_3}(x)| = o(k_3^{-(s_3-1)/2-1/12})$ by again Lemma C.1 in the supplement of Dong et al. (2016). The entire proof is complete.

Proof of Theorem 3.2. Similar to (B.1), we have

$$
\widehat{c}-c = D_n^{-1}\widetilde{U}_k^{-1}D_n^{-1}\widetilde{B}_{nk}^{\intercal}(\widetilde{\gamma}+e)(1+o_P(1)),
$$

where $\widetilde{\gamma}=(\widetilde{\gamma}(1),\ldots,\widetilde{\gamma}(n))^{\intercal}$ with $\widetilde{\gamma}(t)=\gamma_{1k_1}(t/n)+\gamma_{2k_2}(z_{t,n})+\gamma_{3k_3}(x_t)$. Hence, $D_n(\widehat{c}-c)=\widetilde{U}_k^{-1}D_n^{-1}\widetilde{B}_{nk}^{\intercal}(\widetilde{\gamma}+e)$ where the term $o_P(1)$ is omitted for better exposition. Also, note that for any $r\in[0,1]$, $z\in[a_{\min},a_{\max}]$ and $x\in\mathbb{R}$,

$$
\begin{aligned}
\begin{pmatrix}
\frac{\sqrt{n}}{\|\phi_{k_1}(r)\|}[\widehat{\beta}_n(r)-\beta(r)] \\
\frac{\sqrt{n}}{\|a_{k_2}(z)\|}[\widehat{g}_n(z)-g(z)] \\
\frac{\sqrt{n}}{\|b_{k_3}(x)\|\sqrt{d_n}}[\widehat{m}_n(x)-m(x)]
\end{pmatrix}
&= \overline{\Psi}(r,z,x)^{\intercal}D_n(\widehat{c}-c)-
\begin{pmatrix}
\frac{\sqrt{n}}{\|\phi_{k_1}(r)\|}\gamma_{1k_1}(r) \\
\frac{\sqrt{n}}{\|a_{k_2}(z)\|}\gamma_{2k}(z) \\
\frac{\sqrt{n}}{\|b_{k_3}(x)\|\sqrt{d_n}}\gamma_{3k_3}(x)
\end{pmatrix} \\
&= \overline{\Psi}(r,z,x)^{\intercal}\widetilde{U}_k^{-1}D_n^{-1}\widetilde{B}_{nk}^{\intercal}(\widetilde{\gamma}+e)-
\begin{pmatrix}
\frac{\sqrt{n}}{\|\phi_{k_1}(r)\|}\gamma_{1k_1}(r) \\
\frac{\sqrt{n}}{\|a_{k_2}(z)\|}\gamma_{2k}(z) \\
\frac{\sqrt{n}}{\|b_{k_3}(x)\|\sqrt{d_n}}\gamma_{3k_3}(x)
\end{pmatrix}. \qquad \text{(B.6)}
\end{aligned}
$$

The normality will be derived from $\Psi(r,z,x)^{\intercal}\widetilde{U}_k^{-1}D_n^{-1}\widetilde{B}_{nk}^{\intercal}e$. It can be shown exactly in the same fashion as Theorem 3.1 by Cramér–Wold theorem as well as the diagonal block structure of $\widetilde{U}_k$ and $\widetilde{V}_k$. In addition, using the approximation of $z_t(t/n)$ to $z_{t,n}$ [some examples can be found in the proof of the lemmas] it is not hard to demonstrate all the remainder terms are asymptotically negligible. These are omitted for the sake of similarity. The proof thus is finished.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2018.05.007.

References

Andrews, D., Whang, Y., 1990. Additive interaction regression models: Circumvention of the curse of dimensionality. Econometric Theory 6, 466–479.
Belloni, A., Chernozhukov, V., Chetverikov, D., Kato, K., 2015. Some new asymptotic theory for least squares series: Pointwise and uniform results. J. Econometrics 186, 345–366.
Cai, Z., Li, Q., Park, J., 2009. Functional-coefficient cointegration models for nonstationary time series data. J. Econometrics 148, 101–113.
Chang, Y., Park, J.Y., Phillips, P.C.B., 2001. Nonlinear econometric models with cointegrated and deterministically trending regressors. Econom. J. 4, 1–36.
Chen, X., Christensen, T., 2015. Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. J. Econometrics 188, 447–465.
Chen, X., Hansen, L.P., Carrasco, M., 2010. Nonlinearity and temporal dependence. J. Econometrics 155, 155–169.
Chen, X., Shen, X., 1998. Sieve extremum estimates for weakly dependent data. Econometrica 66, 289–314.
Dong, C., Gao, J., Peng, B., 2015. Semiparametric single-index panel data models with cross-sectional dependence. J. Econometrics 188, 301–312.
Dong, C., Gao, J., Tjøstheim, D., 2016. Estimation for single-index and partially linear single-index integrated models. Ann. Statist. 44, 425–453.
Dudley, R.M., 2003. Real Analysis and Probability. Cambridge studies in advanced mathematics 74. Cambridge University Press, Cambridge, U.K..
Fan, J., 1993. Local linear regression smoothers and thier minmax efficiency. Ann. Statist. 21, 196–216.
Gao, J., 2007. Nonlinear Time Series: Semiparametric and Nonparametric Methods. Monographs on statistics and applied probability. Chapman & Hall, New York.
Gao, J., Phillips, P.C.B., 2013. Semiparametric estimation in triangular system equations with nonstationarity. J. Econometrics 176, 59–79.
Gao, J., Tong, H., Wolff, R., 2001. Adaptive orthogonal series estimation in additive stochastic regression models. Statistic Sinica 11, 409–428.
Gatev, E., Goetzmann, W.N., Rouwenhorst, K.G., 2006. Pairs trading: Performance of a relative-value arbitrage rule. Review Finance Stud. 19, 797–827.
Grenander, U., Rosenblatt, M., 1957. Statistical Analysis of Stationary Time Series. Wiley, New York.
Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and Its Application. Academic Press, New York.
Hansen, B.E., 2015. A unified asymptotic distribution theory for parametric and nonparametric least square. Working paper, University of Wisconsin.
Karlsen, H.A., Mykelbust, T., Tjøstheim, D., 2007. Nonparametric estimation in a nonlinear cointegration type model. Ann. Statist. 35, 252–299.
Koo, B., Linton, O., 2012. Estimation of semiparametric locally stationary diffusion models. J. Econometrics 170, 210–233.
Li, D., Phillips, P., Gao, J., 2016. Uniform consistency of nonstationary kernel-weighted sample covariance for nonparametric regressions. Econometric Theory 32, 655–685.
Linton, O., 1997. Efficient estimate of additive nonparametric models. Biometrika 84, 469–474.
Linton, O., Wang, Q., 2016. Nonparametric transformation regression with nonstationary data. Econometric Theory 32, 1–29.
Liu, R.L., Yang, L., Hardle, W.K., 2013. Oracally efficient two-step estimation of generalized additive model. J. Amer. Statist. Assoc. 108, 619–631.
Mammen, E.O., Linton, O., Nielsen, J., 1999. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. Ann. Statist. 27, 1443–1490.
Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. J. Econometrics 79, 147–168.
Park, J., Hahn, S., 1999. Cointegrating regressions with time varying coefficients. Econometric Theory 15, 664–703.
Park, J.Y., Phillips, P.C.B., 1999. Asymptotics for nonlinear transformations of integrated time series. Econometric Theory 15, 269–298.
Park, J.Y., Phillips, P.C.B., 2001. Nonlinear regression with integreted time series. Econometrica 69 (1), 117–161.
Phillips, P.C.B., 2005. HAC estimation by automated regression. Econometric Theory 21, 116–142.
Phillips, P.C.B., 2007. Regression with slowly varying regressors and nonlinear treads. Econometric Theory 23, 557–614.
Phillips, P.C.B., 2010. The mysteries of trend. Discussion paper 1771, Cowles Foundation, Yale University.
Phillips, P.C.B., Li, D., Gao, J., 2017. Estimating smooth structure change in cointegration models. J. Econometrics 196, 180–195.
Phillips, P.C.B., Solo, V., 1992. Asymptotics for linear processes. Ann. Statist. 20 (2), 971–1001.
Revuz, D., Yor, M., 2005. Continuous Martingales and Brownian Motion. In: A Series of Comprehensive Studies in Mathematics, vol. 293, Springer-Verlag.
Schienle, M., 2008. Nonparametric Nonstationary Regression (Doctoral thesis), Mannheim University, Mannheim, Germany.
Stone, C., 1982. Optimal global rates of convergence for nonparametric regression. Ann. Statist. 10, 1040–1053.
Stone, C., 1985. Additive regression and other nonparametric models. Ann. Statist. 13, 689–705.
Su, L., Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. J. Econometrics 169, 34–47.
Szego, G., 1975. Orthogonal Polynomials. In: Colloquium publications XXIII, American Mathematical Association, Providence, Rhode Island.
Vogt, M., 2012. Nonparametric regression for locally stationary time series. Ann. Statist. 46, 2601–2633.
Wang, Q., 2015. Limit Theorems for Nonlinear Cointegrating Regression. World Scientific Press, Singapore.
Wang, Q., Phillips, P.C.B., 2009. Asymptotic theory for local time density estimation and nonparametric cointegreting regression. Econometric Theory 25, 710–738.
Xiao, Z., 2009. Functional-coefficient cointegration models. J. Econometrics 152, 81–92.