UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

Cambridge Working Papers in Economics

Cambridge-INET Working Paper Series No: 2019/18

Cambridge Working Papers in Economics: 1981

SEMIPARAMETRIC SINGLE-INDEX PREDICTIVE REGRESSION

Weilun Zhou (University of Cambridge)

Jiti Gao (Monash University)

David Harris (University of Melbourne)

Hsein Kew (Monash University)

This paper studies a semiparametric single-index predictive regression model with multiple nonstationary predictors that exhibit co-movement behaviour. Orthogonal series expansion is employed to approximate the unknown link function in the model and the estimator is derived from an optimization under constraint. The main finding includes two types of super-consistency rates for the estimators of the index parameter. The central limit theorem is established for a plug-in estimator of the unknown link function. In the empirical studies, we provide ample evidence in favor of nonlinear predictability of the stock return using four pairs of nonstationary predictors.

Semiparametric Single–index Predictive Regression[1]

Weilun Zhou†, Jiti Gao†, David Harris‡ and Hsein Kew†

† *Monash University and* ‡ *University of Melbourne*

September 2, 2019

Abstract

This paper studies a semiparametric single-index predictive regression model with multiple nonstationary predictors that exhibit co-movement behaviour. Orthogonal series expansion is employed to approximate the unknown link function in the model and the estimator is derived from an optimization under constraint. The main finding includes two types of super-consistency rates for the estimators of the index parameter. The central limit theorem is established for a plug-in estimator of the unknown link function. In the empirical studies, we provide ample evidence in favor of nonlinear predictability of the stock return using four pairs of nonstationary predictors.

Keywords: Predictive regression; Single-index model; Hermite orthogonal estimation; Dual super-consistency rates; Co-moving predictors.

JEL classification: C13, C14, C32, C51

[1]The authors would like to thank Prof Chaohua Dong, Prof Oliver Linton, and Prof Peter C.B. Phillips for their constructive comments on earlier versions, and seminar participants of the Monash Department of Econometrics and Business Statistics and several conference participants at the 2018 China Meeting of the Econometric Society in Shanghai and the 2018 North American Summer Meeting of the Econometric Society at UC Davies for their useful comments. The first two authors also acknowledge the financial support by the Australian Research Council Discovery Grants Program under Grant Numbers: DP150101012 and DP170104421.

1 Introduction

Whether stock returns are predictable or not is a fundamental issue in finance. In the study of a standard predictive regression, predictability is examined in the context of a parametrically linear model:

$$y_t = \alpha + \beta \times x_{t-1} + e_t, \tag{1.1}$$

where y_t is the equity premium, x_{t-1} is the lagged financial predictor and e_t is a martingale difference sequence. The earliest method used in the literature to test predictability is to apply the conventional t-test for β. If the estimate of the slope coefficient β is statistically significant, we can conclude that x_{t-1} is a significant predictor.

Although various empirical studies have been conducted to examine stock return predictability (Fama and French, 1988; Goyal and Welch, 2003; Shiller et al., 1984; Welch and Goyal, 2008), this widely used linear predictive regression model may encounter two main problems (see Phillips, 2015 for an overview of certain aspects of predictive regression). The first problem is that several financial predictors are highly persistent or even nonstationary, yet the equity premium behaves like a stationary process. Therefore, a linear predictive regression model can be unbalanced because the time-series properties on both sides of the equation (1.1) are different. The second problem is that the parametrically linear models may not be robust to functional form misspecification. To address these two problems, Kasparis et al. (2015) proposed a nonparametric predictive regression model and estimated it with a kernel-based method. Cai and Gao (2013) estimated this unknown function with Hermite functions—a sieve-based method.

However, practical implementation of these methods presents one major drawback—the methodology is restricted to the case of a scalar predictor only. Research on the multiple predictive regression model is limited in the literature, with one difficulty being the need to cope with multiple degrees of persistence of the predictors. Lamont (1998) suggested using dividend-price ratio (dp) and the payout ratio as predictors based on the conventional t-test. Ang and Bekaert (2007) found the predictability of the equity premium using both dividend yield (dy) and short rates according to the F test, with standard errors adjusted for the overlapping issue. In addition, Chen and Hong (2009) applied a smoothed kernel method on the predictive residuals to capture the potentially nonlinear predictable component. Kostakis et al. (2015) proposed a testing procedure based on IVX estimation (self-generated instrument variables estimation)—which was first studied by Phillips and Magdalinos (2009)—and found some evidence regarding the short-horizon predictability of the equity premium. Recently, Xu and Guo (2019) proposed three new dimensionality-robust tests built on the IVX estimator. Their proposed tests can detect potential spurious predictability driven by existing tests that tend to over-reject the null of no predictability in a finite sample with a large model size. The methods discussed here are all based on parametrically linear models, while the nonlinear predictability of the equity premium using multiple predictors remains unknown.

To make our proposed model more balanced and allow for a potential nonlinear relationship between the liner combination of comoving predictors and the dependent variable, we propose a semiparametric single-index predictive regression model of the form:

$$y_t = g_0(\theta_0^\top x_{t-1}) + e_t, \tag{1.2}$$

where $x_t = (x_{1,t}, ..., x_{d,t})^\top$ is a vector of d-dimensional nonstationary time series, $g_0(.)$ is an unknown univariate link function, θ_0 is the single-index parameter such that $u_{t-1} = \theta_0^\top x_{t-1}$ is stationary, and e_t is a martingale difference sequence. In terms of the identification condition, we impose that $\|\theta_0\|^2 = 1$ with a positive first element.

In a nonparametric multiple regression estimation context, researchers often encounter the curse of dimensionality problem. The single-index model considers a linear combination of predictors that can capture the most information about the potentially nonlinear relationship between the dependent variable and the predictors; hence, this is an efficient way to solve the dimensionality problem.

Dong et al. (2016) (hereafter DGT) assumed that the single-index component $u_t = \theta_0^\top x_t$ was nonstationary based on the nonstationary assumption for x_t. However, we are more interested in the case in which u_{t-1} is stationary, and this is a natural way to cope with the unbalanced issue we mentioned before. From an empirical point of view, many financial predictors exhibit co-movement behaviour (e.g., Figure 3 below shows the co-movement between dp and dy), and our proposed model can potentially consider this characteristic in the context of stock return predictability. In the literature for predictive regression with multiple predictors, Amihud et al. (2008) only considered stationary predictors and Kostakis et al. (2015) assumed predictors with an arbitrary degree of persistence, yet excluded comoving predictors. Recently, Koo et al. (2016) proposed a Least Absolute Shrinkage and Selection Operator (LASSO) estimator in the presence of comoving predictors. In addition, Xu (2017) considered a linear predictive regression model allowing for both highly persistent and comoving predictors, and studied the behaviour of the proposed IVX test. To the best of our knowledge, no study is available for the single-index model when x_{t-1} is nonstationary yet $u_{t-1} = \theta_0^\top x_{t-1}$ is stationary.

In the literature for single-index models, Härdle and Stoker (1989) and Powell et al. (1989) proposed an average derivative estimation for the single-index parameter θ_0. In addition, there have been many papers (Ichimura, 1993; Powell et al., 1989; Xia, 2006) devoted to the estimation of single-index models based on the conventional nonparametric kernel-based method. Alternatively, the nonparametric sieve-based approach has attracted great attention recently in the literature to approximate unknown functions (see Chen, 2007 for a detailed review). Yu and Ruppert (2002) proposed penalised spline estimation for partially linear single-index models. Dong et al. (2015) proposed consistent closed-form estimators for both the single-index parameter and the unknown link function, based on Hermite expansion.

This paper studies the estimation of model (1.2) using Hermite polynomials. Although $u_{t-1} = \theta_0^\top x_{t-1}$ is considered a stationary process, the nonstationarity of each regressor is harder to deal with than the pure stationary case. Some recent work by Park and Phillips (2000) and DGT employed the so-called rotation technique to decompose the estimator into two directions: alongside and orthogonal to the direction of the true parameter θ_0. We adopt the same technique to develop the theory. However, in contrast with these two previous papers, we assume $u_{1t-1} = \theta_0^\top x_{t-1}$ is stationary, rather than nonstationary, and need to ensure that the nonstationary component will not dominate and the stationarity on u_{t-1} will not break down.

To ensure the identification requirements we discussed before, the relevant literature uses the estimate $\widehat{\theta}$ without constraint at first, and then standardises it with the form $\widehat{\theta}/\|\widehat{\theta}\|$. This paper employs the Lagrange optimisation, which adds the constraint $\|\theta_0\|^2 = 1$ directly to the estimation procedure. In addition, we allow for a possible unbounded support of the unknown link function and an unbounded link function itself. In the literature for unbounded issues of nonparametric sieve regression, Chen and Christensen (2015) introduced an indicator function based on the sample size, which reduced the unbounded support to a compact set. Hansen (2015) allowed for an unbounded support by imposing the bound on the moment. Wang et al. (2010) applied a re-parametrisation method that estimated the equation over a restricted region in the Euclidean space $\mathbb{R}^{d-1}$. We will adopt our own method to develop the theory.

In summary, this paper aims to find a pair of (θ_0, g_0), such that $e_t = y_t - g_0(\theta_0^\top x_{t-1})$ is stationary. In contrast to DGT, who considered a pure nonstationary case with integrable function $g_0(w) \in L^2(\mathbb{R})$, we assume $u_{t-1} = \theta_0^\top x_{t-1} \sim I(0)$ with $g_0(w) \in L^2(\mathbb{R}, \exp(-w^2/2))$—a larger Hilbert space. The main contributions of this paper are as follows:

1. The proposed model considers comoving nonstationary predictors, such that $u_{t-1} = \theta_0^\top x_{t-1}$ is stationary.
2. The stationarity on u_{t-1} implies that the model becomes more balanced with the observed $I(0)$ property of the equity premium.
3. The model allows for unbounded support of the regressors and unbounded regression function itself.
4. The proposed estimation method estimates θ_0 under the constraint $\|\theta_0\|^2 = 1$ directly, rather than artificially standardising $\widehat{\theta}$ by the form $\widehat{\theta}/\|\widehat{\theta}\|$. Under our model setting, a n-super-rate of convergence can be achieved for the proposed estimator, $\widehat{\theta}_n$.
5. The model establishes new asymptotic properties for the proposed estimators, including both the NLS estimator of the single-index parameter and the plug-in estimator of the unknown link function.

This paper uncovers some important results. We find that there are dual convergence rates for the estimator of the index parameter in a new coordinate system. They include a type 1 super-consistency rate, $O_P(n^{-2})$, in the direction along θ_0,[2] and a type 2 super-consistency rate, $O_P(n^{-1})$, along all the other directions orthogonal to θ_0. Given that $\widehat{\theta}_n$ is the composite of its coordinates along these two directions in the new system, its behaviour will be dominated by the one with a slower rate of convergence, and then we have $\widehat{\theta}_n - \theta_0 = O_P(n^{-1})$, which is still super-consistent. One factor contributing to this super rate is our constraint, $\|\theta_0\|^2 = 1$. Roughly speaking, the constraint within the estimation procedure can scale $\widehat{\theta}_n$ to the unit ball, so that the norm of the estimate $\widehat{\theta}_n$ always matches that of θ_0. Therefore, it accelerates the convergence rate along θ_0 direction relative to the one without constraint and hence the overall convergence rate.

Given that our model includes multiple regressors and can cope with the unbalance issue naturally, we then apply it in the context of stock return predictability. Considering monthly and quarterly data over the 1927 to 2017 sample period and the 1952 to 2017 sub-period, we examine the predictability of the equity premium using four pairs of nonstationary predictors, and find significant evidence of nonlinear predictability.

The remainder of this paper is organised as follows. Section 2 gives some preliminaries about the Hermite polynomials that will be used in the series expansion and then proposes the estimation procedures. The asymptotic theories for the nonlinear least squares estimator $\widehat{\theta}_n$ as well as the plug-in estimator $\widehat{g}_n(w)$ are discussed in Section 3. In Section 4, computational estimation procedures are introduced and Monte-Carlo simulation experiments are conducted to examine the finite sample performance of the proposed estimators. Section 5 provides an empirical study to examine stock return predictability. Section 6 concludes this paper. Appendix A presents some discussions of the main assumptions in Section 3. Appendix B gives the proof of the main theorems. Appendix C and Appendix D show all the lemmas and their proofs, respectively. An online supplemental document (Zhou et al., 2019) contains Appendices E–G where the remaining proofs of Lemma 8 and Lemma 9 are proven in Appendix E, the additional Monte-Carlo results are placed in Appendix F and additional empirical results are shown in Appendix G.

Throughout this paper, the following notation is used. I_d is the d-dimensional identity matrix; $[a]$ is the maximum integer not exceeding a; $\mathbb{R}$ is the real line; and, for any function

[2]Without the identification condition that $\|\theta_0\|^2 = 1$, $\widehat{\theta}_n$ will degenerate along θ_0 direction.

$f(\cdot)$, $f^{(1)}(x)$, $f^{(2)}(x)$ and $f^{(3)}(x)$ are the derivatives of the first, second and third order of $f(\cdot)$ at x. $\|\cdot\|$ is the Euclidean norm for vectors and element-wise norm for matrices—that is, if $A = (a_{ij})_{nm}$, $\|A\| = (\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}^2)^{1/2}$; Convergence in probability and convergence in distribution are signified as $\to_P$ and $\to_D$.

2 Estimation procedure

Suppose that the link function $g_0(w)$ belongs to the Hilbert space $L^2(\mathbb{R}, \exp(-w^2/2))$, which is a very useful space covering a great deal of functions on $\mathbb{R}$, such as polynomials, power functions, and bounded functions. It is known that Hermite polynomials form a complete orthogonal system in the Hilbert space $L^2(\mathbb{R}, \exp(-w^2/2))$ with each element defined by

$$h_i(w) = (-1)^i \exp(w^2/2) \frac{d^i}{dw^i} \exp(-w^2/2), \quad i = 0, 1, 2, ..., \tag{2.1}$$

and the orthogonality gives $\int h_i(w) h_j(w) \exp(-w^2/2) dw = \sqrt{2\pi} i! \delta_{ij}$, where δ_{ij} is the Kronecker delta. Based on this property, we define the standardized Hermite polynomials as

$$H_i(w) = (\sqrt{2\pi} i!)^{-1/2} h_i(w), \tag{2.2}$$

and hence, $\{H_i(w)\}$ becomes a complete orthonormal basis in $L^2(\mathbb{R}, \exp(-w^2/2))$ satisfying $\int H_i(w) H_j(w) \exp(-w^2/2) dw = \delta_{ij}$. Then we have an orthogonal series expansion for any $g_0(w) \in L^2(\mathbb{R}, \exp(-w^2/2))$ as follows

$$g_0(w) = \sum_{i=0}^{\infty} c_{0,i} H_i(w), \tag{2.3}$$

where $c_{0,i} = \int g_0(w) H_i(w) \exp(-w^2/2) dw$.

The standardized Hermite polynomials can be listed as follows

$$H_0(w) = \frac{1}{\sqrt{\sqrt{2\pi}}} \cdot 1, \qquad H_1(w) = \frac{1}{\sqrt{\sqrt{2\pi}}} \cdot w,$$
$$H_2(w) = \frac{1}{\sqrt{2\sqrt{2\pi}}} \cdot (w^2 - 1), \qquad H_3(w) = \frac{1}{\sqrt{6\sqrt{2\pi}}} \cdot (w^3 - 3w),$$

and so on.

By virtue of (2.3), model (1.2) can be represented as

$$y_t = \mathcal{H}_k(\theta_0^\top x_{t-1})^\top C_{0,k} + \gamma_k(\theta_0^\top x_{t-1}) + e_t, \quad t = 1, ..., n, \tag{2.4}$$

where $x_{t-1} = (x_{1,t-1}, \ldots, x_{d,t-1})^\top$, $\mathcal{H}_k(\cdot) = (H_0(\cdot), ..., H_{k-1}(\cdot))^\top$, $C_{0,k} = (c_{0,0}, ..., c_{0,k-1})^\top$, and $\gamma_k(\cdot) = \sum_{i=k}^{\infty} c_{0,i} H_i(\cdot)$. Throughout this paper, let k be the truncation parameter and $k \to \infty$ as $n \to \infty$. We then define $g_k(w) = \mathcal{H}_k(w)^\top C_{0,k} = \sum_{i=0}^{k-1} c_{0,i} H_i(w)$, which converges to $g_0(w)$ under certain conditions.

Let $Y = (y_1, ..., y_n)^\top$, $Z = (\mathcal{H}_k(\theta_0^\top x_0), ..., \mathcal{H}_k(\theta_0^\top x_{n-1}))^\top$ an $n \times k$ matrix, $\gamma = (\gamma_k(\theta_0^\top x_0), ..., \gamma_k(\theta_0^\top x_{n-1}))^\top$, and $e = (e_1, ..., e_n)^\top$. We have a matrix form equation

$$Y = ZC_{0,k} + \gamma + e \tag{2.5}$$

Since our interests are in both unknown index parameter θ_0 and the unknown link function g_0, we define a 2-fold Cartesian product space by $\mathbb{R}^d$ and $L^2(\mathbb{R}, \exp(-w^2/2))$. Thus, (θ_0, g_0) can be viewed as a point in this infinite-dimensional space and this space is equipped with the norm $\|.\|_2$ given by

$$\big\|(\theta, g)\big\|_2 = \Big(\|\theta\|_2^2 + \|g\|_{L^2}^2\Big)^{1/2}, \tag{2.6}$$

Then it follows from the Parseval's equality that $\|g\|_{L^2}^2 = \int \big(g(w)\big)^2 \exp(-w^2/2) dw = \sum_{i=0}^{\infty} c_i^2$, and hence, the unknown link function $g(w)$ can be identified by its corresponding expansion coefficients $\{c_i, i = 0, 1, 2, ...\}$.

Suppose that $\Theta \subset \mathbb{R}^d$, Θ is compact, and $\theta_0 \in \Theta$. Suppose further that G is a subset of $L^2(\mathbb{R}, \exp(-w^2/2))$ and $g_0 \in G$. After taking into account the identification condition, we introduce the following objective function:

$$W_{n,\lambda}(\theta, g) = \sum_{t=1}^{n} \Big[y_t - g(\theta^\top x_{t-1})\Big]^2 + \lambda(\|\theta\|^2 - 1), \tag{2.7}$$

where $(\theta, g) \in \Theta \times G_k$ and $G_k = G \cap \text{span}\{H_0(.), H_1(.), ..., H_{k-1}(.)\}$. After the truncation, the infinite-dimensional point (θ_0, g_0) can be approximated by the finite dimensional parameter θ and function g.

Using the Hermite expansion, the objective function employed in practice is given by

$$W_{n,\lambda}(\theta, C_k) = \sum_{t=1}^{n} \Big[y_t - C_k^\top \mathcal{H}_k(\theta^\top x_{t-1})\Big]^2 + \lambda(\|\theta\|^2 - 1), \tag{2.8}$$

where $C_k = (c_0, ..., c_{k-1})^\top$, $\mathcal{H}_k(\cdot)$ is defined in equation (2.4), and k is the truncation parameter.

Suppose θ is given for the time being, the estimator of the expansion coefficient can be easily obtained from the matrix form equation (2.5) by ordinary least squares (OLS) method,

$$\bar{C}_k = \bar{C}_k(\theta) = \left(Z(\theta)^\top Z(\theta) \right)^{-1} Z(\theta)^\top Y. \tag{2.9}$$

Then, we obtain the optimum $\widehat{\theta}_n$ such that

$$\widehat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} \; W_{n,\lambda}(\theta, \bar{C}_k(\theta)). \tag{2.10}$$

We then define a plug-in estimator $\widehat{g}_n(w) = \bar{C}_k^\top \mathcal{H}_k(w)$ for any $w \in \mathbb{R}$ with $\bar{C}_k = \bar{C}_k(\widehat{\theta}_n)$.

To study the asymptotic properties of $(\widehat{\theta}_n, \widehat{g}_n)$, we need to introduce some necessary assumptions.

3 Asymptotic theory

The rest of this paper focuses on the case where x_{t-1} is nonstationary but $u_{t-1} = \theta_0^\top x_{t-1}$ is strictly stationary. To show the main results of this paper, we make the following assumptions. Their justifications are available from Appendix A below.

Assumption 1.

1. *(a) There exists a σ-field $\mathcal{F}_{n,t}$, such that $\{e_t, \mathcal{F}_{n,t}\}$ is a martingale difference sequence with $E(e_t|\mathcal{F}_{n,t-1}) = 0$ almost surely (a.s.), $E(e_t^2|\mathcal{F}_{n,t-1}) = \sigma^2$ a.s., and $\sup_{1 \le t \le n} E(e_t^4|\mathcal{F}_{n,t-1}) < \infty$ a.s*

   (b) ϵ_t are d-dimensional independent and identically distributed (i.i.d.) continuous random variables with $E(\epsilon_t) = 0$, $E(\epsilon_1 \epsilon_1^\top) = \Sigma_\epsilon$, and $E\|\epsilon_1\|^p < \infty$ for some $p > 4$.

2. *Let $x_t = x_{t-1} + v_t$ and $x_0 = O_P(1)$, where $v_t = \phi(L)\epsilon_t$ with $\phi(L) = \sum_{j=0}^\infty \phi_j L^j$ and $\{\phi_j\}$ being a sequence of $d \times d$ matrices, such that:*

   (a) $\phi_0 = I_d$

(b) $\sum_{j=0}^{\infty} j\|\phi_j\| < \infty$

(c) $\phi(1)$ *has rank* $d-1$ *and* $\theta_0^\top \phi(1) = 0$

(d) $u_t = \theta_0^\top x_t$ *is a strictly stationary process and has probability density function* $\rho(u)$, *such that* $\exp(u^2/2)\rho(u) < \infty$ *uniformly in* u.

3. *Suppose* $\left\|(\widehat{\theta}_n, \widehat{g}_n) - (\theta_0, g_0)\right\|_2 \to_P 0$ *as* $n \to \infty$.

4. *Suppose that* $g_0(w)$ *is differentiable on* $\mathbb{R}$ *and* $g_0^{(r-i)}(w)w^i \in L^2(\mathbb{R}, \exp(-w^2/2))$ *for* $0 \leq i \leq r$ *and an integer* $r \geq 4$.

5. $k = [a \cdot n^\kappa]$ *with some constant* $a > 0$ *and* $\kappa \in [1/r, 1/4)$ *with* r *as in 4 above.*

6. *Suppose that:*

(a) $\inf_{c \in \mathbb{R}} E\left[g_0(\theta_0^\top x_1) - c\right]^2 > 0$

(b) *The smallest eigenvalue of* $E\left[\mathcal{H}_k(\theta_0^\top x_1)\mathcal{H}_k(\theta_0^\top x_1)^\top\right]$ *is bounded away from zero uniformly in* $k \geq 1$.

7. *Let* $u_t = \theta_0^\top x_t$.

(a) $E\left[g_0^{(1)}(u_1)\right]^4 < \infty$

(b) $\sum_{i=0}^{k-1}\sum_{j=0}^{k-1} E\left[H_i(u_1)H_j(u_1)\right]^2 = o(n)$ *as* $(n,k) \to (\infty, \infty)$

(c) $\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|Cov\left(\left(g_0^{(1)}(u_{t-1})\right)^2, \left(g_0^{(1)}(u_{s-1})\right)^2\right)\right| = o(n^2)$ *as* $n \to \infty$

(d) $\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|Cov\left(H_i(u_{t-1})H_j(u_{t-1}), H_i(u_{s-1})H_j(u_{s-1})\right)\right| = o(n^2)$ *as* $(n,k) \to (\infty, \infty)$.

We assume that $\theta_0 \in \text{int}(\Theta)$ and use the ideas from Wooldridge (1994) to establish the asymptotic normality for the extremum estimator $\widehat{\theta}_n$. From equation (2.8), the Score $S_n(\theta)$ and the Hessian $J_n(\theta)$ are given by:

$$\begin{aligned}
S_n(\theta) &= \frac{\partial}{\partial\theta} W_{n,\widehat{\lambda}(\theta)}\bigg|_{(\theta,C_k)=(\theta,\bar{C}_k(\theta))} \\
&= -2\sum_{t=1}^{n}\left(y_t - \widehat{g}_n(\theta^\top x_{t-1})\right)\frac{\partial\widehat{g}_n(\theta^\top x_{t-1})}{\partial\theta} + 2\theta\widehat{\lambda}(\theta) \\
J_n(\theta) &= \frac{\partial^2}{\partial\theta\partial\theta^\top} W_{n,\widehat{\lambda}(\theta)}\bigg|_{(\theta,C_k)=(\theta,\bar{C}_k(\theta))}
\end{aligned}$$

$$= 2\sum_{t=1}^{n} \frac{\partial \widehat{g}_n(\theta^\top x_{t-1})}{\partial \theta} \frac{\partial \widehat{g}_n(\theta^\top x_{t-1})}{\partial \theta^\top} - 2\sum_{t=1}^{n} \left(y_t - \widehat{g}_n(\theta_0^\top x_{t-1})\right) \frac{\partial^2 \widehat{g}_n(\theta^\top x_{t-1})}{\partial \theta \partial \theta^\top} + 2\widehat{\lambda}(\theta),$$

where $\widehat{\lambda}(\theta) = (\theta^\top \theta)^{-1}\theta^\top \sum_{t=1}^{n} \left(y_t - \widehat{g}_n(\theta^\top x_{t-1})\right) \frac{\partial \widehat{g}_n(\theta^\top x_{t-1})}{\partial \theta}$ with $\theta^\top \theta \neq 0$.

Then, the asymptotic distribution of $\widehat{\theta}_n$ can be obtained by the expansion:

$$0 = S_n(\widehat{\theta}_n) = S_n(\theta_0) + J_n(\theta_n)(\widehat{\theta}_n - \theta_0), \tag{3.1}$$

where $S_n(\widehat{\theta}_n)$ and $S_n(\theta_0)$ are the Scores evaluated at $\widehat{\theta}_n$ and θ_0, respectively. $J_n(\theta_n)$ is the Hessian matrix with the rows evaluated at a point θ_n between $\widehat{\theta}_n$ and θ_0.

To further develop the theory, we need to rotate the coordinate system based on the true parameter θ_0. Let $Q = (\theta_0, Q_2)$ be a $d \times d$ orthogonal matrix. We can represent the single-index model as:

$$y_t = g_0(\theta_0^\top Q Q^\top x_{t-1}) + e_t = g_0(\alpha_0^1 x_{1t-1} + \alpha_0^{2\top} x_{2t-1}) + e_t, \tag{3.2}$$

where $\alpha_0^1 = \|\theta_0\|^2 = 1$, $\alpha_0^2 = Q_2^\top \theta_0 = 0_{d-1}$ is a $(d-1)$-dimensional zero vector, $x_{1t-1} = \theta_0^\top x_{t-1}$ is a stationary scalar process and $x_{2t-1} = Q_2^\top x_{t-1}$ is a $(d-1)$-dimensional nonstationary process. Let $\alpha_0 = (\alpha_0^1, (\alpha_0^2)^\top)^\top = Q^\top \theta_0$, and $\alpha = (\alpha^1, (\alpha^2)^\top)^\top = Q^\top \theta$ for later use. If $\widehat{\alpha}_n$ is the NLS estimator of α_0, then $\widehat{\alpha}_n = Q^\top \widehat{\theta}_n$. In addition, the Score function $S_n(\alpha)$ and the Hessian function $J_n(\alpha)$ can be derived from those for θ, such that $S_n(\alpha) = Q^\top S_n(\theta)$ and $J_n(\alpha) = Q^\top J_n(\theta) Q$. Based on these relationships, we can obtain:

$$0 = S_n(\widehat{\alpha}_n) = S_n(\alpha_0) + J_n(\alpha_n)(\widehat{\alpha}_n - \alpha_0). \tag{3.3}$$

Given that the constraint $\|\theta\|^2 = 1$ is imposed directly within the estimation procedure, a projection matrix $P_{\alpha_0} = I_d - \alpha_0 \alpha_0^\top = \begin{pmatrix} 0 & 0 \\ 0 & I_{d-1} \end{pmatrix}$ will be evolved and will project the Score function into the space orthogonal to α_0, which is a $(d-1)$-dimensional space. The projection matrix P_{α_0} has eigenvalues $0, 1, ..., 1$, where 0 corresponds to the eigenvector α_0. Let $P_1 = (p_1, ..., p_{d-1}) = \begin{pmatrix} p_{1,1} & \cdots & p_{1,d-1} \\ \vdots & \ddots & \vdots \\ p_{d,1} & \cdots & p_{d,d-1} \end{pmatrix}$ with $p_{i+1,i} = 1$ for $1 \leq i \leq d-1$ and zero otherwise. $p_1, ..., p_{d-1}$ are the eigenvectors associated with the eigenvalues 1 of P_{α_0} and are orthogonal to α_0. Therefore, we have $P_{\alpha_0} = P_1 P_1^\top$ and $P_1^\top P_1 = I_{d-1}$.

To establish the asymptotic distribution of $\widehat{\alpha}_n - \alpha_0$, we can obtain the following equation through (3.3):

$$P_1^\top D_n(\widehat{\alpha}_n - \alpha_0) = -\left(P_1^\top D_n^{-1} J_n(\alpha_0) D_n^{-1} P_1\right)^{-1} P_1^\top D_n^{-1} S_n(\alpha_0) + o_P(1), \tag{3.4}$$

where $D_n = diag(\sqrt{n}, nI_{d-1})$ and the asymptotic properties of the Score $S_n(\alpha_0)$ and the Hessian $J_n(\alpha_0)$ are discussed at Lemma 8 in Appendix C.

Given that the leading term in the Score function belongs to a $(d-1)$-dimensional space orthogonal to α_0, P_1 is used to rotate the whole Score function. Also notice that, without P_1 that transforms $\widehat{\alpha}_n - \alpha_0$ into a $(d-1)$-dimensional space, the covariance matrix must be singular.

Let $\widehat{\alpha}_n = \left(\widehat{\alpha}_n^1, (\widehat{\alpha}_n^2)^\top\right)^\top$ be the estimator for $\alpha_0 = (\alpha_0^1, (\alpha_0^2)^\top)^\top$. In view of the structure of D_n and the constraint $\|\alpha\|^2 = \|\theta\|^2 = 1$, we have two limits obtained from (C.3).

Theorem 3.1. *Under Assumption 1, as $n \to \infty$*

$$n^2(\widehat{\alpha}_n^1 - \alpha_0^1) \to_D -\frac{1}{2}\|\xi\|^2, \tag{3.5}$$

and

$$n(\widehat{\alpha}_n^2 - \alpha_0^2) \to_D \xi. \tag{3.6}$$

where $\xi := (\xi_1, ..., \xi_{d-1}) \sim \boldsymbol{MN}(0, \sigma^2 r_0^{-1})$, $\boldsymbol{MN}$ stands for mixture normal distribution, $r_0 = E\left[g_0^{(1)}(\theta_0^\top x_1)\right]^2 \left(\int_0^1 V_2(r)V_2^\top(r)dr - \int_0^1 V_2(r)dr \int_0^1 V_2(r)^\top dr\right)$, and V_2 is Brownian motion of dimension $d-1$ with variance matrix $\Sigma_V = Q_2^\top \phi(1)\Sigma_\epsilon \phi(1)^\top Q_2$.

By using the rotation technique, the estimator $(\widehat{\alpha}_n^1, (\widehat{\alpha}_n^2)^\top)^\top := Q^\top \widehat{\theta}_n = (\theta_0^\top \widehat{\theta}_n, Q_2^\top \widehat{\theta}_n)$ is the coordinates of $\widehat{\theta}_n$ in the system $Q = (\theta_0, Q_2)$ with $\widehat{\alpha}_n^1$ along the θ_0 direction and $\widehat{\alpha}_n^2$ along all the other directions orthogonal to θ_0. As can be seen from Theorem 3.1, there are two types of super-consistency rates: the higher rate of convergence $O_P(n^{-2})$ lying in the direction along θ_0, and the lower rate of convergence $O_P(n^{-1})$, which is still super-consistent, lying along all the other directions orthogonal to θ_0. Also notice that $|\widehat{\alpha}_n^1| = |\theta_0^\top \widehat{\theta}_n| \le \|\theta_0\|\|\widehat{\theta}_n\| = 1$ by Cauchy-Schwarz inequality and the equality holds when $\widehat{\theta}_n = \theta_0$, which implies that $\widehat{\alpha}_n^1$ is an under-estimator for $\alpha_0^1 = 1$.

Therefore, there are dual rates of convergences in our proposed single-index model and the asymptotic distribution for $\widehat{\theta}_n$ in the next theorem can be obtained from $\widehat{\alpha}_n$, more precisely $(\widehat{\theta}_n - \theta_0) = Q(\widehat{\alpha}_n - \alpha_0)$.

Theorem 3.2. *Under Assumption 1, as $n \to \infty$*

$$n(\widehat{\theta}_n - \theta_0) \to_D \boldsymbol{MN}(0, \sigma^2 Q_2 r_0^{-1} Q_2^\top), \tag{3.7}$$

where r_0 is the same as in Theorem 3.1.

Theorem 3.2 indicates that $\widehat{\theta}_n$ converges to θ_0 at rate of $O_P(n^{-1})$ and this is because the slower rate $O_P(n^{-1})$ along Q_2 direction will eventually dominate the faster rate $O_P(n^{-2})$ along θ_0 direction. Intuitively, the constraint $\|\theta\|^2 = 1$ scales the estimator to the surface of the unit ball, so that the norm of $\widehat{\theta}_n$ can always match that of θ_0; therefore, it accelerates the convergence rate along θ_0 direction, and hence the overall convergence rate. This n-super-rate of convergence is not a surprise to us. As has been shown in Park and Phillips (2001), if $g_0^{(1)}$ is H-regular with x_t being nonstationary, $\sqrt{n}\dot{v}(\sqrt{n})(\widehat{\theta}_n - \theta_0) = O_P(1)$. The convergence rate will be faster than $\sqrt{n}$ when $\dot{v}(\sqrt{n})$ is divergent, which is usually the case. The proposed estimation procedure in Section 2 is called the 'profile method' in the literature, and a general discussion on the asymptotic properties of profiled semiparametric estimators for the i.i.d. case can be found in Chen et al. (2003).

Meanwhile, define the estimator for σ^2 and $\mathcal{H}_x = E\left[\mathcal{H}_k(\theta_0^\top x_1)\mathcal{H}_k(\theta_0^\top x_1)^\top\right]$ by:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\left[y_t - \widehat{g}_n(\widehat{\theta}_n^\top x_{t-1})\right]^2 \qquad \widehat{\mathcal{H}}_x = \frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_k(\widehat{\theta}_n^\top x_{t-1})\mathcal{H}_k(\widehat{\theta}_n^\top x_{t-1})^\top. \tag{3.8}$$

We then establish the central limit theorem (CLT) for the plug-in estimator $\widehat{g}_n(w) = \mathcal{H}_k(w)^\top \bar{C}_k(\widehat{\theta}_n)$, given $w \in \mathbb{R}$.

Theorem 3.3. *Under Assumption 1, as $n \to \infty$*

$$\sqrt{n}\widehat{\Sigma}^{-1}(w)\left(\widehat{g}_n(w) - g_0(w) + \gamma_k(w)\right) \to_D N(0,1), \tag{3.9}$$

where $\widehat{\Sigma}^2(w) = \widehat{\sigma}^2\mathcal{H}_k(w)^\top\widehat{\mathcal{H}}_x^{-1}\mathcal{H}_k(w)$ is the estimator of $\Sigma^2(w) = \sigma^2\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)$ and $\gamma_k(w) = g_0(w) - \sum_{i=0}^{k-1} c_{0,i}H_i(w) = \sum_{i=k}^{\infty} c_{0,i}H_i(w)$.

The order involved in the normality is $O_P(1)n^{1/2}k^{-1/2}$ in view of $\left|\Sigma(w)\right|^2 = O(1)k$, and it is not a super rate. This is because we assume that $\theta_0^\top x_{t-1} \sim I(0)$ and (3.9) is a standard result in the literature for the nonparametric series estimation. The term $\gamma_k(w)$ is the bias of the estimator $\widehat{g}_n(w)$ and $\sqrt{n}\widehat{\Sigma}^{-1}(w)\gamma_k(w) = O_P(1)$ under Assumption 1.5.

Before the proofs of Theorem 3.1 - Theorem 3.3 are given in Appendix B, we now discuss how to computationally implement our proposed model and how to construct the confidence interval for $g_0(.)$ in practice, and then evaluate the finite-sample performance of $\widehat{\theta}_n$, $\widehat{\alpha}_n$ and $\widehat{g}_n$ in Section 4 below.

4 Numerical results

In this section, we conduct Monte-Carlo simulations to examine the finite-sample performance of the proposed estimators in the single-index model.

4.1 Computational aspects

To conduct the optimisation of (2.10) in practice, we introduce the estimation procedures using a bivariate case ($x_t = (x_{1,t}, x_{2,t})^\top$) as follows:

1. Conduct a cointegration test on x_t to see whether they are cointegrated or not.

2. Estimate the cointegrated coefficient for x_t from the cointegrated model $x_{1,t} = \theta x_{2,t} + z_t$ and the estimate is denoted as $\tilde{\beta}$. Let $\tilde{\theta} = (1, -\tilde{\beta})$ and we use $\tilde{\theta}_0 = \tilde{\theta}/\|\tilde{\theta}\|$ as the initial value for the NLS estimation algorithm.[3]

3. For given data $\{(x_{t-1}, y_t), 1 \le t \le n\}$, estimate (θ_0, g_0) by our proposed estimation procedure in Section 2 and denote the resulting estimates by $(\widehat{\theta}_n, \widehat{g}_n)$. The value for the truncation parameter k can be chosen by theory driven value $k = [a \cdot n^\kappa]$ with some constants a and κ that satisfy Assumption 1.5. Alternatively, we can consider some statistics to help us determine k. In this paper, we consider two methods to select the optimal truncation parameter. The first is the Generalised Cross-Validation (GCV) method proposed by Gao et al. (2002), which selects an optimal value $\widehat{k}$ such that:

$$\widehat{k} = \underset{k \in \mathcal{K}}{\operatorname{argmin}} \left(1 - \frac{k}{n}\right)^{-2} \widehat{\sigma}_1^2(k), \tag{4.1}$$

   where $\widehat{\sigma}_1^2(k) = \frac{1}{n}\sum_{t=1}^n \left(r_t - \widehat{g}_n(\widehat{\theta}_n^\top x_{t-1})\right)^2$ and $\mathcal{K} = \{2, \cdots, K_0\}$ with K_0 predetermined.

   The second method is a nonparametric version of Akaike information criterion (AIC) (see Cai, 2007) that selects the truncation parameter $\widehat{k}$ such that:

$$\widehat{k} = \underset{k \in \mathcal{K}}{\operatorname{argmin}} \log\left(\widehat{\sigma}_1^2(k)\right) + 2\frac{n_\lambda + 1}{n - n_\lambda - 2}, \tag{4.2}$$

[3] Given that we assume that $\theta_0^\top x_t \sim I(0)$, the estimated cointegrating coefficient is a consistent initial estimate.

where n_λ is the trace of $Z(\widehat{\theta}_n)(Z(\widehat{\theta}_n)^\top Z(\widehat{\theta}_n))^{-1}Z(\widehat{\theta}_n)^\top$, which is called the effective number of parameters or the nonparametric version of degrees of freedom for nonparametric models.

4. For given w, according to the CLT in Theorem 3.3, the 95% confidence interval of $g_0(w)$ is given by:

$$\left[\widehat{g}_n(w) - 1.96 \times \widehat{SD}(\widehat{g}_n(w)), \widehat{g}_n(w) + 1.96 \times \widehat{SD}(\widehat{g}_n(w))\right],$$

where $\widehat{g}_n(w) = \mathcal{H}_k(w)^\top \bar{C}_k$, $\widehat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^n \left(y_t - \widehat{g}_n(\widehat{\theta}_n^\top x_{t-1})\right)^2$ and $\widehat{SD}^2(\widehat{g}_n(w)) = \frac{1}{n}\widehat{\sigma}^2 \mathcal{H}_k(w)^\top \left(\frac{1}{n}\sum_{t=1}^n \mathcal{H}_k(\widehat{\theta}_n^\top x_{t-1})\mathcal{H}_k(\widehat{\theta}_n^\top x_{t-1})^\top\right)^{-1} \mathcal{H}_k(w)$.

4.2 Simulation experiments

Let $d = 2$ and $x_t = (x_{1,t}, x_{2,t})^\top$ be generated by:

$$x_t = x_{t-1} + v_{it}, \quad t = 1, ..., n \quad \text{and} \quad i = 1, 2,$$

$$v_{1t} = \epsilon_t + C_{10}\epsilon_{t-1} \tag{4.3}$$

$$v_{2t} = A_{20}v_{2t-1} + \epsilon_t + C_{20}\epsilon_{t-1} \tag{4.4}$$

where $\epsilon_t \sim iiN(0, \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right))$ and $x_{-500} = 0_2$ surely. In addition, set $C_{10} = \left(\begin{smallmatrix} -1 & 4/3 \\ 0 & 0 \end{smallmatrix}\right)$, $A_{20} = \left(\begin{smallmatrix} 2/5 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ and $C_{20} = \left(\begin{smallmatrix} -1 & 4/5 \\ 0 & 0 \end{smallmatrix}\right)$. Case (4.3) assumes MA(1) process for the innovations of $I(1)$ variables, and case (4.4) considers a VARMA(1,1) that can be rewritten as an infinite MA process according to the Wold representation theorem. Both settings are consistent with Assumption 1.2. Under these two settings, $x_{1,t}$ and $x_{2,t}$ are cointegrated with cointegrating vector $\theta_0 = (0.6, -0.8)^\top$, which satisfies the identification condition $\|\theta_0\|^2 = 1$, and, hence, $Q_2 = (0.8, 0.6)^\top$. The simulation is conducted with sample sizes $n = 100$, 200, 600, 1000, and the Monte-Carlo replication $M = 2000$. The truncation parameter k is determined by the GCV method described in Section 4.1.[4] The initial value for the estimation procedure is set at the standardised estimated cointegrating coefficient and is a consistent initial estimate.

[4]We use the average value of $\widehat{k}$ from another 100 replications. The use of nonparametric AIC produces identical results.

The single-index model is given by $y_t = g_0(\theta_0^\top x_{t-1}) + e_t$ with $e_t \sim iiN(0,1)$. We next consider four options for the link function:

$$(a).\ g_{10}(w) = 1 + w$$
$$(b).\ g_{20}(w) = 1 + w^2$$
$$(c).\ g_{30}(w) = \exp(w)$$
$$(d).\ g_{40}(w) = (1 + w^2)^{-1}.$$

It is clear that the first three link functions are unbounded on $\mathbb{R}$, and the last is bounded on $\mathbb{R}$. More importantly, $g_{i0}(w) \in L^2(\mathbb{R}, \exp(-w^2/2))$ for $i = 1, 2, 3, 4$.

Table 1: Bias and standard deviation for single-index model for case (4.3)

$g_0(u)$		Bias				S.d.			
	n	100	200	600	1000	100	200	600	1000
(a)	$\widehat{\theta}_n^1$	5.6820 $\times 10^{-4}$	-2.1518 $\times 10^{-4}$	-2.6838 $\times 10^{-5}$	8.3123 $\times 10^{-6}$	0.0172	0.0078	0.0026	0.0017
	$\widehat{\theta}_n^2$	7.1917 $\times 10^{-4}$	-1.0116 $\times 10^{-4}$	-1.3459 $\times 10^{-5}$	8.9217 $\times 10^{-6}$	0.0132	0.0059	0.0020	0.0012
	$\widehat{\alpha}_n^1$	-2.3442 $\times 10^{-4}$	-4.8182 $\times 10^{-5}$	-5.3353 $\times 10^{-6}$	-2.1499 $\times 10^{-6}$	7.2664 $\times 10^{-4}$	1.0668 $\times 10^{-5}$	1.0975 $\times 10^{-5}$	4.3762 $\times 10^{-6}$
	$\widehat{\alpha}_n^2$	-2.3442 $\times 10^{-4}$	-4.8182 $\times 10^{-5}$	-5.3353 $\times 10^{-6}$	-2.1499 $\times 10^{-6}$	0.0216	0.0098	0.0033	0.0021
(b)	$\widehat{\theta}_n^1$	2.1118 $\times 10^{-4}$	1.3483 $\times 10^{-4}$	1.4560 $\times 10^{-5}$	2.7273 $\times 10^{-6}$	0.0135	0.0061	0.0019	0.0011
	$\widehat{\theta}_n^2$	3.3963 $\times 10^{-4}$	1.3754 $\times 10^{-4}$	-1.4305 $\times 10^{-5}$	3.2635 $\times 10^{-6}$	0.0104	0.0046	0.0014	8.3773 $\times 10^{-4}$
	$\widehat{\alpha}_n^1$	-1.4499 $\times 10^{-4}$	-2.9132 $\times 10^{-5}$	-2.7077 $\times 10^{-6}$	-9.7440 $\times 10^{-7}$	7.6506 $\times 10^{-4}$	6.2255 $\times 10^{-5}$	5.4454 $\times 10^{-6}$	2.0854 $\times 10^{-6}$
	$\widehat{\alpha}_n^2$	3.7272 $\times 10^{-4}$	1.9039 $\times 10^{-4}$	2.0231 $\times 10^{-5}$	4.1399 $\times 10^{-6}$	0.0170	0.0076	0.0023	0.0014
(c)	$\widehat{\theta}_n^1$	-5.0755 $\times 10^{-4}$	-9.2284 $\times 10^{-5}$	-1.6759 $\times 10^{-5}$	-8.8116 $\times 10^{-6}$	0.0112	0.0053	0.0016	9.8054 $\times 10^{-4}$
	$\widehat{\theta}_n^2$	-2.5732 $\times 10^{-4}$	9.6652 $\times 10^{-5}$	-1.0010 $\times 10^{-5}$	-5.6702 $\times 10^{-6}$	0.0084	0.0040	0.0012	7.3533 $\times 10^{-4}$
	$\widehat{\alpha}_n^1$	-9.8673 $\times 10^{-5}$	-2.1951 $\times 10^{-5}$	-2.0479 $\times 10^{-6}$	-7.5077 $\times 10^{-7}$	1.9668 $\times 10^{-4}$	4.6332 $\times 10^{-5}$	3.7910 $\times 10^{-6}$	1.3660 $\times 10^{-6}$
	$\widehat{\alpha}_n^2$	-5.6043 $\times 10^{-4}$	1.3182 $\times 10^{-4}$	-1.9413 $\times 10^{-5}$	-1.0451 $\times 10^{-5}$	0.0140	0.0066	0.0020	0.0012
(d)	$\widehat{\theta}_n^1$	-0.0332	-0.0130	1.4519 $\times 10^{-4}$	4.0386 $\times 10^{-5}$	0.1756	0.1140	0.0080	0.0038
	$\widehat{\theta}_n^2$	0.0598	0.0248	1.7172 $\times 10^{-4}$	4.4294 $\times 10^{-5}$	0.3164	0.2037	0.0060	0.0028
	$\widehat{\alpha}_n^1$	-0.0678	-0.0276	-5.0264 $\times 10^{-5}$	-1.1204 $\times 10^{-5}$	0.2740	0.1776	2.7537 $\times 10^{-4}$	2.2445 $\times 10^{-5}$
	$\widehat{\alpha}_n^2$	0.0093	0.0045	2.1918 $\times 10^{-4}$	5.8885 $\times 10^{-5}$	0.2364	0.1515	0.0100	0.0047

The aim of this simulated setting is to illustrate the asymptotic results in Theorem 3.1 and Theorem 3.2. Actually, the rotation technique is not necessary in practice because we will never know the value for the true parameter θ_0 and its corresponding rotation matrix Q. It is only used as a tool to develop the asymptotic theory and can help us better understand the theory.

Table 2: Bias and standard deviation for single-index model for case (4.4)

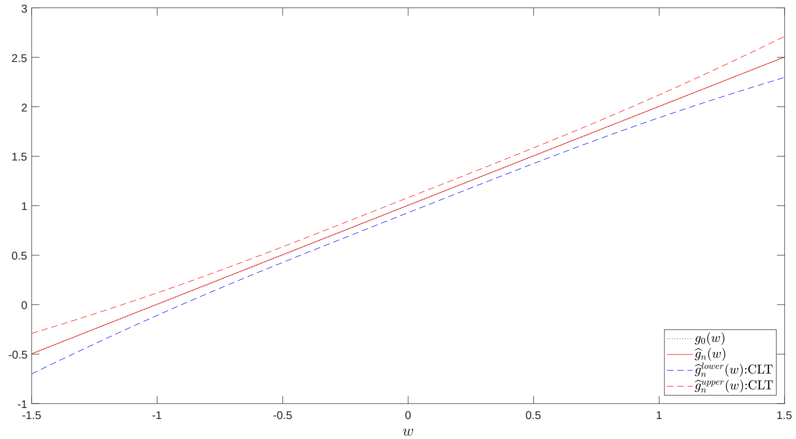
$g_0(u)$		Bias				S.d.			
	n	100	200	600	1000	100	200	600	1000
(a)	$\widehat{\theta}_n^1$	4.4993×10^{-4}	9.2089×10^{-5}	3.7655×10^{-5}	1.3384×10^{-6}	0.0173	0.0083	0.0026	0.0015
	$\widehat{\theta}_n^2$	6.3140×10^{-4}	1.3625×10^{-4}	-3.4817×10^{-5}	3.3417×10^{-6}	0.0131	0.0062	0.0019	0.0012
	$\widehat{\alpha}_n^1$	-2.3516×10^{-4}	-5.3745×10^{-5}	-5.2609×10^{-6}	-1.8704×10^{-6}	5.2595×10^{-4}	1.1703×10^{-4}	1.1005×10^{-5}	3.6561×10^{-6}
	$\widehat{\alpha}_n^2$	7.3878×10^{-4}	1.5542×10^{-4}	5.1014×10^{-5}	3.0757×10^{-6}	0.0217	0.0104	0.0032	0.0019
(b)	$\widehat{\theta}_n^1$	-2.5591×10^{-4}	6.6924×10^{-5}	6.2712×10^{-5}	-2.2750×10^{-5}	0.0124	0.0054	0.0017	9.8457×10^{-4}
	$\widehat{\theta}_n^2$	-4.3268×10^{-5}	7.8797×10^{-5}	4.9923×10^{-5}	-1.6116×10^{-5}	0.0092	0.0041	0.0013	7.3832×10^{-4}
	$\widehat{\alpha}_n^1$	-1.1893×10^{-4}	-2.2883×10^{-5}	-2.3116×10^{-6}	-7.5726×10^{-7}	3.8760×10^{-4}	5.0733×10^{-5}	4.5696×10^{-6}	1.4376×10^{-6}
	$\widehat{\alpha}_n^2$	-2.3069×10^{-4}	1.0082×10^{-4}	-8.0124×10^{-5}	-2.7870×10^{-5}	0.0154	0.0068	0.0021	0.0012
(c)	$\widehat{\theta}_n^1$	1.0929×10^{-4}	4.0062×10^{-5}	2.2476×10^{-5}	-3.9887×10^{-6}	0.0111	0.0051	0.0015	9.0131×10^{-4}
	$\widehat{\theta}_n^2$	2.0266×10^{-4}	5.5099×10^{-5}	1.9142×10^{-5}	-2.1986×10^{-6}	0.0084	0.0038	0.0011	6.7596×10^{-4}
	$\widehat{\alpha}_n^1$	-9.6553×10^{-5}	-2.0042×10^{-5}	-1.8278×10^{-6}	-6.3434×10^{-7}	2.2034×10^{-4}	4.3659×10^{-5}	3.8001×10^{-6}	1.4144×10^{-6}
	$\widehat{\alpha}_n^2$	2.0902×10^{-4}	6.5109×10^{-5}	2.9466×10^{-5}	-4.5102×10^{-6}	0.0139	0.0063	0.0019	0.0011
(d)	$\widehat{\theta}_n^1$	-0.0249	-0.0129	-2.7656×10^{-4}	1.0647×10^{-5}	0.1717	0.1012	0.0109	0.0041
	$\widehat{\theta}_n^2$	0.0678	0.0166	-9.9898×10^{-5}	2.4583×10^{-5}	0.3222	0.1774	0.0074	0.0031
	$\widehat{\alpha}_n^1$	-0.0692	-0.0211	-8.6019×10^{-5}	-1.3278×10^{-5}	0.2696	0.1575	9.7176×10^{-4}	4.9688×10^{-5}
	$\widehat{\alpha}_n^2$	0.0207	-3.2857×10^{-4}	-2.8119×10^{-4}	2.3268×10^{-5}	0.2462	0.1299	0.0131	0.0052

The simulation results of the bias and the standard deviation for $\widehat{\theta}_n = (\widehat{\theta}_n^1, \widehat{\theta}_n^2)^\top$ and $\widehat{\alpha}_n = (\widehat{\alpha}_n^1, \widehat{\alpha}_n^2)^\top$ are summarised in Table 1 and Table 2. We can observe that $\widehat{\theta}_n^1$ and $\widehat{\theta}_n^2$ under all four link functions have similar performance. In general, the biases and standard deviations for $\widehat{\theta}_n$ decrease with the increase of the sample size n and the convergence speed is quite fast,[5] which verifies the asymptotic theory in Theorem 3.2 that $\widehat{\theta}_n - \theta_0 = O_P(n^{-1})$. In terms of the rotated estimator $\widehat{\alpha}_n$, both the biases and standard deviations are approaching zero with the sample size increasing. Moreover, $\widehat{\alpha}_n^1$ converges at a faster rate than $\widehat{\alpha}_n^2$. This is implied by Theorem 3.1 that $\widehat{\alpha}_n^1 - \alpha_0^1 = O_P(n^{-2})$ and $\widehat{\alpha}_n^2 - \alpha_0^2 = O_P(n^{-1})$. It is noteworthy that the biases of $\widehat{\alpha}_n^1$ are always negative, which verifies that $\widehat{\alpha}_n^1$ is an under-estimator for α_0^1.

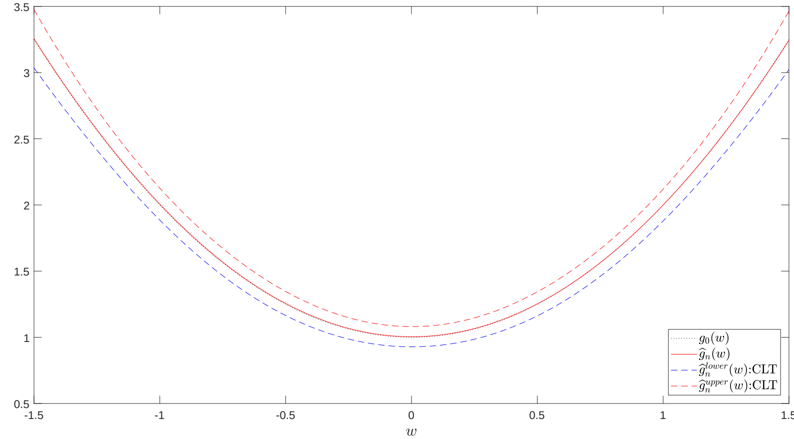
Next, we move on to examine the CLT results in Theorem 3.3. The 95% confidence intervals of $g_0(w)$ are constructed using the procedure described in Section 4.1. In terms

[5] Under the stationary setting, it is well known that $\widehat{\theta}_n - \theta_0 = O_P(n^{-1/2})$. When n = 1000, the magnitude of the s.d. is about $1000^{-1/2} = 0.0316$; however, under our setting, the s.d. is 10 times smaller than the usual case and is of magnitude around $1000^{-1} = 0.001$.

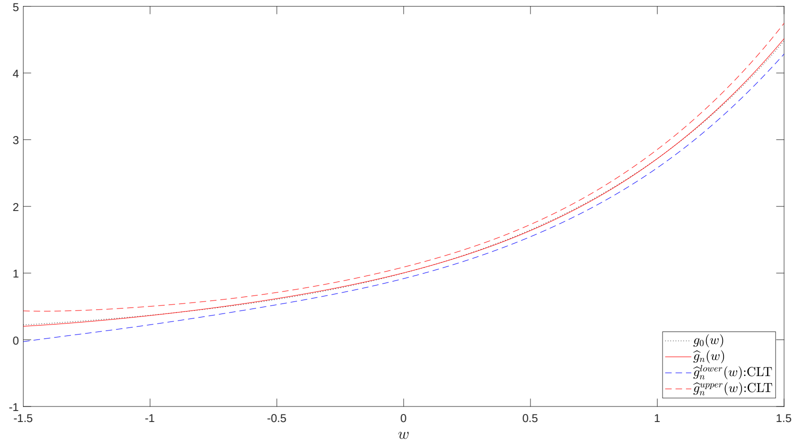
Figure 1: 95% confidence interval for case (4.3) (n=1000)



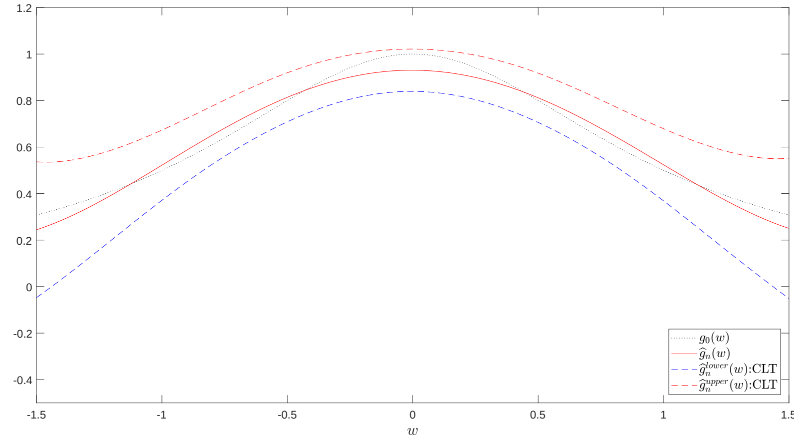
(a) $g_{10}(w) = 1 + w$



(b) $g_{20}(w) = 1 + w^2$

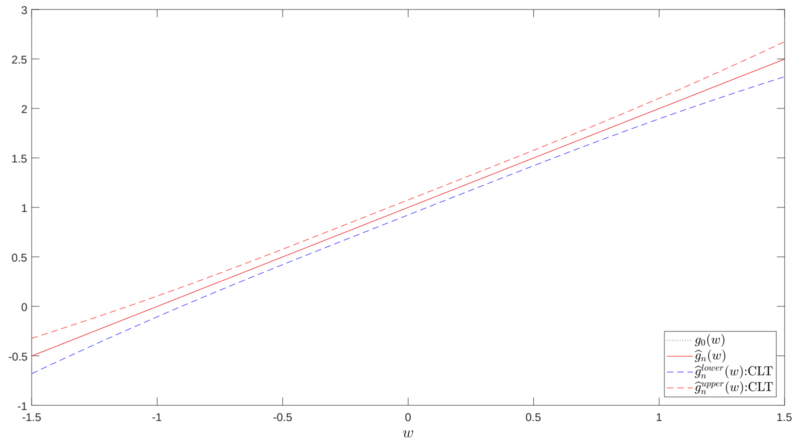


(c) $g_{30}(w) = \exp(w)$

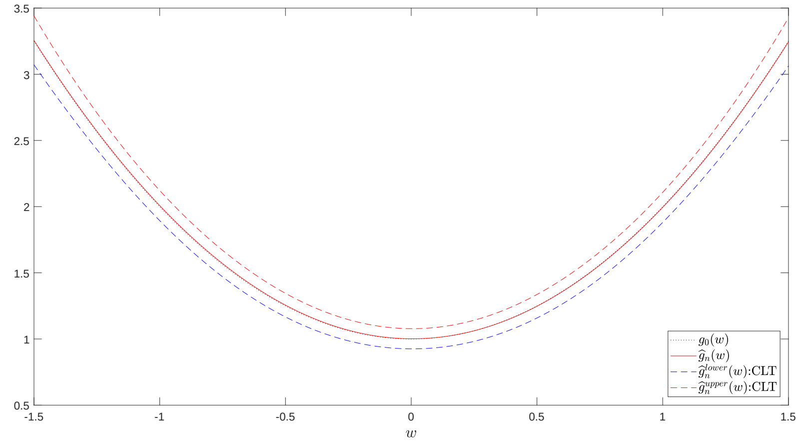


(d) $g_{40}(w) = (1 + w^2)^{-1}$

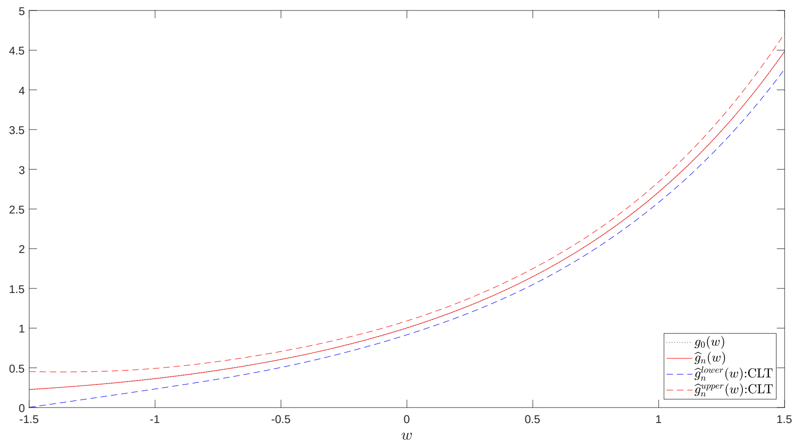
Figure 2: 95% confidence interval for case (4.4) (n=1000)



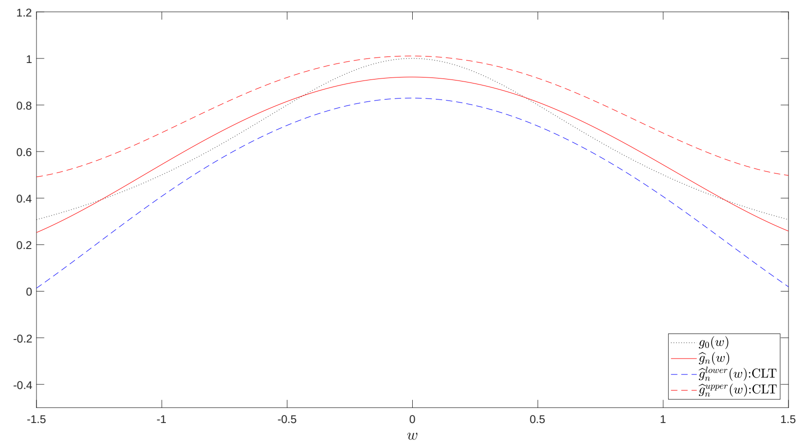
(a) $g_{10}(w) = 1 + w$



(b) $g_{20}(w) = 1 + w^2$



(c) $g_{30}(w) = \exp(w)$



(d) $g_{40}(w) = (1 + w^2)^{-1}$

of the evaluation point w, we use n evenly spaced points between -1.5 and 1.5.

We plot the average values of the estimate $\widehat{g}_n(w)$ and the 95% pointwise confidence interval for each function based on $M = 1000$ replicated data when $n = 1000$ in Figure 1 and Figure 2. All the figures show that the 95% pointwise confidence interval constructed from the asymptotic normality covers $g_0(w)$ very well and the plot of $\widehat{g}_n(w)$ seems to coincide with the plot of $g_0(w)$, which supports the result in Theorem 3.3.

In addition, we also consider an empirical example in Section 5 below.

5 Empirical study

There is now a large quantity of empirical literature examining the predictability of stock returns using a variety of lagged financial and macroeconomic variables, including dividend-price ratio, earning-price ratio, dividend-payout ratio, book-to-market ratio, interest rates, term spreads and default spreads; see, for example, Lettau and Ludvigson (2001), Cochrane (2011) and Rapach and Zhou (2013). Numerous studies, including those by Campbell and Yogo (2006) and Kostakis et al. (2015), have found evidence that many of these predictor variables are highly persistent and are often integrated of order one. If these variables are cointegrated, our semiparametric single-index predictive model can be used to test the predictability of stock returns.[6]

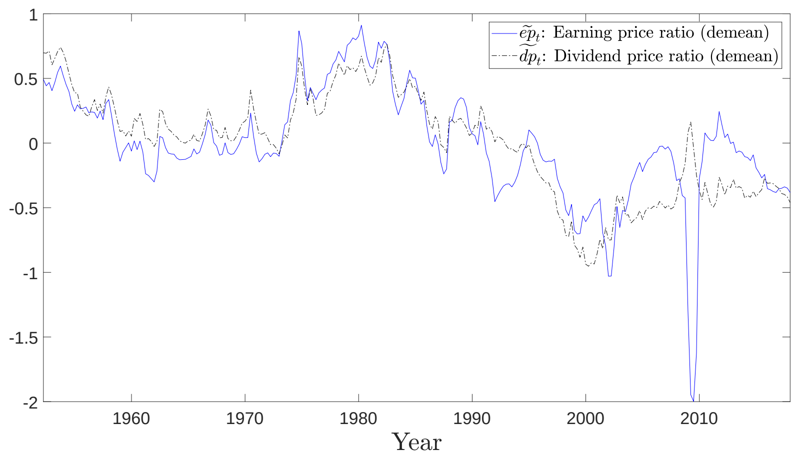
We extend the univariate linear predictive regression model of Welch and Goyal (2008), focusing on predictors that can plausibly be modelled as cointegrated. We use their updated monthly and quarterly data over the 1927 to 2017 sample period.[7] Their dataset is one of the most widely used datasets in empirical finance. The dependent variable is the United States (US) equity premium, which is defined as the log return on the S&P 500 index, including dividends minus the log return on a risk-free bill.

Among the 16 financial and macroeconomic variables used by Welch and Goyal (2008) to predict the equity premium, we consider the following four pairs of $I(1)$ variables for which the two variables in each pair are potentially cointegrated: (a) dividend-price ratio

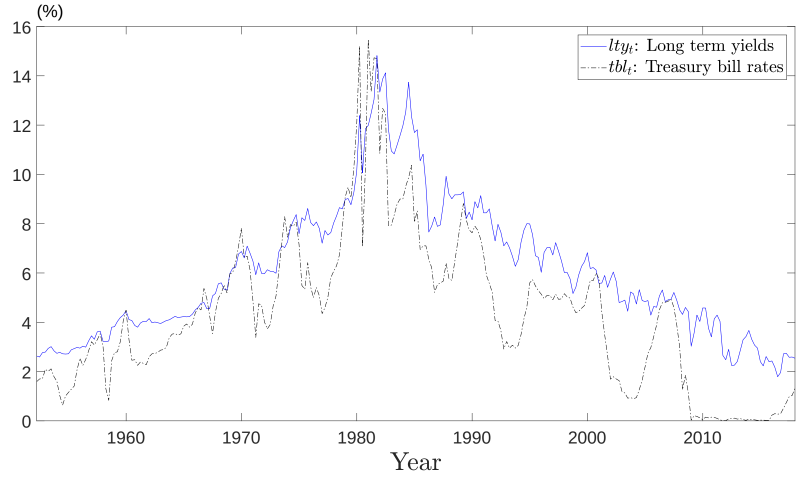
[6] Recent studies by Koo et al. (2016) and Xu (2017) have found evidence that a subset of these integrated predictors are cointegrated.

[7] The dataset was obtained from Amit Goyal's website at `http://www.hec.unil.ch/agoyal/`.

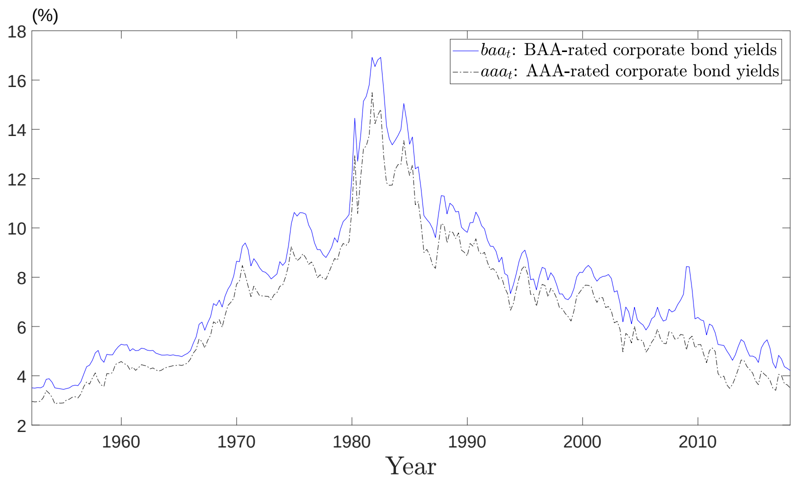
Figure 3: Time-series plots of cointegrated predictors



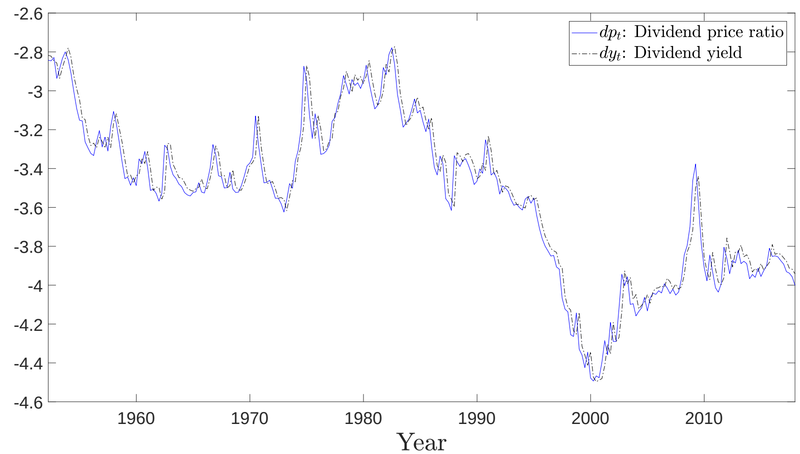
(a) ep and dp (demean)



(b) lty and tbl



(c) baa and aaa



(d) dp and dy

Notes: This figure plots the following four pairs of cointegrated predictors: (a) ep (earning-price ratio) and dp (dividend-price ratio), (b) tbl (T-bill rate) and lty (long-term yield), (c) BAA and AAA (-rated corporate bond yields), and (d) dp and dy (dividend yield). The sample period is 1952:Q1 to 2017:Q4.

(dp) and earning-price ratio (ep); (b) three-month T-bill rate (tbl) and long-term yield (lty); (c) BAA- & AAA- rated corporate bond yields; and (d) dp and dividend yield (dy). Welch and Goyal (2008) provided the definitions and sources of these predictors.

For initial illustration, Figure 3 plots the four pairs of cointegrated variables using quarterly data in the sub-period 1952 to 2017, and demonstrated that each pair of variables appeared to be cointegrated. In addition, visual inspection of Figure 1 in Campbell and Yogo (2006) suggests that cointegration is plausible between dp and ep. Fama and French (1989) used the term spread (which is tbl minus lty) and the default spread (which is BAA minus AAA) to predict the equity premium, and, under the assumption that these spreads are stationary, their paper implies that tbl-lty and BAA-AAA are modelled in cointegrating relationships.

A preliminary unit test indicates that every variable has a unit root, while the Engle-Granger Cointegration test suggests the existence of cointegration in each of the four pairs. These tests provide statistical evidence supporting the impression of co-movement behaviour from visually inspecting Figure 3. We now test the hypothesis that the US equity premium is predictable using a linear combination of a pair of $I(1)$ predictors, via the following semiparametric single-index predictive regression model:

$$y_t = d_0 + d_1 u_{t-1} + d_2 u_{t-1}^2 + ... + d_{k-1} u_{t-1}^{k-1} + e_{k,t}, \tag{5.1}$$

with $e_{k,t} = \gamma_k(u_{t-1}) + e_t$, while the truncation parameter k is determined by the GCV method described in Section 4.1.[8] Under the null hypothesis of no predictability, $d_1 = d_2 = ... = d_{k-1} = 0$; thus, the model (5.1) reduces to the constant expected equity premium model. Given that $u_{t-1} \sim I(0)$, the no-predictability null hypothesis can be tested using F-statistic. The OLS coefficient estimates in (5.1) and their conventional standard errors can be obtained in the standard way from a multiple regression of y_t on $1, u_{t-1}, u_{t-1}^2, ..., u_{t-1}^k$.

Table 3 reports the least-squares estimates of the coefficients in (5.1) and the results of the F-tests under the null hypothesis of no predictability. Numbers in parentheses below the coefficients are t-ratios and below the F-tests are p-values. Panels A and C report

[8]The use of nonparametric AIC produces identical results. We provide the results of both the GCV and nonparametric AIC methods in the online supplemental material Appendix G.

Table 3: Estimates of the single index model parameters and predictive test

Pair of predictors	$\widehat{d_0}$	$\widehat{d_1}$	$\widehat{d_2}$	$\widehat{d_3}$	$\widehat{d_4}$	F-test	$\bar{R}^2$
				Panel A: 1927Q1 - 2017Q4			
$\widetilde{\text{ep}}$ and $\widetilde{\text{dp}}$	0.0221***	0.0685**	−0.2288**	−0.1998***		4.1762	0.0256
	(3.3005)	(2.3507)	(−2.2317)	(−2.9103)		(0.0063)	
lty and tbl	0.0167***	0.5449				2.1357	0.0031
	(2.9890)	(1.4614)				(0.1448)	
BAA and AAA	0.0128*	9.0449**	−1602.0866***	52 371.3111***		11.2738	0.0785
	(1.8883)	(2.5737)	(−4.0084)	(4.9546)		(0.0000)	
dp and dy	0.0223***	−0.4421***	0.7292**	7.0280***		10.2787	0.0714
	(3.5254)	(−3.8853)	(2.2705)	(5.4615)		(0.0000)	
				Panel B: 1952Q1 - 2017Q4			
$\widetilde{\text{ep}}$ and $\widetilde{\text{dp}}$	0.0192***	0.0585***	−0.0942**	−0.0848***		4.1318	0.0346
	(3.3997)	(2.7383)	(−1.9821)	(−2.8597)		(0.0069)	
lty and tbl	0.0035	3.3153***	−30.5827	−3713.8087**		4.8013	0.0417
	(0.5064)	(3.7572)	(−1.1124)	(−2.5512)		(0.0028)	
BAA and AAA	0.0234***	3.0394**				3.9085	0.0110
	(3.6901)	(1.9770)				(0.0491)	
dp and dy	0.1483***	−1.2062***	2.4386***			6.7875	0.0423
	(4.0082)	(−3.4235)	(3.0305)			(0.0013)	
				Panel C: 1927M01 - 2017M12			
$\widetilde{\text{ep}}$ and $\widetilde{\text{dp}}$	0.0078***	0.0280***	−0.0779***	−0.0660***		6.3716	0.0146
	(3.8704)	(3.3305)	(−2.7900)	(−3.5847)		(0.0003)	
lty and tbl	0.0060***	0.1776*				3.0217	0.0019
	(3.4961)	(1.7383)				(0.0824)	
BAA and AAA	−0.0027	4.6983***	−556.7487***	14 845.4336***		8.8212	0.0211
	(−0.6603)	(2.8003)	(−3.6998)	(4.3978)		(0.0000)	
dp and dy	−0.0032	0.1956	0.8806	−26.2959***	75.3145***	9.8392	0.0314
	(−0.4405)	(1.4243)	(1.5099)	(−4.5989)	(4.1201)	(0.0000)	
				Panel D: 1952M01 - 2017M12			
$\widetilde{\text{ep}}$ and $\widetilde{\text{dp}}$	0.0067***	0.0183**	−0.0776**	−0.0646***		4.3291	0.0125
	(3.7773)	(2.0161)	(−2.3023)	(−2.9177)		(0.0049)	
lty and tbl	0.0030	0.8977***	−14.0718*	−844.6374**		4.5656	0.0134
	(1.4755)	(3.4994)	(−1.8033)	(−2.0658)		(0.0035)	
BAA and AAA	0.0079***	1.8151***	−203.2458**			5.6802	0.0117
	(4.6176)	(3.0778)	(−2.4959)			(0.0036)	
dp and dy	0.0244**	−0.0705**				4.2858	0.0041
	(2.5816)	(−2.0702)				(0.0388)	

Notes: This table reports ordinary least squares estimates of the parameters in (5.1). The dependent variable y_t is the US equity premium, while the lagged regressors, $x_{1,t-1}$ and $x_{2,t-1}$, are the cointegrated predictors. Four pairs of cointegrated predictors are considered, as follows: (i) ep (earning-price ratio) and dp (dividend-price ratio), (ii) tbl (T-bill rate) and lty (long-term yield), (iii) BAA and AAA (-rated corporate bond yields), and (iv) dp and dy (dividend yield). We use the GCV method to select the truncation parameter k. The F-tests are computed under the null hypothesis of no predictability—that is, $H_0 : d_1 = d_2 = ... = d_k = 0$. The number in parenthesis below each estimate is t-ratio and below each F-test is p-value. Panels A and B (C and D) report estimation results for the quarterly (monthly) data. *,**,*** indicate significance at the 10%, 5% and 1% levels, respectively.

the results for the whole sample period of 1927 to 2017, based on quarterly and monthly data, respectively. Following Kostakis et al. (2015), we also consider the post-1952 period because the interest rate variables are expected to be linked together after the Federal Reserve abandoned the interest rate pegging policy in 1951. Moreover, Campbell and Yogo (2006) and Kostakis et al. (2015) reported weak or no evidence of stock return predictability in the post-1952 period. Our results for this sub-period are reported in Panels B (quarterly data) and D (monthly data).

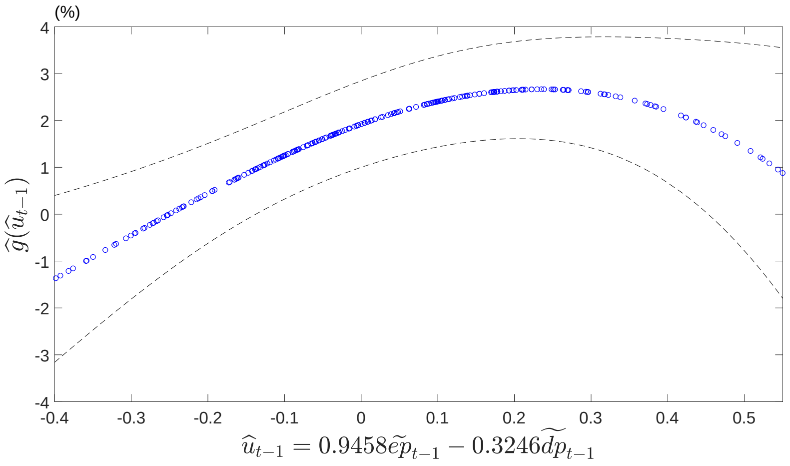
In Table 3, using the F-tests, we reject the null hypothesis of no predictability at the 5% level in both the full sample and the post-1952 sample for all four pairs at quarterly and monthly frequencies, with one exception. The pair of lty and tbl is not a significant predictor of equity premium at quarterly and monthly frequencies in the full sample, yet is a significant predictor in the post-1952 period. This result supports the view that the term-structure variables are closely linked together after 1952, yet not before.

While numerous studies (such as Campbell and Yogo, 2006 and Kostakis et al., 2015) found no or weak evidence of predictability in the post-1952 period using a univariate or multivariate framework, we do find strong evidence using bivariate cointegrated predictors in this sub-period.

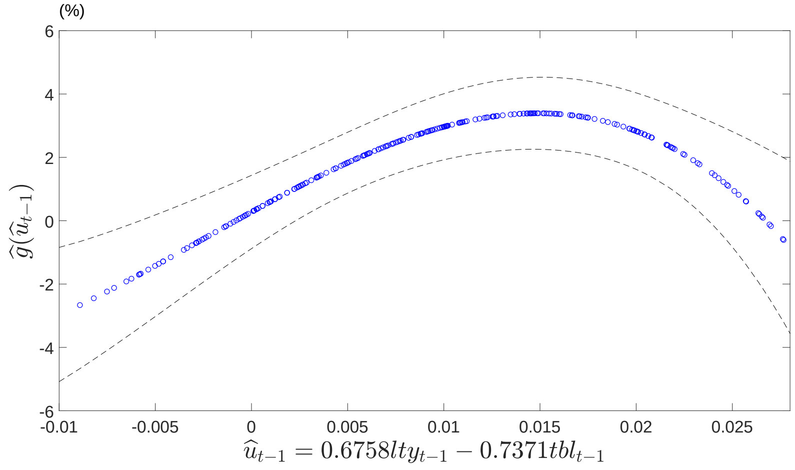
Moreover, the results in Table 3 provide ample evidence in favour of nonlinear predictability of stock returns, since the coefficients on the highest power in the polynomial regression (5.1) are statistically significant at conventional levels. To illustrate the approximate form of nonlinearity, Figure 4 plots predicted value of equity premium, $\widehat{g}_n(\widehat{u}_{t-1})$, against $\widehat{u}_{t-1} = \widehat{\theta}_1 x_{1,t-1} + \widehat{\theta}_2 x_{2,t-1}$, along with the 90% pointwise confidence intervals using the post-1952 quarterly data. The confidence intervals are obtained using the procedure described in Section 4.1. The corresponding plots for the monthly data are given in Figure 5.

Figure 4 and Figure 5 indicate that the pair of lty and tbl and pair of ep and dp exhibit a hump-shaped relationship between $\widehat{g}_n(\widehat{u}_{t-1})$ and $\widehat{u}_{t-1}$ at both quarterly and monthly frequencies. This empirical finding of nonlinear predictability using these two pairs of cointegrated predictors highlights a useful feature of our proposed semiparametric single-index predictive model. Using quarterly lty and tbl as an illustration, Figure 4 shows that the predicted value of equity premium peaks at around $\widehat{u}_{t-1} = 0.6758 lty_{t-1} -$

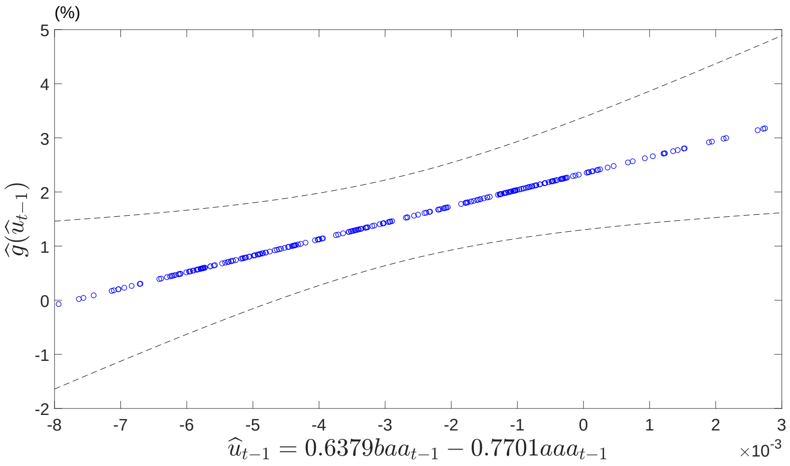
Figure 4: Estimated link function $\widehat{g}\left(\widehat{u}_{t-1}\right)$ at quarterly frequency



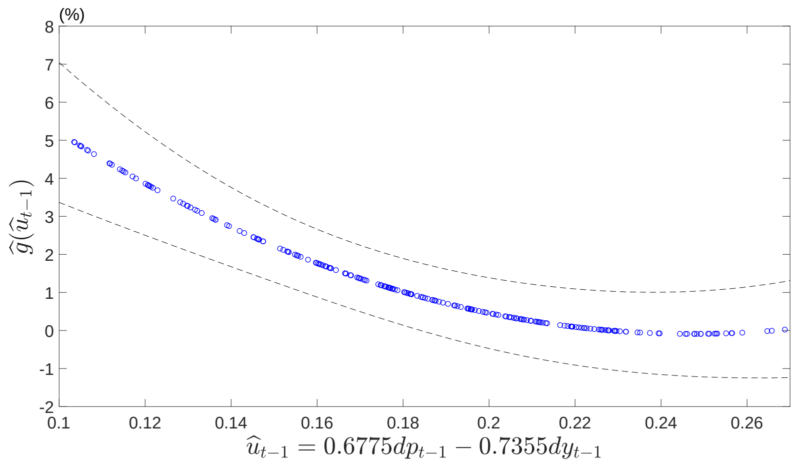
(a) $\widetilde{\text{ep}}$ and $\widetilde{\text{dp}}$ (k = 4)



(b) lty and tbl (k=4)



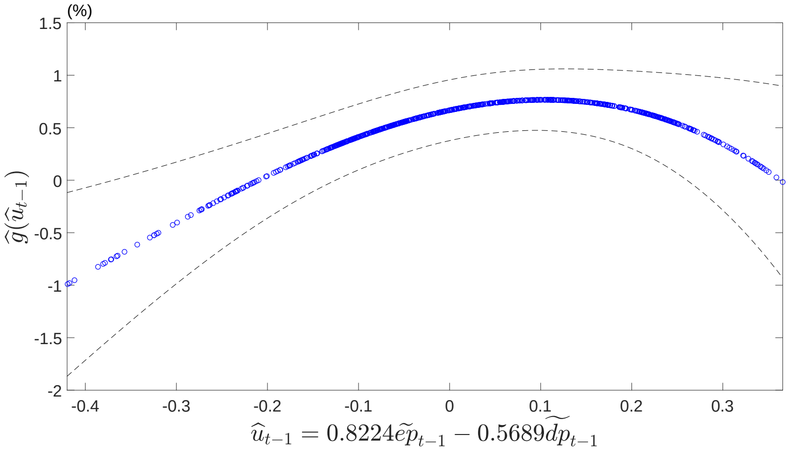
(c) baa and aaa (k = 2)



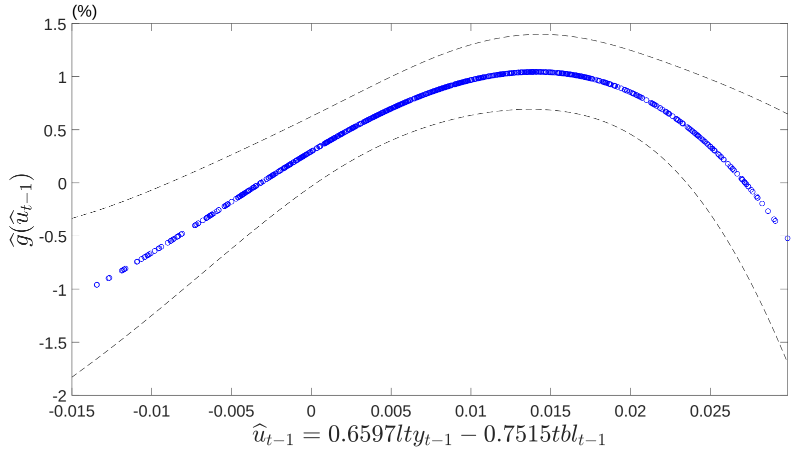
(d) dp and dy (k=3)

Notes: This figure plots the estimated link function of each pair of comoving predictors. The dashed line shows the approximate 90% pointwise confidence interval, and the horizontal line depicts the sample mean of equity premium. The confidence interval is constructed by the procedure described in Section 4.1. The sample period is 1952:Q1 to 2017:Q4.

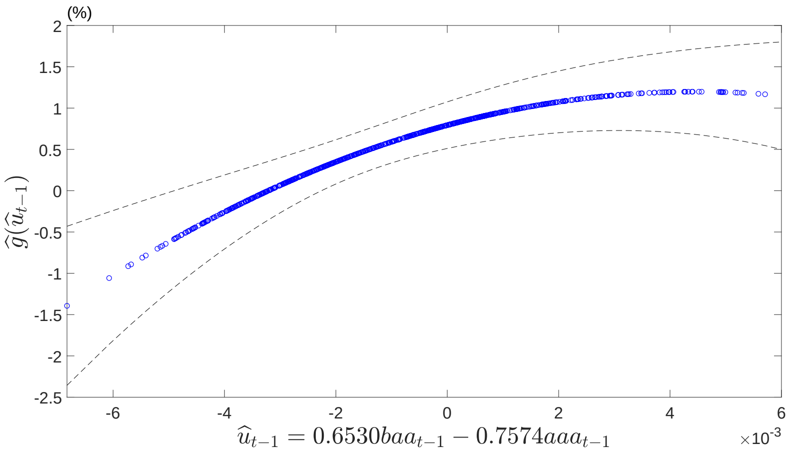
Figure 5: Estimated link function $\widehat{g}\left(\widehat{u}_{t-1}\right)$ at monthly frequency



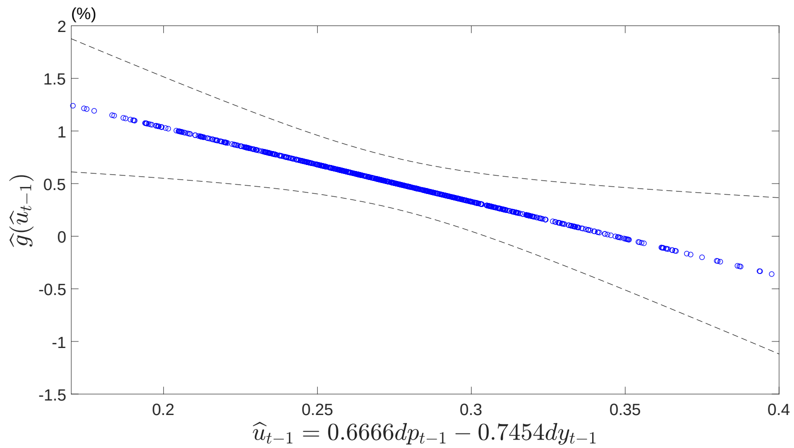
(a) $\widetilde{\text{ep}}$ and $\widetilde{\text{dp}}$ (k = 4)



(b) lty and tbl (k=4)



(c) baa and aaa (k = 3)



(d) dp and dy (k=2)

Notes: This figure plots the estimated link function of each pair of comoving predictors. The dashed line shows the approximate 90% pointwise confidence interval, and the horizontal line depicts the sample mean of equity premium. The confidence interval is constructed by the procedure described in Section 4.1. The sample period is January 1952 to December 2017.

$0.7371 tbl_{t-1} = 0.015$. In contrast, there is a small or negligible amount of nonlinear predictability of equity premium using the pair of BAA and AAA and the pair of dp and dy, at quarterly and monthly frequencies.

6 Conclusion

This paper has proposed estimation procedures for the single-index predictive regression model when the nonstationary predictors exhibit co-movement behaviour. We have studied the two types of super-consistency rates for the estimator of the single-index parameter θ_0 along two orthogonal directions in a new coordinate system, as well as their corresponding asymptotic distributions. This paper has also established the asymptotic normality of the plug-in estimator of the unknown link function. In addition, through Monte-Carlo simulations, we have evaluated the finite-sample properties of $\widehat{\alpha}_n$, $\widehat{\theta}_n$, as well as $\widehat{g}_n$. Further, we have applied the proposed model in the context of stock return predictability, and found nonlinear predictability of the equity premium using four pairs of comoving predictors.

Appendix A Discussion on the assumptions

For Assumption 1.1 (a), similar arguments are widely used in the literature for nonstationary models, such as by Park and Phillips (2000, 2001), and the σ-field sequence $\mathcal{F}_{n,t-1}$ can be taken as $\mathcal{F}_{n,t-1} = \sigma(x_1, \cdots, x_{n-1}; e_1, \cdots, e_{t-1})$. For Assumption 1.2 $(a)-(b)$, suppose that x_t is a d-dimensional integrated process, which is generated by a linear process v_t with i.i.d. sequence $\{\epsilon_j, -\infty < j < \infty\}$ in Assumption 1.1 (b) as building blocks. Assumption 1.2 (c) assumes a cointegration structure for x_t, and more details of cointegration structure have been discussed by Granger and Weiss (1983) and Engle and Granger (1987). Assumption 1.2 (c) also implies that there exists only one cointegration equation among x_t. This is an important assumption to develop the asymptotic theory for $\widehat{\theta}_n - \theta_0$ using the rotation technique because we need to ensure that x_{2t} is a pure $(d-1)$-dimensional nonstationary process. Assumption 1.2 (d) is our main assumption, in which we consider $\theta_0^\top x_t$ to be stationary inside the unknown link function, even though x_t is a d-dimensional integrated process. We also impose some restrictions on the probability density function of $u_t = \theta_0^\top x_t$ to exclude heavy-tailed distributions, and subsequently can control the potentially unbounded support and unbounded function smoothly.

Assumption 1.3 assumes the consistency for $(\widehat{\theta}_n, \widehat{g}_n)$ directly. The consistency is established with respect to the norm $\|.\|_2$ defined in (2.6). This assumption implies that $\widehat{\theta}_n \to_P \theta_0$ and $\|\widehat{g}_n - g_0\|_{L^2} \to_P 0$, respectively. Let $\delta > 0$ and define $\Theta_\delta \times G_\delta = \{\left\|(\theta, g) - (\theta_0, g_0)\right\|_2 \geq \delta\} \subset \Theta \times G$. This assumption can be replaced by the condition $\sum_{t=1}^n \left(g(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2 \to_P \infty$ uniformly in $(\theta, g) \in \Theta_\delta \times G_\delta$. To prove the consistency under this condition, define:

$$
\begin{aligned}
A_n &= \sum_{t=1}^n \left(g(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2, \\
B_n &= \sum_{t=1}^n \left(g(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right) e_t, \\
D_n &= \sum_{t=1}^n \left(y_t - g(\theta^\top x_{t-1})\right)^2 - \sum_{t=1}^n \left(y_t - g_0(\theta_0^\top x_{t-1})\right)^2.
\end{aligned}
$$

Then we can show that:

$$
\begin{aligned}
E\left[A_n^{-1/2} B_n\right]^2 &= E\left[\left(\sum_{t=1}^n \left(g(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2\right)^{-1/2} \sum_{t=1}^n \left(g(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right) e_t\right]^2 \\
=& E\left[\left(\sum_{t=1}^n \left(g(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2\right)^{-1} \sum_{t=1}^n \left(g(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2 E\left[e_t^2 | \mathcal{F}_{n,t-1}\right]\right] = \sigma^2.
\end{aligned}
$$

Therefore, $A_n^{-1/2} B_n = O_P(1)$ uniformly in $(\theta, g) \in \Theta_\delta \times G_\delta$. Then, we have:

$$
D_n = A_n(1 - A_n^{-1} B_n) = A_n(1 + o_P(1)) \to_P \infty,
$$

uniformly in $(\theta, g) \in \Theta_\delta \times G_\delta$. Given that $\Theta_\delta \times G_\delta$ is compact, we may easily deduce that:

$$
\inf_{(\theta, g) \in \Theta_\delta \times G_\delta} D_n \to_P \infty.
$$

This condition is sufficient to ensure the consistency, as shown in earlier work by Wu (1981): for any $\delta > 0$, $\liminf_{n \to \infty} \inf_{\|(\theta, g) - (\theta_0, g_0)\|_2 \geq \delta} D_n > 0$ in probability.

To verity the assumption that $\sum_{t=1}^n \left(g(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2 \to_P \infty$ uniformly in $(\theta, g) \in \Theta_\delta \times G_\delta$, we consider four cases:

(1) Given the point $(\theta_0, g) \in \Theta_\delta \times G_\delta$, by Weak Law of Large Numbers (WLLN), we can show that

$$
\frac{1}{n} \sum_{t=1}^n \left(g(\theta_0^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2 \to_P E\left[g(x_{11}) - g_0(x_{11})\right]^2
$$

uniformly in $(\theta_0, g) \in \Theta_\delta \times G_\delta$ and $E\left[g(x_{11}) - g_0(x_{11})\right]^2 > 0$ is implied by $\|g - g_0\|_{L^2}^2 > \delta^2 > 0$. Then we can obtain that $\sum_{t=1}^n \left(g(\theta_0^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2 \to_P \infty$ uniformly in $(\theta_0, g) \in \Theta_\delta \times G_\delta$.

(2) Given the point $(\theta, g_0) \in \Theta_\delta \times G_\delta$, suppose that g_0 is H-regular such that:

$$g_0(\eta x) = \varkappa(\eta) H(x) + \xi(\eta; x), \quad \left|\xi(\eta; x)\right| \leq a(\eta) P(x),$$

where $H(x)$ and $P(x)$ are both locally integrable, $\limsup_{\eta \to \infty} a(\eta)/\varkappa(\eta) = 0$ and $\varkappa(\sqrt{n}) \to \infty$ as $n \to \infty$.

According to (19) in Phillips and Solo (1992), we have:

$$\sup_r \left| \frac{1}{n^{1/2}} \sum_{t=1}^{\lfloor nr \rfloor} \theta^\top x_{t-1} - \theta^\top \phi(1) \frac{1}{n^{1/2}} \sum_{t=1}^{\lfloor nr \rfloor} \epsilon_{t-1} \right| \to_P 0.$$

We further suppose that, for all $m > 0$, $\int_{|r| \leq m} H(r)^2 dr > 0$. Then we can show that:

$$\frac{1}{n\varkappa(\sqrt{n})^2} \sum_{t=1}^n \left(g_0(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2 \to_P \int_0^1 H(V_\theta(r))^2 dr$$

uniformly in $(\theta, g_0) \in \Theta_\delta \times G_\delta$, where V_θ is Brownian motion of dimension 1 with variance $\Sigma_{V_\theta} = \theta^\top \phi(1) \Sigma_\epsilon \phi(1)^\top \theta$. Define a scaled local time L of V_θ by $L(t, s) = 1/\Sigma_{V_\theta} L_{V_\theta}(t, s)$, where L_{V_θ} is the local time of Brownian motion $V_\theta(r)$. By the occupation formula for Brownian motion:

$$\int_0^1 H(V_\theta(r))^2 dr = \int H(s)^2 L(1, s) ds \geq \int_{|s| \leq m} H(s)^2 L(1, s) ds > 0 \ a.s..$$

Then we can obtain that $\sum_{t=1}^n \left(g_0(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2 \to_P \infty$ uniformly in $(\theta, g_0) \in \Theta_\delta \times G_\delta$.

(3) Given the point $(\theta, g_0) \in \Theta_\delta \times G_\delta$, and suppose that g_0 is I-regular, we can show that:

$$\frac{1}{n} \sum_{t=1}^n \left(g_0(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2 \to_P E\left[g_0(x_{11})\right]^2 > 0,$$

uniformly in $(\theta, g_0) \in \Theta_\delta \times G_\delta$. Then we can obtain that $\sum_{t=1}^n \left(g_0(\theta^\top x_{t-1}) - g_0(\theta_0^\top x_{t-1})\right)^2 \to_P \infty$ uniformly in $(\theta, g_0) \in \Theta_\delta \times G_\delta$.

(4) Given the point $(\theta, g) \in \Theta_\delta \times G_\delta$, following the same ideas in cases (2) and (3), we can show that $\sum_{t=1}^n \left(g(\theta_0^\top x_{t-1}) - g_0(\theta^\top x_{t-1})\right)^2 \to_P \infty$ uniformly in $(\theta, g) \in \Theta_\delta \times G_\delta$ when g is H-regular and I-regular, respectively. For more details about H-regular and I-regular, we refer to Park and Phillips (2001).

Assumption 1.4 assumes a high degree of smoothness for the unknown link function $g_0(w)$, and all polynomials on $\mathbb{R}$ satisfy this condition. Although there is no theory about how to choose the truncation parameter, it must satisfy some conditions in Assumption 1.5 to ensure that the estimators $\widehat{\theta}_n$ and $\widehat{g}_n$ converge with a certain rate. In addition, we also consider the identification condition that $\inf_{c\in\mathbb{R}} E\left[g_0(\theta_0^\top x_1) - c\right]^2 > 0$ in Assumption 1.6 (a). If there exists $c \in \mathbb{R}$ such that $E\left[g_0(\theta_0^\top x_1) - c\right]^2 = 0$, then θ_0 will be unidentifiable. Assumption 1.6 (b) is standard in the literature (Newey, 1997). Suppose that $u_t = \theta_0^\top x_t \sim iiN(0,1)$, we have $E\left[\mathcal{H}_k(u_1)\mathcal{H}_k(u_1)^\top\right] = (2\pi)^{-1/2} I_k$, where I_k is a k-dimensional identity matrix. Then all the eigenvalues of $E\left[\mathcal{H}_k(u_1)\mathcal{H}_k(u_1)^\top\right]$ are $(2\pi)^{-1/2}$, and hence they are bounded away from zero uniformly in $k \geq 1$.

In Assumption 1.7 (a), we require the fourth moment of $g_0(\theta_0^\top x_{t-1})$ to exist, and many functional forms for $g_0(.)$ together with Assumption 1.2 (d) can satisfy this condition. Suppose $g_0(.)$ to be polynomials (e.g. $g_0(w) = 1+w^2$), exponential functions (e.g. $g_0(w) = \exp(w)$) or bounded functions (e.g. $g_0(w) = (1+w^2)^{-1}$), and it is easy to see that $g_0(w) \in L^2(\mathbb{R}.\exp(-w^2/2))$ and $\left(g_0(w)\right)^2 \in L^2(\mathbb{R}.\exp(-w^2/2))$. Then, simple algebra can show that Assumption 1.7 (a) is satisfied. Assumption 1.7 (b) can be replaced by a stronger version of Assumption 1.2 (d), as follows:

- Suppose that $u_t = \theta_0^\top x_t$ is a strictly stationary time series and has probability density function $\rho(u)$ such that $\exp(u^2)\rho(u) < \infty$ uniformly in u.

Follow the truth that $\left|H_i(u)\right| \times \exp(-u^2/4)$ being bounded uniformly, we are able to show that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1} E\left[H_i(u_1)H_j(u_1)\right]^2 = \frac{1}{n}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\int H_i^2(u)H_j^2(u)\rho(u)du\\
=&\frac{1}{n}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\int H_i^2(u)\exp(-u^2/2)H_j^2(u)\exp(-u^2/2)\exp(u^2)\rho(u)du\\
\leq& O(1)\frac{1}{n}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\int H_j^2(u)\exp(-u^2/2)du. = O(n^{-1}k^2) = o(1).
\end{aligned}$$

Alternatively, we can assume $E\left[H_i(u_1)\right]^4$ is uniformly bounded for $1 \leq i \leq k-1$, and then Assumption 1.7 (b) can be easily verified.

Assumption 1.7 (c) and (d) can be replaced by a condition on the density function of ϵ_j and a condition on the coefficients of the linear process for v_t. According to the Beveridge and Nelson (BN) decomposition Beveridge and Nelson, 1981 for x_t, we can write $x_{1t} = \theta_0^\top x_t = \sum_{i=0}^{\infty} d_i\epsilon_{t-i}$ (more details can be found in the proof of Lemma 3 in Appendix D). Suppose that: (1) the innovations $\{\epsilon_j, -\infty < j < \infty\}$ have density $p(x)$ satisfying $\int \left|p(x) - p(x+y)\right| \leq C\left|y\right|$ where $0 < C < \infty$; and (2) $\lim_{j\to\infty} d_j j^\lambda$ exists with $\lambda > 11/4$. Then, using the Corollary 4 in Withers

(1981), we can show that the linear process x_{1t} is a α-mixing process with mixing coefficient $\alpha(\tau)$, such that $\alpha(\tau) = O(\tau^{-1/2})$ and hence $\frac{1}{n}\sum_{\tau=1}^{n-1}\alpha(\tau)^{\nu/(4+\nu)} = O(n^{-1/2})$ for some $\nu > 0$. In addition, we also need to assume that $E\left|g_0^{(1)}(\theta_0^\top x_1)\right|^{4+\nu} < \infty$ and $\max_{0\le i\le k-1} E\left|H_i(\theta_0^\top x_1)\right|^{4+\nu} < \infty$ for the same ν defined before.

Then for Assumption 1.7 (c), we can show that:

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|Cov\left(\left(g_0^{(1)}(\theta_0^\top x_{t-1})\right)^2, \left(g_0^{(1)}(\theta_0^\top x_{s-1})\right)^2\right)\right| \\
=&\frac{1}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|Cov\left(\left(g_0^{(1)}(x_{1t-1})\right)^2, \left(g_0^{(1)}(x_{1s-1})\right)^2\right)\right| \\
=&\frac{1}{n}\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)\left|Cov\left(\left(g_0^{(1)}(x_{11})\right)^2, \left(g_0^{(1)}(x_{1,1+\tau})\right)^2\right)\right| \\
\le& c_\alpha \frac{1}{n}\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)\alpha(\tau)^{\nu/(4+\nu)}\left(E\left[g_0^{(1)}(x_{11})\right]^{(4+\nu)}\right)^{4/(4+\nu)} \\
=&O(1)\frac{1}{n}\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)\alpha(\tau)^{\nu/(4+\nu)} = O(n^{-1/2}) = o(1),
\end{aligned}
$$

where $c_\alpha = 2^{(4+2\nu)/(4+\nu)} \times (4+v)/v$.

Similarly, in terms of Assumption 1.7 (d), we have:

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|Cov\left(H_i(\theta_0^\top x_{t-1})H_j(\theta_0^\top x_{t-1}), H_i(\theta_0^\top x_{s-1})H_j(\theta_0^\top x_{s-1})\right)\right| \\
=&\frac{1}{n^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|Cov\left(H_i(x_{1t-1})H_j(x_{1t-1}), H_i(x_{1s-1})H_j(x_{1s-1})\right)\right| \\
=&\frac{1}{n}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)\left|Cov\left(H_i(x_{11})H_j(x_{11}), H_i(x_{1,1+\tau})H_j(x_{1,1+\tau})\right)\right| \\
\le& c_\alpha\frac{1}{n}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)\alpha(\tau)^{\nu/(4+\nu)}\left(E\left[H_i(x_{11})H_j(x_{11})\right]^{(4+\nu)/2}\right)^{4/(4+\nu)} \\
\le& c_\alpha\frac{1}{n}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)\alpha(\tau)^{\nu/(4+\nu)}\left(E\left[H_i(x_{11})\right]^{4+\nu}E\left[H_j(x_{11})\right]^{4+\nu}\right)^{2/(4+\nu)} \\
=&O(1)\frac{1}{n}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{\tau=1}^{n-1}\left(1-\frac{\tau}{n}\right)\alpha(\tau)^{\nu/(4+\nu)} = O(n^{-1/2}k^2) = o(1).
\end{aligned}
$$

Appendix B Proofs of the theorems

Proof of Theorem 3.1:

According to Lemma 9 in Appendix C, we have as $n \to \infty$

$$P_1^\top D_n(\widehat{\alpha}_n - \alpha_0) \to_D \sigma r_0^{-1/2} W(1).$$

Since $P_1 = \begin{pmatrix} p_{1,1} & \cdots & p_{1,d-1} \\ \vdots & \ddots & \vdots \\ p_{d,1} & \cdots & p_{d,d-1} \end{pmatrix}$ with $p_{i+1,i} = 1$ for $1 \leq i \leq d-1$ and others equal zero, simple algebra shows that

$$n(\widehat{\alpha}_n^2 - \alpha_0^2) \to_D \xi.$$

In addition, notice that

$$\widehat{\theta}_n^\top \theta_0 - 1 = (\widehat{\theta}_n - \theta_0)^\top \theta_0 = (\widehat{\theta}_n - \theta_0)^\top (\theta_0 - \widehat{\theta}_n + \widehat{\theta}_n) = -\|\widehat{\theta}_n - \theta_0\|^2 - (\widehat{\theta}_n^\top \theta_0 - 1).$$

Therefore, $\widehat{\theta}_n^\top \theta_0 - 1 = -\frac{1}{2}\|\widehat{\theta}_n - \theta_0\|^2$.

Consider the orthogonal expansion that $\|\widehat{\theta}_n - \theta_0\|^2 = \|Q_2^\top(\widehat{\theta}_n - \theta_0)\|^2 + \|\theta_0^\top(\widehat{\theta}_n - \theta_0)\|^2$, we can obtain

$$\begin{aligned} n^2(\widehat{\alpha}_n^1 - \alpha_0^1) =& n^2\theta_0^\top(\widehat{\theta}_n - \theta_0) = -\frac{1}{1 + \theta_0^\top \widehat{\theta}_n} \|nQ_2^\top(\widehat{\theta}_n - \theta_0)\|^2 \\ =& -\frac{1}{2}\|nQ_2^\top(\widehat{\theta}_n - \theta_0)\|^2(1 + o_P(1)) \to_D -\frac{1}{2}\|\xi\|^2. \end{aligned}$$ □

Proof of Theorem 3.2:

Since $\widehat{\theta}_n$ is the composite of $\widehat{\alpha}_n^1$ and $\widehat{\alpha}_n^2$, we have

$$\begin{aligned} n(\widehat{\theta}_n - \theta_0) = Qn(\widehat{\alpha}_n - \alpha_0) =& Qn \begin{pmatrix} \widehat{\alpha}_n^1 - 1 \\ \widehat{\alpha}_n^2 \end{pmatrix} \\ =& (\theta_0, \ Q_2) \begin{pmatrix} 0 \\ n\widehat{\alpha}_n^2 \end{pmatrix} + o_P(1) = Q_2 n\widehat{\alpha}_n^2 + o_P(1) \\ \to_D& \ \ \mathrm{MN}(0, \sigma^2 Q_2 r_0^{-1} Q_2^\top). \end{aligned}$$ □

Proof of Theorem 3.3:

We first show the consistency of $\widehat{\mathcal{H}}_x$ and $\widehat{\sigma}^2$. $\widehat{\mathcal{H}}_x \to_P \mathcal{H}_x$ follows from Lemma 5 in Appendix C directly.

For $\widehat{\sigma}^2$, note that

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{t=1}^{n}\left[y_t - \widehat{g}_n(\widehat{\theta}_n^\top x_{t-1})\right]^2$$

$$
\begin{aligned}
=&\frac{1}{n}\sum_{t=1}^{n}\left[e_t+g_0(x_{1t-1})-\widehat{g}_n(\widehat{\eta}_{t-1})\right]^2\\
=&\frac{1}{n}\sum_{t=1}^{n}e_t^2+\frac{1}{n}\sum_{t=1}^{n}\left[g_0(x_{1t-1})-\widehat{g}_n(\widehat{\eta}_{t-1})\right]^2+\frac{2}{n}\sum_{t=1}^{n}e_t\left[g_0(x_{1t-1})-\widehat{g}_n(\widehat{\eta}_{t-1})\right]\\
:=&A_1+A_2+2A_3.
\end{aligned}
$$

It is obvious that $A_1 \to_P \sigma^2$.

Given any $\epsilon > 0$, define for any function $f(x) \in L^2(\mathbb{R}, \exp(-x^2/2))$,

$$
f_{\sup}^{\epsilon}(x) = \sup_{|\alpha-1|<\epsilon}\sup_{|b|<\epsilon}\left|f(ax+b)\right|.
$$

The discussion of its properties can be found in the proof of Lemma 5 in Appendix D.

For A_2, write

$$
\begin{aligned}
A_2 =& \frac{1}{n}\sum_{t=1}^{n}\left[g_0(x_{1t-1})-\widehat{g}_n(\widehat{\eta}_{t-1})\right]^2\\
=&\frac{1}{n}\sum_{t=1}^{n}\left[g_0(x_{1t-1})-g_0(\widehat{\eta}_{t-1})+g_k(\widehat{\eta}_{t-1})-\widehat{g}_n(\widehat{\eta}_{t-1})+\gamma_k(\widehat{\eta}_{t-1})-\gamma_k(x_{1t-1})+\gamma_k(x_{1t-1})\right]^2\\
\leq&O(1)\frac{1}{n}\sum_{t=1}^{n}\left[g_0(x_{1t-1})-g_0(\widehat{\eta}_{t-1})\right]^2+O(1)\frac{1}{n}\sum_{t=1}^{n}\left[g_k(\widehat{\eta}_{t-1})-\widehat{g}_n(\widehat{\eta}_{t-1})\right]^2\\
&+O(1)\frac{1}{n}\sum_{t=1}^{n}\left[\gamma_k(\widehat{\eta}_{t-1})-\gamma_k(x_{1t-1})\right]^2+O(1)\frac{1}{n}\sum_{t=1}^{n}\left[\gamma_k(x_{1t-1})\right]^2\\
\leq&O(1)\left|\widehat{\alpha}_n^1-\alpha_0^1\right|^2\frac{1}{n}\sum_{t=1}^{n}\left[\left(g_0^{(1)}\right)_{\sup}(x_{1t-1})x_{1t-1}\right]^2(1+o_P(1))\\
&+O(1)\left|\widehat{\alpha}_n^2-\alpha_0^2\right|^2\frac{1}{n}\sum_{t=1}^{n}\left[\left(g_0^{(1)}\right)_{\sup}(x_{1t-1})x_{2t-1}\right]^2(1+o_P(1))\\
&+O(1)\left\|\bar{C}_k(\widehat{\alpha}_n)-C_{0,k}\right\|^2\frac{1}{n}\sum_{t=1}^{n}\sum_{i=0}^{k-1}\left[(H_i)_{\sup}(x_{1t-1})\right]^2(1+o_P(1))\\
&+O(1)\left|\widehat{\alpha}_n^1-\alpha_0^1\right|^2\frac{1}{n}\sum_{t=1}^{n}\left[\left(\gamma_k^{(1)}\right)_{\sup}(x_{1t-1})x_{1t-1}\right]^2(1+o_P(1))\\
&+O(1)\left|\widehat{\alpha}_n^2-\alpha_0^2\right|^2\frac{1}{n}\sum_{t=1}^{n}\left[\left(\gamma_k^{(1)}\right)_{\sup}(x_{1t-1})x_{2t-1}\right]^2(1+o_P(1))\\
&+O(1)\frac{1}{n}\sum_{t=1}^{n}\left[\gamma_k(x_{1t-1})\right]^2\\
:=&O(1)A_{2,1}+\cdots+O(1)A_{2,6}.
\end{aligned}
$$

Similar to the proof of (E.3) in online Appendix E, we can show that $\|\bar{C}_k(\widehat{\alpha}_n) - C_{0,k}\| = O_P(n^{-1/2}k^{1/2}) + o_P(k^{-r/2})$. Then, we can obtain that

$$
A_{2,1} = O_P(n^{-4}), \qquad\qquad A_{2,2} = O_P(n^{-1}),
$$

$$A_{2,3} = O_P(n^{-1}k^2) + o_P(k^{-(r-1)}), \qquad A_{2,5} = o_P(n^{-4}k^{-(r-2)}),$$
$$A_{2,6} = o_P(n^{-1}k^{-(r-1)}), \qquad A_{2,7} = o_P(k^{-r}),$$

and hence we have shown that $A_2 = o_P(1)$.

For the proof of the normality, in view of the consistency of $\widehat{\sigma}$ and $\widehat{\mathcal{H}}_x$, we show the result with the replacement of σ and $\mathcal{H}_x$. Let $\widehat{Z} = Z(\widehat{\theta}_n) = Z(\widehat{\alpha}_n)$ and write

$$\begin{aligned}
&\widehat{g}_n(w) - g_0(w) + \gamma_k(w) = \mathcal{H}_k(w)^\top \left(\bar{C}_k(\widehat{\theta}_n) - C_{0,k}\right)\\
=&\mathcal{H}_k(w)^\top \left(\widehat{Z}^\top \widehat{Z}\right)^{-1} \widehat{Z}^\top (\gamma + e) + \mathcal{H}_k(w)^\top \left(\widehat{Z}^\top \widehat{Z}\right)^{-1} \widehat{Z}^\top \left(Z - \widehat{Z}\right) C_{0,k}\\
=&\frac{1}{n}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} \widehat{Z}^\top e(1 + o_P(1)) + \frac{1}{n^{1/2}}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1/2} \left(Z^\top Z\right)^{-1/2} \widehat{Z}^\top \gamma(1 + o_P(1))\\
&+ \frac{1}{n}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} \widehat{Z}^\top \left(Z - \widehat{Z}\right) C_{0,k}(1 + o_P(1))\\
=&\frac{1}{n}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} Ze(1 + o_P(1)) + \frac{1}{n}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} \left(\widehat{Z} - Z\right)^\top e(1 + o_P(1))\\
&+ \frac{1}{n^{1/2}}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1/2} \left(Z^\top Z\right)^{-1/2} Z^\top \gamma(1 + o_P(1)) + \frac{1}{n}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} \left(\widehat{Z} - Z\right)^\top \gamma(1 + o_P(1))\\
&+ \frac{1}{n}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} Z^\top \left(Z - \widehat{Z}\right) C_{0,k}(1 + o_P(1))\\
&+ \frac{1}{n}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} \left(\widehat{Z} - Z\right)^\top \left(Z - \widehat{Z}\right) C_{0,k}(1 + o_P(1))
\end{aligned}$$

Then it follows that

$$\begin{aligned}
&\sqrt{n}\Sigma^{-1}(w) \left(\widehat{g}_n(w) - g_0(w) + \gamma_k(w)\right)\\
=&n^{-1/2}\sigma^{-1} \left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2} \mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} Z^\top e(1 + o_P(1))\\
&+ n^{-1/2}\sigma^{-1} \left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2} \mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} \left(\widehat{Z} - Z\right)^\top e(1 + o_P(1))\\
&+ \sigma^{-1} \left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2} \mathcal{H}_k(w)^\top \mathcal{H}_x^{-1/2} \left(Z^\top Z\right)^{-1/2} Z^\top \gamma(1 + o_P(1))\\
&+ n^{-1/2}\sigma^{-1} \left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2} \mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} \left(\widehat{Z} - Z\right)^\top \gamma(1 + o_P(1))\\
&+ n^{-1/2}\sigma^{-1} \left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2} \mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} Z^\top \left(Z - \widehat{Z}\right) C_{0,k}(1 + o_P(1))\\
&+ n^{-1/2}\sigma^{-1} \left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2} \mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} \left(\widehat{Z} - Z\right)^\top \left(Z - \widehat{Z}\right) C_{0,k}(1 + o_P(1))\\
=&F_1(1 + o_P(1)) + \cdots + F_6(1 + o_P(1))
\end{aligned}$$

By Assumption 1.1, F_1 is a martingale array and we shall use Corollary 3.1 of Hall and Heyde (1980) to show that $F_1 \to_D N(0, 1)$.

The conditional variance process is given by

$$\frac{1}{n}\sigma^{-2} \left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1} \sum_{t=1}^{n} \left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(x_{1t-1})\right)^2 E\left[e_t^2 | \mathcal{F}_{n,t-1}\right]$$

$$=\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\left(\frac{1}{n}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\mathcal{H}_k(x_{1t-1})^\top\right)\mathcal{H}_x^{-1}\mathcal{H}_k(w)$$
$$=\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_x\mathcal{H}_x^{-1}\mathcal{H}_k(w)(1+o_P(1))=1+o_P(1)$$

Moreover, to make the conditional Lindeberg's condition fulfilled, we have

$$\frac{1}{n^2}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-2}\sum_{t=1}^n E\left[\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(x_{1t-1})e_t\right)^4\middle|\mathcal{F}_{n,t-1}\right]$$
$$=O(1)\frac{1}{n^2}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-2}\sum_{t=1}^n\left[\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(x_{1t-1})\right]^4$$
$$\leq O(1)\frac{1}{n^2}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-2}\left\|\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\right\|^4\sum_{t=1}^n\left\|\mathcal{H}_k(x_{1t-1})\right\|^4$$
$$=O(1)\frac{1}{n^2}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-2}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-2}\mathcal{H}_k(w)\right)^2\sum_{t=1}^n\left\|\mathcal{H}_k(x_{1t-1})\right\|^4$$
$$\leq O(1)\lambda_{min}^{-2}(\mathcal{H}_x)\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-2}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^2\frac{1}{n^2}\sum_{t=1}^n\left\|\mathcal{H}_k(x_{1t-1})\right\|^4$$
$$=o_P(1)$$

To show that $F_2=o_P(1)$, by mean value theorem

$$F_2=n^{-1/2}\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\left(\widehat{Z}-Z\right)^\top e$$
$$=n^{-1/2}\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\sum_{t=1}^n\mathcal{H}_k^{(1)}(\widehat{\eta}_{t-1}^*)\left(\left(\widehat{\alpha}_n^1-\alpha_0^1\right)x_{1t-1}+\left(\widehat{\alpha}_n^2-\alpha_0^2\right)x_{2t-1}\right)e_t$$
$$=\left(\widehat{\alpha}_n^1-\alpha_0^1\right)\sigma^{-1}n^{-1/2}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\sum_{t=1}^n\mathcal{H}_k^{(1)}(\widehat{\eta}_{t-1}^*)x_{1t-1}e_t$$
$$+\left(\widehat{\alpha}_n^2-\alpha_0^2\right)\sigma^{-1}n^{-1/2}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\sum_{t=1}^n\mathcal{H}_k^{(1)}(\widehat{\eta}_{t-1}^*)x_{2t-1}e_t$$
$$=\left(\widehat{\alpha}_n^1-\alpha_0^1\right)\sigma^{-1}n^{-1/2}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\sum_{t=1}^n\mathcal{H}_k^{(1)}(x_{1t-1})x_{1t-1}e_t(1+o_P(1))$$
$$+\left(\widehat{\alpha}_n^2-\alpha_0^2\right)\sigma^{-1}n^{-1/2}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\sum_{t=1}^n\mathcal{H}_k^{(1)}(x_{1t-1})x_{2t-1}e_t(1+o_P(1))$$
$$:=F_{2,1}(1+o_P(1))+F_{2,2}(1+o_P(1)).$$

Then for the stationary component

$$E\left[n^{-1/2}\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\sum_{t=1}^n\mathcal{H}_k^{(1)}(x_{1t-1})x_{1t-1}e_t\right]^2$$
$$=\frac{1}{n}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1}\sum_{t=1}^n E\left[\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k^{(1)}(x_{1t-1})x_{1t-1}\right]^2$$

$$
\begin{aligned}
\leq & \frac{1}{n}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1}\left\|\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\right\|^2 \sum_{t=1}^n E\left\|\mathcal{H}_k^{(1)}(x_{1t-1})x_{1t-1}\right\|^2 \\
= & \frac{1}{n}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-2}\mathcal{H}_k(w)\right) \sum_{t=1}^n E\left\|\mathcal{H}_k^{(1)}(x_{1t-1})x_{1t-1}\right\|^2 \\
\leq & \lambda_{min}^{-1}(\mathcal{H}_x)\frac{1}{n}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right) \sum_{t=1}^n E\left\|\mathcal{H}_k^{(1)}(x_{1t-1})x_{1t-1}\right\|^2 \\
= & O(k^3).
\end{aligned}
$$

Since $\left|\widehat{\alpha}_n^1 - \alpha_0^1\right| = O_P(n^{-2})$ from Theorem 3.1, we have $F_{2,1} = O_P(n^{-2}k^{3/2}) = o_P(1)$.

Regarding the nonstationary component, consider

$$
\begin{aligned}
& E\left[\sigma^{-1}n^{-1/2}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\sum_{t=1}^n \mathcal{H}_k^{(1)}(x_{1t-1})x_{2t-1}e_t\right]^2 \\
= & \frac{1}{n}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1}\sum_{t=1}^n E\left[\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k^{(1)}(x_{1t-1})x_{2t-1}\right]^2 \\
\leq & \frac{1}{n}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1}\left\|\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\right\|^2 \sum_{t=1}^n E\left\|\mathcal{H}_k^{(1)}(x_{1t-1})x_{2t-1}\right\|^2 \\
= & \frac{1}{n}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-2}\mathcal{H}_k(w)\right) \sum_{t=1}^n E\left\|\mathcal{H}_k^{(1)}(x_{1t-1})x_{2t-1}\right\|^2 \\
\leq & \lambda_{min}^{-1}(\mathcal{H}_x)\frac{1}{n}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right) \sum_{t=1}^n E\left\|\mathcal{H}_k^{(1)}(x_{1t-1})x_{2t-1}\right\|^2 \\
= & O(nk^2).
\end{aligned}
$$

Since $\left|\widehat{\alpha}_n^2 - \alpha_0^2\right| = o_P(n^{-1})$ from Theorem 3.1, we have $F_{2,2} = O_P(n^{-1/2}k) = o_P(1)$.

Then we move on to F_3, write

$$
\begin{aligned}
|F_3| = & \left|\sigma^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1/2}\left(Z^\top Z\right)^{-1/2}Z^\top \gamma\right| \\
\leq & \sigma^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\left\|\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1/2}\right\|\left\|\left(Z^\top Z\right)^{-1/2}Z^\top \gamma\right\| \\
= & \sigma^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{1/2}\left(\gamma^\top Z\left(Z^\top Z\right)^{-1}Z^\top \gamma\right)^{1/2} \\
\leq & O(1)\left\|\gamma\right\| = o_P(n^{1/2}k^{-r/2}) = o(1).
\end{aligned}
$$

In terms of F_4, by mean value theorem, we have

$$
\begin{aligned}
|F_4| = & \left|n^{-1/2}\sigma^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\left(\widehat{Z}-Z\right)^\top \gamma\right| \\
= & \Bigg|n^{-1/2}\sigma^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1} \\
& \times \sum_{t=1}^n \mathcal{H}_k^{(1)}(\widehat{\eta}_{t-1}^*)\left(\left(\widehat{\alpha}_n^1 - \alpha_0^1\right)x_{1t-1} + \left(\widehat{\alpha}_n^2 - \alpha_0^2\right)x_{2t-1}\right)\gamma_k(x_{1t-1})\Bigg|
\end{aligned}
$$

$$
\begin{aligned}
\leq & \left|\widehat{\alpha}_n^1-\alpha_0^1\right| n^{-1/2}\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\left\|\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\right\|\sum_{t=1}^n\left(\left\|\mathcal{H}_k^{(1)}\gamma_k x\right\|\right)_{\sup}(x_{1t-1})(1+o_P(1))\\
&+\left|\widehat{\alpha}_n^2-\alpha_0^2\right| n^{-1/2}\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\left\|\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\right\|\sum_{t=1}^n\left(\left\|\mathcal{H}_k^{(1)}\gamma_k\right\|\right)_{\sup}(x_{1t-1})\left|x_{2t-1}\right|(1+o_P(1))\\
\leq & O(1)\left|\widehat{\alpha}_n^1-\alpha_0^1\right| n^{-1/2}\sum_{t=1}^n\left(\left\|\mathcal{H}_k^{(1)}\gamma_k x\right\|\right)_{\sup}(x_{1t-1})(1+o_P(1))\\
&+O(1)\left|\widehat{\alpha}_n^2-\alpha_0^2\right| n^{-1/2}\sum_{t=1}^n\left(\left\|\mathcal{H}_k^{(1)}\gamma_k\right\|\right)_{\sup}(x_{1t-1})\left|x_{2t-1}\right|(1+o_P(1))\\
= & O(1)O_P(n^{-2})n^{-1/2}o_P(nk^{-(r-3)/2})+O(1)O_P(n^{-1})n^{-1/2}o_P(n^{3/2}k^{-(r-2)/2})\\
= & o_P(n^{-3/2}k^{-(r-3)/2})+o_P(k^{-(r-2)/2})=o_P(1).
\end{aligned}
$$

Regarding F_5, we can show that

$$
\begin{aligned}
F_5 &= n^{-1/2}\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}Z^\top\left(Z-\widehat{Z}\right)C_{0,k}\\
&=\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\left(g_0(x_{1t-1})-g_0(\widehat{\eta}_{t-1})+\gamma_k(\widehat{\eta}_{t-1})-\gamma_k(x_{1t-1})\right)\\
&=\left(\alpha_0^1-\widehat{\alpha}_n^1\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})g_0^{(1)}(\widehat{\eta}_{t-1}^*)x_{1t-1}\\
&\quad+\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})g_0^{(1)}(\widehat{\eta}_{t-1}^*)x_{2t-1}\\
&\quad+\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\left(\gamma_k(\widehat{\eta}_{t-1})-\gamma_k(x_{1t-1})\right)\\
&:=F_{5,1}+F_{5,2}+F_{5,3}
\end{aligned}
$$

In terms of $F_{5,1}$, consider

$$
\begin{aligned}
\left|F_{5,1}\right| &= \left|\left(\alpha_0^1-\widehat{\alpha}_n^1\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})g_0^{(1)}(\widehat{\eta}_{t-1}^*)x_{1t-1}\right|\\
&\leq\left|\widehat{\alpha}_n^1-\alpha_0^1\right|\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\left\|\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\right\|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\|\mathcal{H}_k(x_{1t-1})g_0^{(1)}(\widehat{\eta}_{t-1}^*)x_{1t-1}\right\|\\
&\leq O(1)\left|\widehat{\alpha}_n^1-\alpha_0^1\right|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left(\left\|\mathcal{H}_k g_0^{(1)}x\right\|\right)_{\sup}(x_{1t-1})(1+o_P(1))=O_P(n^{-3/2}k^{1/2})=o_P(1).
\end{aligned}
$$

For $F_{5,2}$, we have

$$
\begin{aligned}
&\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})g_0^{(1)}(\widehat{\eta}_{t-1}^*)x_{2t-1}\\
=&\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\mathcal{H}_k(\widehat{\eta}_{t-1}^*)^\top B_k^\top C_{0,k}x_{2t-1}\\
&+\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\gamma_k^{(1)}(\widehat{\eta}_{t-1}^*)x_{2t-1}\\
=&\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\mathcal{H}_k(x_{1t-1})^\top B_k^\top C_{0,k}x_{2t-1}
\end{aligned}
$$

$$+\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\left(\mathcal{H}_k(\widehat{\eta}_{t-1}^*)-\mathcal{H}_k(x_{1t-1})\right)^\top B_k^\top C_{0,k}x_{2t-1}$$

$$+\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\gamma_k^{(1)}(\widehat{\eta}_{t-1}^*)x_{2t-1}$$

$$:=F_{5,2,1}+F_{5,2,2}+F_{5,2,3},$$

where $B_k=\begin{pmatrix}0&0&\dots&\dots&0\\1&0&\dots&\dots&0\\0&\sqrt{2}&\dots&\dots&0\\\vdots&\vdots&\ddots&\dots&0\\0&0&\dots&\sqrt{k-1}&0\end{pmatrix}$.

For $F_{5,2,1}$, by Lemma 3 and Lemma 5 in Appendix C

$$F_{5,2,1}=\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\mathcal{H}_k(x_{1t-1})^\top B_k^\top C_{0,k}x_{2t-1}$$

$$=n\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_xB_k^\top C_{0,k}\frac{1}{n\sqrt{n}}\sum_{t=1}^n x_{2t-1}(1+o_P(1))$$

$$=n\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\left(g_0^{(1)}(w)-\gamma_k^{(1)}(w)\right)\frac{1}{n\sqrt{n}}\sum_{t=1}^n x_{2t-1}(1+o_P(1)).$$

Since $n\left(\widehat{\alpha}_n^2-\alpha_0^2\right)=O_P(1)$, $\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)=O(k)$, $\gamma_k^{(1)}(w)=o(k^{-(r-1)/2})$, and $\frac{1}{n\sqrt{n}}\sum_{t=1}^n x_{2t-1}\to_D\int_0^1 V_2(r)dr$, it is easy to see that $F_{5,2,1}=O_P(k^{-1/2})+o_P(k^{-(r-2)/2})=o_P(1)$. More proof details about the first equality can be found in (E.1) in online Appendix E.

In terms of $F_{5,2,2}$, by mean value theorem again, we have

$$\begin{aligned}&F_{5,2,2}\\=&\left|\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\left(\mathcal{H}_k(\widehat{\eta}_{t-1}^*)-\mathcal{H}_k(x_{1t-1})\right)^\top B_k^\top C_{0,k}x_{2t-1}\right|\\\leq&\left|\widehat{\alpha}_n^2-\alpha_0^2\right|\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\left\|\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\right\|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\|\mathcal{H}_k(x_{1t-1})\left(\mathcal{H}_k(\widehat{\eta}_{t-1}^*)-\mathcal{H}_k(x_{1t-1})\right)^\top B_k^\top C_{0,k}x_{2t-1}\right\|\\\leq&O(1)\left|\widehat{\alpha}_n^1-\alpha_0^1\right|\left|\widehat{\alpha}_n^2-\alpha_0^2\right|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left(\left\|\mathcal{H}_k\mathcal{H}_k^{(2)}x\right\|\right)_{\sup}(x_{1t-1})\left|x_{2t-1}\right|(1+o_P(1))\\&+O(1)\left|\widehat{\alpha}_n^2-\alpha_0^2\right|^2\frac{1}{\sqrt{n}}\sum_{t=1}^n\left(\left\|\mathcal{H}_k\mathcal{H}_k^{(2)}\right\|\right)_{\sup}(x_{1t-1})\left|x_{2t-1}\right|^2(1+o_P(1))\\=&O_P(n^{-2}k^{5/2})+O_P(n^{-1/2}k^2)=o_P(1).\end{aligned}$$

For $F_{5,2,3}$, we have

$$\begin{aligned}\left|F_{5,2,3}\right|=&\left|\left(\alpha_0^2-\widehat{\alpha}_n^2\right)\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^n\mathcal{H}_k(x_{1t-1})\gamma_k^{(1)}(\widehat{\eta}_{t-1}^*)x_{2t-1}\right|\\\leq&\left|\widehat{\alpha}_n^2-\alpha_0^2\right|\sigma^{-1}\left(\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\left\|\mathcal{H}_k(w)^\top\mathcal{H}_x^{-1}\right\|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left\|\mathcal{H}_k(x_{1t-1})\gamma_k^{(1)}(\widehat{\eta}_{t-1}^*)x_{2t-1}\right\|\\\leq&O(1)\left|\widehat{\alpha}_n^2-\alpha_0^2\right|\frac{1}{\sqrt{n}}\sum_{t=1}^n\left(\left\|\mathcal{H}_k\gamma_k^{(1)}\right\|\right)_{\sup}(x_{1t-1})\left|x_{2t-1}\right|(1+o_P(1))=o_P(k^{-(r-2)/2})=o_P(1).\end{aligned}$$

Regarding $F_{5,3}$, by mean value theorem

$$
\begin{aligned}
\left|F_{5,3}\right| =& \left|\sigma^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\mathcal{H}_k(x_{1t-1})\left(\gamma_k(\widehat{\eta}_{t-1})-\gamma_k(x_{1t-1})\right)\right| \\
\le& \sigma^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\left\|\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\right\|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\mathcal{H}_k(x_{1t-1})\left(\gamma_k(\widehat{\eta}_{t-1})-\gamma_k(x_{1t-1})\right)\right\| \\
\le& O(1)\left|\widehat{\alpha}_n^1-\alpha_0^1\right|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\mathcal{H}_k(x_{1t-1})\gamma_k^{(1)}(\widehat{\eta}_{t-1}^*)x_{1t-1}\right\| + O(1)\left|\widehat{\alpha}_n^2-\alpha_0^2\right|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left\|\mathcal{H}_k(x_{1t-1})\gamma_k^{(1)}(\widehat{\eta}_{t-1}^*)x_{2t-1}\right\| \\
\le& O(1)\left|\widehat{\alpha}_n^1-\alpha_0^1\right|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k\gamma_k^{(1)}x\right\|\right)_{\sup}(x_{1t-1})(1+o_P(1)) \\
&+O(1)\left|\widehat{\alpha}_n^2-\alpha_0^2\right|\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k\gamma_k^{(1)}\right\|\right)_{\sup}(x_{1t-1})\left|x_{2t-1}\right|(1+o_P(1)) \\
=& o_P(n^{-3/2}k^{-(r-3)/2})+o_P(k^{-(r-2)/2})=o_P(1).
\end{aligned}
$$

Therefore, we have shown that $F_5=o_P(1)$.

For F_6, we write

$$
\begin{aligned}
|F_6| =& \left|n^{-1/2}\sigma^{-1}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\left(\widehat{Z}-Z\right)^\top\left(Z-\widehat{Z}\right)C_{0,k}\right| \\
\le& O(1)n^{-1/2}\left(\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\mathcal{H}_k(w)\right)^{-1/2}\left\|\mathcal{H}_k(w)^\top \mathcal{H}_x^{-1}\right\| \\
&\times\left|\sum_{t=1}^{n}\left(\mathcal{H}_k(\widehat{\eta}_{t-1})-\mathcal{H}_k(x_{1t-1})\right)\left(g_0(x_{1t-1})-g_0(\widehat{\eta}_{t-1})+\gamma_k(\widehat{\eta}_{t-1})-\gamma_k(x_{1t-1})\right)\right| \\
\le& O(1)n^{-1/2}\left|\sum_{t=1}^{n}\left(\mathcal{H}_k(\widehat{\eta}_{t-1})-\mathcal{H}_k(x_{1t-1})\right)\left(g_0(x_{1t-1})-g_0(\widehat{\eta}_{t-1})+\gamma_k(\widehat{\eta}_{t-1})-\gamma_k(x_{1t-1})\right)\right| \\
\le& O(1)\left|\widehat{\alpha}_n^1-\alpha_0^1\right|^2 n^{-1/2}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k^{(1)}g_0^{(1)}x^2\right\|\right)_{\sup}(x_{1t-1})(1+o_P(1)) \\
&+O(1)\left|\widehat{\alpha}_n^2-\alpha_0^2\right|^2 n^{-1/2}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k^{(1)}g_0^{(1)}\right\|\right)_{\sup}(x_{1t-1})(x_{2t-1})^2(1+o_P(1)) \\
&+O(1)\left|\widehat{\alpha}_n^1-\alpha_0^1\right|^2 n^{-1/2}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k^{(1)}\gamma_k^{(1)}x^2\right\|\right)_{\sup}(x_{1t-1})(1+o_P(1)) \\
&+O(1)\left|\widehat{\alpha}_n^2-\alpha_0^2\right|^2 n^{-1/2}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k^{(1)}\gamma_k^{(1)}\right\|\right)_{\sup}(x_{1t-1})(x_{2t-1})^2(1+o_P(1)) \\
=& O_P(n^{-7/2}k)+O_P(n^{-1/2}k)+o_P(n^{-7/2}k^{-(r-5)/2})+o_P(n^{-1/2}k^{-(r-3)/2})=o_P(1). \qquad \square
\end{aligned}
$$

Appendix C Lemmas

Lemma 1. *Suppose that $g(u)$ is differentiable on $\mathbb{R}$ and $g^{r-i}(u)u^i \in L^2(\mathbb{R}, \exp(-u^2/2))$ for $0 \le i \le r$ and $r \ge 4$, and $u_t = \theta_0^\top x_t$. Then the following holds:*

(1) (i) $E|\gamma_k(u_1)|^2 = o(k^{-r})$; (ii) $E|\gamma_k^{(1)}(u_1)|^2 = o(k^{-(r-1)})$; (iii) $E|\gamma_k(u_1)u_1|^2 = o(k^{-(r-1)})$;

(iv) $E|\gamma_k^{(1)}(u_1)u_1|^2 = o(k^{-(r-2)})$.

(2) (i) $E\|\mathcal{H}_k(u_1)\|^2 = O(k)$; (ii) $E\|\mathcal{H}_k(u_1)u_1\|^2 = O(k^2)$; (iii) $E\|\mathcal{H}_k^{(1)}(u_1)\|^2 = O(k^2)$; (iv) $E\|\mathcal{H}_k^{(1)}(u_1)u_1\|^2 = O(k^3)$; (v) $E\|\mathcal{H}_k^{(2)}(u_1)\|^2 = O(k^3)$; (vi) $E\|\mathcal{H}_k^{(2)}(u_1)u_1\|^2 = O(k^4)$.

Lemma 2. *The following assertions hold:*

(1) $\left(x_{1t}, \frac{1}{\sqrt{t}}x_{2t}^\top\right)$ *has a joint density* $\psi_t(x, w^\top)$; $\left(x_{1t}, x_{1s}, \frac{1}{\sqrt{t}}x_{2t}^\top, \frac{1}{\sqrt{s}}x_{2s}^\top\right)$ *has a joint probability density* $\psi_{ts}(x, y, w^\top, z^\top)$ *where* $t > s$. *Meanwhile, these functions are bounded uniformly in* $(x, w^\top)$ *and* $(x, y, w^\top, z^\top)$ *as well as* t *and* (t, s)*, respectively.*

(2) *For large* t *and* s*, we have* $\psi_t(x, w^\top) = \rho(x)q_t(w)(1+o(1))$ *and* $\psi_{ts}(x, y, w^\top, z^\top) = \rho_{ts}(x, y)q_{ts}(w^\top, z^\top)(1+o(1))$ *where* $\rho(x)$ *is the marginal density of* x_{1t}, $\rho_{ts}(x, y)$ *is the joint density of* (x_{1t}, x_{1s}), $q_t(w)$ *is the marginal density of* $\frac{1}{\sqrt{t}}x_{2t}$ *and* $q_{ts}(w^\top, z^\top)$ *is the joint density of* $\left(\frac{1}{\sqrt{t}}x_{2t}^\top, \frac{1}{\sqrt{s}}x_{2s}^\top\right)$.

Lemma 3. *Let Assumption 1 hold. If* $\frac{1}{n^2}\sum_{t=2}^n\sum_{s=1}^{t-1}\left|Cov(f(x_{1t}), f(x_{1s}))\right| \to 0$ *as* $n \to \infty$*, where* $f(w) \in L^2(\mathbb{R}, \exp(-w^2/2))$*, then as* $n \to \infty$:

(a) $\frac{1}{n}\sum_{t=1}^n f(x_{1t}) \to_p E[f(x_{11})]$,

(b) $\frac{1}{n\sqrt{n}}\sum_{t=1}^n f(x_{1t})x_{2t} \to_d E[f(x_{11})]\int_0^1 V_2(r)dr$,

(c) $\frac{1}{n^2}\sum_{t=1}^n f(x_{1t})x_{2t}x_{2t}^\top \to_d E[f(x_{11})]\int_0^1 V_2(r)V_2^\top(r)dr$,

(d) $\frac{1}{n^2}\sum_{t=1}^n f(x_{1t})\left(x_{2t} - \bar{x}_2\right)\left(x_{2t} - \bar{x}_2\right)^\top \to_d E[f(x_{11})]\left[\int_0^1 V_2(r)V_2^\top(r)dr - \int_0^1 V_2(r)dr\int_0^1 V_2(r)^\top dr\right]$

where V_2 *is the Brownian motion with variance matrix* $\Sigma_V = Q_2^\top\phi(1)\Sigma_\epsilon\phi(1)^\top Q_2$ *and* $\bar{x}_2 = \frac{1}{n}\sum_{t=1}^n x_{2t}$.

Lemma 4. *Let Assumption 1 hold, as* $n \to \infty$

$$\frac{1}{n}\sum_{t=1}^n g_0^{(1)}(x_{1t-1})\left(x_{2t-1} - \frac{1}{n}\sum_{s=1}^n x_{2s-1}\right)e_t \to_D \sigma r_0^{1/2}W(1), \tag{C.1}$$

where $r_0 = E\left[g_0^{(1)}(x_{11})\right]^2\left(\int_0^1 V_2(r)V_2^\top(r)dr - \int_0^1 V_2(r)dr\int_0^1 V_2(r)^\top dr\right)$ *and* $W(1)$ *is an (d-1)-dimensional standard normal vector independent of* V_2.

Lemma 5. *(1). Let Assumption 1 hold, as* $n \to \infty$

$$\left\|\frac{1}{n}\sum_{t=1}^n \mathcal{H}_k(x_{1t})\mathcal{H}_k(x_{1t})^\top - E\left[\mathcal{H}_k(x_{11})\mathcal{H}_k(x_{11})^\top\right]\right\| \to_P 0.$$

(2) Let Assumption 1 hold, as $n \to \infty$

$$\left\|\frac{1}{n}\sum_{t=1}^n \mathcal{H}_k(\widehat{\eta}_t)\mathcal{H}_k(\widehat{\eta}_t)^\top - \frac{1}{n}\sum_{t=1}^n \mathcal{H}_k(x_{1t})\mathcal{H}_k(x_{1t})^\top\right\| \to_P 0,$$

where $\widehat{\eta}_t = \widehat{\theta}_n^\top x_t$.

Lemma 6. *Let Assumption 1 hold, as* $n \to \infty$

$$\left\|\bar{C}_n(\alpha_0) - C_{0,k}\right\| = O_p(n^{-1/2}k^{1/2}) + o_P(k^{-r/2}). \tag{C.2}$$

Lemma 7. *For any function* $f(x)$ *defined on* $\mathbb{R}$*, define*

$$T^{i,j} = \sum_{t=1}^{n} f(x_{1t-1})x_{it-1}x_{jt-1}\left[\left(\alpha^1 - \alpha_0^1\right)x_{1t-1} + \left(\alpha^2 - \alpha_0^2\right)^\top x_{2t-1}\right],$$

where $\alpha = (\alpha^1, (\alpha^2)^\top)^\top \in \Phi_n^\delta = \{\alpha = (\alpha^1, \alpha^2)^\top : \left|\alpha^1 - \alpha_0^1\right| < n^{-1/2+\delta}, \left\|\alpha^2 - \alpha_0^2\right\| < n^{-1+\delta}\}$ *and* $i, j \in \{1, 2\}$ *for some small* $\delta > 0$*. Then*

$$T^{i,j} = \begin{cases} O_P(1)n^{1/2+\delta}E\left|f(x_{11})\left(x_{11}\right)^3\right|, & \text{when } i = j = 1 \\ O_P(1)n^{1+\delta}E\left|f(x_{11})\left(x_{11}\right)^2\right|, & \text{when } i = 1, j = 2 \text{ or } i = 2, j = 1 \\ O_P(1)n^{3/2+\delta}E\left|f(x_{11})x_{11}\right|, & \text{when } i = j = 2. \end{cases}$$

Lemma 8. *Under Assumption 1, as* $n \to \infty$*,*

$$P_1^\top D_n^{-1}S_n(\alpha_0) \to_D -2\sigma r_0^{1/2}W(1) \quad \text{and} \quad P_1^\top D_n^{-1}J_n(\alpha_0)D_n^{-1}P_1 \to_D 2r_0, \tag{C.3}$$

where $D_n = diag(\sqrt{n}, nI_{d-1})$*,* $r_0 = E\left[g_0^{(1)}(x_{11})\right]^2\left(\int_0^1 V_2(r)V_2^\top(r)dr - \int_0^1 V_2(r)dr\int_0^1 V_2(r)^\top dr\right)$*, W(1) is a standard normal vector of dimension* $d-1$ *independent of* $V_2(r)$*, which is Brownian motion of dimension* $d-1$ *with variance matrix* $\Sigma_V = Q_2^\top\phi(1)\Sigma_\epsilon\phi(1)^\top Q_2$*.*

Lemma 9. *Under Assumption 1, as* $n \to \infty$

$$P_1^\top\left(D_n(\widehat{\alpha}_n - \alpha_0)\right) \to_D \sigma r_0^{-1/2}W(1), \tag{C.4}$$

where the notations are the same as those in Lemma 8.

Appendix D Proofs of lemmas

Proof of Lemma 1:

(1). According to definition, $h_i(u) = (-1)^i\exp(u^2/2)\frac{d^i}{du^i}\exp(-u^2/2)$. By integration by parts, we have

$$\begin{aligned} c_i(g) &= \int g(u)H_i(u)e^{-u^2/2}du = \frac{(-1)^i}{b_i}\int g(u)d(e^{-u^2/2})^{(i-1)} = -\frac{(-1)^i}{b_i}\int g^{(1)}(u)(e^{-u^2/2})^{(i-1)}du \\ &= \frac{b_{i-1}}{b_i}\int g^{(1)}(u)H_{i-1}(u)e^{-u^2/2}du = \frac{1}{\sqrt{i}}c_{i-1}(g^{(1)}), \end{aligned}$$

where $b_i = \sqrt{\sqrt{\pi}i!}$ and $c_{i-1}(g^{(1)})$ is the $(i-1)$-th Hermite polynomial expansion coefficient of $g^{(1)}(u)$.

Iterate the previous procedure, for $i \geq r-1$ we have

$$c_i(g) = \frac{1}{\sqrt{i(i-1)...(i-r+1)}}c_{i-r}(g^{(r)}).$$

By orthogonality, for $k \geq r$,

$$
\begin{aligned}
E|\gamma_k(u_1)|^2 &= \int \left(\gamma_k(u)\right)^2 \rho(u)\exp(u^2/2)\exp(-u^2/2)du \leq O(1)\int \left(\gamma_k(u)\right)^2 \exp(-u^2/2)du \\
&= O(1)\sum_{i=k}^{\infty} c_i^2(g) = O(1)\sum_{i=k}^{\infty} \frac{1}{i(i-1)...(i-r+1)} c_{i-r}^2(g^{(r)}) \\
&= O(k^{-r})\sum_{i=k}^{\infty} c_{i-r}^2(g^{(r)}) = o(k^{-r}),
\end{aligned}
$$

where $\sum_{i=k}^{\infty} c_{i-r}^2(g^{(r)}) = o(1)$ as $n \to \infty$ is due to Parseval's equality $\sum_{i=r}^{\infty} c_{i-r}^2(g^{(r)}) = \|g^{(r)}(u)\|_{L^2}^2 < \infty$[9].

The following formulae are necessary for the development.

$$
H_0^{(1)}(u) = 0, \qquad H_i^{(1)}(u) = \sqrt{i}H_{i-1}(u) \tag{D.1}
$$

$$
H_0(u)u = H_1(u), \qquad H_i(u)u = \sqrt{i+1}H_{i+1}(u) + \sqrt{i}H_{i-1}(u). \tag{D.2}
$$

For $E|\gamma_k^{(1)}(u)|^2$, we have

$$
\begin{aligned}
&E|\gamma_k^{(1)}(u_t)|^2 = \int \left(\gamma_k^{(1)}(u)\right)^2 \rho(u)\exp(u^2/2)\exp(-u^2/2)dw \\
\leq& O(1)\int \left(\gamma_k^{(1)}(u)\right)^2 \exp(-u^2/2)du = O(1)\sum_{i=k}^{\infty} i c_i^2(g) \\
=& O(1)\sum_{i=k}^{\infty} \frac{i}{i(i-1)...(i-r+1)} c_{i-r}^2(g^{(r)}) = O(k^{-(r-1)})\sum_{i=k}^{\infty} c_{i-r}^2(g^{(r)}) \\
=& o(k^{-(r-1)}).
\end{aligned}
$$

(2). The assertion (i) is obvious by orthogonality. To prove (ii), it follows from (D.1) that

$$
\begin{aligned}
&E\left\|\mathcal{H}_k(u_t)u_t\right\|^2 = \int \|\mathcal{H}_k(u)u\|^2 \rho(u)\exp(u^2/2)\exp(-u^2/2)du \\
\leq& O(1)\int \|\mathcal{H}_k(u)u\|^2 \exp(-u^2/2)du = \int \sum_{i=0}^{k-1} H_i^2(u)u^2\exp(-u^2/2)du \\
=& O(1)\sum_{i=0}^{k-1}\int (i+1)H_{i+1}^2(u)\exp(-u^2/2)dx + O(1)\sum_{i=1}^{k-1}\int iH_{i-1}^2(u)\exp(-u^2/2)du = O(k^2)
\end{aligned}
$$

To prove (iii), it follows from (D.2) that

$$
\begin{aligned}
&E\left\|\mathcal{H}_k^{(1)}(u_t)\right\|^2 = \int \left\|\mathcal{H}_k^{(1)}(u)\right\|^2 \rho(u)\exp(u^2/2)\exp(-u^2/2)du \\
\leq& O(1)\int \left\|\mathcal{H}_k^{(1)}(u)\right\|^2 \exp(-u^2/2)du = O(1)\int \sum_{i=0}^{k-1}\left(H_i^{(1)}(u)\right)^2 \exp(-u^2/2)dw \\
=& O(1)\sum_{i=1}^{k-1}\int iH_{i-1}^2(u)\exp(-u^2/2)du = O(k^2).
\end{aligned}
$$

Similarly, we can show that other assertions hold. □

[9] We can further assume that $\sum_{i=k}^{\infty} c_i^2 = O(k^{-\nu})$ for a constant $0 < \nu \leq r$, then $E|\gamma_k(u_t)|^2 = O(k^{-2\nu})$

Proof of Lemma 2:

(1): Due to BN decomposition for linear process, x_{1t} is stationary process and x_{2t} is integrated process (more details can be found in the proof of Lemma 3 below). Similar to the proof of Corollary 2.2 in Wang and Phillips (2009, p. 729), we can show that $\left(x_{1t}, \frac{1}{\sqrt{t}}x_{2t}^\top\right)$ has a joint density $\psi_t(x, w^\top)$ and $\left(x_{1t}, x_{1s}, \frac{1}{\sqrt{t}}x_{2t}^\top, \frac{1}{\sqrt{s}}x_{2s}^\top\right)$ has a joint probability density $\psi_{ts}(x, y, w^\top, z^\top)$ where $t > s$. Meanwhile, these functions are bounded uniformly in $(x, w^\top)$ and $(x, y, w^\top, z^\top)$ as well as t and (t, s), respectively.

(2): $\psi_t(x, w^\top) = \nu_t(x|w)q_t(w)$, where $\nu_t(x|w)$ is the conditional density of x_{1t} given $\frac{1}{\sqrt{t}}x_{2t}$. Meanwhile, x_{1t} and x_{2t} are asymptotically independent (see Remark 1 of Park and Phillips (2000, p. 1257), Lemma A.3 of Dong and Gao (2014) and Lemma A.6 of Dong et al. (2017)). According to the proof of Lemma A.5 in Cai et al. (2015), we can get that $\sup_{x,w}\left|\psi_t(x, w^\top) - \rho(x)q_t(w)\right| \to 0$ as $t \to \infty$. Similarly, we can show that $\sup_{x,y,w,z}\left|\psi_{ts}(x, y, w^\top, z^\top) - \rho_{ts}(x, y)q_{ts}(w^\top, z^\top)\right| \to 0$ as $(t, s) \to (\infty, \infty)$. Therefore, we have $q_t(x, w^\top) = \rho(x)\rho_t(w)(1 + o(1))$ for large t and $\psi_{ts}(x, y, w^\top, z^\top) = \rho_{ts}(x, y)q_{ts}(w^\top, z^\top)(1 + o(1))$ for large t and s. □

Proof of Lemma 3:

We consider the Beveridge and Nelson (BN) decomposition (Beveridge and Nelson, 1981) for x_t. Without loss of generality, in what follows let $x_0 = 0$ almost surely. It follows that

$$
\begin{aligned}
(1 - L)x_t = \phi(L)\epsilon_t &= \left(\phi(1) - (1 - L)\widetilde{\phi}(L)\right)\epsilon_t \\
x_t &= \sum_{i=1}^{t}\phi(1)\epsilon_i - \sum_{i=1}^{t}\widetilde{\phi}(L)\left(\epsilon_i - \epsilon_{i-1}\right) = \sum_{i=1}^{t}\phi(1)\epsilon_i + \widetilde{\phi}(L)\left(\epsilon_0 - \epsilon_t\right),
\end{aligned}
$$

where $\phi(L) = \sum_{j=0}^{\infty}\phi_j L^j$ with $\{\phi_j\}$ being a $d \times d$ matrix such that $\phi_0 = I_d$, $\sum_{j=0}^{\infty} j\|\phi_j\| < \infty$, $\phi(1) = \sum_{j=0}^{\infty}\phi_j$, and $\widetilde{\phi}(L) = \sum_{j=0}^{\infty}\widetilde{\phi}_j L^j$ with $\widetilde{\phi}_j = \sum_{k=j+1}^{\infty}\phi_k$. Then based on lemma 2.1 in Phillips and Solo (1992), we have $\sum_{j=0}^{\infty}\|\widetilde{\phi}_j\|^2 < \infty$.

Since θ_0 is the standardized cointegrated coefficient, it is obvious that $\theta_0^\top\phi(1) = 0_{1\times d}$. Therefore, we can rewrite $x_{1t} = \theta_0^\top x_t$ as follows

$$
x_{1t} = \theta_0^\top x_t = \theta_0^\top\left(\sum_{i=1}^{t}\phi(1)\epsilon_i + \widetilde{\phi}(L)\left(\epsilon_0 - \epsilon_t\right)\right) = \theta_0^\top\widetilde{\phi}(L)(\epsilon_0 - \epsilon_t).
$$

In terms of x_{2t}, we have

$$
x_{2t} = Q_2^\top\left(\sum_{i=1}^{t}\phi(1)\epsilon_i + \widetilde{\phi}(L)\left(\epsilon_0 - \epsilon_t\right)\right) = Q_2^\top\phi(1)\sum_{i=1}^{t}\epsilon_i + \zeta_t,
$$

where $\zeta_t = Q_2^\top\widetilde{\phi}(L)\left(\epsilon_0 - \epsilon_t\right)$ is a stationary process.

After simple algebra, we can show that

$$
\begin{aligned}
x_{1t} &= \sum_{i=0}^{t-1} -\theta_0^\top\widetilde{\phi}_i\epsilon_{t-i} + \sum_{i=t}^{\infty}\theta_0^\top\left(\widetilde{\phi}_{i-t} - \widetilde{\phi}_i\right)\epsilon_{t-i} := \sum_{i=0}^{\infty} d_i\epsilon_{t-i} \\
\zeta_t &= \sum_{i=0}^{t-1} -Q_2^\top\widetilde{\phi}_i\epsilon_{t-i} + \sum_{i=t}^{\infty} Q_2^\top\left(\widetilde{\phi}_{i-t} - \widetilde{\phi}_i\right)\epsilon_{t-i} := \sum_{i=0}^{\infty} b_i\epsilon_{t-i}.
\end{aligned}
$$

Then, we can show that

$$\sum_{i=0}^{\infty}\|d_i\|^2=\sum_{i=0}^{t-1}\left\|\theta_0^\top\widetilde{\phi}_i\right\|^2+\sum_{i=t}^{\infty}\left\|\theta_0^\top\left(\widetilde{\phi}_{i-t}-\widetilde{\phi}_i\right)\right\|^2\leq 5\|\theta_0\|^2\sum_{i=0}^{\infty}\left\|\widetilde{\phi}_i\right\|^2=5\sum_{i=0}^{\infty}\left\|\widetilde{\phi}_i\right\|^2<\infty,$$

where d_i is a $1\times d$-dimensional matrix. Similarly, we can also show that $\sum_{i=0}^{\infty}\|b_i\|^2<\infty$, where b_i is a $(d-1)\times d$-dimensional matrix.

We set $d=2$ for all the following proofs in the supplementary material, this is just for notational simplicity. The proof for the general case is essentially identical. For part (a), we have

$$\begin{aligned}
&E\left|\frac{1}{n}\sum_{t=1}^{n}f(x_{1t})-E[f(x_{11})]\right|^2\\
=&\frac{1}{n^2}\sum_{t=1}^{n}E\left[f(x_{1t})-E\left[f(x_{1t})\right]\right]^2+\frac{2}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}E\left[\left(f(x_{1t})-E\left[f(x_{1t})\right]\right)\left(f(x_{1s})-E\left[f(x_{1s})\right]\right)\right]\\
\leq&O(n^{-1})+\frac{2}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|Cov\left(f(x_{1t}),f(x_{1s})\right)\right|=O(n^{-1})+o(1)=o(1)
\end{aligned}$$

For part (b), we consider the following expression

$$\begin{aligned}
&\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}f(x_{1t})x_{2t}\\
=&E\left[f(x_{11})\right]\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}x_{2t}+\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\left(f(x_{1t})-E\left[f(x_{1t})\right]\right)x_{2t}\\
=&E\left[f(x_{11})\right]\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}Q_2^\top\phi(1)\sum_{i=1}^{t}\epsilon_i+E\left[f(x_{11})\right]\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\zeta_t+\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\left(f(x_{1t})-E\left[f(x_{1t})\right]\right)x_{2t}\\
=&C_1+C_2+C_3
\end{aligned}$$

It is known that

$$C_1=E\left[f(x_{11})\right]\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}Q_2^\top\phi(1)\sum_{i=1}^{t}\epsilon_i\to_D E[f(x_{11})]\int_0^1 V_2(r)dr,$$

where V_2 is Brownian motion with variance $\Sigma_V=Q_2^\top\phi(1)\Sigma_\epsilon\phi(1)^\top Q_2$.

For C_2, consider

$$\begin{aligned}
&E\left[\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\zeta_t\right]^2\leq\frac{1}{n^2}\sum_{t=1}^{n}E\left[\zeta_t\right]^2=\frac{1}{n^2}\sum_{t=1}^{n}E\left[Q_2^\top\tilde{\phi}(L)(\epsilon_0-\epsilon_t)\right]^2\\
\leq&\frac{2}{n^2}\sum_{t=1}^{n}E\left[\sum_{i=0}^{\infty}b_i\epsilon_{-i}\right]^2+\frac{2}{n^2}\sum_{t=1}^{n}E\left[\sum_{i=0}^{\infty}b_i\epsilon_{t-i}\right]^2=\frac{4}{n}\sum_{i=0}^{\infty}b_i\Sigma_\epsilon b_i^\top\leq O(1)\frac{1}{n}\sum_{i=0}^{\infty}\|b_i\|^2=o(1)
\end{aligned}$$

In terms of C_3, we have

$$\begin{aligned}
&C_3^2=\left(\frac{1}{n\sqrt{n}}\sum_{t=1}^{n}\left(f(x_{1t})-E\left[f(x_{1t})\right]\right)x_{2t}\right)^2\\
=&\frac{1}{n^3}\sum_{t=1}^{n}\left[\left(f(x_{1t})-E\left[f(x_{1t})\right]\right)x_{2t}\right]^2+\frac{2}{n^3}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left(f(x_{1t})-E\left[f(x_{1t})\right]\right)\left(f(x_{1s})-E\left[f(x_{1s})\right]\right)x_{2t}x_{2s}
\end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{n^3}\sum_{t=1}^{a_n}\left[\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)x_{2t}\right]^2+\frac{1}{n^3}\sum_{t=a_n+1}^{n}\left[\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)x_{2t}\right]^2\\
&+\frac{2}{n^3}\sum_{t=2}^{b_n}\sum_{s=1}^{t-1}\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)x_{2t}x_{2s}\\
&+\frac{2}{n^3}\sum_{t=b_n+1}^{n}\sum_{s=1}^{b_n}\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)x_{2t}x_{2s}\\
&+\frac{2}{n^3}\sum_{t=b_n+2}^{n}\sum_{s=b_n+1}^{t-1}\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)x_{2t}x_{2s}\\
=&C_{3,1}+C_{3,2}+2C_{3,3}+2C_{3,4}+2C_{3,5}
\end{aligned}
$$

where $a_n/n^2 \to 0$, $b_n/n \to 0$, and $a_n \to \infty$, $b_n \to \infty$ as $n \to \infty$.

For $C_{3,1}$, by the fact that $\sup_{1\le t\le n}|x_{2t}|/\sqrt{n} = O_P(1)$, we have

$$
\begin{aligned}
C_{3,1}=&\frac{1}{n^3}\sum_{t=1}^{a_n}\left[\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)x_{2t}\right]^2=\frac{1}{n^2}\sum_{t=1}^{a_n}\left[\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\frac{x_{2t}}{\sqrt{n}}\right]^2\\
=&O_P(1)\frac{1}{n^2}\sum_{t=1}^{a_n}\Big[f(x_{1t})-E\left[f(x_{1t})\right]\Big]^2=O_P(n^{-2}a_n)=o_P(1).
\end{aligned}
$$

For $C_{3,2}$, since t is large enough, by Lemma 2, write

$$
\begin{aligned}
EC_{3,2}=&E\left[\frac{1}{n^3}\sum_{t=a_n+1}^{n}\left[\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)x_{2t}\right]^2\right]\\
=&\frac{1}{n^3}\sum_{t=a_n+1}^{n}tE\Big[f(x_{1t})-E\left[f(x_{1t})\right]\Big]^2E\left[\frac{x_{2t}}{\sqrt{t}}\right]^2(1+o(1))\\
=&O(1)\frac{1}{n^3}\sum_{t=a_n+1}^{n}t=O(n^{-1})=o(1).
\end{aligned}
$$

For $C_{3,3}$, note that

$$
\begin{aligned}
|C_{3,3}|=&\left|\frac{1}{n^3}\sum_{t=2}^{b_n}\sum_{s=1}^{t-1}\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)x_{2t}x_{2s}\right|\\
\le&\frac{1}{n^2}\sum_{t=2}^{b_n}\sum_{s=1}^{t-1}\left|\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)\right|\left|\frac{x_{2t}}{\sqrt{n}}\frac{x_{2s}}{\sqrt{n}}\right|\\
=&O_P(1)\frac{1}{n^2}\sum_{t=2}^{b_n}\sum_{s=1}^{t-1}\left|\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)\right|
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left[\frac{1}{n^2}\sum_{t=2}^{b_n}\sum_{s=1}^{t-1}\left|\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)\right|\right]\\
\le&\frac{1}{n^2}\sum_{t=2}^{b_n}\sum_{s=1}^{t-1}\left[E\Big[f(x_{1t})-E\left[f(x_{1t})\right]\Big]^2E\Big[f(x_{1s})-E\left[f(x_{1s})\right]\Big]^2\right]^{1/2}=O(n^{-2}b_n^2)=o(1)
\end{aligned}
$$

For $C_{3,4}$, consider

$$
|C_{3,4}|=\left|\frac{1}{n^3}\sum_{t=b_n+1}^{n}\sum_{s=1}^{b_n}\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)x_{2t}x_{2s}\right|
$$

$$
\begin{aligned}
&\leq \frac{1}{n^2}\sum_{t=b_n+1}^{n}\sum_{s=1}^{b_n}\left|\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)\frac{x_{2t}}{\sqrt{n}}\frac{x_{2s}}{\sqrt{n}}\right| \\
&= O_P(1)\frac{1}{n^2}\sum_{t=b_n+1}^{n}\sum_{s=1}^{b_n}\left|\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)\right|
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left[\frac{1}{n^2}\sum_{t=b_n+1}^{n}\sum_{s=1}^{b_n}\left|\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)\right|\right] \\
\leq &\frac{1}{n^2}\sum_{t=b_n+1}^{n}\sum_{s=1}^{b_n}\left[E\Big[f(x_{1t})-E\left[f(x_{1t})\right]\Big]^2 E\Big[f(x_{1s})-E\left[f(x_{1s})\right]\Big]^2\right]^{1/2} = O(n^{-1}b_n)=o(1)
\end{aligned}
$$

In terms of $C_{3,5}$, since t and s are large enough, by Lemma 2, we have

$$
\begin{aligned}
&\left|EC_{3,5}\right| = \left|E\left[\frac{1}{n^3}\sum_{t=b_n+2}^{n}\sum_{s=b_n+1}^{t-1}\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)x_{2t}x_{2s}\right]\right| \\
=&\left|\frac{1}{n^3}\sum_{t=b_n+2}^{n}\sum_{s=b_n+1}^{t-1}\sqrt{t}\sqrt{s}E\left[\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s})-E\left[f(x_{1s})\right]\Big)\right]E\left[\frac{x_{2t}}{\sqrt{t}}\frac{x_{2s}}{\sqrt{s}}\right](1+o(1))\right| \\
\leq&\frac{1}{n^3}\sum_{t=b_n+2}^{n}\sum_{s=b_n+1}^{t-1}\sqrt{t}\sqrt{s}\left|Cov(f(x_{1t}),f(x_{1s}))\right|\left|E\left[\frac{x_{2t}}{\sqrt{t}}\frac{x_{2s}}{\sqrt{s}}\right]\right|(1+o(1)) \\
\leq&O(1)\frac{1}{n^2}\sum_{t=b_n+2}^{n}\sum_{s=b_n+1}^{t-1}\left|Cov(f(x_{1t}),f(x_{1s}))\right| = o(1).
\end{aligned}
$$

Without loss of generality, in what follows we abuse the density by neglecting the argument on a_n and b_n as we did before.

For part (c), we consider the following expression

$$
\begin{aligned}
&\frac{1}{n^2}\sum_{t=1}^{n}f(x_{1t})\left(x_{2t}\right)^2 \\
=&E\left[f(x_{11})\right]\frac{1}{n^2}\sum_{t=1}^{n}\left(x_{2t}\right)^2+\frac{1}{n^2}\sum_{t=1}^{n}\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\left(x_{2t}\right)^2 \\
=&E\left[f(x_{11})\right]\frac{1}{n^2}\sum_{t=1}^{n}\left(\sum_{i=1}^{t}Q_2^\top\phi(1)\epsilon_i\right)^2+E\left[f(x_{11})\right]\frac{1}{n^2}\sum_{t=1}^{n}\zeta_t^2+E\left[f(x_{11})\right]\frac{2}{n^2}\sum_{t=1}^{n}\sum_{i=1}^{t}Q_2^\top\phi(1)\epsilon_i\zeta_t \\
&+\frac{1}{n^2}\sum_{t=1}^{n}\Big(f(x_{1t})-E\left[f(x_{1t})\right]\Big)\left(x_{2t}\right)^2 \\
=&C_4+C_5+C_6+C_7
\end{aligned}
$$

It is known that $C_4 = E\left[f(x_{11})\right]\frac{1}{n^2}\sum_{t=1}^{n}\left(\sum_{i=1}^{t}Q_2^\top\phi(1)\epsilon_i\right)^2 \to_D E[f(x_{11})]\int_0^1 V_2(r)^2dr$, where V_2 is Brownian motion with variance $\Sigma_V = Q_2^\top\phi(1)\Sigma_\epsilon\phi(1)^\top Q_2$.

For C_5, consider

$$
E\left[\frac{1}{n^2}\sum_{t=1}^{n}\zeta_t^2\right] = \frac{1}{n^2}\sum_{t=1}^{n}E\left[\tilde{\phi}(L)(\epsilon_0-\epsilon_t)\right]^2
$$

$$\leq \frac{2}{n^2}\sum_{t=1}^{n} E\left[\sum_{i=0}^{\infty} b_i\epsilon_{-i}\right]^2 + \frac{2}{n^2}\sum_{t=1}^{n} E\left[\sum_{i=0}^{\infty} b_i\epsilon_{t-i}\right]^2 = \frac{4}{n}\sum_{i=0}^{\infty} b_i\Sigma_\epsilon b_i^\top \leq O(1)\frac{1}{n}\sum_{i=0}^{\infty}\|b_i\|^2 = o(1)$$

Then by Cauchy-Schwartz inequality, we can immediately obtain that $C_6 = o_P(1)$.

In term of C_7, we have

$$\begin{aligned}
EC_7^2 =& E\left[\frac{1}{n^2}\sum_{t=1}^{n}\Big(f(x_{1t}) - E\left[f(x_{1t})\right]\Big)(x_{2t})^2\right]^2 \\
=&\frac{1}{n^4}\sum_{t=1}^{n} E\left[\Big(f(x_{1t}) - E\left[f(x_{1t})\right]\Big)(x_{2t})^2\right]^2 \\
&+ \frac{2}{n^4}\sum_{t=2}^{n}\sum_{s=1}^{t-1} E\left[\Big(f(x_{1t}) - E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s}) - E\left[f(x_{1s})\right]\Big)(x_{2t})^2(x_{2s})^2\right] \\
=&\frac{1}{n^4}\sum_{t=1}^{n} t^2 E\left[\Big(f(x_{1t}) - E\left[f(x_{1t})\right]\Big)\right]^2 E\left[\frac{x_{2t}}{\sqrt{t}}\right]^4 (1+o(1)) \\
&+ \frac{2}{n^4}\sum_{t=2}^{n}\sum_{s=1}^{t-1} tsE\left[\Big(f(x_{1t}) - E\left[f(x_{1t})\right]\Big)\Big(f(x_{1s}) - E\left[f(x_{1s})\right]\Big)\right] E\left[\frac{x_{2t}}{\sqrt{t}}\frac{x_{2s}}{\sqrt{s}}\right]^2 (1+o(1)) \\
\leq& O(n^{-1}) + O(1)\frac{1}{n^2}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|Cov(f(x_{1t}), f(x_{1s}))\right| = O(n^{-1}) + o(1) = o(1)
\end{aligned}$$

For part (d), according to Lemma 3. (a), (b) and (c), we can immediately obtain that

$$\begin{aligned}
&\frac{1}{n^2}\sum_{t=1}^{n} f(x_{1t})(x_{2t} - \bar{x}_2)^2 \\
=&\frac{1}{n^2}\sum_{t=1}^{n} f(x_{1t})x_{2t}^2 - \frac{2}{n^2}\sum_{t=1}^{n} f(x_{1t})x_{2t}\bar{x}_2 + \frac{1}{n^2}\sum_{t=1}^{n} f(x_{1t})\bar{x}_2^2 \\
=&\frac{1}{n^2}\sum_{t=1}^{n} f(x_{1t})x_{2t}^2 - \frac{2}{n\sqrt{n}}\sum_{t=1}^{n} f(x_{1t})x_{2t}\frac{1}{n\sqrt{n}}\sum_{s=1}^{n} x_{2s} + \frac{1}{n}\sum_{t=1}^{n} f(x_{1t})\left(\frac{1}{n\sqrt{n}}\sum_{s=1}^{n} x_{2s}\right)^2 \\
=&\frac{1}{n^2}\sum_{t=1}^{n} f(x_{1t})x_{2t}^2 - E\left[f(x_{11})\right]\left(\frac{1}{n\sqrt{n}}\sum_{t=1}^{n} x_{2t}\right)^2 + o_P(1) \\
&\to_D E\left[f(x_{11})\right]\left[\int_0^1 V_2(r)^2 dr - \left(\int_0^1 V_2(r)dr\right)^2\right]. \qquad \square
\end{aligned}$$

Proof of Lemma 4: According to Lemma 3 (c), the conditional variance process is given by

$$\begin{aligned}
&\sum_{t=1}^{n} E\left[\left(\frac{1}{n}g_0^{(1)}(x_{1t-1})\left(x_{2t-1} - \frac{1}{n}\sum_{s=1}^{n} x_{2s-1}\right)e_t\right)^2 |\mathcal{F}_{n,t-1}\right] \\
=&\sigma^2\frac{1}{n^2}\sum_{t=1}^{n}\left(g_0^{(1)}(x_{1t-1})\right)^2\left(x_{2t-1} - \frac{1}{n}\sum_{s=1}^{n} x_{2s-1}\right)^2 \to_D \sigma^2 E\left[g_0^{(1)}(x_{1t-1})\right]^2\left[\int_0^1 V_2(r)^2 dr - \left(\int_0^1 V_2(r)dr\right)^2\right].
\end{aligned}$$

To make the conditional Lindeberg's condition fulfilled, we have

$$\sum_{t=1}^{n} E\left[\left(\frac{1}{n}g_0^{(1)}(x_{1t-1})\left(x_{2t-1}-\frac{1}{n}\sum_{s=1}^{n}x_{2s-1}\right)e_t\right)^4|\mathcal{F}_{n,t-1}\right]$$

$$=E\left[e_t^4|\mathcal{F}_{n,t-1}\right]\frac{1}{n^4}\sum_{t=1}^{n}\left(g_0^{(1)}(x_{1t-1})\right)^4\left(x_{2t-1}-\frac{1}{n}\sum_{s=1}^{n}x_{2s-1}\right)^4$$

$$\leq 8E\left[e_t^4|\mathcal{F}_{n,t-1}\right]\frac{1}{n^2}\sum_{t=1}^{n}\left(g_0^{(1)}(x_{1t-1})\right)^4\left(\frac{x_{2t-1}}{\sqrt{n}}\right)^4+8E\left[e_t^4|\mathcal{F}_{n,t-1}\right]\frac{1}{n^2}\sum_{t=1}^{n}\left(g_0^{(1)}(x_{1t-1})\right)^4\left(\frac{1}{n}\sum_{s=1}^{n}\frac{x_{2s-1}}{\sqrt{n}}\right)^4$$

$$=O_P(1)\frac{1}{n^2}\sum_{t=1}^{n}\left(g_0^{(1)}(x_{1t-1})\right)^4=O_P(n^{-1})=o_P(1).$$

Then, the stated result follows from Corollary 3.1 of Hall and Heyde (1980). □

Proof of Lemma 5:

To prove the the result (1), we consider

$$E\left\|\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_k(x_{1t})\mathcal{H}_k(x_{1t})^\top-E\left[\mathcal{H}_k(x_{11})\mathcal{H}_k(x_{11})^\top\right]\right\|^2$$

$$=\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}E\left[\frac{1}{n}\sum_{t=1}^{n}H_i(x_{1t})H_j(x_{1t})-E\left[H_i(x_{11})H_j(x_{11})\right]\right]^2$$

$$=\frac{1}{n^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{t=1}^{n}E\left[H_i(x_{1t})H_j(x_{1t})-E\left[H_i(x_{1t})H_j(x_{1t})\right]\right]^2$$

$$+\frac{2}{n^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1}E\left[\left(H_i(x_{1t})H_j(x_{1t})-E\left[H_i(x_{1t})H_j(x_{1t})\right]\right)\left(H_i(x_{1s})H_j(x_{1s})-E\left[H_i(x_{1s})H_j(x_{1s})\right]\right)\right]$$

$$=C_8+2C_9.$$

For the first term C_8, according to Assumption 1.7 (b), we have

$$C_8=\frac{1}{n^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{t=1}^{n}E\left[H_i(x_{1t})H_j(x_{1t})-E\left[H_i(x_{1t})H_j(x_{1t})\right]\right]^2$$
$$\leq\frac{1}{n^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{t=1}^{n}E\left[H_i(x_{1t})H_j(x_{1t})\right]^2=o(1).$$

In terms of C_9, according to Assumption 1.7 (d), write

$$|C_9|=\left|\frac{1}{n^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1}E\left[\left(H_i(x_{1t})H_j(x_{1t})-E\left[H_i(x_{1t})H_j(x_{1t})\right]\right)\left(H_i(x_{1s})H_j(x_{1s})-E\left[H_i(x_{1s})H_j(x_{1s})\right]\right)\right]\right|$$

$$=\left|\frac{1}{n^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1}Cov\left(H_i(x_{1t})H_j(x_{1t}),H_i(x_{1s})H_j(x_{1s})\right)\right|$$

$$\leq\frac{1}{n^2}\sum_{i=0}^{k-1}\sum_{j=0}^{k-1}\sum_{t=2}^{n}\sum_{s=1}^{t-1}\left|Cov\left(H_i(x_{1t})H_j(x_{1t}),H_i(x_{1s})H_j(x_{1s})\right)\right|=o(1).$$

Therefore it is obvious that

$$\left\| \frac{1}{n}\sum_{t=1}^{n} \mathcal{H}_k(x_{1t})\mathcal{H}_k(x_{1t})^\top - E\left[\mathcal{H}_k(x_{11})\mathcal{H}_k(x_{11})^\top\right] \right\| \to_P 0.$$

To prove the the result (2), given any $\epsilon > 0$, define for any function $f(x) \in L^2(\mathbb{R}, \exp(-x^2/2))$,

$$f^{\epsilon}_{\sup}(x) = \sup_{|\alpha-1|<\epsilon} \sup_{|b|<\epsilon} \left|f(ax+b)\right|$$
$$\tilde{f}^{\epsilon}_{\sup}(x) = \sup_{|\alpha-1|<\epsilon} \sup_{|b|<\epsilon} \left|f(ax+b)\rho^{1/2}(ax+b)\right|,$$

where $\rho(u)$ is the density function for $u_t = \theta_0^\top x_t$. And it is obvious that $f^{\epsilon}_{\sup}(x)\rho^{1/2}(x) \leq \tilde{f}^{\epsilon}_{\sup}(x)$.

Then it is easy to show that $f(x)\rho^{1/2}(x) \in L^2(\mathbb{R})$:

$$\int f(x)^2\rho(x)dx = \int f(x)^2\rho(x)\exp(x^2/2)\exp(-x^2/2)dx \leq O(1)\int f(x)^2\exp(-x^2/2)dx = O(1).$$

Similar to the proof of Lemma A1 in Park and Phillips (2000), we can show that $\tilde{f}^{\epsilon}_{\sup}(x) \in L^2(\mathbb{R})$. Because of square integrability, we may assume without loss of generality that for sufficient large $|x|$, say $|x| > M$, $|f(x)\rho^{1/2}(x)|$ is monotone. Therefore, we have $\tilde{f}^{\epsilon}_{\sup}(x) = \left|f(x-\epsilon)\rho^{1/2}(x-\epsilon)\right|$ for $x > M+\epsilon$ and $\tilde{f}^{\epsilon}_{\sup}(x) = \left|f(x+\epsilon)\rho^{1/2}(x+\epsilon)\right|$ for $x < -M-\epsilon$. Then we can obtain

$$\int_{|x|>M+\epsilon}\left(\tilde{f}^{\epsilon}_{\sup}(x)\right)^2 dx = \int_{|x|>M+\epsilon}\left(f(x\pm\epsilon)\rho^{1/2}(x\pm\epsilon)\right)^2 dx = \int_{|x|>M}\left(f(x)\rho^{1/2}(x))\right)^2 dx.$$

Meanwhile, on the interval $[-M-\epsilon, M+\epsilon]$, the function $\tilde{f}^{\epsilon}_{\sup}(x)$ can be approximated by $|f(x)\rho^{1/2}(x)|$ as accurate as we wish due to continuity as long as ϵ is sufficiently small. Therefore, we can conclude

$$\int \left(f^{\epsilon}_{\sup}(x)\right)^2\rho(x)dx \leq \int\left(\tilde{f}^{\epsilon}_{\sup}(x)\right)^2 dx = \int\left(f(x)\right)^2\rho(x)dx(1+o(1)).$$

More details have been discussed in Park and Phillips (2000) and Dong et al. (2016).

Since $\sup_{1\leq t\leq n}|x_{2t}|/\sqrt{n} = O_P(1)$, $|\widehat{\alpha}^1 - \alpha^1_0| = O_P(n^{-2})$, and $|\widehat{\alpha}^2 - \alpha^2_0| = O_P(n^{-1})$, we have for any $\epsilon > 0$ and large n,

$$\left|f(\widehat{\eta}_t)\right| = \left|f\left(\widehat{\alpha}^1 x_{1t} + \widehat{\alpha}^2 x_{2t}\right)\right| \leq f^{\epsilon}_{\sup}(x_{1t})(1+o_P(1)),$$

uniformly in t.

Then, by mean value theorem we have

$$\begin{aligned}
&\left\|\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_k(\widehat{\eta}_t)\mathcal{H}_k(\widehat{\eta}_t)^\top - \frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_k(x_{1t})\mathcal{H}_k(x_{1t})^\top\right\| \\
\leq&\frac{1}{n}\sum_{t=1}^{n}\left\|\left(\mathcal{H}_k(\widehat{\eta}_t) - \mathcal{H}_k(x_{1t})\right)\mathcal{H}_k(x_{1t})^\top\right\| + \frac{1}{n}\sum_{t=1}^{n}\left\|\mathcal{H}_k(x_{1t})\left(\mathcal{H}_k(\widehat{\eta}_t) - \mathcal{H}_k(x_{1t})\right)^\top\right\| \\
&+ \frac{1}{n}\sum_{t=1}^{n}\left\|\left(\mathcal{H}_k(\widehat{\eta}_t) - \mathcal{H}_k(x_{1t})\right)\left(\mathcal{H}_k(\widehat{\eta}_t) - \mathcal{H}_k(x_{1t})\right)^\top\right\| \\
=&\frac{1}{n}\sum_{t=1}^{n}\left\|\mathcal{H}^{(1)}_k(\widehat{\eta}^*_t)\mathcal{H}_k(x_{1t})^\top\left(\left(\widehat{\alpha}^1_n - \alpha^1_0\right)x_{1t} + \left(\widehat{\alpha}^2_n - \alpha^2_0\right)x_{2t}\right)\right\|
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{n}\sum_{t=1}^{n}\left\|\mathcal{H}_k(x_{1t})\mathcal{H}_k^{(1)}(\widehat{\eta}_t^*)^\top\left(\left(\widehat{\alpha}_n^1-\alpha_0^1\right)x_{1t}+\left(\widehat{\alpha}_n^2-\alpha_0^2\right)x_{2t}\right)\right\| \\
&+\frac{1}{n}\sum_{t=1}^{n}\left\|\mathcal{H}_k^{(1)}(\widehat{\eta}_t^*)\mathcal{H}_k^{(1)}(\widehat{\eta}_t^*)^\top\left(\left(\widehat{\alpha}_n^1-\alpha_0^1\right)x_{1t}+\left(\widehat{\alpha}_n^2-\alpha_0^2\right)x_{2t}\right)^2\right\| \\
\leq & O_P(1)\frac{1}{n^3}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k^{(1)}\right\|\|\mathcal{H}_k x\|\right)_{\sup}(x_{1t})(1+o_P(1))+O_P(1)\frac{1}{n^{3/2}}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k^{(1)}\right\|\|\mathcal{H}_k\|\right)_{\sup}(x_{1t})(1+o_P(1)) \\
&+O_P(1)\frac{1}{n^5}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k^{(1)}\right\|\left\|\mathcal{H}_k^{(1)}x^2\right\|\right)_{\sup}(x_{1t})(1+o_P(1))+O_P(1)\frac{1}{n^2}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k^{(1)}\right\|^2\right)_{\sup}(x_{1t})(1+o_P(1)) \\
=&O_P(1)C_{10}+\cdots+O_P(1)C_{13}
\end{aligned}
$$

For C_{10}, write

$$
\begin{aligned}
&E\left[\frac{1}{n^3}\sum_{t=1}^{n}\left(\left\|\mathcal{H}_k^{(1)}\right\|\|\mathcal{H}_k x\|\right)_{\sup}(x_{1t})\right]=\frac{1}{n^3}\sum_{t=1}^{n}E\left[\left\|\mathcal{H}_k^{(1)}(x_{1t-1})\right\|\left\|\mathcal{H}_k(x_{1t-1})x_{1t-1}\right\|\right](1+o(1)) \\
\leq&\frac{1}{n^3}\sum_{t=1}^{n}\left[E\left\|\mathcal{H}_k^{(1)}(x_{1t-1})\right\|^2 E\left\|\mathcal{H}_k(x_{1t-1})x_{1t-1}\right\|^2\right]^{1/2}(1+o(1))=O(n^{-2}k^2)
\end{aligned}
$$

Similarly, we can show that $C_{11}=O_P(n^{-1/2}k^{3/2})$, $C_{12}=O_P(n^{-4}k^3)$, and $C_{13}=O_P(n^{-1}k^2)$. Therefore, $\left\|\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_k(\widehat{\eta}_t)\mathcal{H}_k(\widehat{\eta}_t)^\top-\frac{1}{n}\sum_{t=1}^{n}\mathcal{H}_k(x_{1t})\mathcal{H}_k(x_{1t})^\top\right\|\to_P 0$. □

Proof of Lemma 6:

According to Hermite expansion, we have

$$
\begin{aligned}
\bar{C}_k(\alpha_0)-C_{0,k}=&\left(Z(\alpha_0)^\top Z(\alpha_0)\right)^{-1}Z(\alpha_0)^\top\gamma(\alpha_0)+\left(Z(\alpha_0)^\top Z(\alpha_0)\right)^{-1}Z(\alpha_0)^\top e \\
=&C_{14}+C_{15}
\end{aligned}
$$

Regarding C_{14}, it follows from Lemma 5 that

$$
\begin{aligned}
\|C_{14}\|^2=&\left\|(Z(\alpha_0)^\top Z(\alpha_0))^{-1}Z(\alpha_0)^\top\gamma(\alpha_0)\right\|^2 \\
=&\gamma(\alpha_0)^\top Z(\alpha_0)\left(Z(\alpha_0)^\top Z(\alpha_0)\right)^{-1}\left(Z(\alpha_0)^\top Z(\alpha_0)/n\right)^{-1}Z(\alpha_0)^\top\gamma(\alpha_0)/n \\
=&\gamma(\alpha_0)^\top Z(\alpha_0)\left(Z(\alpha_0)^\top Z(\alpha_0)\right)^{-1}\mathcal{H}_x^{-1}Z(\alpha_0)^\top\gamma(\alpha_0)/n(1+o_P(1)) \\
\leq&\lambda_{min}^{-1}(\mathcal{H}_x)\cdot\gamma(\alpha_0)^\top Z(\alpha_0)\left(Z(\alpha_0)^\top Z(\alpha_0)\right)^{-1}Z(\alpha_0)^\top\gamma(\alpha_0)/n(1+o_P(1)) \\
\leq&\lambda_{min}^{-1}(\mathcal{H}_x)\cdot\lambda_{max}\left(Z(\alpha_0)(Z(\alpha_0)^\top Z(\alpha_0))^{-1}Z(\alpha_0)^\top\right)\cdot\left\|\gamma(\alpha_0)\right\|^2/n(1+o_P(1)),
\end{aligned}
$$

where the first inequality is due to Magnus and Neudecker (2007, exercise 5 on P. 267). Since $Z(\alpha_0)\left(Z(\alpha_0)^\top Z(\alpha_0)\right)^{-1}Z(\alpha_0$ is a symmetric and idempotent matrix, the max eigenvalue is $\lambda_{max}=1$. Also note that, according to Lemma 1.(1), we have $\frac{1}{n}\left\|\gamma(\alpha_0)\right\|^2=o_P(k^{-r})$.

Similarly, for C_{15}, we consider

$$
\begin{aligned}
\|C_{15}\|^2=&\left\|(Z(\alpha_0)^\top Z(\alpha_0))^{-1}Z(\alpha_0)^\top e\right\|^2 \\
=&e^\top Z(\alpha_0)\left(Z(\alpha_0)^\top Z(\alpha_0)\right)^{-1}\left(Z(\alpha_0)^\top Z(\alpha_0)/n\right)^{-1}Z(\alpha_0)^\top e/n
\end{aligned}
$$

$$
\begin{aligned}
&= e^{\top} Z(\alpha_0) \left(Z(\alpha_0)^{\top} Z(\alpha_0) \right)^{-1} \mathcal{H}_x^{-1} Z(\alpha_0)^{\top} e/n(1+o_P(1)) \\
&\leq \lambda_{min}^{-1} \left(\mathcal{H}_x\right) \cdot e^{\top} Z(\alpha_0)/n \left(Z(\alpha_0)^{\top} Z(\alpha_0)/n \right)^{-1} Z(\alpha_0)^{\top} e/n(1+o_P(1)) \\
&\leq \lambda_{min}^{-2} \left(\mathcal{H}_x\right) \cdot \left\| \frac{1}{n} Z(\alpha_0)^{\top} e \right\|^2 (1+o_P(1)),
\end{aligned}
$$

where $E \left\| \frac{1}{n} Z(\alpha_0)^{\top} e \right\|^2 = \frac{1}{n^2} \sum_{i=0}^{k-1} \sum_{t=1}^{n} E \left[H_i(x_{1t-1}) \right]^2 = O(n^{-1}k)$.

According to the above derivations, we can obtain

$$
\left\| \bar{C}_k(\alpha_0) - C_{0,k} \right\| = O_p(n^{-1/2} k^{1/2}) + o_P(k^{-r/2}).
$$

□

Proof of Lemma 7: When $i = j = 1$, we have

$$
\begin{aligned}
\left| T^{1,1} \right| &= \left| \sum_{t=1}^{n} f(x_{1t-1}) \left(x_{1t-1}\right)^2 \left[\left(\alpha^1 - \alpha_0^1 \right) x_{1t-1} + \left(\alpha^2 - \alpha_0^2 \right) x_{2t-1} \right] \right| \\
\leq & \left| \alpha^1 - \alpha_0^1 \right| \sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^3 \right| + \left| \alpha^2 - \alpha_0^2 \right| \sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^2 x_{2t-1} \right| \\
\leq & n^{-1/2+\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^3 \right| + n^{-1/2+\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^2 \right| \left| \frac{x_{2t-1}}{\sqrt{n}} \right| \\
= & n^{-1/2+\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^3 \right| + O_P(1) n^{-1/2+\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^2 \right|,
\end{aligned}
$$

where

$$
\begin{aligned}
E \left[\sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^3 \right| \right] &= nE \left| f(x_{1t-1}) \left(x_{11}\right)^3 \right| \\
E \left[\sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^2 \right| \right] &= nE \left| f(x_{1t-1}) \left(x_{11}\right)^2 \right|
\end{aligned}
$$

Therefore, it follows that $T^{1,1} = O_P(n^{1/2+\delta}) E \left| f(x_{11}) \left(x_{11}\right)^3 \right|$.

For $i = 1$, $j = 2$, write

$$
\begin{aligned}
\left| T^{1,2} \right| &= \left| \sum_{t=1}^{n} f(x_{1t-1}) x_{1t-1} x_{2t-1} \left[\left(\alpha^1 - \alpha_0^1 \right) x_{1t-1} + \left(\alpha^2 - \alpha_0^2 \right) x_{2t-1} \right] \right| \\
\leq & \left| \alpha^1 - \alpha_0^1 \right| \sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^2 x_{2t-1} \right| + \left| \alpha^2 - \alpha_0^2 \right| \sum_{t=1}^{n} \left| f(x_{1t-1}) x_{1t-1} \left(x_{2t-1}\right)^2 \right| \\
\leq & n^{-1/2+\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^2 x_{2t-1} \right| + n^{-1+\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) x_{1t-1} \left(x_{2t-1}\right)^2 \right| \\
= & O_P(1) n^{\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) \left(x_{1t-1}\right)^2 \right| + O_P(1) n^{\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) x_{1t-1} \right| \\
= & O_P(n^{1+\delta}) E \left| f(x_{11}) \left(x_{11}\right)^2 \right|.
\end{aligned}
$$

For $i = 2$, $j = 2$, notice that

$$
\left| T^{2,2} \right| = \left| \sum_{t=1}^{n} f(x_{1t-1}) \left(x_{2t-1}\right)^2 \left[\left(\alpha^1 - \alpha_0^1 \right) x_{1t-1} + \left(\alpha^2 - \alpha_0^2 \right) x_{2t-1} \right] \right|
$$

$$\leq \left|\alpha^1 - \alpha_0^1\right| \sum_{t=1}^{n} \left| f(x_{1t-1}) x_{1t-1} \left(x_{2t-1}\right)^2 \right| + \left|\alpha^2 - \alpha_0^2\right| \left| \sum_{t=1}^{n} f(x_{1t-1}) \left(x_{2t-1}\right)^3 \right|$$

$$\leq n^{-1/2+\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) x_{1t-1} \left(x_{2t-1}\right)^2 \right| + n^{-1+\delta} \left| \sum_{t=1}^{n} f(x_{1t-1}) \left(x_{2t-1}\right)^3 \right|$$

$$= O_P(1) n^{1/2+\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) x_{1t-1} \right| + O_P(1) n^{1/2+\delta} \sum_{t=1}^{n} \left| f(x_{1t-1}) \right|$$

$$= O_P(n^{3/2+\delta}) E \left| f(x_{11}) x_{11} \right| \qquad \square$$

Proof of Lemma 8:

By definition of (2.10), we have

$$0 = \frac{\partial}{\partial \theta} W_{n,\widehat{\lambda}}(\widehat{\theta}_n, \bar{C}_k), \qquad 0 = \frac{\partial}{\partial \lambda} W_{n,\widehat{\lambda}}(\widehat{\theta}_n, \bar{C}_k).$$

The condition $0 = \frac{\partial}{\partial \lambda} W_{n,\widehat{\lambda}}(\widehat{\theta}_n, \bar{C}_k)$ gives that $\|\widehat{\theta}_n\|^2 - 1 = 0$, which satisfies the identification condition for the single-index model. Given θ such that $\theta^\top \theta \neq 0$, multiplying $\theta^\top$ on both sides of $0 = \frac{\partial}{\partial \theta} W_{n,\widehat{\lambda}}(\theta, \bar{C}_k(\theta))$ gives

$$\widehat{\lambda}(\theta) = (\theta^\top \theta)^{-1} \theta^\top \sum_{t=1}^{n} \left(y_t - \widehat{g}_n(\theta^\top x_{t-1}) \right) \frac{\partial \widehat{g}_n(\theta^\top x_{t-1})}{\partial \theta}. \tag{D.3}$$

For $\theta = \theta_0$, we have $\theta_0^\top \theta_0 = 1$ and

$$\widehat{\lambda}(\theta_0) = \theta_0^\top \sum_{t=1}^{n} \left(y_t - \widehat{g}_n(\theta_0^\top x_{t-1}) \right) \left. \frac{\partial \widehat{g}_n(\theta^\top x_{t-1})}{\partial \theta} \right|_{\theta = \theta_0}.$$

Denote $D_n = diag(n^{1/2}, n)$, $\xi_{n,t-1} = D_n^{-1} Q^\top x_{t-1} = (\frac{1}{\sqrt{n}} x_{1t-1}, \frac{1}{n} x_{2t-1})^\top$, and $\mathcal{H}_{n,x} = \sum_{t=1}^{n} \mathcal{H}_k(\theta_0^\top x_{t-1}) \mathcal{H}_k(\theta_0^\top x_{t-1})^\top$ for brevity. It follows that

(a). The score:

$$D_n^{-1} \frac{\partial}{\partial \alpha} W_{n,\widehat{\lambda}(\alpha_0))} \Big|_{(\alpha, C_k) = (\alpha_0, \bar{C}_k(\alpha_0))} = D_n^{-1} Q^\top \frac{\partial}{\partial \theta} W_{n,\widehat{\lambda}(\theta_0)} \Big|_{(\theta, C_k) = (\theta_0, \bar{C}_k(\theta_0))}$$

$$= -2 D_n^{-1} Q^\top \sum_{t=1}^{n} \left(y_t - \widehat{g}_n(\theta_0^\top x_{t-1}) \right) \left. \frac{\partial \widehat{g}_n(\theta^\top x_{t-1})}{\partial \theta} \right|_{\theta = \theta_0} + 2 D_n^{-1} Q^\top \theta_0 \widehat{\lambda}(\theta_0)$$

$$= 2(-I + \alpha_0 \alpha_0^\top) \sum_{t=1}^{n} \left(y_t - \widehat{g}_n(x_{1t-1}) \right) \left(\widehat{g}_n^{(1)}(x_{1t-1}) \xi_{n,t-1} - \mathcal{H}_k(x_{1t-1})^\top \mathcal{H}_{n,x}^{-1} \sum_{s=1}^{n} \mathcal{H}_k(x_{1s-1}) \widehat{g}_n^{(1)}(x_{1s-1}) \xi_{n,s-1} \right)$$

$$+ 2(-I + \alpha_0 \alpha_0^\top) \sum_{t=1}^{n} \left(y_t - \widehat{g}_n(x_{1t-1}) \right) \left(\mathcal{H}_k(x_{1t-1})^\top \mathcal{H}_{n,x}^{-1} \sum_{s=1}^{n} \mathcal{H}_k^{(1)}(x_{1s-1}) \left(y_s - \widehat{g}_n(x_{1s-1}) \right) \xi_{n,s-1} \right)$$

$$= 2(-I + \alpha_0 \alpha_0^\top) \sum_{t=1}^{n} e_t \left(\widehat{g}_n^{(1)}(x_{1t-1}) \xi_{n,t-1} - \mathcal{H}_k(x_{1t-1})^\top \mathcal{H}_{n,x}^{-1} \sum_{s=1}^{n} \mathcal{H}_k(x_{1s-1}) \widehat{g}_n^{(1)}(x_{1s-1}) \xi_{n,s-1} \right)$$

$$- 2(-I + \alpha_0 \alpha_0^\top) \sum_{t=1}^{n} \left(\widehat{g}_n(x_{1t-1}) - g_0(x_{1t-1}) \right)$$

$$\times \left(\widehat{g}_n^{(1)}(x_{1t-1}) \xi_{n,t-1} - \mathcal{H}_k(x_{1t-1})^\top \mathcal{H}_{n,x}^{-1} \sum_{s=1}^{n} \mathcal{H}_k(x_{1s-1}) \widehat{g}_n^{(1)}(x_{1s-1}) \xi_{n,s-1} \right)$$

$$
\begin{aligned}
&+2(-I+\alpha_0\alpha_0^\top)\sum_{t=1}^{n} e_t \mathcal{H}_k(x_{1t-1})^\top \mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k^{(1)}(x_{1s-1})\left(y_s-\widehat{g}_n(x_{1s-1})\right)\xi_{n,s-1}\\
&-2(-I+\alpha_0\alpha_0^\top)\sum_{t=1}^{n}\left(\widehat{g}_n(x_{1t-1})-g_0(x_{1t-1})\right)\mathcal{H}_k(x_{1t-1})^\top \mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k^{(1)}(x_{1s-1})\left(y_s-\widehat{g}_n(x_{1s-1})\right)\xi_{n,s-1}\\
=&2(-I+\alpha_0\alpha_0^\top)\left(S_1-S_2+S_3-S_4\right),
\end{aligned}
$$

It follows from Lemma 4 and the proofs in the online Appendix E that

$$
\begin{aligned}
S_1=&\sum_{t=1}^{n} e_t\left(\widehat{g}_n^{(1)}(x_{1t-1})\xi_{n,t-1}-\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k(x_{1s-1})\widehat{g}_n^{(1)}(x_{1s-1})\xi_{n,s-1}\right)\\
=&\begin{pmatrix}0\\ \frac{1}{n}\sum_{t=1}^{n} g_0^{(1)}(x_{1t-1})\left(x_{2t-1}-\frac{1}{n}\sum_{s=1}^{n}x_{2s-1}\right)e_t\end{pmatrix}+o_P(1)\to_D\begin{pmatrix}0\\ \sigma r_0^{1/2}W(1)\end{pmatrix},
\end{aligned}
$$

and therefore,

$$
(-I+\alpha_0\alpha_0^\top)S_1\to_D\begin{pmatrix}0\\ -\sigma r_0^{1/2}W(1)\end{pmatrix}.
$$

Denote $P_{\alpha_0}=I_d-\alpha_0\alpha_0^\top$ and it has eigenvalues $0,1,...,1$, where 0 corresponds to the eigenvector α_0. Thus, to make sure the asymptotic covariance matrix non-singular, we need to rotate the Score function. Let $P_1=(p_1,...,p_{d-1})$, where $p_1,...,p_{d-1}$ are the eigenvectors associated with the eigenvalues 1 of P_{α_0} and they are orthogonal to α_0. Therefore, we have $P_{\alpha_0}=P_1P_1^\top$ and $P_1^\top P_1=I_{d-1}$. In addition, the detailed proofs of S_2, S_3, and S_4 to be $o_P(1)$ are given in the online Appendix E. Then, we can obtain that

$$
P_1^\top D_n^{-1}S_n(\alpha_0)\to_D -2\sigma r_0^{1/2}W(1),
$$

for $\kappa\in[1/r,1/4)$.

(b). The hessian:

$$
\begin{aligned}
&D_n^{-1}\left.\frac{\partial^2}{\partial\alpha\partial\alpha^\top}W_{n,\widehat{\lambda}(\alpha_0)}D_n^{-1}\right|_{(\alpha,C_k)=(\alpha_0,\bar{C}_k(\alpha_0))}=D_n^{-1}Q^\top\left.\frac{\partial^2}{\partial\theta\partial\theta^\top}W_{n,\widehat{\lambda}(\theta_0)}QD_n^{-1}\right|_{(\theta,C_k)=(\theta_0,\bar{C}_k(\theta_0))}\\
=&2D_n^{-1}Q^\top\sum_{t=1}^{n}\left.\frac{\partial\widehat{g}_n(\theta^\top x_{t-1})}{\partial\theta}\right|_{\theta=\theta_0}\left.\frac{\partial\widehat{g}_n(\theta^\top x_{t-1})}{\partial\theta^\top}\right|_{\theta=\theta_0}QD_n^{-1}\\
&-2D_n^{-1}Q^\top\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(\theta_0^\top x_{t-1})\right)\left.\frac{\partial^2\widehat{g}_n(\theta^\top x_{t-1})}{\partial\theta\partial\theta^\top}\right|_{\theta=\theta_0}QD_n^{-1}+2D_n^{-1}Q^\top\widehat{\lambda}(\theta_0)QD_n^{-1}\\
=&2\sum_{t=1}^{n}\left(\widehat{g}_n^{(1)}(x_{1t-1})\xi_{n,t-1}-\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k(x_{1s-1})\widehat{g}_n^{(1)}(x_{1s-1})\xi_{n,s-1}\right)\\
&\times\left(\widehat{g}_n^{(1)}(x_{1t-1})\xi_{n,t-1}^\top-\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k(x_{1s-1})\widehat{g}_n^{(1)}(x_{1s-1})\xi_{n,s-1}^\top\right)
\end{aligned}
$$

$$
\begin{aligned}
&+4\sum_{t=1}^{n}\left(\widehat{g}_n^{(1)}(x_{1t-1})\xi_{n,t-1}-\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k(x_{1s-1})\widehat{g}_n^{(1)}(x_{1s-1})\xi_{n,s-1}\right)\\
&\times\left(\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k^{(1)}(x_{1s-1})\left(y_s-\widehat{g}_n(x_{1s-1})\right)\xi_{n,s-1}^\top\right)\\
&+2\sum_{t=1}^{n}\left(\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k^{(1)}(x_{1s-1})\left(y_s-\widehat{g}_n(x_{1s-1})\right)\xi_{n,s-1}\right)\\
&\times\left(\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k^{(1)}(x_{1s-1})\left(y_s-\widehat{g}_n(x_{1s-1})\right)\xi_{n,s-1}^\top\right)\\
&-2\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(x_{1t-1})\right)\widehat{g}_n^{(2)}(x_{1t-1})\xi_{n,t-1}\xi_{n,t-1}^\top\\
&-2\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(x_{1t-1})\right)\mathcal{H}_k^{(1)}(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k^{(1)}(x_{1s-1})\left(y_s-\widehat{g}_n(x_{1s-1})\right)\left(\xi_{n,t-1}\xi_{n,s-1}^\top+\xi_{n,s-1}\xi_{n,t-1}^\top\right)\\
&+2\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(x_{1t-1})\right)\mathcal{H}_k^{(1)}(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k(x_{1s-1})\widehat{g}_n^{(1)}(x_{1s-1})\left(\xi_{n,t-1}\xi_{n,s-1}^\top+\xi_{n,s-1}\xi_{n,t-1}^\top\right)\\
&-2\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(x_{1t-1})\right)\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k^{(2)}(x_{1s-1})\left(y_s-\widehat{g}_n(x_{1s-1})\right)\xi_{n,s-1}\xi_{n,s-1}^\top\\
&+2\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(x_{1t-1})\right)\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{l=1}^{n}\mathcal{H}_k^{(1)}(x_{1l-1})\mathcal{H}_k(x_{1l-1})^\top\mathcal{H}_{n,x}^{-1}\\
&\times\sum_{s=1}^{n}\mathcal{H}_k^{(1)}(x_{1s-1})\left(y_s-\widehat{g}_n(x_{1s-1})\right)\left(\xi_{n,s-1}\xi_{n,l-1}^\top+\xi_{n,l-1}\xi_{n,s-1}^\top\right)\\
&+2\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(x_{1t-1})\right)\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{l=1}^{n}\mathcal{H}_k(x_{1l-1})\mathcal{H}_k^{(1)}(x_{1l-1})^\top\mathcal{H}_{n,x}^{-1}\\
&\times\sum_{s=1}^{n}\mathcal{H}_k^{(1)}(x_{1s-1})\left(y_s-\widehat{g}_n(x_{1s-1})\right)\left(\xi_{n,s-1}\xi_{n,l-1}^\top+\xi_{n,l-1}\xi_{n,s-1}^\top\right)\\
&-2\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(x_{1t-1})\right)\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{l=1}^{n}\mathcal{H}_k^{(1)}(x_{1l-1})\mathcal{H}_k(x_{1l-1})^\top\mathcal{H}_{n,x}^{-1}\\
&\times\sum_{s=1}^{n}\mathcal{H}_k(x_{1s-1})\widehat{g}_n^{(1)}(x_{1s-1})\left(\xi_{n,s-1}\xi_{n,l-1}^\top+\xi_{n,l-1}\xi_{n,s-1}^\top\right)\\
&-2\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(x_{1t-1})\right)\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{l=1}^{n}\mathcal{H}_k(x_{1l-1})\mathcal{H}_k^{(1)}(x_{1l-1})^\top\mathcal{H}_{n,x}^{-1}\\
&\times\sum_{s=1}^{n}\mathcal{H}_k(x_{1s-1})\widehat{g}_n^{(1)}(x_{1s-1})\left(\xi_{n,s-1}\xi_{n,l-1}^\top+\xi_{n,l-1}\xi_{n,s-1}^\top\right)\\
&+4\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(x_{1t-1})\right)\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k^{(1)}(x_{1s-1})\widehat{g}_n^{(1)}(x_{1s-1})\xi_{n,s-1}\xi_{n,s-1}^\top\\
&+2\sum_{t=1}^{n}\left(y_t-\widehat{g}_n(x_{1t-1})\right)\mathcal{H}_k(x_{1t-1})^\top\mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k(x_{1s-1})\widehat{g}_n^{(2)}(x_{1s-1})\xi_{n,s-1}\xi_{n,s-1}^\top\\
&+2D_n^{-1}Q^\top\widehat{\lambda}(\theta_0)QD_n^{-1}\\
:=&2J_1+\cdots+2J_{14}
\end{aligned}
$$

It follows from Lemma 3 and the proofs in the online Appendix E that

$$
\begin{aligned}
J_1 &= \sum_{t=1}^{n}\left(\widehat{g}_n^{(1)}(x_{1t-1})\xi_{n,t-1} - \mathcal{H}_k(x_{1t-1})^\top \mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k(x_{1s-1})\widehat{g}_n^{(1)}(x_{1s-1})\xi_{n,s-1}\right) \\
&\times\left(\widehat{g}_n^{(1)}(x_{1t-1})\xi_{n,t-1}^\top - \mathcal{H}_k(x_{1t-1})^\top \mathcal{H}_{n,x}^{-1}\sum_{s=1}^{n}\mathcal{H}_k(x_{1s-1})\widehat{g}_n^{(1)}(x_{1s-1})\xi_{n,s-1}^\top\right) \\
&= \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{n}\sum_{t=1}^{n}\left[g_0^{(1)}(x_{1t-1})\left(x_{2t-1} - \frac{1}{n}\sum_{s=1}^{n}x_{2s-1}\right)\right]^2 \end{pmatrix} + o_P(1) \to_D \begin{pmatrix} 0 & 0 \\ 0 & r_0 \end{pmatrix}.
\end{aligned}
$$

The detailed proofs of all the other terms to be $o_P(1)$ are given in the online Appendix E. Then, we can obatin that $P_1^\top D_n^{-1} J_n(\alpha_0) D_n^{-1} P_1 \to_D 2r_0$ for $\kappa \in [1/r, 1/4)$. $\square$

Proof of Lemma 9:

We use Theorem 10.1 of Wooldridge (1994) to show the asymptotic normality in this paper. The first condition of this theorem is satisfied according to the assumption that $\theta_0 \in \text{int}(\Theta)$, and hence, $\alpha_0 \in \text{int}(\Phi)$. The second condition is achieved by the Assumption 1.4 on the smoothness of $g_0(.)$ function. To verify the third condition, rewrite (3.3) as

$$
S_n(\widehat{\alpha}_0) + J_n(\alpha_0)\left(\widehat{\alpha}_n - \alpha_0\right) + \left[J_n(\alpha_n) - J_n(\alpha_0)\right]\left(\widehat{\alpha}_n - \alpha_0\right) = 0
$$

Define $C_n = n^{-\delta} D_n$ for some $\delta > 0$ such that $C_n D_n^{-1} = o(1)$ as $n \to \infty$. Then we have

$$
\begin{aligned}
0 &= D_n^{-1}S_n(\alpha_0) + D_n^{-1}J_n(\alpha_0)D_n^{-1}D_n\left(\widehat{\alpha}_n - \alpha_0\right) + D_n^{-1}\left[J_n(\alpha_n) - J_n(\alpha_0)\right]D_n^{-1}D_n\left(\widehat{\alpha}_n - \alpha_0\right) \\
&= D_n^{-1}S_n(\alpha_0) + D_n^{-1}J_n(\alpha_0)D_n^{-1}D_n\left(\widehat{\alpha}_n - \alpha_0\right) + n^{-2\delta}C_n^{-1}\left[J_n(\alpha_n) - J_n(\alpha_0)\right]C_n^{-1}D_n\left(\widehat{\alpha}_n - \alpha_0\right).
\end{aligned}
$$

The condition (iii) of Theorem 10.1 in Wooldridge (1994) will be satisfied if we can show

$$
\sup_{\{\alpha:\|C_n(\alpha-\alpha_0)\|\le 1\}}\left\|C_n^{-1}\left[J_n(\alpha) - J_n(\alpha_0)\right]C_n^{-1}\right\| = o_P(1)
$$

According to the previous calculation, the hessian matrix with $\alpha = (\alpha^1, \alpha^2)^\top$ is given by

$$
J_n(\alpha) = 2\begin{pmatrix} J_n^{1,1}(\alpha) & J_n^{1,2}(\alpha) \\ J_n^{2,1}(\alpha) & J_n^{2,2}(\alpha) \end{pmatrix} + 2J_{n,\lambda}(\alpha),
$$

where

$$
\begin{aligned}
\begin{pmatrix} J_n^{1,1}(\alpha) & J_n^{1,2}(\alpha) \\ J_n^{2,1}(\alpha) & J_n^{2,2}(\alpha) \end{pmatrix} &= Q^\top\sum_{t=1}^{n}\frac{\partial\widehat{g}_n(\theta^\top x_{t-1})}{\partial\theta}\frac{\partial\widehat{g}_n(\theta^\top x_{t-1})}{\partial\theta^\top}Q - Q^\top\sum_{t=1}^{n}\left(y_t - \widehat{g}_n(\theta^\top x_{t-1})\right)\frac{\partial^2\widehat{g}_n(\theta^\top x_{t-1})}{\partial\theta\partial\theta^\top}Q \\
J_{n,\lambda}(\alpha) &= Q^\top\widehat{\lambda}(\theta)Q
\end{aligned}
$$

Then we need to show that

$$
\begin{aligned}
n^{-1+2\delta}\left|J_n^{1,1}(\alpha)-J_n^{1,1}(\alpha_0)\right| &= o_P(1)\\
n^{-3/2+2\delta}\left|J_n^{1,2}(\alpha)-J_n^{1,2}(\alpha_0)\right| &= o_P(1)\\
n^{-2+2\delta}\left|J_n^{2,2}(\alpha)-J_n^{2,2}(\alpha_0)\right| &= o_P(1)\\
n^{-1+2\delta}\left|J_{n,\lambda}(\alpha)-J_{n,\lambda}(\alpha_0)\right| &= o_P(1)
\end{aligned}
\tag{D.4}
$$

uniformly in α^1 and α^2 satisfying

$$
\left|\alpha^1-\alpha_0^1\right| < n^{-1/2+\delta} \quad \text{and} \quad \left|\alpha^2-\alpha_0^2\right| < n^{-1+\delta}
$$

for some $\delta > 0$, $\alpha^\top\alpha \neq 0$, $\alpha_0^1 = 1$, and $\alpha_0^2 = 0$.

Then we have shown in the online Appendix E that

$$
\left|J_n^{1,1}(\alpha)-J_n^{1,1}(\alpha_0)\right| = O_P(\max(n^{1/2+\delta+5\kappa/2}, n^{-1/2+\delta+7\kappa})), \quad \left|J_n^{1,2}(\alpha)-J_n^{1,2}(\alpha_0)\right| = O_P(n^{1+\delta+5\kappa/2}),
$$
$$
\left|J_n^{2,2}(\alpha)-J_n^{2,2}(\alpha_0)\right| = O_P(n^{3/2+\delta+2\kappa}), \quad \left|J_{n,\lambda}(\alpha)-J_{n,\lambda}(\alpha_0)\right| = O_P(n^{1/2+\delta}k^2).
$$

To fulfill (D.4), we may choose $\delta : 0 < \delta < \min(1/6 - 5\kappa/6, \kappa/2)$ with $1/r \leq \kappa < 1/5$ stipulated in Assumption 1.5.

Now, we have proved that $D_n^{-1}\left[J_n(\alpha) - J_n(\alpha_0)\right] D_n^{-1} = o_P(1)$ uniformly in α^1 and α^2 satisfying

$$
\left|\alpha^1-\alpha_0^1\right| < n^{-1/2+\delta} \quad \text{and} \quad \left|\alpha^2-\alpha_0^2\right| < n^{-1+\delta}.
$$

Using the same argument in the proof of Theorem 8.1 in Wooldridge (1994), we can show that $D_n(\widehat{\alpha}_n - \alpha_0) = O_P(1)$. Then, we can write

$$
\begin{aligned}
0 &= S_n(\alpha_0) + J_n(\alpha_0)(\widehat{\alpha}_n - \alpha_0) + \left(J_n(\alpha_n) - J_n(\alpha_0)\right)(\widehat{\alpha}_n - \alpha_0)\\
0 &= D_n^{-1}S_n(\alpha_0) + D_n^{-1}J_n(\alpha_0)D_n^{-1}D_n(\widehat{\alpha}_n - \alpha_0) + n^{-2\delta}C_n^{-1}\left[J_n(\alpha_n) - J_n(\alpha_0)\right]C_n^{-1}D_n\left(\widehat{\alpha}_n - \alpha_0\right)\\
0 &= P_1^\top D_n^{-1}S_n(\alpha_0) + P_1^\top D_n^{-1}J_n(\alpha_0)D_n^{-1}\left(P_1P_1^\top + \alpha_0\alpha_0^\top\right)D_n(\widehat{\alpha}_n - \alpha_0) + o_P(1)\\
0 &= P_1^\top D_n^{-1}S_n(\alpha_0) + P_1^\top D_n^{-1}J_n(\alpha_0)D_n^{-1}P_1P_1^\top D_n(\widehat{\alpha}_n - \alpha_0) + P_1^\top D_n^{-1}J_n(\alpha_0)D_n^{-1}\alpha_0\alpha_0^\top D_n(\widehat{\alpha}_n - \alpha_0) + o_P(1)\\
0 &= P_1^\top D_n^{-1}S_n(\alpha_0) + P_1^\top D_n^{-1}J_n(\alpha_0)D_n^{-1}P_1P_1^\top D_n(\widehat{\alpha}_n - \alpha_0) + o_P(1),
\end{aligned}
\tag{D.5}
$$

Then we can immediately obtain from (D.5) that

$$
P_1^\top D_n(\widehat{\alpha}_n - \alpha_0) = -\left(P_1^\top D_n^{-1}J_n(\alpha_0)D_n^{-1}P_1\right)^{-1}P_1^\top D_n^{-1}S_n(\alpha_0) + o_P(1). \tag{D.6}
$$

In Lemma 8, we have already shown that

$$
P_1^\top D_n^{-1}S_n(\alpha_0) \to_D -2\sigma r_0^{1/2}W(1) \quad \textit{and} \quad P_1^\top D_n^{-1}J_n(\alpha_0)D_n^{-1}P_1 \to_D 2r_0,
$$

where r_0 is given in Lemma 8 and it is positive definite with probability one, which indicates the condition (iv) of Theorem 10.1 of Wooldridge (1994) holds. Then the limit distribution follows from (D.6) and Lemma 8 that

$$P_1^\top D_n(\widehat{\alpha}_n - \alpha_0) \to_D \sigma {r_0}^{-1/2} W(1). \qquad \square$$

References

Amihud, Y., Hurvich, C. M., and Wang, Y. (2008). Multiple-predictor regressions: Hypothesis testing. *Review of Financial Studies*, 22(1):413–434.

Ang, A. and Bekaert, G. (2007). Stock return predictability: Is it there? *Review of Financial Studies*, 20(3):651–707.

Beveridge, S. and Nelson, C. R. (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle '. *Journal of Monetary Economics*, 7(2):151–174.

Cai, B., Dong, C., and Gao, J. (2015). Orthogonal series estimation in nonlinear cointegrating models with endogeneity. Working paper available at https://www.monash.edu/__data/assets/pdf_file/0010/925876/wp18-15.pdf.

Cai, B. and Gao, J. (2013). Hermite series estimation in nonlinear cointegrating models. Working paper available at https://ssrn.com/abstract=2305873.

Cai, Z. (2007). Trending time-varying coefficient time series models with serially correlated errors. *Journal of Econometrics*, 136(1):163–188.

Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81(1):27–60.

Chen, Q. and Hong, Y. (2009). Predictability of equity returns over different time horizons: A nonparametric approach. Working Paper available at https://economics.ucr.edu/seminars_colloquia/2009/econometrics/Chen_Hong_UCR_Seminar_2009.pdf.

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics*, volume 6B, pages 5549–5632. New York: Elsevier.

Chen, X. and Christensen, T. M. (2015). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics*, 188(2):447–465.

Chen, X., Linton, O., and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71(5):1591–1608.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, 66(4):1047–1108.

Dong, C. and Gao, J. (2014). Specification testing in structural nonparametric cointegration. Working paper available at https://www.monash.edu/business/econometrics-and-business-statistics/research/publications/ebs/wp02-14.pdf.

Dong, C., Gao, J., and Peng, B. (2015). Semiparametric single-index panel data models with cross-sectional dependence. *Journal of Econometrics*, 188(1):301–312.

Dong, C., Gao, J., and Tjøstheim, D. (2016). Estimation for single-index and partially linear single-index integrated models. *Annals of Statistics*, 44(1):425–453.

Dong, C., Gao, J., Tjøstheim, D., and Yin, J. (2017). Specification testing for nonlinear multivariate cointegrating regressions. *Journal of Econometrics*, 200(1):104–117.

Engle, R. F. and Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55(2):251–276.

Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25.

Fama, E. F. and French, K. R. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25(1):23–49.

Gao, J., Tong, H., and Wolff, R. (2002). Model specification tests in nonparametric stochastic regression models. *Journal of Multivariate Analysis*, 83(2):324–359.

Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios. *Management Science*, 49(5):639–654.

Granger, C. W. and Weiss, A. A. (1983). Time series analysis of error-correction models. In Karlin, S., Amemiya, T., and Goodman, L. A., editors, *Studies in Econometrics, Time series, and Multivariate Statistics*, pages 255–278. New York: Academic Press.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press.

Hansen, B. E. (2015). A unified asymptotic distribution theory for parametric and non-parametric least squares. Working paper available at https://www.ssc.wisc.edu/~bhansen/preliminary/rnormal4.pdf.

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association*, 84(408):986–995.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *Journal of Econometrics*, 58(1):71–120.

Kasparis, I., Andreou, E., and Phillips, P. C. (2015). Nonparametric predictive regression. *Journal of Econometrics*, 185(2):468–494.

Koo, B., Anderson, H. M., Seo, M. H., and Yao, W. (2016). High-dimensional predictive regression in the presence of cointegration. Working paper available at https://ssrn.com/abstract=2851677.

Kostakis, A., Magdalinos, T., and Stamatogiannis, M. P. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies*, 28(5):1506–1553.

Lamont, O. (1998). Earnings and expected returns. *Journal of Finance*, 53(5):1563–1587.

Lettau, M. and Ludvigson, S. (2001). Consumption, aggregate wealth, and expected stock returns. *Journal of Finance*, 56(3):815–849.

Magnus, J. R. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons.

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79(1):147–168.

Park, J. Y. and Phillips, P. C. (2000). Nonstationary binary choice. *Econometrica*, 68(5):1249–1280.

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

Phillips, P. C. (2015). Halbert White Jr. Memorial JFEC Lecture: Pitfalls and possibilities in predictive regression. *Journal of Financial Econometrics*, 13(3):521–555.

Phillips, P. C. and Magdalinos, T. (2009). Econometric inference in the vicinity of unity. Working paper available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.555.8852&rep=rep1&type=pdf.

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

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57(6):1403–1430.

Rapach, D. and Zhou, G. (2013). Forecasting Stock Returns. In Elliott, G. and Timmermann, A., editors, *Handbook of Economic Forecasting*, volume 2, pages 328–383. Amsterdam: Elsevier.

Shiller, R. J., Fischer, S., and Friedman, B. M. (1984). Stock prices and social dynamics. *Brookings Papers on Economic Activity*, 1984(2):457–498.

Wang, J.-L., Xue, L., Zhu, L., and Chong, Y. S. (2010). Estimation for a partial-linear single-index model. *Annals of statistics*, 38(1):246–274.

Wang, Q. and Phillips, P. C. (2009). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25(3):710–738.

Welch, I. and Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21(4):1455–1508.

Withers, C. (1981). Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4):477–480.

Wooldridge, J. M. (1994). Estimation and Inference for Dependent Processes. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, pages 2639–2738. Amsterdam: Elsevier.

Wu, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Annals of Statistics*, 9(3):501–513.

Xia, Y. (2006). Asymptotic distributions for two estimators of the single-index model. *Econometric Theory*, 22(6):1112–1137.

Xu, K.-L. (2017). Testing for return predictability with co-moving predictors of unknown form. Working paper available at http://www.cireqmontreal.com/wp-content/uploads/2017/05/xu.pdf.

Xu, K.-L. and Guo, J. (2019). A dimensionality-robust test in multiple predictive regression. Working paper available at https://drive.google.com/file/d/1PNSkUNbsiYPAMIZ55gzxB7wF9lFdLe9Y/view.

Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, 97(460):1042–1054.

Zhou, W., Gao, J., Harris, D., and Kew, H. (2019). Online supplemental material for "semiparametric single-index predictive regression". Supplemental material available at personal website https://sites.google.com/site/weilunzhou514/.