



**MONASH** University

**Australia**

Department of Econometrics and Business Statistics

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

**Estimation for Single-index and Partially Linear  
Single-index Nonstationary Time Series Models**

**Chaohua Dong, Jiti Gao and Dag Tjøstheim**

**February 2014**

**Working Paper 07/14**

# Estimation for Single-index and Partially Linear Single-index Nonstationary Time Series Models \*

CHAOHUA DONG<sup>†</sup>, JITI GAO<sup>†</sup> AND DAG TJØSTHEIM

*Monash University<sup>†</sup> and University of Bergen*

February 19, 2014

## Abstract

Estimation in two classes of popular models, single-index models and partially linear single-index models, is studied in this paper. Such models feature nonstationarity. Orthogonal series expansion is used to approximate the unknown integrable link function in the models and a profile approach is used to derive the estimators. The findings include dual convergence rates of the estimators for the single-index models and a trio of convergence rates for the partially linear single-index models. More precisely, the estimators for single-index model converge along the direction of the true parameter vector at rate of  $n^{-1/4}$ , while at rate of  $n^{-3/4}$  along all directions orthogonal to the true parameter vector; on the other hand, the estimators of the index vector for the partially single-index model retain the dual convergence rates as in the single-index model but the estimators of the coefficients in the linear part of the model possess rate  $n^{-1}$ . Monte Carlo simulation verifies these theoretical results. An empirical study on the dataset of aggregate disposable income, consumption, investment and real interest rate in the United States between 1960:1-2009:3 furnishes an application of the proposed estimation procedures in practice.

**Keywords:** Nonstationarity, orthogonal series expansion, single-index models, partially linear single-index models, dual convergence rates, a trio of convergence rates.

*JEL classification:* C13; C14; C32

---

\*The first two authors thank the Australian Research Council Discovery Grants Program for its support under Grant numbers: DP1096374 and DP130104229. Corresponding author: Dag Tjøstheim, Department of Mathematics, University of Bergen, 5020 Bergen, Norway. Email: dag.tjostheim@math.uib.no

# 1 Introduction

In the last decade or so, nonlinear (nonparametric or semiparametric) and nonstationary time series models have been studied extensively and improved dramatically as witnessed by the literature, such as those based on the nonparametric kernel approach by Karlsen and Tjøstheim (2001); Karlsen et al. (2007), Gao et al. (2009a,b), Phillips (2009), Wang and Phillips (2009a,b), Gao (2012), Wang and Phillips (2012), Gao and Phillips (2013a,b) and Phillips et al. (2013), among others. The main development in the field is the establishment of new estimation and specification testing procedures as well as the resulting asymptotic properties. In recent years, the conventional nonparametric kernel-based estimation and specification testing theory has been extended to the nonparametric series based approach, such as, Dong and Gao (2012, 2013).

This paper considers the estimation of partially linear single-index models of the form

$$y_t = \beta_0^\top x_t + g(\theta_0^\top x_t) + e_t, \quad t = 1, \dots, n, \quad (1.1)$$

where  $y_t$  is a scalar process,  $g(\cdot)$ , the so-called link function, is an unknown nonlinear function from  $\mathbb{R}$  to  $\mathbb{R}$ ,  $\beta_0$  and  $\theta_0$  are the true but unknown  $d$ -dimensional column vectors of parameters, the superscript  $\top$  signifies the transpose of a vector ( or matrix, hereafter),  $x_t$  is a  $d$ -dimensional nonstationary process,  $e_t$  is an error process and  $n$  is sample size.

The motivations of this study are as follows. In a fully nonparametric estimation context, researchers often suffer from the so-called 'curse of dimensionality', and hence dimensionality reduction is particularly of importance in such a situation. One efficient way of doing so is to use index models like model (1.1). Moreover, model (1.1) is also an extension of linear parametric models, since it would become a linear model under the particular choice of the link function. Taking the curse of dimensionality into account, models formulated by (1.1) are often used as a reasonable compromise between fully parametric and fully nonparametric modelling. See, for example, Robinson (1988), Härdle et al. (1993), Carroll et al. (1997), Xia et al. (1999, 2002), Yu and Ruppert (2002), Zhu and Xue (2006), Liang et al. (2010), Wang et al. (2010) and Ma and Zhu (2013). Nevertheless, most researchers only focus on the stationary covariate case so that their theoretical results are not applicable for practitioners who use partially linear single-index model to deal with nonstationary time series data. For example, in macroeconomic context practitioners may be concerned with inflation, unemployment rates, and other economic indicators. Many of these variables exhibit nonstationary characteristics. Therefore, it is desirable in such circumstances to develop estimation theory for the partially linear single-index models.

Furthermore, recent studies by Gao and Phillips (2013a,b) have pointed out, for multivariate  $I(1)$  processes the conventional kernel regression estimation method may not be

workable because the limit theory may break down in such circumstances. This gives rise to a challenge of seeking alternative estimation methods. In addition, model (1.1) is so flexible that it accommodates many models as special cases, such as,

1. (Single-index model) When  $\beta_0 = 0$ , model (1.1) becomes

$$y_t = g(\theta_0^\top x_t) + e_t, \quad t = 1, \dots, n. \quad (1.2)$$

2. (Generalized linear model) When  $\beta_0 = 0$  and  $g(x) = G(x)$  is known, model (1.1) becomes

$$y_t = G(\theta_0^\top x_t) + e_t, \quad t = 1, \dots, n. \quad (1.3)$$

3. (Linear parametric model) When  $g \equiv 0$ , model (1.1) becomes a linear parametric model.

$$y_t = \beta_0^\top x_t + e_t, \quad t = 1, \dots, n. \quad (1.4)$$

4. (Linear parametric model with constant) When  $\theta_0 = 0$ , model (1.1) becomes a linear parametric model with constant (whose true value is  $g(0)$ ).

$$y_t = \alpha_0 + \beta_0^\top x_t + e_t, \quad t = 1, \dots, n. \quad (1.5)$$

5. (Semilinear model) When  $\theta_0 = (0, \dots, 0, 1)^\top$  and  $\beta_0^\top = (\beta_{10}^\top, 0)$ , model (1.1) becomes

$$y_t = \beta_{10}^\top z_{1t} + g(z_{2t}) + e_t, \quad t = 1, \dots, n, \quad (1.6)$$

where  $z_{1t} = (x_{1,t}, \dots, x_{d-1,t})^\top$  and  $z_{2t} = x_{d,t}$ .

6. (Separable partially linear single-index model) Let  $\beta_0 = (\beta_{10}^\top, 0, \dots, 0)^\top$ ,  $\theta_0 = (0, \dots, 0, \theta_{10}^\top)^\top$ ,  $x_t = (z_{1t}^\top, z_{2t}^\top)^\top$  partitioned conformably with  $\theta_0$  and  $\beta_0$ . Model (1.1) becomes

$$y_t = \beta_{10}^\top z_{1t} + g(\theta_{10}^\top z_{2t}) + e_t, \quad t = 1, \dots, n. \quad (1.7)$$

In view of all aforementioned reasons, this paper will study the estimation of the models (1.1) and (1.2) within the framework of integrated time series using orthogonal series method. We shall first consider model (1.2) and then extend the method to model (1.1). Because of the nonstationarity of the regressor, the link function may behave completely different in the case where it is integrable from the case where it is non-integrable on the real line, as suggested by the existing literature (Park and Phillips, 1999, 2001; Wang and Phillips, 2009a). This paper shall focus on the former case only. As a result, model (1.1) may be regarded as the combination of a linear model with a small departure.

Two unknown quantities are involved in model (1.2). One is the infinite dimensional unknown link function  $g(\cdot)$  and the other is the  $d$ -dimensional index vector  $\theta_0$ . They are

hard to estimate simultaneously. A common idea is first to use an estimator  $\widehat{g}_n(\cdot)$  for the link function pretending  $\theta$  is known at this stage such that  $\widehat{g}_n(\cdot)$  is close enough to  $g(\cdot)$ , and then the estimator  $\widehat{\theta}_n$  of  $\theta_0$  is obtained by the following minimisation with respect to  $\theta$ ,

$$\widehat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} \frac{1}{2} \sum_{t=1}^n (y_t - \widehat{g}_n(\theta^\top x_t))^2. \quad (1.8)$$

This is the rationale of profile method (Severini and Wong, 1992; Liang et al., 2010).

In model (1.1),  $\beta_0$  and  $\theta_0$  can be estimated adopting the similar idea. Unlike the literature where local linear functions and splines are highly prevalent, in this paper  $\widehat{g}_n(\cdot)$  are constructed by the Hermite orthogonal series. In doing so, it is a large challenge to deal with the regression matrix (signal matrix in the sequel) in the equation (1.2) after parameterisation of the link function by orthogonal series. Due to the nonstationarity of the regressor, this problem is much harder than that in the stationary case which, for instance, is handled in Gao et al. (2002). In the nonstationary context, Cai and Gao (2013) give a useful result about the convergence of the signal matrix that requires a truncation parameter to be relatively smaller than that used in the stationary time series case. We shall adopt this result and consequently, in this paper the smoothness of the link function is required to be relatively higher. See the discussion in the remark of Assumption B below.

Note also that there are a variety of papers studying the identifiability of models (1.1) and (1.2). See, for example, Lin and Kulasekera (2007). Following the existing identifiability condition we assume for both models (1.1) and (1.2) that  $\|\theta_0\| = 1$  and the first nonzero component of  $\theta_0$  is positive. Notice that there is no extra condition needed for  $\beta_0$  to make (1.1) identifiable, as discussed in Section 2.2 below. To facilitate the theoretical development in the following sections, we assume also that  $\theta_0$  is an interior point located within a compact and convex parameter space  $\Theta$ , which is also a usual assumption in a parameter estimation context. This means that  $\Theta$  contains the unit ball  $\{\theta : \|\theta\| \leq 1\}$  as a subset.

The findings of this paper include that: (1) In the single-index models there are dual convergence rates, one of them, along the direction orthogonal to  $\theta_0$ , is as fast as  $n^{-3/4}$  whereas it is as slow as  $n^{-1/4}$  along the direction of  $\theta_0$ . This originates from the integrability of the link function. (2) In the partially linear single-index models, there is a trio of convergence rates. Apart from  $\widehat{\theta}_n$  retaining the dual rates converging to  $\theta_0$ , the convergence rate of  $\widehat{\beta}_n$  to  $\beta_0$  is  $n^{-1}$ , identical with that of the linear parametric model (1.4). (3) The normalisation of  $\widehat{\theta}_n$ ,  $\frac{1}{\|\widehat{\theta}_n\|} \widehat{\theta}_n$ , converges to  $\theta_0$  with a faster rate than  $\widehat{\theta}_n$  in both models. These phenomena are verified with finite sample experiments below.

Two relevant papers related to this study are Chang and Park (2003) and Guerre and Moon (2006). Chang and Park (2003) establish econometric theory for index models with integrated time series. Although the index models may cover single-index models and par-

tially linear single-index models, the authors impose that their link function is a smooth distribution function-like transformation and they are not interested in the estimation of the unknown link function. Guerre and Moon (2006) point out that their method may be used for the estimation of single-index models where the link function  $g(x) \rightarrow \infty$  when  $|x| \rightarrow \infty$ . Clearly, both of them are quite different from the setting of this study.

The organisation of the rest of the paper is as follows. Section 2 gives some preliminaries on orthogonal expansion, estimation procedures as well as assumptions for models (1.1) and (1.2). Asymptotic theory is established in Section 3 for the estimator  $\widehat{\theta}_n$  of  $\theta_0$  in the single-index model and the estimator  $(\widehat{\beta}_n, \widehat{\theta}_n)$  of  $(\beta_0, \theta_0)$  in the partially linear single-index model. Monte Carlo simulation experiments are conducted in Section 4 to examine the theoretical results with finite sample size. Section 5 shows the implementation of the proposed estimation schedules for partially linear single-index model and single-index model with an empirical dataset. Appendix A presents some technical lemmas. All proofs (or outline of the proof) of important results are given in Appendix B. Appendices C and D contain all proofs of the Lemmas in Appendix A and the detailed proofs for the outline proofs in Appendix B, respectively.

Throughout the paper, the following notation is used.  $\|\cdot\|$  is Euclidean norm for vector and element-wise norm for matrices, i.e.  $\|A\| = (\sum_{i=1}^n \sum_{j=1}^m a_{ij}^2)^{1/2}$  if  $A = (a_{ij})_{nm}$ ;  $I_k$  is  $k$ -dimensional identity matrix;  $[a]$  is the maximum integer not exceeding  $a$ ;  $\mathbb{R}$  is the real line; for any function  $f(\cdot)$ ,  $\dot{f}(x)$ ,  $\ddot{f}(x)$  and  $\overset{\cdot\cdot}{f}(x)$  are the derivatives of the first, second and third order of  $f(\cdot)$  at  $x$ . Here, when  $f(x)$  is a vector-valued function its derivatives would be understood as element-wise. Furthermore,  $\phi(\cdot)$  stands for multivariate standard normal density; integral  $\int f(w)dw$  means  $\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(w_1, \cdots, w_k)dw_1 \cdots dw_k$  where  $w = (w_1, \cdots, w_k)$ . Convergence in probability and distribution are signified as  $\rightarrow_P$  and  $\rightarrow_D$ , respectively.

## 2 Estimation procedure and assumptions

Suppose that the link function  $g(\cdot)$  belongs to  $L^2(\mathbb{R}) = \{f(x) : \int f^2(x)dx < \infty\}$ . It is known that Hermite function sequence  $\{\mathcal{H}_i(x)\}$  is an orthonormal basis in  $L^2(\mathbb{R})$  where by definition

$$\mathcal{H}_i(x) = (\sqrt{\pi}2^i i!)^{-1/2} H_i(x) \exp\left(-\frac{x^2}{2}\right), \quad (2.1)$$

and

$$H_i(x) = (-1)^i \exp(x^2) \frac{d^i}{dx^i} \exp(-x^2), \quad i \geq 0, \quad (2.2)$$

are Hermite polynomials which are orthogonal with density  $\exp(-x^2)$ . The orthogonality gives  $\int \mathcal{H}_i(x)\mathcal{H}_j(x)dx = \delta_{ij}$ , the Kronecker delta. In addition, the orthonormal basis  $\{\mathcal{H}_i(x)\}$  is bounded uniformly in both  $i$  and  $x \in \mathbb{R}$  (see Szego, 1975, p. 242).

Thus, function  $g(\cdot) \in L^2(\mathbb{R})$  may be expanded in terms of  $\mathcal{H}_i(x)$  into an orthogonal series

$$g(x) = \sum_{i=0}^{\infty} c_i \mathcal{H}_i(x), \quad (2.3)$$

where  $c_i = \int g(x) \mathcal{H}_i(x) dx$ . Throughout, let  $k$  be a positive integer and define  $g_k(x) = \sum_{i=0}^{k-1} \beta_i \mathcal{H}_i(x)$  the truncation series with truncation parameter  $k$ . It is known that the convergence of  $g_k(x) \rightarrow g(x)$  on the real line in different sense (pointwise or in norm) relies on the property of  $g(x)$ .

## 2.1 Estimation procedure for single-index models

For the time being we suppose  $\theta$  is given. Then, by virtue of (2.3) for  $t = 1, \dots, n$ ,

$$y_t = Z_k(\theta^\top x_t)^\top c + \gamma_k(\theta^\top x_t) + e_t$$

where  $k$  is truncation parameter which will be defined below,  $Z_k(\cdot)^\top = (\mathcal{H}_0(\cdot), \dots, \mathcal{H}_{k-1}(\cdot))$ ,  $c^\top = (c_0, \dots, c_{k-1})$  and  $\gamma_k(\cdot) = \sum_{i=k}^{\infty} c_i \mathcal{H}_i(\cdot)$  the residue after truncation.

Let  $Y = (y_1, \dots, y_n)^\top$ ,  $Z = (Z_k(\theta^\top x_1), \dots, Z_k(\theta^\top x_n))^\top$  an  $n \times k$  matrix,  $\gamma = (\gamma_k(\theta^\top x_1), \dots, \gamma_k(\theta^\top x_n))^\top$  and  $e = (e_1, \dots, e_n)^\top$ . We have a matrix form equation

$$Y = Zc + \gamma + e$$

and hence by OLS (Ordinary Least Squares),

$$\hat{c} = \hat{c}(\theta) = (Z^\top Z)^{-1} Z^\top Y = c + (Z^\top Z)^{-1} Z^\top (\gamma + e), \quad (2.4)$$

is an estimator of  $c$  given the sample  $\{(y_t, x_t), t = 1, \dots, n\}$  and the parameter  $\theta$ . Define, for any  $u \in \mathbb{R}$ ,  $\hat{g}_n(u) \equiv \hat{g}_n(u; \theta) = Z_k(u)^\top \hat{c}(\theta)$  as an estimator of  $g(u)$  given the sample  $\{(y_t, x_t), t = 1, \dots, n\}$  and the parameter  $\theta$ .

Then, we choose the optimum  $\hat{\theta}_n$  such that

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}} L_n(\theta) = \frac{1}{2} \sum_{t=1}^n [y_t - \hat{g}_n(\theta^\top x_t)]^2, \quad (2.5)$$

as an estimator of  $\theta$ .

Notice that at empirical level  $\hat{g}_n(u; \theta)$  fails to work because of unavailability of  $\theta$ . Now that  $\hat{\theta}_n$  is available, we have  $\hat{c}(\hat{\theta}_n)$ , the final version estimator of  $c$  given the sample  $\{(y_t, x_t), t = 1, \dots, n\}$  with  $\theta$  being replaced by  $\hat{\theta}_n$  in (2.4). Therefore, the final version of the estimator of  $g(u)$  given sample  $\{(y_t, x_t), t = 1, \dots, n\}$  is  $\hat{g}_n(u; \hat{\theta}_n) = Z_k(u)^\top \hat{c}(\hat{\theta}_n)$ .

## 2.2 Estimation procedure for partially linear single-index models

This subsection discusses the procedure of estimation for the partially single-index model (1.1). Usually, researchers, such as Xia et al. (1999), impose an identification condition that  $\beta_0$  is perpendicular to  $\theta_0$  on the partially single-index models. This is because when  $\beta_0$  is not perpendicular to  $\theta_0$ , a new vector  $\beta_0 - (\beta_0^\top \theta_0)\theta_0$  would be used in the place of  $\beta_0$  and the  $g$  function would be replaced by  $g(u) + (\beta_0^\top \theta_0)u$ . However, in model (1.1) we would not be able to do so due to the integrability of  $g(\cdot)$  and the nonstationarity of the regressor, as  $g(\theta_0^\top x_t) + (\beta_0^\top \theta_0)\theta_0^\top x_t$  and  $g(\theta_0^\top x_t)$  have completely different behavior.

Nevertheless, lack of orthogonality between  $\beta_0$  and  $\theta_0$  does not affect the identifiability of the model at all. Indeed, suppose that there are  $\beta_1, \beta_2$  and  $\theta_1$  and  $\theta_2$  ( $\|\theta_1\| = \|\theta_2\| = 1$ ) as well as integrable functions  $g$  and  $h$  such that

$$\beta_1^\top x + g(\theta_1^\top x) = \beta_2^\top x + h(\theta_2^\top x)$$

for all  $x \in \mathbb{R}^d$ . Then we have  $(\beta_1 - \beta_2)^\top x = h(\theta_2^\top x) - g(\theta_1^\top x)$ , which, by integrability of  $g$  and  $h$ , implies  $\beta_1 = \beta_2$  and  $h(\theta_2^\top x) = g(\theta_1^\top x)$  for any  $x$ . The identifiability of partially linear single-index models is therefore reduced to that of single-index models. Following the proof of Theorem 1 in Lin and Kulasekera (2007), we may show that  $\theta_1 = \theta_2$  and  $g \equiv h$ .

The estimation procedure in partially linear single-index models is as follows. First, suppose that  $\beta$  and  $\theta$  are given. By virtue of (2.3), for each  $t$  rewrite (1.1) into the following

$$y_t - \beta^\top x_t = Z_k(\theta^\top x_t)^\top c + \gamma_k(\theta^\top x_t) + e_t,$$

where  $Z_k(\cdot)$ ,  $c$  and  $\gamma_k(\cdot)$  are defined as before.

Denote  $X = (x_1, x_2, \dots, x_n)^\top$  an  $n \times d$  matrix, and  $Y, Z, \gamma, e$  remain the same as in the last subsection. We have the following matrix form equation,

$$Y - X\beta = Zc + \gamma + e. \quad (2.6)$$

Then, OLS gives  $\tilde{c} = \tilde{c}(\beta, \theta) = (Z^\top Z)^{-1}Z^\top(Y - X\beta)$ , the estimator of  $c$ . It follows from (2.6) that

$$\tilde{c} = (Z^\top Z)^{-1}Z^\top(Y - X\beta) = c + (Z^\top Z)^{-1}Z^\top(\gamma + e), \quad (2.7)$$

which has the same expression as  $\hat{c}$  in (2.4).

Define, for any real  $u \in \mathbb{R}$ ,  $\hat{g}_n(u) \equiv \hat{g}_n(u; \beta, \theta) = Z_k(u)^\top \tilde{c}$ . The estimator of  $(\beta_0, \theta_0)$  is given by

$$\begin{pmatrix} \hat{\beta}_n \\ \hat{\theta}_n \end{pmatrix} = \underset{\theta \in \Theta, \beta}{\operatorname{argmin}} L_n(\beta, \theta) = \frac{1}{2} \sum_{t=1}^n [y_t - \beta^\top x_t - \hat{g}_n(\theta^\top x_t)]^2. \quad (2.8)$$



## 2.3 Assumptions

The following assumptions are stipulated for the theoretical development below.

### Assumption A

- (a) Let  $\{\varepsilon_j, -\infty < j < \infty\}$  be a sequence of  $d$ -dimensional identical independent distributed random vector variables with  $E\varepsilon_1 = 0$ ,  $E[\varepsilon_1\varepsilon_1^\top] = \sigma^2 I_d$  and  $E\|\varepsilon_1\|^p < \infty$  for some  $p > 2$ . The characteristic function of  $\varepsilon_1$  is integrable, that is,  $\int |E \exp(iu\varepsilon_1)| du < \infty$ .
- (b) Let, for  $t \geq 1$ ,  $x_t = x_{t-1} + v_t$  with  $x_0 = O_P(1)$ , where  $\{v_t\}$  is a linear process defined by  $v_t = \sum_{j=0}^{\infty} \rho_j \varepsilon_{t-j}$  satisfying that  $\{\rho_j\}$  is a scalar sequence such that  $\sum_{j=0}^{\infty} |\rho_j| < \infty$  and  $\rho = \sum_{j=0}^{\infty} \rho_j \neq 0$ .
- (c) There is a  $\sigma$ -field sequence  $\mathcal{F}_t$  such that  $(e_t, \mathcal{F}_t)$  is a martingale sequence, that is, for all  $1 \leq t \leq n$ ,  $E(e_t | \mathcal{F}_{t-1}) = 0$  almost surely (a.s.). In addition,  $E(e_t^2 | \mathcal{F}_{t-1}) = \sigma_e^2$  a.s. and  $\mu_4 := \sup_{1 \leq t \leq n} E(e_t^4 | \mathcal{F}_{t-1}) < \infty$  a.s.
- (d)  $x_t$  is adapted with  $\mathcal{F}_{t-1}$ .
- (e) Let  $V_n(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nr]} v_i$  and  $U_n(r) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nr]} e_i$ . Suppose that  $(U_n(r), V_n(r)) \rightarrow_D (U(r), V(r))$  as  $n \rightarrow \infty$ , where  $(U(r), V(r))$  is a  $(d+1)$ -vector of Brownian motions.

*Remark 2.1.* All conditions in Assumption A are routine requirements in the nonstationary model estimation context. Conditions (a) and (b) stipulate that the regressor  $x_t$  is an integrated process generated by a linear process  $v_t$  which has the i.i.d. sequence  $\{\varepsilon_j, -\infty < j < \infty\}$  as building blocks. Meanwhile, (c), (d) and (e) are extensively used in related papers such as Park and Phillips (2000), Wang and Phillips (2009a,b, 2012), Gao et al. (2009a,b, 2012), among others. The  $\sigma$ -fields sequence  $\mathcal{F}_t$  may be taken as  $\mathcal{F}_t = \sigma(\cdots, \varepsilon_t, \varepsilon_{t+1}; e_1, \cdots, e_t)$ .

It is known that by Skorohod representation theorem (p. 71 Pollard, 1984) there exists  $(U_n^0(r), V_n^0(r))$  in a richer probability space such that  $(U_n(r), V_n(r)) =_D (U_n^0(r), V_n^0(r))$  for which  $(U_n^0(r), V_n^0(r)) \rightarrow_{a.s.} (U(r), V(r))$  uniformly on  $[0, 1]^{d+1}$ . To avoid repetitious embedding of  $(U_n(r), V_n(r))$  to the richer probability space where  $(U_n^0(r), V_n^0(r))$  is defined, we simply write  $(U_n(r), V_n(r)) = (U_n^0(r), V_n^0(r))$  rather than  $(U_n(r), V_n(r)) =_D (U_n^0(r), V_n^0(r))$ . Since the crucial Lemmas A.2 and A.3 below are derived in this richer probability space, all proofs in the paper is understood in the richer space as well. We do not mention this again.

### Assumption B

- (a) Suppose that  $g(x)$  is differentiable on  $\mathbb{R}$  and  $g^{(m-\ell)}(x)x^\ell \in L^2(\mathbb{R})$  for  $\ell = 0, 1, \cdots, m$  with an integer  $m$ .

(b)  $k = \lceil a \cdot n^\kappa \rceil$  with some constant  $a > 0$ ,  $\kappa \in (0, 1/8)$  and  $\kappa(m-3) \geq \frac{1}{2}$  with  $m$  as in (a) above.

*Remark 2.2.* Since differentiability and integrability on the real line imply the boundedness of a function, condition (a) ensures that all integrable functions in the form of  $g^{(m-\ell)}(x)x^\ell$  used in this paper are bounded. Regarding condition (b), although it is stringent for  $\kappa$ , we may choose, for example,  $\kappa \in [\frac{5}{44}, \frac{5}{41}]$  and  $m = 8$  in practice. Note also that condition (b) makes Lemma A.2 in Appendix A applicable since Lemma A.2 requires  $0 < \kappa < 3/22$ . Also because of condition (b),  $m$  has to be relatively large in this study such that the orthogonal series expansion for the link function converges so fast that the increase of the nonstationary sample mean can be offset. See, for example, the proof of  $T_2 = o_P(1)$  in Lemma D.1 of Appendix D in the supplemental material.

Additionally, in view of (1) in Lemma A.1 below, this assumption guarantees termwise derivative operation on (2.3) up to the third order, viz.,  $g^{(p)}(x) = \sum_{i=0}^{\infty} c_i \mathcal{H}_i^{(p)}(x)$  with  $p = 1, 2, 3$ .

**Theorem 2.1.** *Under Assumptions A and B, we have*

- (1)  $\|\widehat{c}(\theta) - c\|^2 \rightarrow_P 0$  as  $n \rightarrow \infty$ ; more precisely,  $\|\widehat{c}(\theta) - c\|^2 = O_P(1) \frac{k}{\sqrt{n}}$  independent of  $\theta$ .  
This is also true for  $\widetilde{c}(\beta, \theta)$ , viz.,  $\|\widetilde{c}(\beta, \theta) - c\|^2 = O_P(1) \frac{k}{\sqrt{n}}$  independent of  $(\beta, \theta)$ .
- (2)  $\max_{u \in \mathbb{R}} |\widehat{g}_n(u; \theta) - g(u)| = O_P(1) n^{-1/4+\kappa}$  independent of  $\theta$ . This is also true for  $\widehat{g}_n(u; \beta, \theta)$ .

This theorem gives the convergence rate of  $\|\widehat{c}(\theta) - c\|^2$  ( $\|\widetilde{c}(\beta, \theta) - c\|^2$ , respectively), and more importantly the uniform rate of  $\widehat{g}_n(u; \theta)$  ( $\widehat{g}_n(u; \beta, \theta)$ , respectively) converging to  $g(u)$  by virtue of the uniform boundedness of the orthonormal basis used. As a result, all convergence is independent of  $\theta$  ( $(\beta, \theta)$ , respectively).

## 3 Asymptotic theory

### 3.1 Asymptotic theory for single-index models

There is a routine schedule of investigating the asymptotics for the extremum estimator  $\widehat{\theta}_n$  that we shall follow (e.g. Wooldridge, 1994). From equation (2.5), the score  $S_n(\theta)$  and hessian  $J_n(\theta)$  are

$$S_n(\theta) = \frac{\partial L_n(\theta)}{\partial \theta} = - \sum_{t=1}^n [y_t - \widehat{g}_n(\theta^\top x_t)] \dot{\widehat{g}}_n(\theta^\top x_t) x_t,$$

$$J_n(\theta) = \frac{\partial^2 L_n(\theta)}{\partial \theta \partial \theta^\top} = \sum_{t=1}^n \dot{\widehat{g}}_n^2(\theta^\top x_t) x_t x_t^\top - \sum_{t=1}^n [y_t - \widehat{g}_n(\theta^\top x_t)] \ddot{\widehat{g}}_n(\theta^\top x_t) x_t x_t^\top.$$

To facilitate our development, we need to rotate our parameter vector, namely, reparameterising the model. Let  $Q = (\theta_0, Q_2)$  be a  $d \times d$  orthogonal matrix. We can write the single-index model as

$$y_t = g(\theta_0^\top Q Q^\top x_t) + e_t = g(\alpha_0^1 x_{1t} + \alpha_0^2 x_{2t}) + e_t,$$

where  $\alpha_0^1 = \|\theta_0\|^2 = 1$ ,  $\alpha_0^2 = Q_2^\top \theta_0 = 0$ , a  $(d-1)$ -dimensional zero vector by orthogonality,  $x_{1t} = \theta_0^\top x_t$  a scalar process,  $x_{2t} = Q_2^\top x_t$  a  $(d-1)$ -dimensional vector 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. Conformable with this rotation, define for  $V(r)$  given in Assumption A,

$$V_1(r) = \theta_0^\top V(r), \quad \text{and} \quad V_2(r) = Q_2^\top V(r). \quad (3.1)$$

It is noteworthy that the rotation is not necessary in practice, as shown in the section of simulation, and it is also logically impossible since  $\theta_0$  is unknown. The rotation is only a measure to derive an asymptotic theory for the proposed estimator.

Let  $\hat{\theta}_n$  be the estimator of  $\theta_0$  given by (2.5). As usual, we have expansion

$$0 = S_n(\hat{\theta}_n) = S_n(\theta_0) + J_n(\theta_n)(\hat{\theta}_n - \theta_0), \quad (3.2)$$

where  $S_n(\hat{\theta}_n)$  and  $S_n(\theta_0)$  are the scores at  $\hat{\theta}_n$  and  $\theta_0$ , respectively, and  $J_n(\theta_n)$  is the hessian matrix with the rows evaluated at a point  $\theta_n$  between  $\hat{\theta}_n$  and  $\theta_0$ .

If  $\hat{\alpha}_n$  is the nonlinear least square estimator of  $\alpha_0$ , then  $\hat{\alpha}_n = Q^\top \hat{\theta}_n$  or  $\hat{\theta}_n = Q \hat{\alpha}_n$ . Moreover, the score function  $S_n(\alpha)$  and the hessian  $J_n(\alpha)$  for the parameter  $\alpha$  can be obtained from those for  $\theta$ . More precisely,  $S_n(\alpha) = Q^\top S_n(\theta)$  and  $J_n(\alpha) = Q^\top J_n(\theta) Q$ . Premultiplying equation (3.2) by  $Q^\top$ , we have

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

In order to obtain an asymptotic distribution for  $\hat{\alpha}_n - \alpha_0$ , the relationship (3.3) motivates us to investigate the asymptotic distributions for the score  $S_n(\alpha_0)$  and for the hessian  $J_n(\alpha_0)$  first. Denote  $D_n = \text{diag}(n^{1/4}, n^{3/4} I_{d-1})$ .

**Theorem 3.1.** *Under Assumptions A and B, as  $n \rightarrow \infty$*

$$D_n^{-1} S_n(\alpha_0) \rightarrow_D R^{1/2} W(1) \quad \text{and} \quad D_n^{-1} J_n(\alpha_0) D_n^{-1} \rightarrow_P R \quad (3.4)$$

where  $W(1)$  is a standard normal vector of dimension  $d$  independent of  $V(r)$  and the symmetric block matrix  $R = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$  is given by

$$r_{11} = L_1(1, 0) \int s^2 \dot{g}^2(s) ds, \quad r_{12} = \int_0^1 V_2^\top(r) dL_1(r, 0) \int s \dot{g}^2(s) ds,$$

$$r_{21} = r_{12}^\top, \quad r_{22} = \int_0^1 V_2(r) V_2^\top(r) dL_1(r, 0) \int \dot{g}^2(s) ds.$$

in which  $V_1$  and  $V_2$  given by (3.1) are Brownian motions of dimension 1 and  $d - 1$  respectively,  $L_1(r, 0)$  denotes the local time process of the Brownian motion  $V_1(\cdot)$  standing for the sojourn time of  $V_1$  at zero over  $[0, r]$ .

Given Theorem 3.1 and equation (3.3), it is expected that

$$D_n(\hat{\alpha}_n - \alpha_0) = -[D_n^{-1} J_n(\alpha_0) D_n^{-1}]^{-1} D_n^{-1} S_n(\alpha_0) + o_P(1),$$

to deduce the asymptotical distribution of  $\hat{\alpha}_n$ . Indeed, this is established in the following theorem.

**Theorem 3.2.** *Under Assumptions A and B,  $\hat{\alpha}_n \rightarrow_P \alpha_0$  and*

$$D_n(\hat{\alpha}_n - \alpha_0) \rightarrow_D R^{-1/2} W(1), \quad (3.5)$$

as  $n \rightarrow \infty$  with the same notation in Theorem 3.1.

In view of the structure of  $D_n$ , we may have two different limits from (3.5),

$$\sqrt[4]{n}(\hat{\alpha}_{1n} - \alpha_0^1) \rightarrow_D \mathbf{MN}(0, \rho_{11}), \quad \text{and} \quad \sqrt[4]{n}^3(\hat{\alpha}_{2n} - \alpha_0^2) \rightarrow_D \mathbf{MN}(0, \rho_{22}), \quad (3.6)$$

where  $\hat{\alpha}_n = (\hat{\alpha}_{1n}, \hat{\alpha}_{2n}^\top)^\top$ ,  $\mathbf{MN}$  stands for mixed normal distribution,  $\rho_{11}$  and  $\rho_{22}$  are diagonal blocks on the matrix  $R^{-1} = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix}$ ,

$$\rho_{11} = (r_{11} - r_{12} r_{22}^{-1} r_{21})^{-1}, \quad \text{and} \quad \rho_{22} = (r_{22} - r_{21} r_{11}^{-1} r_{12})^{-1}. \quad (3.7)$$

Hence,  $\hat{\alpha}_n$  has two different convergence rates for its components. We shall discuss this phenomenon later.

The limit distribution of  $\hat{\theta}_n - \theta_0$  may be derived from that of  $\hat{\alpha}_n - \alpha_0$ . To this end, (3.5) is equivalently written as

$$D_n Q^\top (\hat{\theta}_n - \theta_0) \rightarrow_D R^{-1/2} W(1). \quad (3.8)$$

Using the definition of  $D_n$  and  $Q$ , we can state:

**Theorem 3.3.** *Under Assumptions A and B, we have as  $n \rightarrow \infty$ ,*

$$\sqrt[4]{n}(\hat{\theta}_n - \theta_0) \rightarrow_D \mathbf{MN}(0, \rho_{11} \theta_0 \theta_0^\top), \quad (3.9)$$

where  $\rho_{11}$  is given by (3.7).

*Remark 3.1.* It is interesting to understand the implication of the limit forms of (3.8) and (3.9). Observe that  $Q = (\theta_0, Q_2)$  is a coordinate system in  $\mathbb{R}^d$  possibly other than the usual one consisting of all columns of  $I_d$ . Thus, (3.8) means that in the system of  $Q$ , estimator  $\hat{\theta}_n$  has dual convergence rates: along  $\theta_0$  is  $n^{-1/4}$ , while on all directions orthogonal to  $\theta_0$  is as fast as  $n^{-3/4}$ . This is because the estimators as minimizer of (2.5) can capture much more information along the directions orthogonal to  $\theta_0$  than on that of  $\theta_0$ . In other words, it is the integrability of the link function that causes the strong I(1)-type signals from the linear form  $\theta_0^\top x_t$  to be attenuated by  $g(\theta_0^\top x_t)$ . This takes place in both the score and the hessian functions via (3.3) which determines the asymptotic behaviour of the estimators. On the other hand, (3.9) indicates under the usual coordinate system  $\hat{\theta}_n$  converges to  $\theta_0$  in probability at rate of  $n^{-1/4}$ . This happens because the slower rate eventually will dominate the faster one, so that the overall rate is determined by the one on the slower direction.

By Theorem 3.3 we have  $\hat{\theta}_n \sim \mathbf{MN}(\theta_0, \frac{1}{\sqrt{n}}\rho_{11}\theta_0\theta_0^\top)$ . The expected estimate of the covariance matrix of  $\hat{\theta}_n$  is the inverse hessian  $[J_n(\hat{\theta}_n)]^{-1}$  or even  $[\tilde{J}_n(\hat{\theta}_n)]^{-1}$ , where  $\tilde{J}_n(\theta) = \sum_{t=1}^n \hat{g}_n^2(\theta^\top x_t)x_t x_t^\top$  is the leading term of  $J_n(\theta)$ .

We then have the following results.

**Theorem 3.4.** *Under Assumptions A and B, we have as  $n \rightarrow \infty$ ,*

$$\sqrt{n}[J_n(\hat{\theta}_n)]^{-1} \rightarrow_P \rho_{11}\theta_0\theta_0^\top \quad \text{and} \quad \sqrt{n}[\tilde{J}_n(\hat{\theta}_n)]^{-1} \rightarrow_P \rho_{11}\theta_0\theta_0^\top. \quad (3.10)$$

In empirical context, in order to keep the identifiability condition met, we should normalize the estimator  $\hat{\theta}_n$ . Meanwhile, from theoretical point of view we are also interested in the asymptotics of the normalisation of  $\hat{\alpha}_n = (\hat{\alpha}_{1n}, \hat{\alpha}_{2n}^\top)^\top = Q^\top \hat{\theta}_n$ . To proceed further, define

$$\hat{\theta}_{n,\text{emp}} = \frac{1}{\|\hat{\theta}_n\|} \hat{\theta}_n, \quad \text{and} \quad \hat{\alpha}_{n,\text{unit}} = (\hat{\alpha}_{n,\text{unit}}^1, (\hat{\alpha}_{n,\text{unit}}^2)^\top)^\top = \frac{1}{\|\hat{\alpha}_n\|} \hat{\alpha}_n. \quad (3.11)$$

Since  $\hat{\theta}_{n,\text{emp}}$  always have the same norm as  $\theta_0$  ( $\|\theta_0\| = 1$ ), intuitively  $\hat{\theta}_{n,\text{emp}}$  might have a quicker convergence rate than  $\hat{\theta}_n$ . On the other hand, the first element of  $\hat{\alpha}_{n,\text{unit}}$  is  $\hat{\alpha}_{n,\text{unit}}^1 = \frac{1}{\|\hat{\alpha}_n\|} \hat{\alpha}_{1n}$ , which is always less than one, implying that  $\hat{\alpha}_{n,\text{unit}}^1$  is an under estimator for  $\alpha_0^1 = 1$ . The next theorem explores the asymptotical properties for  $\hat{\theta}_{n,\text{emp}}$  and  $\hat{\alpha}_{n,\text{unit}}$ .

**Theorem 3.5.** *Under Assumptions A and B, we have as  $n \rightarrow \infty$ ,*

$$\sqrt[4]{n}^3 (\hat{\theta}_{n,\text{emp}} - \theta_0) \rightarrow_D \mathbf{MN}(0, Q_2 \rho_{22} Q_2^\top), \quad (3.12)$$

where  $\rho_{22}$  is defined by (3.7). In addition,

$$\sqrt{n}^3 (\hat{\alpha}_{n,\text{unit}}^1 - \alpha_0^1) \rightarrow_D -\frac{1}{2} \|\xi\|^2 \quad \text{and} \quad \sqrt[4]{n}^3 \hat{\alpha}_{n,\text{unit}}^2 \rightarrow_D \xi, \quad (3.13)$$

where  $\xi := (\xi_1, \dots, \xi_{d-1})^\top \sim \mathbf{MN}(0, \rho_{22})$  is the limit of  $\sqrt[4]{n}^3 \hat{\alpha}_{2n}$  given by (3.6).

This theorem shows that the convergence of  $\widehat{\theta}_{n,\text{emp}}$  to  $\theta_0$  is faster than that of  $\widehat{\theta}_n$  to  $\theta_0$ . Roughly speaking, the normalisation of  $\widehat{\theta}_n$  scales the estimator to the surface of the unit ball. This accelerates its convergence to the unit vector  $\theta_0$  by compensating for the slow convergence of  $\widehat{\theta}_n$  along  $\theta_0$ . Indeed, the coordinates of  $\widehat{\theta}_{n,\text{emp}}$  under  $Q$  are  $\theta_0^\top \widehat{\theta}_{n,\text{emp}} = \widehat{\alpha}_{n,\text{unit}}^1$  and  $Q_2^\top \widehat{\theta}_{n,\text{emp}} = \widehat{\alpha}_{n,\text{unit}}^2$ . Thus, in view of (3.13) we may expect that  $\widehat{\theta}_{n,\text{emp}}$  converges to  $\theta_0$  at a rate of  $n^{-3/4}$ .

### 3.2 Asymptotic theory for partially linear single-index models

We start as usual with the score and the hessian of the minimisation problem (2.8). The score  $\mathfrak{S}_n(\beta, \theta)$  and the hessian  $\mathfrak{J}_n(\beta, \theta)$  are

$$\mathfrak{S}_n(\beta, \theta) = \begin{pmatrix} \frac{\partial L_n(\beta, \theta)}{\partial \beta} \\ \frac{\partial L_n(\beta, \theta)}{\partial \theta} \end{pmatrix}, \quad \mathfrak{J}_n(\beta, \theta) = \begin{pmatrix} \frac{\partial^2 L_n(\beta, \theta)}{\partial \beta \partial \beta^\top} & \frac{\partial^2 L_n(\beta, \theta)}{\partial \beta \partial \theta^\top} \\ \frac{\partial^2 L_n(\beta, \theta)}{\partial \theta \partial \beta^\top} & \frac{\partial^2 L_n(\beta, \theta)}{\partial \theta \partial \theta^\top} \end{pmatrix}$$

in which

$$\begin{aligned} \frac{\partial L_n(\beta, \theta)}{\partial \beta} &= - \sum_{t=1}^n [y_t - \beta^\top x_t - \widehat{g}_n(\theta^\top x_t)] x_t \\ \frac{\partial L_n(\beta, \theta)}{\partial \theta} &= - \sum_{t=1}^n [y_t - \beta^\top x_t - \widehat{g}_n(\theta^\top x_t)] \dot{\widehat{g}}_n(\theta^\top x_t) x_t \\ \frac{\partial^2 L_n(\beta, \theta)}{\partial \beta \partial \beta^\top} &= \sum_{t=1}^n x_t x_t^\top, \quad \frac{\partial^2 L_n(\beta, \theta)}{\partial \beta \partial \theta^\top} = \sum_{t=1}^n \dot{\widehat{g}}_n(\theta^\top x_t) x_t x_t^\top, \quad \frac{\partial^2 L_n(\beta, \theta)}{\partial \theta \partial \beta^\top} = \frac{\partial^2 L_n(\beta, \theta)}{\partial \beta \partial \theta^\top}, \\ \frac{\partial^2 L_n(\beta, \theta)}{\partial \theta \partial \theta^\top} &= \sum_{t=1}^n \dot{\widehat{g}}_n^2(\theta^\top x_t) x_t x_t^\top - \sum_{t=1}^n [y_t - \beta^\top x_t - \widehat{g}_n(\theta^\top x_t)] \ddot{\widehat{g}}_n(\theta^\top x_t) x_t x_t^\top. \end{aligned}$$

We also need to rotate our index vectors in model (1.1), namely, reparametrizing the model, in order to derive the asymptotics. Using orthogonal matrix  $Q = (\theta_0, Q_2)$  again, we can write the model as

$$y_t = \beta_0^\top x_t + g(\theta_0^\top x_t) + e_t = \lambda_0^1 x_{1t} + (\lambda_0^2)^\top x_{2t} + g(\alpha_0^1 x_{1t} + (\alpha_0^2)^\top x_{2t}) + e_t,$$

where  $\lambda_0^1 = \theta_0^\top \beta_0$  a scalar,  $\lambda_0^2 = Q_2^\top \beta_0$  a  $(d-1)$ -dimensional vector,  $\alpha_0^1 = \|\theta_0\|^2 = 1$ ,  $\alpha_0^2 = Q_2^\top \theta_0 = 0$  a zero  $(d-1)$ -dimensional vector,  $x_{1t} = \theta_0^\top x_t$  and  $x_{2t} = Q_2^\top x_t$  defined as before. Let  $\lambda_0 = (\lambda_0^1, (\lambda_0^2)^\top)^\top = Q^\top \beta_0$ ,  $\alpha_0 = (\alpha_0^1, (\alpha_0^2)^\top)^\top = Q^\top \theta_0$ , and  $\lambda = (\lambda^1, (\lambda^2)^\top)^\top = Q^\top \beta$ ,  $\alpha = (\alpha^1, (\alpha^2)^\top)^\top = Q^\top \theta$ . Also, group them by  $\mu_0 = (\lambda_0^\top, \alpha_0^\top)^\top$  and  $\mu = (\lambda^\top, \alpha^\top)^\top$ .

Denote  $\vartheta_0 = (\beta_0^\top, \theta_0^\top)^\top$ . Let  $\widehat{\vartheta}_n$  be the estimator of  $\vartheta_0$  given by (2.8). As usual, we have the expansion

$$0 = \mathfrak{S}_n(\widehat{\vartheta}_n) = \mathfrak{S}_n(\vartheta_0) + \mathfrak{J}_n(\vartheta_0)(\widehat{\vartheta}_n - \vartheta_0), \quad (3.14)$$

where  $\mathfrak{S}_n(\widehat{\vartheta}_n)$  and  $\mathfrak{S}_n(\vartheta_0)$  are the scores at  $\widehat{\vartheta}_n$  and  $\vartheta_0$ , respectively, and  $\mathfrak{J}_n(\vartheta_n)$  is the hessian matrix with the rows evaluated at a point  $\vartheta_n$  between  $\widehat{\vartheta}_n$  and  $\vartheta_0$ .

If  $\widehat{\mu}_n$  is the estimator of  $\mu_0$  given by the minimizer of  $L_n(\mu)$ , the counterpart of  $L_n(\beta, \theta)$  after reparameterisation, then  $\widehat{\mu}_n = \text{diag}(Q^\top, Q^\top)\widehat{\vartheta}_n$  or  $\widehat{\vartheta}_n = \text{diag}(Q, Q)\widehat{\mu}_n$ . Moreover, the score function  $\mathfrak{S}_n(\mu)$  and the hessian  $\mathfrak{J}_n(\mu)$  for the parameter  $\mu$  can be obtained from those for  $\vartheta$ . Namely,  $\mathfrak{S}_n(\mu) = \text{diag}(Q^\top, Q^\top)\mathfrak{S}_n(\vartheta)$  and  $\mathfrak{J}_n(\mu) = \text{diag}(Q^\top, Q^\top)\mathfrak{J}_n(\vartheta)\text{diag}(Q, Q)$ . Premultiplying equation (3.14) by  $\text{diag}(Q^\top, Q^\top)$ , we have

$$0 = \mathfrak{S}_n(\widehat{\mu}_n) = \mathfrak{S}_n(\mu_0) + \mathfrak{J}_n(\mu_n)(\widehat{\mu}_n - \mu_0), \quad (3.15)$$

from which the following theorem is derived.

**Theorem 3.6.** *Under Assumptions A and B,  $\widehat{\mu}_n \rightarrow_P \mu_0$ . Moreover,*

$$n(\widehat{\lambda}_n - \lambda_0) \rightarrow_D Q^\top \left( \int_0^1 V(r)V^\top(r)dr \right)^{-1} \int_0^1 V(r)dU(r) \quad (3.16)$$

and

$$D_n(\widehat{\alpha}_n - \alpha_0) \rightarrow_D R^{-1/2}W(1), \quad (3.17)$$

as  $n \rightarrow \infty$ , where  $(U(r), V(r))$  is given in Assumption A,  $D_n$ ,  $R$  and  $W$  are the same as in Theorem 3.1.

Theorem 3.6 shows that in partially linear single-index models, the estimator of the parameters in the linear part has the same convergence rate as a linear model like model (1.4), while the estimator of the index vector retain the dual rates in the system of  $Q$ . Hence, there is a trio of convergence rates accommodated in the partially linear single-index models. From the results (3.16) and (3.17), we may derive the asymptotic distribution for  $\widehat{\beta}_n$  and  $\widehat{\theta}_n$ . The result is summarized in the following theorem.

**Theorem 3.7.** *Under Assumptions A and B, for  $(\widehat{\beta}_n, \widehat{\theta}_n)$  given by (2.8) we have*

$$n(\widehat{\beta}_n - \beta_0) \rightarrow_D \left( \int_0^1 V(r)V^\top(r)dr \right)^{-1} \int_0^1 V(r)dU(r), \quad (3.18)$$

and

$$\sqrt[4]{n}(\widehat{\theta}_n - \theta_0) \rightarrow_D \mathbf{MN}(0, \rho_{11}\theta_0\theta_0^\top), \quad (3.19)$$

as  $n \rightarrow \infty$ , where all notations remain the same meanings as in Theorem 3.3.

Furthermore, in view of (3.17), for  $\widehat{\theta}_n$  estimated from the partially linear single-index model, we may still expect the results of Theorems 3.4 and 3.5 with  $\widehat{\theta}_{n,\text{emp}}$  defined in the same way as before.

**Theorem 3.8.** *Under Assumptions A and B, all results in Theorems 3.4 and 3.5 remain true for the estimator  $\widehat{\theta}_n$  given by (2.8).*

Before we prove the theorems, we evaluate the finite-sample performance of the proposed estimators and the resulting theory in Sections 4 and 5 below.

## 4 Simulation experiments

This section is to investigate the performance of the estimation procedures proposed in Section 2 for both single-index models and partially linear single-index models in finite sample circumstances. Let  $d = 2$  and  $x_t$  be generated by

$$x_t = x_{t-1} + v_t \quad \text{with} \quad v_t = \alpha_0 v_{t-1} + \varepsilon_t, \quad (4.1)$$

for  $t = 1, \dots, n$ , where  $\alpha_0 = 0.1$ ,  $\varepsilon_t \sim iiN(0, \sigma^2 I_2)$ ,  $x_0 = 0$  surely. Consider the sample sizes of  $n = 400, 600$  and  $1000$ . The number of Monte Carlo replications is  $M = 2000$ . The truncation parameter  $k = [a \cdot n^\kappa]$  with  $\kappa = \frac{5}{44}$  and  $a = 3.65$ , satisfying the theoretical requirements. We shall then use two examples.

**Example 4.1.** Consider a single-index model  $y_t = g(\theta_0^\top x_t) + e_t$ ,  $t = 1, \dots, n$ . There are two parts in the simulation, according as  $\theta_0^\top = (0.6, -0.8)$  and  $\theta_0^\top = (1, 0)$  that all satisfy  $\|\theta_0\| = 1$ .

We calculate the bias and standard deviation for  $\widehat{\theta}_n = (\widehat{\theta}_n^1, \widehat{\theta}_n^2)^\top$ :

$$\text{Bias} = \widehat{\theta}_n - \theta_0, \quad \text{S.d.} = \left( \frac{1}{M} \sum_{\ell=1}^M (\widehat{\theta}_{n\ell} - \widehat{\theta}_n)^{\otimes 2} \right)^{\otimes 1/2}, \quad (4.2)$$

where  $\otimes$  denotes element-wise operation,  $\widehat{\theta}_n = \frac{1}{M} \sum_{\ell=1}^M \widehat{\theta}_{n\ell}$  is the sample mean and  $\widehat{\theta}_{n\ell}$  stands for the estimation of the  $\ell$ -th Monte Carlo experiment.

In order to verify the asymptotic results in Theorem 3.5 we also calculate the bias and standard deviation using (4.2) for  $\widehat{\theta}_{n,\text{emp}} = \widehat{\theta}_n / \|\widehat{\theta}_n\|$ .

**Part I.** Set  $\theta_0^\top = (-0.6, -0.8)$  (hence  $\|\theta_0\| = 1$ ) and  $g(u) = (1 + u^2)e^{-u^2}$ . We use the proposed procedure in Section 2.1 to estimate  $\theta_0$ . All results are reported in Table 1.

As can be seen from Table 1, both the biases and standard deviations for  $\widehat{\theta}_n$  are approaching zero with the sample size increasing, and  $\widehat{\theta}_n^1$  and  $\widehat{\theta}_n^2$  have similar performance. This verifies the asymptotic result in Theorem 3.3 that  $\widehat{\theta}_n - \theta_0 = O_P(n^{-1/4})$ .

However, although the biases and standard deviations of  $\widehat{\theta}_{n,\text{emp}}$  still converge with the increase of the sample size  $n$ ,  $\widehat{\theta}_{n,\text{emp}}$  performs differently from  $\widehat{\theta}_n$  in both bias and standard deviation, in terms of the rate of the convergence. As indicated by Theorem 3.5,  $\widehat{\theta}_{n,\text{emp}} - \theta_0 = O_P(n^{-3/4})$ , a faster rate than  $\widehat{\theta}_n$ .



Table 1: Bias and standard deviation for single-index model

$n$	Bias			S.d.		
	400	600	1000	400	600	1000
$\hat{\theta}_n^1$	-0.0647	-0.0519	-0.0388	0.2678	0.2507	0.2042
$\hat{\theta}_n^2$	-0.0832	-0.0684	-0.0453	0.3461	0.3285	0.2586
$\hat{\theta}_{n,\text{emp}}^1$	0.0043	0.0024	-0.0016	0.1005	0.0820	0.0679
$\hat{\theta}_{n,\text{emp}}^2$	0.0063	0.0066	0.0050	0.0717	0.0659	0.0515

$$g(u) = (1 + u^2) \exp(-u^2), \theta_0 = (-0.6, -0.8)^\top \text{ and } \sigma = 0.6.$$

**Part II.** Put  $\theta_0^\top = (1, 0)$  and  $g(u) = (1 + u^2) \exp(-u^2)$ . As pointed in Section 3, the rotation of the parameter is only for the derivation of the asymptotic theory. To verify the result in Theorem 3.2, we directly take  $\theta_0 = \alpha_0 = (1, 0)^\top$  so that  $\hat{\alpha}_n = \hat{\theta}_n = (\hat{\theta}_n^1, \hat{\theta}_n^2)^\top$  in this experiment.

Table 2: Bias and standard deviation for single-index model

$n$	Bias			S.d.		
	400	600	1000	400	600	1000
$\hat{\alpha}_n^1$	0.0866	0.0768	0.0340	0.3803	0.3748	0.3338
$\hat{\alpha}_n^2$	0.0013	-0.0008	-0.0006	0.1388	0.1186	0.0898
$\hat{\alpha}_{n,\text{unit}}^1$	-0.0073	-0.0061	-0.0031	0.0246	0.0237	0.0128
$\hat{\alpha}_{n,\text{unit}}^2$	0.0011	-0.0018	-0.0003	0.1186	0.1080	0.0779

$$g(u) = (1 + u^2) \exp(-u^2), \alpha_0 = \theta_0 = (1, 0)^\top \text{ and } \sigma = 0.6.$$

As can be seen from Table 2, both the biases and standard deviations of  $\hat{\alpha}_n^1$  and  $\hat{\alpha}_n^2$  perform well in the sense that as the sample size increases they decrease. Particularly, the decrease for  $\hat{\alpha}_n^2$  is much faster than that for  $\hat{\alpha}_n^1$ . This verifies our theoretical result in Theorem 3.2 that  $\hat{\alpha}_n^2 - \alpha_0^2 = O_P(n^{-3/4})$ , while  $\hat{\alpha}_n^1 - \alpha_0^1 = O_P(n^{-1/4})$ .

Nonetheless, shown by the standard deviation,  $\hat{\alpha}_{n,\text{unit}}^1$  converges significantly quicker than  $\hat{\alpha}_{n,\text{unit}}^2$ . This is implied by the theoretical results in Theorem 3.5 that  $\hat{\alpha}_{n,\text{unit}}^1 - \alpha_0^1 = O_P(n^{-3/2})$  and  $\hat{\alpha}_{n,\text{unit}}^2 - \alpha_0^2 = O_P(n^{-3/4})$ . Notice also that the biases of  $\hat{\alpha}_{n,\text{unit}}^1$  are always negative (by definition,  $\hat{\alpha}_{n,\text{unit}}^1 < \alpha_0^1 = 1$ ) for each Monte Carlo experiment. As a result, the bias of  $\hat{\alpha}_{n,\text{unit}}^1$  converges relatively slower than that of  $\hat{\alpha}_{n,\text{unit}}^2$ .

**Example 4.2.** In this example a partially linear single-index model  $y_t = \beta_0^\top x_t + g(\theta_0^\top x_t) + e_t$ ,  $t = 1, \dots, n$ , is examined with  $g(u) = \frac{1}{1+u^2}$ . Set  $\beta_0^\top = (0.3, 0.5)$  and  $\theta_0^\top = (-0.6, 0.8)$  (hence  $\|\theta_0\| = 1$ ).

Table 3: Bias and standard deviation for partial linear single-index model

$n$	Bias			S.d.		
	400	600	1000	400	600	1000
$\hat{\theta}_n^1$	-0.0495	-0.0470	-0.0324	0.2652	0.2494	0.1991
$\hat{\theta}_n^2$	0.0676	0.0645	0.0435	0.3433	0.3340	0.2572
$\hat{\theta}_{n,\text{emp}}^1$	0.0038	0.0031	0.0023	0.0934	0.0798	0.0597
$\hat{\theta}_{n,\text{emp}}^2$	-0.0062	-0.0041	-0.0019	0.0761	0.0621	0.0475
$\hat{\beta}_n^1$	-0.0010	-0.0002	0.0001	0.0106	0.0068	0.0038
$\hat{\beta}_n^2$	-0.0007	0.0001	0.0001	0.0118	0.0067	0.0037

$g(u) = 1/(1 + u^2)$ ,  $\theta_0 = (-0.6, 0.8)^\top$ ,  $\beta_0 = (0.3, 0.5)^\top$  and  $\sigma = 0.8$ .

Formulae in (4.2) are used for  $\hat{\theta}_n$ ,  $\hat{\theta}_{n,\text{emp}}$  and  $\hat{\beta}_n$ . All results in the simulations with sample size  $n = 400, 600, 1000$  and  $\sigma = 0.8$  are given in Table 3. As can be seen, all biases and standard deviations decrease as the sample size increases. However, comparing with  $\hat{\theta}_n$ ,  $\hat{\theta}_{n,\text{emp}}$  converges to the true value of the parameter with a much faster rate. This is the realisation of the theoretical result in Theorem 3.7 and Theorem 3.8, viz.  $\hat{\theta}_n - \theta_0 = O_P(n^{-1/4})$  and  $\hat{\theta}_{n,\text{emp}} - \theta_0 = O_P(n^{-3/4})$ .

On the other hand, since  $\hat{\beta}_n = (\hat{\beta}_n^1, \hat{\beta}_n^2)^\top$  possesses the fastest convergence rate  $n^{-1}$  by Theorem 3.7 among the estimators, both the biases and standard deviations of  $\hat{\beta}_n$  support the large sample behaviour. The key asymptotical results established in Section 3 are verified in these examples.

## 5 Empirical example

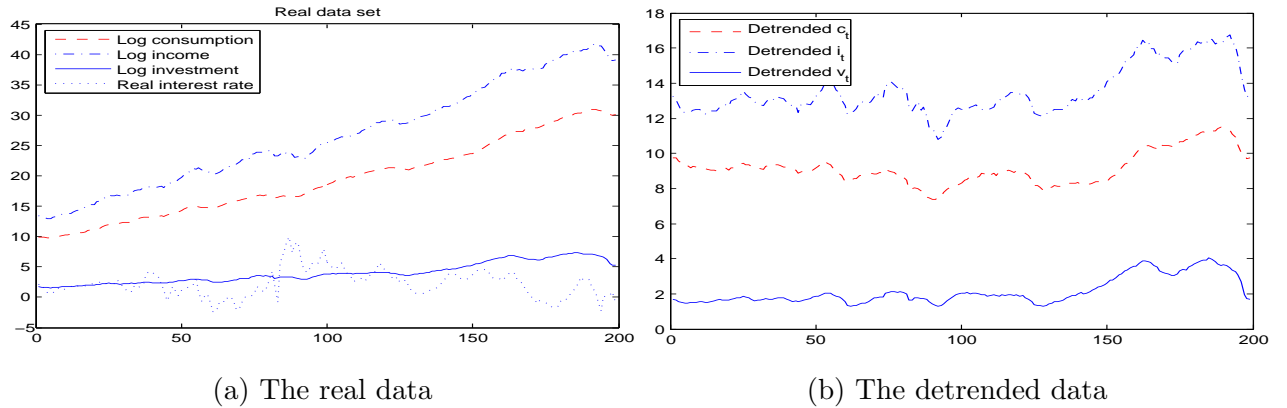
This section is an empirical study. Using an empirical data set, we shall compare three models to see which one is the most suitable model for the given data.

**The data.** The aggregate US data on consumption, income, investment and interest rate are obtained from *Federal Reserve Economic Data (FRED)*. We consider a quarterly data set over 1960:1-2009:3 with 199 observations. Let  $r_t$  stand for the real interest rates, and  $c_t = \log(C_t)$ ,  $i_t = \log(I_t)$  and  $v_t = \log(V_t)$  where  $C_t, I_t$  and  $V_t$  are the consumption expenditures,

disposable incomes and investments, respectively, for  $t = 1, \dots, 199$ . The data of  $c_t, i_t, v_t$  and  $r_t$  are plotted in (a) of Figure 1. It can be seen that all of them have trend except  $r_t$ . To meet the theoretical assumptions, we detrend the data for  $c_t, i_t$  and  $v_t$ . More precisely, suppose that  $c_t = \mu_1 + c_{t-1} + u_{1t}$ ,  $i_t = \mu_2 + i_{t-1} + u_{2t}$  and  $v_t = \mu_3 + v_{t-1} + u_{3t}$  for  $t = 2, \dots, 199$ , where  $u_{it}$ ,  $i = 1, 2, 3$ , are error terms. Then,  $\mu_i$  are estimated as:  $\hat{\mu}_1 = \frac{1}{198} \sum_{i=2}^{199} (c_t - c_{t-1}) = 0.1022$ ,  $\hat{\mu}_2 = \frac{1}{198} \sum_{i=2}^{199} (i_t - i_{t-1}) = 0.1302$ ,  $\hat{\mu}_3 = \frac{1}{198} \sum_{i=2}^{199} (v_t - v_{t-1}) = 0.0181$ .

Define for each  $t$ ,  $\tilde{c}_t = c_t - \hat{\mu}_1 t$ ,  $\tilde{i}_t = i_t - \hat{\mu}_2 t$ ,  $\tilde{v}_t = v_t - \hat{\mu}_3 t$ . They are the detrended data, plotted in (b) of Figure 1.

Figure 1: The real data and the detrended data



Unit root tests are conducted for  $\tilde{c}_t, \tilde{i}_t$ , and  $\tilde{v}_t$ , respectively, by the augmented Dickey Fuller test with nonzero mean. According to the report, all tests fail to reject the null of possessing unit root with  $p$ -values 0.7595, 0.6293 and 0.7637, respectively. In addition, it is known that  $r_t$  is an  $I(1)$  process (Gao et al., 2009b). To visualise the  $I(1)$  processes, the plots of the difference for each variable are put in Figure 2.

**The models.** Three models, linear parametric model, single-index model and partially linear single-index model, are investigated for the processed data  $\tilde{c}_t, \tilde{i}_t, \tilde{v}_t$  as well as  $r_t$  in the following forms:

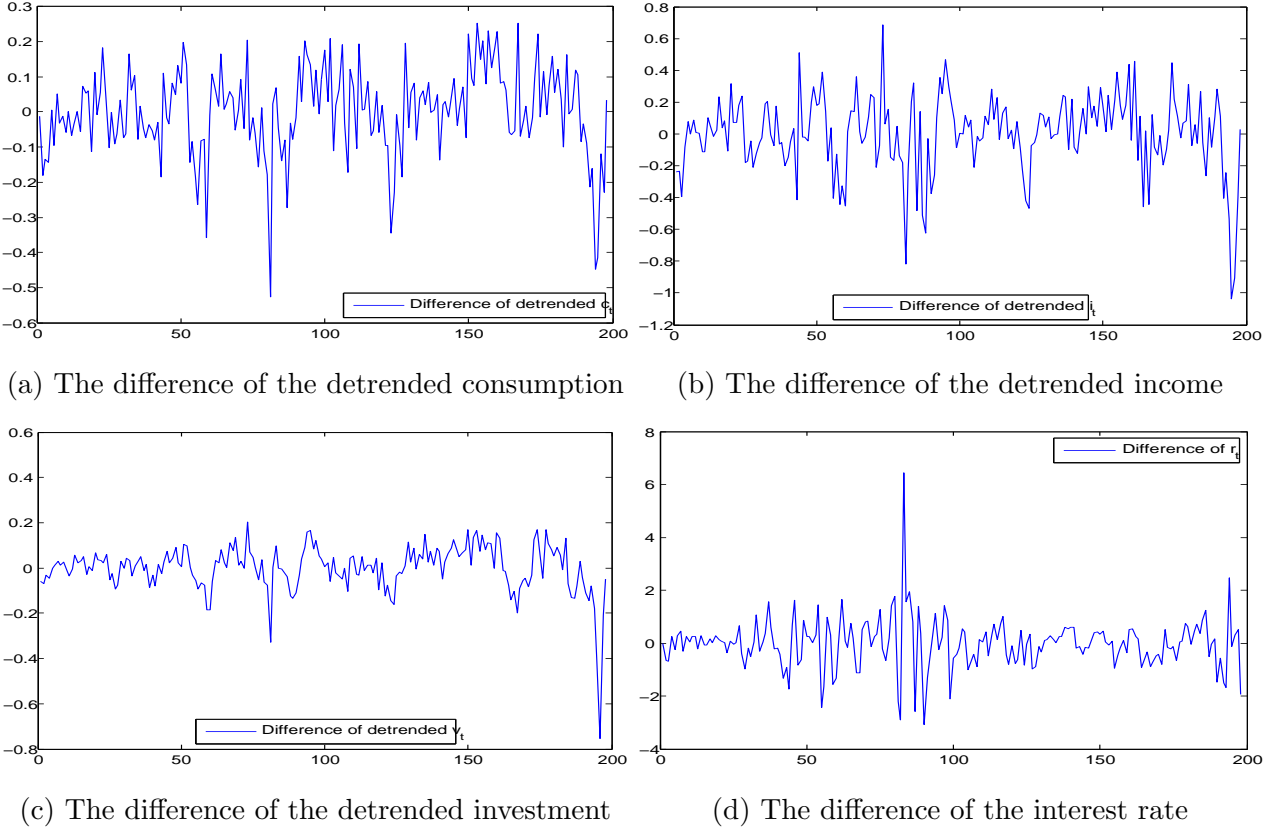
$$y_t = \beta_{\text{linear}}^\top x_t + e_{1t}, \quad (5.1)$$

$$y_t = g_1(\theta_{10}^\top x_t) + e_{2t}, \quad (5.2)$$

$$y_t = \beta_0^\top x_t + g_2(\theta_{20}^\top x_t) + e_{3t}, \quad (5.3)$$

where  $t = 2, \dots, 199$ ,  $y_t = \tilde{c}_t$  and  $x_t^\top = (x_{1t}, x_{2t}, x_{3t}, x_{4t}, x_{5t})$  in which  $x_{1t} = \tilde{i}_t$ ,  $x_{2t} = x_{1,t-1}$ ,  $x_{3t} = \tilde{v}_t$ ,  $x_{4t} = x_{3,t-1}$ ,  $x_{5t} = r_t$ , and  $g_i(\cdot)$  are unknown integrable functions for  $i = 1, 2$ ,  $e_{\ell t}$  are error terms with  $\ell = 1, 2, 3$ . Note that we only include the first lagged information in the discussion, and other lags can also be included.

Figure 2: The difference of each dataset



**Estimation.** To begin with, model (5.1) is estimated by OLS giving  $\widehat{\beta}_{\text{linear}} = (0.7952, -0.0628, -1.2315, 0.9414, -0.0644)^\top$ .

Before implementing our proposed procedures to estimate models (5.2) and (5.3), one question is to determine a suitable truncation parameter  $k$  so that the functions  $g_1(\cdot)$  and  $g_2(\cdot)$  can be better approximated with the first  $k$  terms in  $\{\mathcal{H}_i(x)\}$  defined by (2.1) and (2.2). Towards this end, we propose using the Generalised Cross-Validation (GCV) method (see Gao et al., 2002) to select an optimal value  $k$  for the models (5.2) and (5.3). Note that while there is no theory for such selection in the nonstationary time series case, the selection method works numerically in this example. Let  $\widehat{k}$  denote the optimal value such that

$$\widehat{k} = \underset{k \in K_n}{\operatorname{argmin}} \left( 1 - \frac{k}{n} \right)^{-2} \widehat{\sigma}_\ell^2(k), \quad (5.4)$$

where  $K_n = \{1, 2, \dots, 8\}$ ,  $\ell = 2, 3$ , corresponding to models (5.2) and (5.3), respectively. For model (5.2) (with  $\ell = 2$ ),  $\widehat{\sigma}_2^2(k) = \frac{1}{n} \sum_{t=1}^n \left( y_t - \widehat{g}_k(\widehat{\theta}^\top x_t) \right)^2$ , in which  $\widehat{g}_k(x) = \sum_{i=0}^{k-1} \widehat{c}_i \mathcal{H}_i(x) = Z_k(x)^\top \widehat{c}$ ; for model (5.3) (with  $\ell = 3$ ),  $\widehat{\sigma}^2(k) = \frac{1}{n} \sum_{t=1}^n \left( y_t - \widehat{\beta}^\top x_t - \widehat{g}_k(\widehat{\theta}^\top x_t) \right)^2$ , in which  $\widehat{g}_k(x) = \sum_{i=0}^{k-1} \widetilde{c}_i \mathcal{H}_i(x) = Z_k(x)^\top \widetilde{c}$ .

In model (5.2) we have  $\widehat{k} = 3$  by GCV. The estimate of  $\theta_{10}$  is

$$\widehat{\theta}_{1n} = (0.0152, -0.0014, -0.0229, 0.0176, -0.0016)^\top.$$

Meanwhile,  $\hat{c} = \hat{c}(\hat{k}, \hat{\theta}_{1n}) = (237.05, -61.92, 315.32)^\top$ , implying

$$\hat{g}_{1,\hat{k}}(x) \equiv \hat{g}_{1,3}(x) = 237.05\mathcal{H}_0(x) - 61.92\mathcal{H}_1(x) + 315.32\mathcal{H}_2(x). \quad (5.5)$$

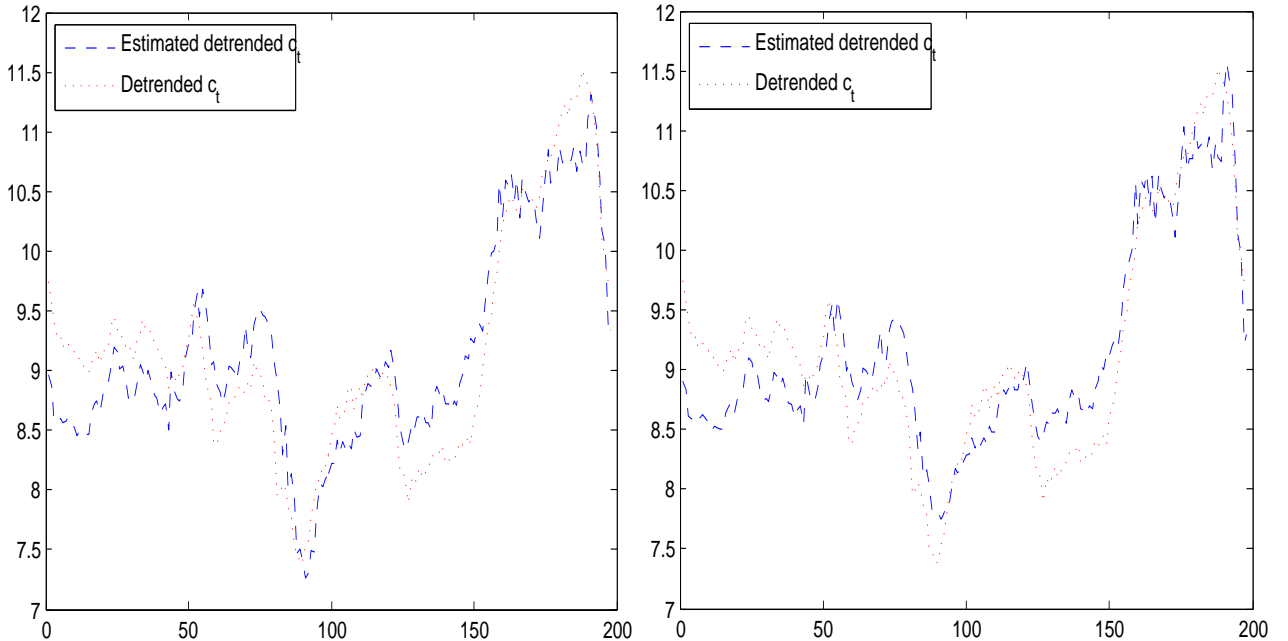
In model (5.3) we have  $\hat{k} = 5$  by GCV. The estimate of  $\beta_0$  is  $\hat{\beta}_n = (0.5701, -0.0479, -1.1689, 1.8685, -0.1223)^\top$  and the estimate of  $\theta_0$  is  $\hat{\theta}_{2n} = (-0.3452, 0.2110, 0.0835, 2.6095, -0.2022)^\top$ . Meanwhile,  $\hat{c} = \hat{c}(\hat{k}, \hat{\beta}_n, \hat{\theta}_{2n}) = (-89.64, 112.54, -74.65, 28.94, -3.33)^\top$ , implying

$$\begin{aligned} \hat{g}_{2,\hat{k}}(x) &\equiv \hat{g}_{2,5}(x) \\ &= -89.64\mathcal{H}_0(x) + 112.54\mathcal{H}_1(x) - 74.65\mathcal{H}_2(x) + 28.94\mathcal{H}_3(x) - 3.33\mathcal{H}_4(x). \end{aligned} \quad (5.6)$$

Given  $x_t$  and  $\hat{\theta}_{2n}$  we may find the link function  $\hat{g}_{2,5}(x)$  is defined on  $[0.8256, 8.2396]$ .

In comparison, the detrended log consumptions  $y_t = \tilde{c}_t$  are plotted respectively with the estimated detrended log consumptions by the linear model  $\hat{y}_t = \hat{\beta}_{\text{linear}}^\top x_t$ , the single-index model  $\hat{y}_t = \hat{g}_{1,3}(\hat{\theta}_{1n}^\top x_t)$  and by the partially linear single-index model  $\hat{y}_t = \hat{\beta}_n^\top x_t + \hat{g}_{2,5}(\hat{\theta}_{2n}^\top x_t)$  in Figure 3 and Figure 4. As can be seen, there are some deviations in Figure 3 but in Figure 4 all departures of the estimated values from the real values are relatively smaller.

Figure 3: Comparison with the real data and the estimated data by (5.1) and (5.2)



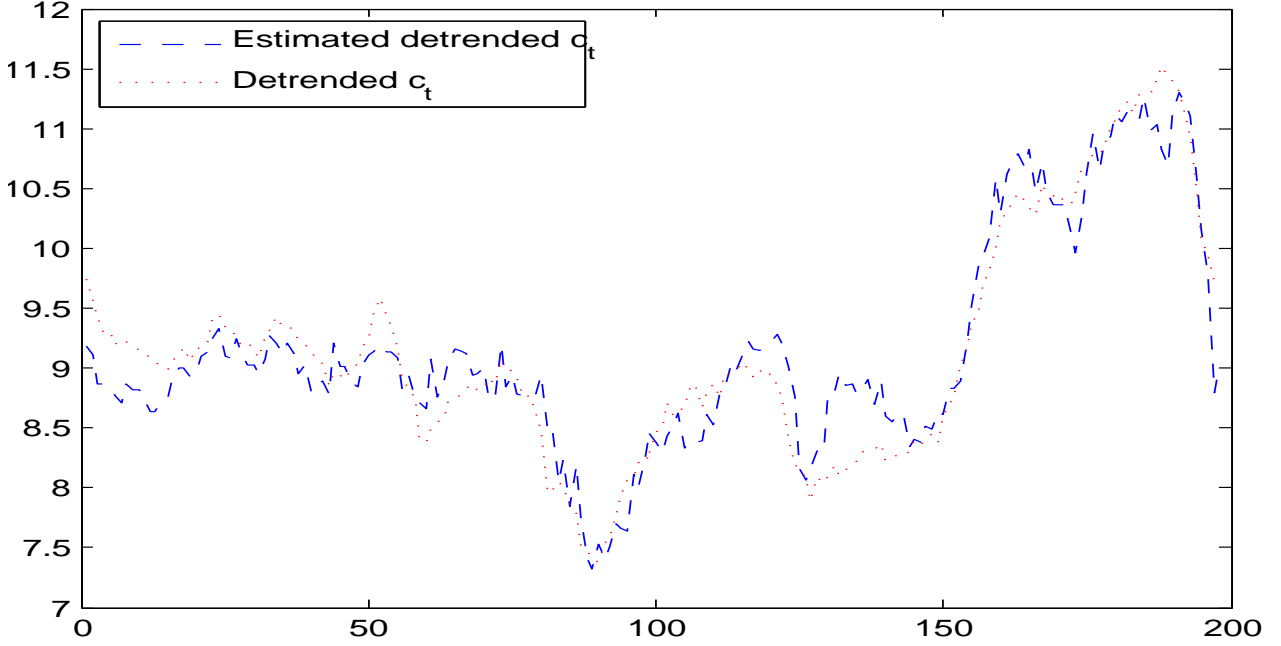
(a) The real and the estimated data by (5.1)

(b) The real and the estimated data by (5.2)

To proceed further, we compare the so-called in-sample and out-of-sample mean square errors among the three models.

(1) In-sample mean square error ( $\text{MSE}_{in}$ ). As above, all unknown parameters and functions in the three models (5.1)-(5.3) are estimated based on the whole observations  $(x_t, y_t)$ ,

Figure 4: Comparison with the real data and the estimated data by (5.3)



$t = 2, \dots, 199$ . Once these have been done, we shall have estimated  $\hat{y}_t^\ell$  with  $\ell = 1, 2, 3$  corresponding to models (5.1)- (5.3) for  $t = 2, \dots, 199$ ,

$$\hat{y}_t^1 = \hat{\beta}_{\text{linear}}^\top x_t, \quad \hat{y}_t^2 = \hat{g}_{1,3}(\hat{\theta}_{1n}^\top x_t), \quad \text{and} \quad \hat{y}_t^3 = \hat{\beta}_n^\top x_t + \hat{g}_{2,5}(\hat{\theta}_{2n}^\top x_t).$$

Then, the in-sample mean square errors are calculated, for  $\ell = 1, 2, 3$ , by

$$\text{MSE}_{in}(\ell) = \frac{1}{198} \sum_{t=2}^{199} (y_t - \hat{y}_t^\ell)^2. \quad (5.7)$$

Meanwhile, to verify the choice of GCV for both model (5.2) and (5.3), the in-sample mean square errors for model (5.2) with  $k = 2, 4$  respectively and for model (5.3) with  $k = 4, 6$  respectively are calculated as well.

(2) Out-of-sample mean square error ( $\text{MSE}_{out}$ ). Each time a part of observations is used to estimate all unknown parameters and functions in the three models (5.1)-(5.3), then the next value of the dependent variable is forecasted using the estimated parameters. More precisely, letting  $j = 1, 2, \dots, 10$ , we use the observations  $\{(y_t, x_t) : 2 \leq t \leq 178 + 2j\}$  to estimate the unknown parameters and functions (with fixed  $\hat{k} = 3$  for model (5.2) and fixed  $\hat{k} = 5$  for model (5.3)) in the three models, then the next  $y_{179+2j}$  is forecasted by three models with estimated parameters,

$$\begin{aligned} \hat{y}_{179+2j}^1 &= \hat{\beta}_{j,\text{linear}}^\top x_{179+2j}, \quad \hat{y}_{179+2j}^2 = \hat{g}_{1,3}^j(\hat{\theta}_{j,1}^\top x_{179+2j}), \quad \text{and} \\ \hat{y}_{179+2j}^3 &= \hat{\beta}_j^\top x_{179+2j} + \hat{g}_{2,5}^j(\hat{\theta}_{j,2}^\top x_{179+2j}). \end{aligned}$$

The out-of-sample mean square errors are evaluated, for  $\ell = 1, 2, 3$ , by

$$\text{MSE}_{out}(\ell) = \frac{1}{10} \sum_{j=1}^{10} (y_t - \widehat{y}_{179+2j}^\ell)^2. \quad (5.8)$$

Both the in-sample and out-sample results are given in Table 4 below.

Table 4: The MSEs for the models (5.1)-(5.3)

	Model (5.1)	Model (5.2)			Model (5.3)		
$k =$		2	3	4	4	5	6
$\text{MSE}_{in}$	0.1666	0.1641	0.1544	0.7709	0.1018	0.0946	0.1460
$\text{MSE}_{out}$	0.2598	0.2733	0.2607	1.9060	0.1761	0.1232	0.2146

In addition, to verify the choice of GCV for both model (5.2) and (5.3), the out-of-sample mean square errors for model (5.2) with  $k = 2, 4$  respectively and for model (5.3) with  $k = 4, 6$  respectively are computed.

The results of the MSE's are reported in Table 4. Among the three models, the partially linear single-index model (5.3) performs much better than the other two, in the sense that both its in-sample and out-of-sample mean square errors are the minimum within the models. Particularly, model (5.3) outperforms model (5.2) over all choices of the truncation parameter which whether or not are chosen by GCV method. However, with  $\widehat{k} = 5$  model (5.3) has the best performance. Therefore, model (5.3) with  $\widehat{k} = 5$  is the most favorable one to describe the relationship of the processed consumptions, incomes, investments and real interest rate for the US customers from 1960 to 2009.

## 6 Conclusion

The estimation procedures for both single-index models and partially single-index models in the presence of nonstationarity and integrability have been proposed, and the asymptotic properties of the estimators have been studied. The dual convergence rates for the single-index models and the trio of convergence rates for partially linear single-index models highlight the findings of the paper. Other results, such as the limit of the normalised estimators, are also of interest. Moreover, Monte Carlo simulations have verified the key theoretical results. Furthermore, the empirical study has shown that the partially linear single-index model outperforms both the linear and the single-index models, and is the most suitable one for the aggregate US data on consumption, income, investment and interest rate.

To broaden the application of these two classes of models in practice, two related further topics are worthy to investigate. (i) We are very much interested in studying the two classes of models within the context of penal data; and (ii) models exhibiting some type of endogeneity are also our future research direction.

## Appendix A Lemmas

Six lemmas are given in this appendix. They are crucial for the proofs of the main results.

**Lemma A.1.** *Suppose that  $g(x)$  is differentiable on  $\mathbb{R}$  and  $g^{(m-\ell)}(x)x^\ell \in L^2(\mathbb{R})$  for  $\ell = 0, 1, \dots, m$  and  $m \geq 1$ . Then the following holds:*

- (1) *When  $m = 1$ , the orthogonal expansion  $g(x) = \sum_{i=0}^{\infty} c_i \mathcal{H}_i(x)$  with  $c_i = \int g(x) \mathcal{H}_i(x) dx$  converges uniformly and absolutely on  $\mathbb{R}$ ; moreover, when  $m \geq 4$ , we have  $g^{(p)}(x) = \sum_{i=0}^{\infty} c_i \mathcal{H}_i^{(p)}(x)$  where  $p = 1, 2, 3$ .*
- (2)  *$\gamma_k(x) = g(x) - g_k(x) = \sum_{i=k}^{\infty} c_i \mathcal{H}_i(x) = o(1)k^{-(m-1)/2-1/12}$  uniformly on  $\mathbb{R}$ ; moreover,  $|\dot{\gamma}_k(x)| = o(1)k^{-(m-2)/2-1/12}$  and  $|\ddot{\gamma}_k(x)| = o(1)k^{-(m-3)/2-1/12}$  uniformly on  $\mathbb{R}$ .*
- (3) *(i)  $\int \gamma_k^2(x) dx = o(1)k^{-m}$ ; (ii)  $\int \gamma_k^2(x) x^2 dx = o(1)k^{-(m-1)}$ ; (iii)  $\int \dot{\gamma}_k^2(x) x^2 dx = o(1)k^{-(m-2)}$ ; (iv)  $\int \ddot{\gamma}_k^2(x) x^4 dx = o(1)k^{-(m-4)}$ .*
- (4) *(i)  $\int \|Z_k(x)\|^2 dx = k$ ; (ii)  $\int x^2 \|Z_k(x)\|^2 dx = O(1)k^2$ ; (iii)  $\int \|\dot{Z}_k(x)\|^2 dx = O(1)k^2$ ; (iv)  $\int x^2 \|\dot{Z}_k(x)\|^2 dx = O(1)k^3$ ; (v)  $\int \|\ddot{Z}_k(x)\|^2 dx = O(1)k^3$ ; (vi)  $\int x^4 \|\ddot{Z}_k(x)\|^2 dx = O(1)k^5$ ; (vii)  $\int x^4 \|\ddot{\ddot{Z}}_k(x)\|^2 dx = O(1)k^6$ ; (viii)  $\int x^6 \|\ddot{\ddot{Z}}_k(x)\|^2 dx = O(1)k^7$ .*
- (5) *In addition, (i)  $\|Z_k(x)\|^2 \leq O(1)k$ ; (ii)  $\|\dot{Z}_k(x)\|^2 \leq O(1)k^2$ ; (iii)  $\|\ddot{Z}_k(x)\|^2 \leq O(1)k^3$  uniformly on  $\mathbb{R}$ .*
- (6) *Furthermore, (i)  $\int \|Z_k(x)\| dx = O(1)k^{11/12}$ ; (ii)  $\int |x| \|Z_k(x)\| dx = O(1)k^{17/12}$ ; (iii)  $\int \|\dot{Z}_k(x)\| dx = O(1)k^{17/12}$ ; (iv)  $\int \|\dot{Z}_k(x)\| |x| dx = O(1)k^{23/12}$ ; (v)  $\int \|\dot{Z}_k(x)\| x^2 dx = O(1)k^{29/12}$ ; (vi)  $\int \|\ddot{Z}_k(x)\| |x| dx = O(1)k^{29/12}$ ; (vii)  $\int \|\ddot{Z}_k(x)\| x^2 dx = O(1)k^{35/12}$ ; (viii)  $\int \|\ddot{\ddot{Z}}_k(x)\| |x|^3 dx = O(1)k^{41/12}$ .*
- (7) *(i)  $\int |\gamma_k(x)| dx = o(1)k^{-m/2+11/12}$ ; (ii)  $\int |x \gamma_k(x)| dx = o(1)k^{-m/2+17/12}$ ; (iii)  $\int |\dot{\gamma}_k(x)| dx = o(1)k^{-m/2+17/12}$ ; (iv)  $\int |x \dot{\gamma}_k(x)| dx = o(1)k^{-m/2+23/12}$ ; (v)  $\int |\dot{\gamma}_k(x)| x^2 dx = o(1)k^{-m/2+29/12}$ ; (vi)  $\int |\ddot{\gamma}_k(x)| x^2 dx = o(1)k^{-m/2+29/12}$ ; (vii)  $\int |\ddot{\gamma}_k(x)| x^2 dx = o(1)k^{-m/2+35/12}$ ; (viii)  $\int |\ddot{\ddot{\gamma}}_k(x)| |x|^3 dx = o(1)k^{-m/2+41/12}$ .*

*Proof.* Let  $\varphi(x) = \exp(-x^2)$  and  $b_i^2 = \sqrt{\pi} 2^i i!$ . For  $i$  large, integration by parts gives

$$c_i(g) = \int g(x) \mathcal{H}_i(x) dx = \frac{1}{b_i} \int g(x) H_i(x) e^{-x^2/2} dx$$



$$\begin{aligned}
&= (-1)^i \frac{1}{b_i} \int g(x) \varphi^{(i)}(x) e^{x^2/2} dx = (-1)^i \frac{1}{b_i} \int g(x) e^{x^2/2} d\varphi^{(i-1)}(x) \\
&= (-1)^i \frac{1}{b_i} g(x) e^{x^2/2} \varphi^{(i-1)}(x) \Big|_{-\infty}^{\infty} - (-1)^i \frac{1}{b_i} \int \varphi^{(i-1)}(x) [g(x) e^{x^2/2}]' dx \\
&= (-1)^{i-1} \frac{1}{b_i} \int \varphi^{(i-1)}(x) [g(x) e^{x^2/2}]' dx \\
&= \frac{1}{b_i} \int [g(x) e^{x^2/2}]' H_{i-1}(x) \varphi(x) dx = \frac{b_{i-1}}{b_i} \int [g(x) e^{x^2/2}]' e^{-x^2/2} \mathcal{H}_{i-1}(x) dx \\
&= \frac{1}{\sqrt{2i}} c_{i-1}(\tilde{g}_1).
\end{aligned}$$

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

Clearly, for  $m > 1$ , we can iterate the procedure to derive that

$$c_i = \frac{1}{\sqrt{2^m} \sqrt{i(i-1)\cdots(i-m+1)}} c_{i-m}(\tilde{g}_m) = O(1) i^{-m/2} c_{i-m}(\tilde{g}_m) \quad (\text{A.1})$$

for  $i > m$ .

The following formulae are also necessary for the development.

$$\dot{\mathcal{H}}_0(x) = -\frac{1}{\sqrt{2}} \mathcal{H}_1(x), \quad \dot{\mathcal{H}}_i(x) = \frac{1}{\sqrt{2}} (\sqrt{i} \mathcal{H}_{i-1}(x) - \sqrt{i+1} \mathcal{H}_{i+1}(x)), \quad (\text{A.2})$$

and by recursion relation for Hermite function

$$x \mathcal{H}_0(x) = \frac{1}{\sqrt{2}} \mathcal{H}_1(x), \quad x \mathcal{H}_i(x) = \frac{1}{\sqrt{2}} (\sqrt{i} \mathcal{H}_{i-1}(x) + \sqrt{i+1} \mathcal{H}_{i+1}(x)). \quad (\text{A.3})$$

Moreover, simple algebra gives

$$\begin{aligned}
\ddot{\mathcal{H}}_0(x) &= -\frac{1}{2} (\mathcal{H}_0(x) - \sqrt{2} \mathcal{H}_1(x)), \quad \ddot{\mathcal{H}}_1(x) = \frac{1}{2} (-3\mathcal{H}_1(x) + \sqrt{6} \mathcal{H}_3(x)) \\
\ddot{\mathcal{H}}_i(x) &= \frac{1}{2} (\sqrt{i(i-1)} \mathcal{H}_{i-2}(x) - (2i+1) \mathcal{H}_i(x) + \sqrt{(i+1)(i+2)} \mathcal{H}_{i+2}(x)),
\end{aligned} \quad (\text{A.4})$$

and we may have similar formulae for  $\ddot{\mathcal{H}}_i(x)$ .

In addition, by Christoffel-Darboux formula we have

$$\begin{aligned}
\|Z_k(x)\|^2 &= e^{-x^2} \sum_{i=0}^{k-1} \frac{1}{b_i^2} H_i^2(x) = e^{-x^2} \frac{1}{2b_{k-1}^2} (H_{k-1}(x)H'_k(x) - H'_{k-1}(x)H_k(x)) \\
&= e^{-x^2} \frac{1}{2b_{k-1}^2} (2kH_{k-1}^2(x) - 2(k-1)H_{k-2}(x)H_k(x)) \\
&= k\mathcal{H}_{k-1}^2(x) - \sqrt{(k-1)k} \mathcal{H}_{k-2}(x)\mathcal{H}_k(x),
\end{aligned} \quad (\text{A.5})$$

which implies

$$\|Z_k(x)\| \leq \sqrt{k} |\mathcal{H}_{k-1}(x)| + \sqrt[4]{(k-1)k} \sqrt{|\mathcal{H}_{k-2}(x)\mathcal{H}_k(x)|}. \quad (\text{A.6})$$

(1) Notice that  $\max_{x \in \mathbb{R}} |\mathcal{H}_i^2(x)|$  is uniformly bounded in  $i$ . We even have  $\max_{x \in \mathbb{R}} |\mathcal{H}_i^2(x)| = O(1)i^{-1/6}$  for  $i \geq 1$ . See Szego (1975, p. 242). It follows from Cauchy-Schwarz inequality that

$$\begin{aligned} |g(x) - g_k(x)| &= \left| \sum_{i=k}^{\infty} c_i \mathcal{H}_i(x) \right| \leq \sum_{i=k}^{\infty} \frac{1}{\sqrt{2i}} |c_{i-1}(\tilde{g}_1) \mathcal{H}_i(x)| \leq o(1) \left( \sum_{i=k}^{\infty} \frac{1}{2i} \mathcal{H}_i^2(x) \right)^{1/2} \\ &\leq o(1) \left( \sum_{i=k}^{\infty} \frac{1}{2i^{7/6}} \right)^{1/2} = o(1)k^{-1/12}, \end{aligned}$$

due to the Parseval equality of  $\sum_{i \geq 1} |c_{i-1}(\tilde{g}_1)|^2 = \|\tilde{g}_1\|^2$ . Thus,  $g_k(x)$  converges to  $g(x)$  uniformly and absolutely on  $\mathbb{R}$ .

When  $m \geq 4$ , in view of (A.1), (A.2) and (A.4), similar to the above derivation,  $\sum_{i=0}^{\infty} c_i \mathcal{H}_i^{(p)}(x)$  converges uniformly, so the sum is  $g^{(p)}(x)$  with  $p = 1, 2, 3$ .

(2) Using the relation (A.1) for the coefficients, we obtain that

$$\begin{aligned} |\gamma_k(x)| &= \left| \sum_{i=k}^{\infty} c_i \mathcal{H}_i(x) \right| = O(1) \left| \sum_{i=k}^{\infty} i^{-m/2} c_{i-m}(\tilde{g}_m) \mathcal{H}_i(x) \right| \leq o(1) \left( \sum_{i=k}^{\infty} i^{-m} \mathcal{H}_i^2(x) \right)^{1/2} \\ &= o(1) \left( \sum_{i=k}^{\infty} i^{-m-1/6} \right)^{1/2} = o(1)k^{-(m-1)/2-1/12}, \end{aligned}$$

hence,  $|\gamma_k(x)| = o(1)k^{-(m-1)/2-1/12}$  uniformly in  $x$ . Moreover, in view of (A.2) and (A.4), a similar derivation gives that  $|\dot{\gamma}_k(x)| = o(1)k^{-(m-2)/2-1/12}$  and  $|\ddot{\gamma}_k(x)| = o(1)k^{-(m-3)/2-1/12}$ .

(3) We prove (i) first. By orthogonality,

$$\int \gamma_k^2(x) dx = \sum_{i=k}^{\infty} c_i^2 = O(1) \sum_{i=k}^{\infty} i^{-m} c_{i-m}^2(\tilde{g}_m) \leq O(1)k^{-m} \sum_{i=k}^{\infty} c_{i-m}^2(\tilde{g}_m) = o(1)k^{-m}.$$

To prove (ii), note that, it follows from the recursion relation (A.3) and the orthogonality that  $\int x^2 \mathcal{H}_i^2(x) dx = i + 1/2$  and for  $i \neq j$ ,

$$\begin{aligned} \int x^2 \mathcal{H}_i(x) \mathcal{H}_j(x) dx &= \frac{1}{2} \int (\sqrt{i} \mathcal{H}_{i-1}(x) + \sqrt{i+1} \mathcal{H}_{i+1}(x)) (\sqrt{j} \mathcal{H}_{j-1}(x) + \sqrt{j+1} \mathcal{H}_{j+1}(x)) dx \\ &= \frac{1}{2} \int \sqrt{i(j+1)} \mathcal{H}_{i-1}(x) \mathcal{H}_{j+1}(x) dx + \frac{1}{2} \int \sqrt{j(i+1)} \mathcal{H}_{i+1}(x) \mathcal{H}_{j-1}(x) dx \end{aligned}$$

which assumes  $\frac{1}{2} \sqrt{i(i-1)}$  when  $j = i-2$ ,  $\frac{1}{2} \sqrt{(i+1)(i+2)}$  when  $j = i+2$ , and 0 otherwise. Thus,

$$\begin{aligned} \int \gamma_k^2(x) x^2 dx &= \sum_{i=k}^{\infty} \sum_{j=k}^{\infty} c_i c_j \int x^2 \mathcal{H}_i(x) \mathcal{H}_j(x) dx = \sum_{i=k}^{\infty} c_i^2 \int x^2 \mathcal{H}_i^2(x) dx \\ &+ \sum_{i=k}^{\infty} \sum_{j=k, \neq i}^{\infty} c_i c_j \int x^2 \mathcal{H}_i \mathcal{H}_j(x) dx = \sum_{i=k}^{\infty} c_i^2 (i + 1/2) + \frac{1}{2} \sum_{i=k}^{\infty} c_i c_{i-2} \sqrt{i(i-1)} \end{aligned}$$

$$+ \frac{1}{2} \sum_{i=k}^{\infty} c_i c_{i+2} \sqrt{(i+1)(i+2)} = O(1) \sum_{i=k}^{\infty} c_i^2 i = O(1) \sum_{i=k}^{\infty} i i^{-m} c_{i-m}^2(\tilde{g}_m) = o(1) k^{-(m-1)},$$

using the coefficient relation (A.1).

To prove (iii), it follows from (A.2) and (ii) that

$$\begin{aligned} \int \dot{\gamma}^2(x) x^2 dx &= \int \left( \sum_{i=k}^{\infty} c_i \dot{\mathcal{H}}_i(x) \right)^2 x^2 dx = \frac{1}{2} \int \left( \sum_{i=k}^{\infty} c_i (\sqrt{i} \mathcal{H}_{i-1}(x) - \sqrt{i+1} \mathcal{H}_{i+1}(x)) \right)^2 x^2 dx \\ &\leq \int \left( \sum_{i=k}^{\infty} \sqrt{i} c_i \mathcal{H}_{i-1}(x) \right)^2 x^2 dx + \int \left( \sum_{i=k}^{\infty} \sqrt{i+1} c_i \mathcal{H}_{i+1}(x) \right)^2 x^2 dx = o(1) k^{-(m-2)}. \end{aligned}$$

It is quite clear that (iv) follows in a similar fashion as (iii).

(4) The assertion (i) is obvious by orthogonality. As shown before,  $\int x^2 \mathcal{H}_i^2(x) dx = i+1/2$ . (ii) then follows straightforward. To prove (iii), noting (A.2)–(A.4), it is readily seen that  $\int \dot{\mathcal{H}}_i^2(x) dx = O(1)i$ ,  $\int x^2 \dot{\mathcal{H}}_i^2(x) dx = O(1)i^2$ ,  $\int \ddot{\mathcal{H}}_i^2(x) dx = O(1)i^2$  and  $\int x^4 \ddot{\mathcal{H}}_i^2(x) dx = O(1)i^4$ . Therefore,  $\int \|\dot{Z}_k(x)\|^2 dx = O(1)k^2$ ,  $\int x^2 \|\dot{Z}_k(x)\|^2 dx = O(1)k^3$ ,  $\int \|\ddot{Z}_k(x)\|^2 dx = O(1)k^3$  and  $\int x^4 \|\ddot{Z}_k(x)\|^2 dx = O(1)k^5$ . The last two assertions follow similarly.

(5). All the assertions follows immediately from the facts in (A.2) and (A.4) and the boundedness of the orthonormal basis.

(6). (i) Because of (A.6) and  $\int |\mathcal{H}_{k-1}(x)| dx = O(1)k^{5/12}$  by (c) of Lemma B.7 in Cai and Gao (2013), the assertion holds by Cauchy-Schwarz inequality.

To prove (ii), notice again by (A.6) that  $|x| \|Z_k(x)\| \leq \sqrt{k} |x \mathcal{H}_{k-1}(x)| + \sqrt[4]{(k-1)k} \times \sqrt{x^2 |\mathcal{H}_{k-2}(x) \mathcal{H}_k(x)|}$ , and we only need to tackle the first term since the second one will have the same order by Cauchy-Schwarz inequality. From (A.3),  $\sqrt{k} |x \mathcal{H}_{k-1}(x)| = \sqrt{k} |\sqrt{k-1} \mathcal{H}_{k-2}(x) + \sqrt{k} \mathcal{H}_k(x)| \leq \sqrt{k(k-1)} |\mathcal{H}_{k-2}(x)| + k |\mathcal{H}_k(x)|$ , which gives the result by noting  $\int |\mathcal{H}_k(x)| dx = O(1)k^{5/12}$  again.

In view of (A.3) and (A.5), all the rest assertions can be derived in the same fashion.

(7) To prove (i), by (A.1) and once again  $\int |\mathcal{H}_k(x)| dx = O(1)k^{5/12}$ ,

$$\begin{aligned} \int |\gamma_k(x)| dx &\leq \sum_{i=k}^{\infty} |c_i| \int |\mathcal{H}_i(x)| dx = O(1) \sum_{i=k}^{\infty} i^{-m/2} c_i(\tilde{g}) i^{5/12} \\ &\leq o(1) \left( \sum_{i=k}^{\infty} i^{-m+5/6} \right)^{1/2} = o(1) k^{-m/2+11/12}. \end{aligned}$$

In view of (A.3) and (A.5), all the rest assertions can be derived in the same way.  $\square$

**Lemma A.2.** *Under Assumptions A and B, we have as  $n \rightarrow \infty$*

$$\left\| \frac{1}{\sqrt{n}} Z^\top Z - L_1(1, 0) I_k \right\| = o_P(1)$$

*in a richer probability space.*

*Proof.* This is exactly Theorem 4 in Cai and Gao (2013).  $\square$

**Lemma A.3.** *Let Assumption A hold. Assume that functions  $g$  and  $f$  defined on  $\mathbb{R}$  are bounded and integrable. Then,*

$$\left( \begin{array}{c} \frac{1}{\sqrt[4]{n}} \sum_{t=1}^n g(x_{1t})e_t \\ \frac{1}{n^{3/4}} \sum_{t=1}^n f(x_{1t})x_{2t}e_t \end{array} \right) \rightarrow_D M^{1/2}W(1), \quad (\text{A.7})$$

where  $W(1)$  is an  $d$ -dimensional standard normal vector independent of  $V$  and  $M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$  is a symmetric block matrix with

$$\begin{aligned} m_{11} &= L_1(1, 0) \int g^2(s)ds, & m_{12} &= \int_0^1 V_2^\top(r) dL_1(r, 0) \int g(s)f(s)ds, \\ m_{21} &= m_{12}^\top, & m_{22} &= \int_0^1 V_2(r)V_2^\top(r) dL_1(r, 0) \int f^2(s)ds, \end{aligned}$$

in which Brownian motion processes  $V_1(r)$  and  $V_2(r)$  are given in (3.1), and  $L_1(r, s)$  stands for the local time process of the Brownian motion  $V_1(\cdot)$  at  $s$  over  $[0, r]$ .

*Proof.* The assertion (A.7) is a simplified version of Lemma 3 in Park and Phillips (2000, p. 1256).  $\square$

**Lemma A.4.** *The following assertions hold:*

- (1)  $\frac{1}{\sqrt{t}}(x_{1t}, x_{2t}^\top)$  has a joint probability density  $\psi_t(x, w^\top)$ ; and given  $\mathcal{F}_s$  (defined in Assumption A),  $\frac{1}{\sqrt{t-s}}(x_{1t} - x_{1s}, x_{2t}^\top - x_{2s}^\top)$  has a joint density  $\psi_{ts}(x, w^\top)$  where  $t > s + 1$ . Meanwhile, these functions are bounded uniformly in  $(x, w)$  as well as  $t$  and  $(t, s)$ , respectively.
- (2) For large  $t$  and  $t-s$ , we have  $\psi_t(x, w) = \phi(w)f_t(x)(1+o(1))$  and  $\psi_{ts}(x, w) = \phi(w)f_{ts}(x)(1+o(1))$  where  $\phi(w)$  is the density of an  $(d-1)$ -dimensional normal distribution,  $f_t(x)$  is the marginal density of  $\frac{1}{\sqrt{t}}x_{1t}$  and  $f_{ts}(x)$  is the marginal density of  $\frac{1}{\sqrt{t-s}}(x_{1t} - x_{1s})$ .

*Proof.* (1) By virtue of Assumption A, it follows from Wang et al. (2003) that  $E[x_t x_t^\top] = t\rho^2 I_d$ . By the orthogonality of  $Q$ ,  $x_{1t}$  and  $x_{2t}$  are uncorrelated, viz.,  $x_{1t}$  is uncorrelated with every entry of  $x_{2t}$ . In addition, since  $\Delta x_{1t} = \theta_0^\top \Delta x_t = \theta_0^\top v_t$  and  $\Delta x_{2t} = Q_2^\top \Delta x_t = Q_2^\top v_t$ ,  $x_{1t}$  and  $x_{2t}$  are still integrated processes. Thus, similar to the proof of Corollary 2.2 in Wang and Phillips (2009a, p. 729), we may show that  $\frac{1}{\sqrt{t}}(x_{1t}, x_{2t}^\top)$  has a joint probability density  $\psi_t(x, w^\top)$ ; and given  $\mathcal{F}_s$ ,  $\frac{1}{\sqrt{t-s}}(x_{1t} - x_{1s}, x_{2t}^\top - x_{2s}^\top)$  has a joint density  $\psi_{ts}(x, w^\top)$ . Meanwhile, these functions are bounded uniformly in  $(x, w)$  as well as  $t$  and  $(t, s)$ .

(2) Because  $\psi_t(x, w) = \nu_t(w|x)f_t(x)$  where  $\nu_t(w|x)$  is the conditional density of  $\frac{1}{\sqrt{t}}x_{2t}$  given  $\frac{1}{\sqrt{t}}x_{1t} = x$ , and  $x_{1t}$  and  $x_{2t}$  are asymptotically independent (see Remark 1 of Park and Phillips (2000, p. 1257)), we have  $\psi_t(x, w) = \nu_t(w)f_t(x)(1+o(1))$  for large  $t$ , where  $\nu_t(w)$  is the

marginal density of  $\frac{1}{\sqrt{t}}x_{2t}$ . Again, by the proof of Corollary 2.2 in Wang and Phillips (2009a, p. 729),  $\sup_w |\nu_t(w) - \phi(w)| = o(1)$ . Thus, for large  $t$ , we have  $\psi_t(x, w) = \phi(w)f_t(x)(1 + o(1))$ , and similarly, for large  $t - s$ ,  $\psi_{ts}(x, w) = \phi(w)f_{ts}(x)(1 + o(1))$ .  $\square$

**Lemma A.5.** *For any function  $f(x)$  defined on  $\mathbb{R}$  such that  $x^3 f(x)$  is bounded and integrable, define  $\Pi_n^{i,j} = \sum_{t=1}^n f(x_{1t})x_{it}x_{jt}^\top[(\alpha^1 - 1)x_{1t} + (\alpha^2)^\top x_{2t}]$  where  $i, j \in \{1, 2\}$  and  $|\alpha^1 - 1| < n^{-1/4+\delta}$  and  $\|\alpha^2\| < n^{-3/4+\delta}$  for some small  $\delta > 0$ . Then,*

$$\Pi_n^{i,j} = \begin{cases} O_P(1)n^{1/4+\delta} \int |x^3 f(x)| dx & \text{when } i = j = 1, \\ O_P(1)n^{3/4+\delta} \int |x^2 f(x)| dx & \text{when } i = 1, j = 2, \\ O_P(1)n^{5/4+\delta} \int |x f(x)| dx & \text{when } i = j = 2. \end{cases}$$

*Proof.* Notice that due to the boundedness of  $x^3 f(x)$  and  $\sup_{1 \leq t \leq n} \|x_{2t}\|/\sqrt{n} = O_P(1)$ , in the summation of  $\Pi_n^{i,j}$  we only need consider the sufficient large  $t$ , say  $t \geq a_n$  where  $a_n \rightarrow \infty$  and  $a_n/n \rightarrow 0$  with  $n$  increasing. Notice also that  $\|\Pi_n^{i,j}\| \leq \sum_{t=1}^n |f(x_{1t})| \|x_{it}\| \|x_{jt}\| [|\alpha^1 - 1| \|x_{1t}\| + \|\alpha^2\| \|x_{2t}\|] \leq n^{-1/4+\delta} \sum_{t=1}^n |f(x_{1t})| \|x_{it}\| \|x_{jt}\| \|x_{1t}\| + n^{-3/4+\delta} \sum_{t=1}^n |f(x_{1t})| \|x_{it}\| \|x_{jt}\| \|x_{2t}\|$ . Using the joint densities  $\psi_t(x, w)$  of  $(\frac{1}{\sqrt{t}}x_{1t}, \frac{1}{\sqrt{t}}x_{2t})$  satisfying  $\psi_t(x, w) = f_t(x)\phi(w)(1 + o(1))$  by Lemma A.4, we have for  $i = j = 1$ ,

$$\begin{aligned} E|\Pi_n^{1,1}| &\leq n^{-1/4+\delta} E \sum_{t=a_n}^n |f(x_{1t})| \|x_{1t}\|^3 + n^{-3/4+\delta} E \sum_{t=a_n}^n |f(x_{1t})| \|x_{1t}\|^2 \|x_{2t}\| \\ &\leq n^{-1/4+\delta} \sum_{t=a_n}^n \int |f(\sqrt{t}x)| |\sqrt{t}x|^3 f_t(x) dx \\ &\quad + n^{-3/4+\delta} \sum_{t=a_n}^n \int \int |f(\sqrt{t}x)| |\sqrt{t}x|^2 \|\sqrt{t}w\| \psi_t(x, w) dx dw \\ &= n^{-1/4+\delta} \sum_{t=a_n}^n \frac{1}{\sqrt{t}} \int |x^3 f(x)| f_t\left(\frac{1}{\sqrt{t}}x\right) dx \\ &\quad + n^{-3/4+\delta} \sum_{t=a_n}^n \int \int |f(\sqrt{t}x)| |\sqrt{t}x|^2 \|\sqrt{t}w\| f_t(x)\phi(w)(1 + o(1)) dx dw \\ &\leq O(1)n^{-1/4+\delta} \sqrt{n} \int |x^3 f(x)| dx + O(1)n^{-3/4+\delta} n \int |x^2 f(x)| dx \\ &= O(1)n^{1/4+\delta} \int |x^3 f(x)| dx. \end{aligned}$$

Similarly, for  $i = 1, j = 2$ ,

$$\begin{aligned} E\|\Pi_n^{1,2}\| &\leq n^{-1/4+\delta} E \sum_{t=a_n}^n |f(x_{1t})| \|x_{1t}\|^2 \|x_{2t}\| + n^{-3/4+\delta} E \sum_{t=a_n}^n |f(x_{1t})| \|x_{1t}\| \|x_{2t}\|^2 \\ &\leq n^{-1/4+\delta} \sum_{t=a_n}^n \int \int |f(\sqrt{t}x)| |\sqrt{t}x|^2 \|\sqrt{t}w\| \psi_t(x, w) dx dw \end{aligned}$$

$$\begin{aligned}
& + n^{-3/4+\delta} \sum_{t=a_n}^n \int \int |f(\sqrt{t}x)| |\sqrt{t}x| \|\sqrt{t}w\|^2 \psi_t(x, w) dx dw \\
& = n^{-1/4+\delta} \sum_{t=a_n}^n \int \int |f(\sqrt{t}x)| |\sqrt{t}x|^2 \|\sqrt{t}w\| f_t(x) \phi(w) (1 + o(1)) dx dw \\
& \quad + n^{-3/4+\delta} \sum_{t=a_n}^n \int \int |f(\sqrt{t}x)| |\sqrt{t}x| \|\sqrt{t}w\|^2 f_t(x) \phi(w) (1 + o(1)) dx dw \\
& \leq n^{-1/4+\delta} \sum_{t=a_n}^n \int |x^2 f(x)| f_t \left( \frac{1}{\sqrt{t}} x \right) dx \int \|w\| \phi(w) dw \\
& \quad + n^{-3/4+\delta} \sum_{t=a_n}^n \sqrt{t} \int |x f(\sqrt{t}x)| f_t \left( \frac{1}{\sqrt{t}} x \right) dx \int \|w\|^2 \phi(w) dw \\
& \leq O(1) n^{-1/4+\delta} n \int |x^2 f(x)| dx + O(1) n^{-3/4+\delta} \sqrt{n}^3 \int |x f(x)| dx \\
& = O(1) n^{3/4+\delta} \int |x^2 f(x)| dx,
\end{aligned}$$

and for  $i = j = 2$ ,

$$\begin{aligned}
E \|\Pi_n^{2,2}\| & \leq n^{-1/4+\delta} E \sum_{t=a_n}^n |f(x_{1t})| |x_{1t}| \|x_{2t}\|^2 + n^{-3/4+\delta} E \sum_{t=a_n}^n |f(x_{1t})| \|x_{2t}\|^3 \\
& \leq n^{-1/4+\delta} \sum_{t=a_n}^n \int \int |f(\sqrt{t}x)| |\sqrt{t}x| \|\sqrt{t}w\|^2 \psi_t(x, w) dx dw \\
& \quad + n^{-3/4+\delta} \sum_{t=a_n}^n \int \int |f(\sqrt{t}x)| \|\sqrt{t}w\|^3 \psi_t(x, w) dx dw \\
& = n^{-1/4+\delta} \sum_{t=a_n}^n \int \int |f(\sqrt{t}x)| |\sqrt{t}x| \|\sqrt{t}w\|^2 f_t(x) \phi(w) (1 + o(1)) dx dw \\
& \quad + n^{-3/4+\delta} \sum_{t=a_n}^n \int \int |f(\sqrt{t}x)| \|\sqrt{t}w\|^3 f_t(x) \phi(w) (1 + o(1)) dx dw \\
& = n^{-1/4+\delta} \sum_{t=a_n}^n \sqrt{t} \int |x f(x)| f_t \left( \frac{1}{\sqrt{t}} x \right) dx \int \|w\| \phi(w) dw \\
& \quad + n^{-3/4+\delta} \sum_{t=a_n}^n t \int |f(\sqrt{t}x)| f_t \left( \frac{1}{\sqrt{t}} x \right) dx \int \|w\|^3 \phi(w) dw \\
& \leq O(1) n^{-1/4+\delta} \sqrt{n}^3 \int |x f(x)| dx + O(1) n^{-3/4+\delta} n^2 \int |f(x)| dx \\
& = O(1) n^{5/4+\delta} \int |x f(x)| dx.
\end{aligned}$$

□

**Lemma A.6.** Define for any  $x, y \in \mathbb{R}$ ,

$$\mathcal{U}_k(x, y) = Z_k(x)^\top Z_k(y) = \sum_{i=0}^{k-1} \mathcal{H}_i(x) \mathcal{H}_i(y). \tag{A.8}$$

Then  $\lim_{k \rightarrow \infty} \iint_{x \neq y} |\mathcal{U}_k(x, y)| dx dy = 0$ .

*Proof.* Mehler's formula (see Szego, 1975, p. 380) gives, for  $|r| < 1$ ,

$$\sum_{i=0}^{\infty} \mathcal{H}_i(x) \mathcal{H}_i(y) r^i = \frac{1}{\sqrt{1-r^2}} \exp\left(\frac{2r}{1+r}xy - \frac{r^2}{1-r^2}(x-y)^2\right) \quad (\text{A.9})$$

and it follows from Example 2 of Gel'fand and Shilov (1964, P. 36) that

$$\lim_{r \rightarrow 1^-} \frac{1}{\sqrt{1-r^2}} \exp\left(\frac{2r}{1+r}xy - \frac{r^2}{1-r^2}(x-y)^2\right) = \delta(x-y)$$

where  $\delta(\cdot)$  is the Dirac delta function rigorously defined by a measure  $\delta(A) = 1$  if  $0 \in A$  and 0 otherwise, or in the generalized function terminology,  $\delta(x) = 0$  for all  $x \neq 0$  but  $\delta(0) = \infty$ .

Observe that

$$\begin{aligned} & \lim_{k \rightarrow \infty} \iint_{x \neq y} |\mathcal{U}_k(x, y)| dx dy = \lim_{k \rightarrow \infty} \iint_{x \neq y} \left| \sum_{i=0}^{k-1} \mathcal{H}_i(x) \mathcal{H}_i(y) \right| dx dy \\ &= \lim_{k \rightarrow \infty} \iint_{x \neq y} \lim_{r \rightarrow 1^-} \left| \sum_{i=0}^{k-1} \mathcal{H}_i(x) \mathcal{H}_i(y) r^i \right| dx dy = \iint_{x \neq y} \lim_{k \rightarrow \infty} \lim_{r \rightarrow 1^-} \left| \sum_{i=0}^{k-1} \mathcal{H}_i(x) \mathcal{H}_i(y) r^i \right| dx dy \\ &= \iint_{x \neq y} \lim_{r \rightarrow 1^-} \lim_{k \rightarrow \infty} \left| \sum_{i=0}^{k-1} \mathcal{H}_i(x) \mathcal{H}_i(y) r^i \right| dx dy = \iint_{x \neq y} \lim_{r \rightarrow 1^-} \left| \sum_{i=0}^{\infty} \mathcal{H}_i(x) \mathcal{H}_i(y) r^i \right| dx dy \\ &= \iint_{x \neq y} \lim_{r \rightarrow 1^-} \frac{1}{\sqrt{1-r^2}} \exp\left(\frac{2r}{1+r}xy - \frac{r^2}{1-r^2}(x-y)^2\right) dx dy \\ &= \iint_{x \neq y} \delta(x-y) dx dy = 0. \end{aligned}$$

Taking limit under the integration for  $k \rightarrow \infty$  is due to the dominated convergence theorem for the series is dominated by an integrable function  $\sum_{i=0}^{\infty} |\mathcal{H}_i(y) \mathcal{H}_i(x)| r^i$  with  $0 < r < 1$ . Exchange of two limits for  $k$  and  $r$  is because the series is convergent uniformly in  $r$  in any bounded interval. In fact, noting that Hermite polynomials have an integral expression  $H_j(x) = (-1)^j \frac{1}{2^j \sqrt{\pi}} e^{x^2} \int (iu)^j e^{iux - u^2/4} du$  where  $i$  is imaginary unit,

$$\begin{aligned} & \sum_{j=0}^{k-1} \mathcal{H}_j(y) \mathcal{H}_j(x) r^j = \sum_{j=0}^{k-1} \frac{r^j}{2^j j! \sqrt{\pi}} H_j(y) H_j(x) e^{-(x^2+y^2)/2} \\ &= \frac{e^{(x^2+y^2)/2}}{4\pi \sqrt{\pi}} \iint \left( \sum_{j=0}^{k-1} \frac{1}{2^j j!} (-ruv)^j \right) e^{i(ux+vy) - u^2/4 - v^2/4} dudv \\ &\rightarrow \frac{e^{(x^2+y^2)/2}}{4\pi \sqrt{\pi}} \iint e^{-ruv/2} e^{i(ux+vy) - u^2/4 - v^2/4} dudv \end{aligned}$$

where the sum converges to  $e^{-ruv/2}$  uniformly in  $r$  in any fixed finite interval, giving the Mehler's formula by calculating the last double integral.  $\square$

## Appendix B Proofs of the main results

*Proof of Theorem 2.1.* (1) Note that

$$\begin{aligned}\|\widehat{c}(\theta) - c\|^2 &= \|(Z^\top Z)^{-1}Z^\top(\gamma + e)\|^2 = (\gamma + e)^\top Z(Z^\top Z)^{-2}Z^\top(\gamma + e) \\ &= e^\top Z(Z^\top Z)^{-2}Z^\top e + \gamma^\top Z(Z^\top Z)^{-2}Z^\top \gamma + 2\gamma^\top Z(Z^\top Z)^{-2}Z^\top e.\end{aligned}$$

Note also that  $e^\top Z(Z^\top Z)^{-2}Z^\top e = \frac{1}{nL_B^2(1,0)}e^\top ZZ^\top e(1 + o_P(1))$  by Lemma A.2. Using martingale difference structure and  $\theta^\top x_t = \theta^\top QQ^\top x_t = \alpha^1 x_{1t} + (\alpha^2)^\top x_{2t}$  as well as the joint density for  $\frac{1}{\sqrt{t}}(x_{1t}, x_{2t})$  for  $t$  large by Lemma A.4 we have

$$\begin{aligned}E[e^\top ZZ^\top e] &= E \sum_{i=0}^{k-1} \left( \sum_{t=1}^n \mathcal{H}_i(\theta^\top x_t) e_t \right)^2 = \sigma_e^2 E \sum_{i=0}^{k-1} \sum_{t=1}^n \mathcal{H}_i^2(\theta^\top x_t) = \sigma_e^2 \sum_{t=1}^n E \|Z_k(\theta^\top x_t)\|^2 \\ &= \sigma_e^2 \sum_{t=1}^{a_n} E \|Z_k(\alpha^1 x_{1t} + (\alpha^2)^\top x_{2t})\|^2 + \sigma_e^2 \sum_{t=a_n+1}^n E \|Z_k(\alpha^1 x_{1t} + (\alpha^2)^\top x_{2t})\|^2 \\ &= O(1)ka_n + \sigma_e^2 \sum_{t=a_n+1}^n \iint \|Z_k(\sqrt{t}\alpha^1 x + \sqrt{t}(\alpha^2)^\top w)\|^2 f_t(x)\phi(w) dx dw (1 + o(1)) \\ &= O(1)ka_n + O(1) \sum_{t=a_n+1}^n \frac{1}{\sqrt{t}} \iint \|Z_k(x)\|^2 f_t \left( \frac{x - \sqrt{t}(\alpha^2)^\top w}{\alpha^1 \sqrt{t}} \right) \phi(w) dx dw (1 + o(1)) \\ &\leq O(1)ka_n + O(1) \sum_{t=a_n+1}^n \frac{1}{\sqrt{t}} \iint \|Z_k(x)\|^2 \phi(w) dx dw (1 + o(1)) \\ &= O(1)ka_n + O(1)\sqrt{n} \int \|Z_k(x)\|^2 dx = O(1)ka_n + O(1)\sqrt{nk} = O(1)\sqrt{nk}\end{aligned}$$

provided  $a_n/\sqrt{n} \rightarrow 0$  and  $a_n \rightarrow \infty$  as  $n \rightarrow \infty$ .

Meanwhile,  $\gamma^\top Z(Z^\top Z)^{-2}Z^\top \gamma = \frac{1}{nL_B^2(1,0)}\gamma^\top ZZ^\top \gamma(1 + o_P(1))$  again by Lemma A.2 and

$$\begin{aligned}\frac{1}{\sqrt{nk}}E[\gamma^\top ZZ^\top \gamma] &= \frac{1}{\sqrt{nk}}E \sum_{i=0}^{k-1} \left( \sum_{t=1}^n \mathcal{H}_i(\theta^\top x_t) \gamma_k(\theta^\top x_t) \right)^2 \\ &= \frac{1}{\sqrt{nk}}E \sum_{i=0}^{k-1} \sum_{t=1}^n \mathcal{H}_i^2(\theta^\top x_t) \gamma_k^2(\theta^\top x_t) + 2\frac{1}{\sqrt{nk}}E \sum_{i=0}^{k-1} \sum_{t=2}^n \sum_{s=1}^{t-1} \mathcal{H}_i(\theta^\top x_t) \gamma_k(\theta^\top x_t) \mathcal{H}_i(\theta^\top x_s) \gamma_k(\theta^\top x_s) \\ &= \frac{1}{\sqrt{nk}}E \sum_{t=1}^n \|Z_k(\theta^\top x_t)\|^2 \gamma_k^2(\theta^\top x_t) + 2\frac{1}{\sqrt{nk}}E \sum_{t=2}^n \sum_{s=1}^{t-1} \mathcal{U}_k(\theta^\top x_t, \theta^\top x_s) \gamma_k(\theta^\top x_t) \gamma_k(\theta^\top x_s) \\ &= o(1)k^{-(m-1)-1/6} \frac{1}{\sqrt{nk}}E \sum_{t=1}^n \|Z_k(\theta^\top x_t)\|^2 + 2\frac{1}{\sqrt{nk}}E \sum_{t=2}^n \sum_{s=1}^{t-1} \mathcal{U}_k(\theta^\top x_t, \theta^\top x_s) \gamma_k(\theta^\top x_t) \gamma_k(\theta^\top x_s),\end{aligned}$$

by (2) of Lemma A.1, where  $\mathcal{U}_k(x, y)$  is defined by (A.8).

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

$$\frac{1}{\sqrt{nk}}E \sum_{t=1}^n \|Z_k(\theta^\top x_t)\|^2 = \frac{1}{\sqrt{nk}} \sum_{t=1}^n \iint \|Z_k(\sqrt{t}\alpha^1 x + \sqrt{t}(\alpha^2)^\top w)\|^2 f_t(x)\phi(w) dx dw (1 + o(1))$$



$$= \frac{1}{\sqrt{nk}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|Z_k(x)\|^2 f_t \left( \frac{x - \sqrt{t}(\alpha^2)^\top w}{\alpha^1 \sqrt{t}} \right) \phi(w) dx dw \leq C \frac{1}{k} \int \|Z_k(x)\|^2 dx = C.$$

Moreover, notice that  $\frac{1}{\sqrt{t-s}} \theta^\top (x_t - x_s) = \frac{1}{\sqrt{t-s}} \theta^\top Q(Q^\top x_t - Q^\top x_s) = \frac{1}{\sqrt{t-s}} (\alpha^1 (x_{1t} - x_{1s}) + (\alpha^2)^\top (x_{2t} - x_{2s}))$  and  $\frac{1}{\sqrt{t-s}} ((x_{1t} - x_{1s}), (x_{2t} - x_{2s}))$  has joint density  $\psi_{ts}(x, w) = f_{ts}(x) \phi(w) (1 + o(1))$  when  $t - s$  is large by Lemma A.4. Once again we ignore the argument on whether  $t - s$  is large or not with a bit abuse of the density  $\psi_{ts}(x, w) = f_{ts}(x) \phi(w) (1 + o(1))$ . This is because  $|\mathcal{U}_k(x, y)| \leq Ck$  and  $\gamma_k(x) = o(1)$  uniformly in  $(x, y)$  and  $x$ , respectively. Whence,

$$\begin{aligned} & \frac{1}{\sqrt{nk}} \left| E \sum_{t=2}^n \sum_{s=1}^{t-1} \mathcal{U}_k(\theta^\top x_t, \theta^\top x_s) \gamma_k(\theta^\top x_t) \gamma_k(\theta^\top x_s) \right| \\ & \leq \frac{1}{\sqrt{nk}} E \sum_{t=2}^n \sum_{s=1}^{t-1} |\mathcal{U}_k(\theta^\top x_t, \theta^\top x_s) \gamma_k(\theta^\top x_t) \gamma_k(\theta^\top x_s)| \\ & \leq \frac{1}{\sqrt{nk}} E \sum_{t=2}^n \sum_{s=1}^{t-1} \|Z_k(\theta^\top x_t)\| \|Z_k(\theta^\top x_s)\| |\gamma_k(\theta^\top x_t) \gamma_k(\theta^\top x_s)| \\ & = \frac{1}{\sqrt{nk}} \sum_{t=2}^n \sum_{s=1}^{t-1} \int \cdots \int \\ & \quad \times \|Z_k(\sqrt{t-s}(\alpha^1 x + (\alpha^2)^\top w_1) + \sqrt{s}(\alpha^1 y + (\alpha^2)^\top w_2))\| \|Z_k(\sqrt{s}(\alpha^1 y + (\alpha^2)^\top w_2))\| \\ & \quad \times |\gamma_k(\sqrt{t-s}(\alpha^1 x + (\alpha^2)^\top w_1) + \sqrt{s}(\alpha^1 y + (\alpha^2)^\top w_2)) \gamma_k(\sqrt{s}(\alpha^1 y + (\alpha^2)^\top w_2))| \\ & \quad \times f_{ts}(x) f_s(y) \phi(w_1) \phi(w_2) dx dy dw_1 dw_2 \\ & \leq O(1) \frac{1}{\sqrt{nk}} \sum_{t=2}^n \sum_{s=1}^{t-1} \frac{1}{\sqrt{t-s}} \frac{1}{\sqrt{s}} \iint \|Z_k(x)\| \|Z_k(y)\| |\gamma_k(x) \gamma_k(y)| dx dy \\ & = O(1) \frac{1}{k} \sqrt{n} \left( \int \|Z_k(x)\| |\gamma_k(x)| dx \right)^2 \leq O(1) \frac{1}{k} \sqrt{n} \int \|Z_k(x)\|^2 dx \int |\gamma_k(x)|^2 dx \\ & = o(1) \sqrt{nk}^{-m} = n^{1/2-m\kappa} = o(1), \end{aligned}$$

using Cauchy-Schwarz inequality, (3) of Lemma A.1 and Assumption B.

We have shown so far that  $e^\top Z(Z^\top Z)^{-2} Z^\top e = O_P(1) \frac{k}{\sqrt{n}}$  and  $\gamma^\top Z(Z^\top Z)^{-2} Z^\top \gamma = o_P(1) \frac{k}{\sqrt{n}}$ . Therefore,  $\|\hat{c}(\theta) - c\|^2 = O_P(1) \frac{k}{\sqrt{n}}$  by Cauchy-Schwarz inequality.

(2) Notice that  $\hat{g}_n(u; \theta) - g(u) = Z_k(u)^\top (\hat{c}(\theta) - c) - \gamma_k(u)$ . Thus,

$$\begin{aligned} & \max_{u \in \mathbb{R}} |\hat{g}_n(u; \theta) - g(u)| \leq \max_{u \in \mathbb{R}} |Z_k(u)^\top (\hat{c}(\theta) - c)| + \max_{u \in \mathbb{R}} |\gamma_k(u)| \\ & \leq \max_{u \in \mathbb{R}} \|Z_k(u)\| \|\hat{c}(\theta) - c\| + \max_{u \in \mathbb{R}} |\gamma_k(u)| \leq O_P(1) \sqrt{k} \frac{\sqrt{k}}{\sqrt[4]{n}} + o(1) k^{-(m-1)/2-1/12} \\ & = O_P(1) n^{-1/4+\kappa} + o(1) n^{-(m-5/6)\kappa/2} = O_P(1) n^{-1/4+\kappa} \end{aligned}$$

by (2) and (5) of Lemma A.1 and Assumption B.

It is evident all assertions remain true for  $\tilde{c}(\beta, \theta)$  and  $\hat{g}_n(u; \beta, \theta)$ .  $\square$

*Proof of Theorem 3.1.* (a) The score. Notice that  $\widehat{g}_n(\theta^\top x_t) - g(\theta^\top x_t) = Z_k(\theta^\top x_t)^\top (\widehat{c} - c) - \gamma_k(\theta^\top x_t)$  and thereby  $y_t - \widehat{g}_n(x_{1t}) = e_t + \gamma_k(x_{1t}) - Z_k(x_{1t})^\top (\widehat{c} - c)$ . In what follows, denote  $\xi_{n,t} = D_n^{-1} Q^\top x_t = (\frac{1}{\sqrt[4]{n}} x_{1t}, \frac{1}{\sqrt[4]{n^3}} x_{2t}^\top)^\top$  for brevity. It follows that

$$\begin{aligned}
D_n^{-1} S_n(\alpha_0) &= D_n^{-1} Q^\top S_n(\theta_0) = D_n^{-1} Q^\top \sum_{t=1}^n [y_t - \widehat{g}_n(\theta_0^\top x_t)] \dot{\widehat{g}}_n(\theta_0^\top x_t) x_t \\
&= \sum_{t=1}^n [y_t - \widehat{g}_n(x_{1t})] [\dot{g}(x_{1t}) + \dot{Z}_k(x_{1t})^\top (\widehat{c} - c) - \dot{\gamma}_k(x_{1t})] \xi_{n,t} \\
&= \sum_{t=1}^n [y_t - \widehat{g}_n(x_{1t})] \dot{g}(x_{1t}) \xi_{n,t} + \sum_{t=1}^n [y_t - \widehat{g}_n(x_{1t})] [\dot{Z}_k(x_{1t})^\top (\widehat{c} - c) - \dot{\gamma}_k(x_{1t})] \xi_{n,t} \\
&= \sum_{t=1}^n \dot{g}(x_{1t}) \xi_{n,t} e_t + \sum_{t=1}^n \dot{g}(x_{1t}) \gamma_k(x_{1t}) \xi_{n,t} - \sum_{t=1}^n \dot{g}(x_{1t}) Z_k(x_{1t})^\top (\widehat{c} - c) \xi_{n,t} \\
&\quad + \sum_{t=1}^n [y_t - \widehat{g}_n(x_{1t})] [\dot{Z}_k(x_{1t})^\top (\widehat{c} - c) - \dot{\gamma}_k(x_{1t})] \xi_{n,t} \\
&:= T_1 + T_2 - T_3 + T_4, \quad \text{say.}
\end{aligned}$$

It follows from Lemma A.3 that

$$T_1 = \begin{pmatrix} n^{-1/4} \sum_{t=1}^n \dot{g}(x_{1t}) x_{1t} e_t \\ n^{-3/4} \sum_{t=1}^n \dot{g}(x_{1t}) x_{2t} e_t \end{pmatrix} \rightarrow_D R^{1/2} W(1),$$

as  $n \rightarrow \infty$ .

To fulfil the assertion about the score, it suffices to show the remaining terms of  $T_2$ ,  $T_3$  and  $T_4$  are infinitesimal in probability. The proofs of these are relegated to Lemma B.1 below.

(b) The hessian. Notice that

$$\begin{aligned}
D_n^{-1} J_n(\alpha_0) D_n^{-1} &= D_n^{-1} Q^\top J_n(\theta_0) Q D_n^{-1} \\
&= D_n^{-1} Q^\top \sum_{t=1}^n \dot{\widehat{g}}_n^2(x_{1t}) x_t x_t^\top Q D_n^{-1} - D_n^{-1} Q^\top \sum_{t=1}^n [y_t - \widehat{g}_n(x_{1t})] \ddot{\widehat{g}}_n(x_{1t}) x_t x_t^\top Q D_n^{-1} \\
&= D_n^{-1} Q^\top \sum_{t=1}^n [\dot{g}(x_{1t})]^2 x_t x_t^\top Q D_n^{-1} + 2 D_n^{-1} Q^\top \sum_{t=1}^n \dot{g}(x_{1t}) [\dot{\widehat{g}}_n(x_{1t}) - \dot{g}(x_{1t})] x_t x_t^\top Q D_n^{-1} \\
&\quad + D_n^{-1} Q^\top \sum_{t=1}^n [\dot{\widehat{g}}_n(x_{1t}) - \dot{g}(x_{1t})]^2 x_t x_t^\top Q D_n^{-1} \\
&\quad - D_n^{-1} Q^\top \sum_{t=1}^n [y_t - \widehat{g}_n(x_{1t})] \ddot{\widehat{g}}_n(x_{1t}) x_t x_t^\top Q D_n^{-1} \\
&= \sum_{t=1}^n [\dot{g}(x_{1t})]^2 \xi_{n,t} \xi_{n,t}^\top + 2 \sum_{t=1}^n \dot{g}(x_{1t}) [\dot{\widehat{g}}_n(x_{1t}) - \dot{g}(x_{1t})] \xi_{n,t} \xi_{n,t}^\top \\
&\quad + \sum_{t=1}^n [\dot{\widehat{g}}_n(x_{1t}) - \dot{g}(x_{1t})]^2 \xi_{n,t} \xi_{n,t}^\top - \sum_{t=1}^n [y_t - \widehat{g}_n(x_{1t})] \ddot{\widehat{g}}_n(x_{1t}) \xi_{n,t} \xi_{n,t}^\top
\end{aligned}$$

$$:= I_1 + 2I_2 + I_3 - I_4, \quad \text{say.}$$

It follows that

$$I_1 = \begin{pmatrix} n^{-1/2} \sum_{t=1}^n \dot{g}^2(x_{1t}) x_{1t}^2 & n^{-1} \sum_{t=1}^n \dot{g}^2(x_{1t}) x_{1t} x_{2t}^\top \\ n^{-1} \sum_{t=1}^n \dot{g}^2(x_{1t}) x_{1t} x_{2t} & n^{-3/2} \sum_{t=1}^n \dot{g}^2(x_{1t}) x_{2t} x_{2t}^\top \end{pmatrix} \quad (\text{B.1})$$

converges to  $R$  in probability as  $n \rightarrow \infty$  by Theorem 1 of Phillips (2009). Hence, what we need to show is that all remaining terms ( $I_j$ ,  $j = 2, 3, 4$ ) are  $o_P(1)$ , which are relegated to Lemma B.1 below as well.  $\square$

**Lemma B.1** (Continuation of the proof of Theorem 3.1). *Under Assumptions A and B, all remaining terms in the proof of Theorem 3.1 are  $o_P(1)$  as  $n \rightarrow \infty$ .*

*Proof.* (a) The score. We shall show the  $T_2, T_3, T_4$  are all  $o_P(1)$ . Let  $a_n$  be a positive integer sequence such that  $a_n \rightarrow \infty$  and  $a_n/\sqrt[4]{n} \rightarrow 0$  as  $n \rightarrow \infty$ . Note that  $T_2 = \sum_{t=1}^n \dot{g}(x_{1t}) \gamma_k(x_{1t}) \xi_{n,t} = \sum_{t=1}^{a_n} \dot{g}(x_{1t}) \gamma_k(x_{1t}) \xi_{n,t} + \sum_{t=a_n+1}^n \dot{g}(x_{1t}) \gamma_k(x_{1t}) \xi_{n,t}$  and the first part is  $o_P(1)$ . In fact,

$$\begin{aligned} & \left\| \sum_{t=1}^{a_n} \dot{g}(x_{1t}) \gamma_k(x_{1t}) \xi_{n,t} \right\| \leq \sum_{t=1}^{a_n} |\dot{g}(x_{1t}) \gamma_k(x_{1t})| \|\xi_{n,t}\| \\ & \leq \frac{1}{\sqrt[4]{n}} \sum_{t=1}^{a_n} |\dot{g}(x_{1t}) \gamma_k(x_{1t})| |x_{1t}| + \frac{1}{\sqrt[4]{n^3}} \sum_{t=1}^{a_n} |\dot{g}(x_{1t}) \gamma_k(x_{1t})| \|x_{2t}\| \\ & = o_{a.s.}(1) \frac{1}{\sqrt[4]{n}} a_n + \frac{1}{\sqrt[4]{n}} \sum_{t=1}^{a_n} |\dot{g}(x_{1t}) \gamma_k(x_{1t})| \|x_{2t,n}\| = o_{a.s.}(1) + o_P(1) \frac{1}{\sqrt[4]{n}} a_n = o_P(1), \end{aligned}$$

because  $x\dot{g}(x)$  is bounded and  $\gamma_k(\cdot) = k^{-(m-1)/2-1/12} \rightarrow 0$  uniformly, while  $x_{2t,n} = x_{2t}/\sqrt{n} \rightarrow_D V_2(r)$  a  $(d-1)$ -dimensional Brownian motion so that  $\gamma_k(x_{1t}) x_{2t,n} \rightarrow_P 0$  independent of  $t$ .

Thus, to show  $T_2 = o_P(1)$  we only need to calculate the following expectation. Using the density  $f_t(x)$  for  $\frac{1}{\sqrt{t}} x_{1t}$  and the joint density  $\psi_t(x, w)$  for  $(\frac{1}{\sqrt{t}} x_{1t}, \frac{1}{\sqrt{t}} x_{2t})$ , (2) of Lemma A.1,

$$\begin{aligned} & E \left\| \sum_{t=a_n+1}^n \dot{g}(x_{1t}) \gamma_k(x_{1t}) \xi_{n,t} \right\| \leq E \sum_{t=a_n+1}^n |\dot{g}(x_{1t}) \gamma_k(x_{1t})| \|\xi_{n,t}\| \\ & \leq \frac{1}{\sqrt[4]{n}} k^{-(m-1)/2-1/12} E \sum_{t=a_n+1}^n |\dot{g}(x_{1t}) x_{1t}| + \frac{1}{\sqrt[4]{n^3}} k^{-(m-1)/2-1/12} E \sum_{t=a_n+1}^n |\dot{g}(x_{1t})| \|x_{2t}\| \\ & = \frac{1}{\sqrt[4]{n}} k^{-(m-1)/2-1/12} \sum_{t=a_n+1}^n \int |\dot{g}(\sqrt{t}x)| \sqrt{t} |x| f_t(x) dx \\ & \quad + \frac{1}{\sqrt[4]{n^3}} k^{-(m-1)/2-1/12} \sum_{t=a_n+1}^n \iint |\dot{g}(\sqrt{t}x)| \|\sqrt{t}w\| \psi_t(x, w) dx dw \\ & \leq O(1) \frac{1}{\sqrt[4]{n}} k^{-(m-1)/2-1/12} \sum_{t=a_n+1}^n \frac{1}{\sqrt{t}} \int |\dot{g}(x)x| dx \end{aligned}$$

$$\begin{aligned}
& + O(1) \frac{1}{\sqrt[4]{n^3}} k^{-(m-1)/2-1/12} \sum_{t=a_n+1}^n \int |\dot{g}(x)| dx \int \|w\| \phi(w) dw (1 + o(1)) \\
& = O(1) \sqrt[4]{n} k^{-(m-1)/2-1/12} = O(1) n^{1/4-(m-1)\kappa/2-\kappa/12} = o(1),
\end{aligned}$$

by Assumption B, implying  $T_2 = o_P(1)$ .

This argument on  $a_n$  should be engaged throughout whenever we calculate expectations using densities  $\psi_t(x, w)$ ,  $f_t(x)$  and  $\psi_{ts}(x, w)$ ,  $f_{ts}(x)$ . Nevertheless, in order to save space and avoid the repetitious statement, we abuse a bit the densities aforementioned for all  $t$  and  $s$  throughout the proof.

Observe by Lemma A.2 that,

$$\begin{aligned}
T_3 &= \sum_{t=1}^n \dot{g}(x_{1t}) Z_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} = \sum_{t=1}^n \dot{g}(x_{1t}) Z_k(x_{1t})^\top (Z^\top Z)^{-1} Z^\top (\gamma + e) \xi_{n,t} \\
&= \frac{1 + o_P(1)}{\sqrt{n} L_1(1, 0)} \sum_{t=1}^n \dot{g}(x_{1t}) Z_k(x_{1t})^\top Z^\top (\gamma + e) \xi_{n,t} \\
&= \frac{1 + o_P(1)}{\sqrt{n} L_1(1, 0)} \sum_{t=1}^n \dot{g}(x_{1t}) Z_k(x_{1t})^\top Z^\top e \xi_{n,t} + \frac{1 + o_P(1)}{\sqrt{n} L_1(1, 0)} \sum_{t=1}^n \dot{g}(x_{1t}) Z_k(x_{1t})^\top Z^\top \gamma \xi_{n,t} \\
&:= \frac{1 + o_P(1)}{L_1(1, 0)} T_{31} + \frac{1 + o_P(1)}{L_1(1, 0)} T_{32}, \quad \text{say.}
\end{aligned}$$

Using the notation  $\mathcal{U}_k(x_{1t}, x_{1s})$  in (A.8),

$$\begin{aligned}
E \|T_{31}\|^2 &= \frac{1}{n} E \left\| \sum_{t=1}^n \dot{g}(x_{1t}) Z_k(x_{1t})^\top Z^\top e \xi_{n,t} \right\|^2 = \frac{1}{n} E \left\| \sum_{t=1}^n \dot{g}(x_{1t}) \sum_{s=1}^n \mathcal{U}_k(x_{1t}, x_{1s}) e_s \xi_{n,t} \right\|^2 \\
&= \frac{1}{n} E \sum_{s=1}^n \sum_{s_1=1}^n e_s e_{s_1} \sum_{t=1}^n \sum_{t_1=1}^n \dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1s}) \dot{g}(x_{1t_1}) \mathcal{U}_k(x_{1t_1}, x_{1s_1}) \xi_{n,t}^\top \xi_{n,t_1}.
\end{aligned}$$

To make use of the martingale difference structure for  $e_t$  we need to consider the relation among  $s, t, s_1$  and  $t_1$ . If  $s$  and/or  $s_1$  is bigger than  $t$  and  $t_1$ , we would directly use the martingale difference structure. On the other hand, if  $t$  and/or  $t_1$  is larger than  $s$  and  $s_1$ , e.g. both  $t$  and  $t_1$  ( $t \geq t_1$ ) are larger than  $s$  and  $s_1$  where  $s \geq s_1$ , we may use the notation  $\mathcal{U}_k(x, y) = \sum_{i=0}^{k-1} \mathcal{H}_i(x) \mathcal{H}_i(y)$  such that the related terms are equal to the sum of product in the form of  $f_1(x_{1t}) f_2(x_{1t_1}) f_3(x_{1s}) f_4(x_{1s_1}) e_s e_{s_1}$ . To compute the expectation, write  $x_{1t} = (x_{1t} - x_{1t_1}) + (x_{1t_1} - x_{1s}) + x_{1s}$  and  $x_{1t_1} = (x_{1t_1} - x_{1s}) + x_{1s}$ . Recall that all increments after normalization have corresponding densities by Lemma A.4, and therefore

$$\begin{aligned}
& E[f_1(x_{1t}) f_2(x_{1t_1}) f_3(x_{1s}) f_4(x_{1s_1}) e_s e_{s_1}] = E[E[f_1(x_{1t}) f_2(x_{1t_1}) | \mathcal{F}_s] f_3(x_{1s}) f_4(x_{1s_1}) e_s e_{s_1}]] \\
& = E[E[f_1(x_{1t}) f_2(x_{1t_1}) | \mathcal{F}_s] f_3(x_{1s}) f_4(x_{1s_1}) E[e_s | \mathcal{F}_{s-1}] e_{s_1}]
\end{aligned}$$

in the case that  $s > s_1$ , otherwise  $E[e_s | \mathcal{F}_{s-1}] e_{s_1}$  would be substituted by  $E[e_s^2 | \mathcal{F}_{s-1}]$ . This is because the conditional expectation  $E[f_1(x_{1t}) f_2(x_{1t_1}) | \mathcal{F}_s]$  is a function of  $x_{1s}$ , which is adapted

with respect to  $\mathcal{F}_{s-1}$  by Assumption A. Hence, we can use the property of the structure for  $e_t$  in each term and thus,

$$\begin{aligned}
E\|T_{31}\|^2 &= \frac{1}{n} E \left\| \sum_{t=1}^n \dot{g}(x_{1t}) Z_k(x_{1t})^\top Z^\top e \xi_{n,t} \right\|^2 = \frac{1}{n} E \left\| \sum_{t=1}^n \dot{g}(x_{1t}) \sum_{s=1}^n \mathcal{U}_k(x_{1t}, x_{1s}) e_s \xi_{n,t} \right\|^2 \\
&= \frac{1}{n} E \left\| \sum_{s=1}^n e_s \sum_{t=1}^n \dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1s}) \xi_{n,t} \right\|^2 = \sigma_e^2 \frac{1}{n} E \sum_{s=1}^n \left\| \sum_{t=1}^n \dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1s}) \xi_{n,t} \right\|^2 \\
&\leq O(1) \frac{1}{n} E \sum_{t=1}^n \|\dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1t}) \xi_{n,t}\|^2 + O(1) \frac{1}{n} E \sum_{s=2}^n \left\| \sum_{t=1}^{s-1} \dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1s}) \xi_{n,t} \right\|^2 \\
&\quad + O(1) \frac{1}{n} E \sum_{s=1}^{n-1} \left\| \sum_{t=s+1}^n \dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1s}) \xi_{n,t} \right\|^2,
\end{aligned}$$

in which the first term can be calculated as

$$\begin{aligned}
&\frac{1}{n} E \sum_{t=1}^n \|\dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1t}) \xi_{n,t}\|^2 = \frac{1}{n} E \sum_{t=1}^n |\dot{g}(x_{1t})|^2 \|Z_k(x_{1t})\|^4 \|\xi_{n,t}\|^2 \\
&= \frac{1}{n\sqrt{n}} E \sum_{t=1}^n |\dot{g}(x_{1t})|^2 \|Z_k(x_{1t})\|^4 |x_{1t}|^2 + \frac{1}{n^2\sqrt{n}} E \sum_{t=1}^n |\dot{g}(x_{1t})|^2 \|Z_k(x_{1t})\|^4 \|x_{2t}\|^2 \\
&\leq O(1) \frac{1}{n\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\dot{g}(x)|^2 \|Z_k(x)\|^4 |x|^2 dx \\
&\quad + O(1) \frac{1}{n^2\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\dot{g}(x)|^2 \|Z_k(x)\|^4 \|\sqrt{t}w\|^2 \phi(w) dx dw (1 + o(1)) \\
&= O(1) \frac{1}{n} \int |\dot{g}(x)|^2 \|Z_k(x)\|^4 |x|^2 dx \leq O(1) \frac{1}{n} k^2 \int |\dot{g}(x)|^2 |x|^2 dx = o(1),
\end{aligned}$$

as  $\|Z_k(x)\|^2 \leq O(1)k$  uniformly in  $x$ . Next, Lemma A.6 is used to deal with the second term. Notice that as the probabilities of  $x_{1t} = x_{1s}$  and  $x_{1\ell} = x_{1s}$  are all zero whenever  $t \neq s$  and  $\ell \neq s$ . The expectation will be computed excluding the regions  $x_{1t} = x_{1s}$  and  $x_{1\ell} = x_{1s}$ . In a usual situation, such exclusion does not make any difference, however, this time it really matters. The detail is as follows.

Using the dominated convergence theorem and the notation in (A.8), we have

$$\begin{aligned}
&\frac{1}{n} E \sum_{s=2}^n \left\| \sum_{t=1}^{s-1} \dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1s}) \xi_{n,t} \right\|^2 \leq \frac{1}{n} E \sum_{s=2}^n \left( \sum_{t=1}^{s-1} |\dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1s})| \|\xi_{n,t}\| \right)^2 \\
&= \frac{1}{n} E \sum_{s=2}^n \sum_{t=1}^{s-1} |\dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1s})|^2 \|\xi_{n,t}\|^2 \\
&\quad + 2 \frac{1}{n} E \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{\ell=1}^{t-1} |\dot{g}(x_{1t}) \mathcal{U}_k(x_{1t}, x_{1s})| \|\xi_{n,t}\| |\dot{g}(x_{1\ell}) \mathcal{U}_k(x_{1\ell}, x_{1s})| \|\xi_{n,\ell}\| \\
&\leq O(1) \frac{1}{n} k E \sum_{s=2}^n \sum_{t=1}^{s-1} |\dot{g}(x_{1t})|^2 \|Z_k(x_{1s})\|^2 \|\xi_{n,t}\|^2
\end{aligned}$$

$$\begin{aligned}
& + 2\frac{1}{n}E \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{\ell=1}^{t-1} |\dot{g}(x_{1t})| |\mathcal{U}_k(x_{1t}, x_{1s})| \|\xi_{n,t}\| |\dot{g}(x_{1\ell})| |\mathcal{U}_k(x_{1\ell}, x_{1s})| \|\xi_{n,\ell}\| \\
& = \frac{O(1)k}{n^{3/2}} E \sum_{s=2}^n \sum_{t=1}^{s-1} |\dot{g}(x_{1t})|^2 \|Z_k(x_{1s})\|^2 |x_{1t}|^2 + \frac{O(1)k}{n^{5/2}} E \sum_{s=2}^n \sum_{t=1}^{s-1} |\dot{g}(x_{1t})|^2 \|Z_k(x_{1s})\|^2 \|x_{2t}\|^2 \\
& \quad + \frac{2}{n} E \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{\ell=1}^{t-1} |\dot{g}(x_{1t})| |\mathcal{U}_k(x_{1t}, x_{1s})| \|\xi_{n,t}\| |\dot{g}(x_{1\ell})| |\mathcal{U}_k(x_{1\ell}, x_{1s})| \|\xi_{n,\ell}\| \\
& \leq O(1) \frac{k}{n^{3/2}} \sum_{s=2}^n \sum_{t=1}^{s-1} \frac{1}{\sqrt{s-t}} \frac{1}{\sqrt{t}} \iint |\dot{g}(y)|^2 \|Z_k(x)\|^2 |y|^2 dx dy \\
& \quad + O(1) \frac{k}{n^{5/2}} \sum_{s=2}^n \sum_{t=1}^{s-1} \frac{1}{\sqrt{s-t}} \frac{1}{\sqrt{t}} \int \cdots \int |\dot{g}(y)|^2 \|Z_k(x)\|^2 \|\sqrt{t-sw} + \sqrt{sw'}\|^2 \phi(w)\phi(w') dx dy dw dw' \\
& \quad + 2\frac{1}{n\sqrt{n}} E \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{\ell=1}^{t-1} |\dot{g}(x_{1t})| |\mathcal{U}_k(x_{1t}, x_{1s})| |x_{1t}| |\dot{g}(x_{1\ell})| |\mathcal{U}_k(x_{1\ell}, x_{1s})| |x_{1\ell}| \\
& \quad + 2\frac{1}{n^2\sqrt{n}} E \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{\ell=1}^{t-1} |\dot{g}(x_{1t})| |\mathcal{U}_k(x_{1t}, x_{1s})| \|x_{2t}\| |\dot{g}(x_{1\ell})| |\mathcal{U}_k(x_{1\ell}, x_{1s})| \|x_{2\ell}\| \\
& \leq O(1) \frac{1}{\sqrt{n}} k \int \|Z_k(x)\|^2 dx \int |\dot{g}(y)|^2 |y|^2 dy \\
& \quad + O(1) \frac{1}{n\sqrt{n}} \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{\ell=1}^{t-1} \frac{1}{\sqrt{s-t}} \frac{1}{\sqrt{t-\ell}} \frac{1}{\sqrt{\ell}} \\
& \quad \times \iiint_{x \neq y \neq z} |\dot{g}(y)| |\mathcal{U}_k(x, y)| |y| |\dot{g}(z)| |\mathcal{U}_k(x, z)| |z| dx dy dz \\
& \quad + O(1) \frac{1}{n^2\sqrt{n}} \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{\ell=1}^{t-1} \frac{1}{\sqrt{s-t}} \frac{1}{\sqrt{t-\ell}} \frac{1}{\sqrt{\ell}} \int \cdots \int_{x \neq y \neq z} \\
& \quad \times |\dot{g}(y)| |\mathcal{U}_k(x, y)| \|\sqrt{t-\ell}w + \sqrt{\ell}v\| |\dot{g}(z)| |\mathcal{U}_k(x, z)| \|\sqrt{\ell}v\| \phi(v)\phi(w) dx dy dz dw dv \\
& = O(1) \frac{1}{\sqrt{n}} k^2 + O(1) \iiint_{x \neq y \neq z} |y\dot{g}(y)| |\mathcal{U}_k(x, y)| |z\dot{g}(z)| |\mathcal{U}_k(x, z)| dx dy dz \\
& \quad + O(1) \iiint_{x \neq y \neq z} |\dot{g}(y)| |\mathcal{U}_k(x, y)| |\dot{g}(z)| |\mathcal{U}_k(x, z)| dx dy dz \\
& = o(1) + O(1) \iiint_{x \neq y \neq z} |y\dot{g}(y)| |\mathcal{U}_k(x, y)| |z\dot{g}(z)| |\mathcal{U}_k(x, z)| dx dy dz \\
& \xrightarrow{k \rightarrow \infty} O(1) \iiint_{x \neq y \neq z} |y\dot{g}(y)| \delta(x-y) |z\dot{g}(z)| \delta(x-z) dx dy dz = 0
\end{aligned}$$

by Lemma A.6 and the integrability of  $|\dot{g}(y)y|$ . Similarly, the third term in  $T_{31}$  can be shown to be  $o_P(1)$ .

Regarding of  $T_{32}$ , note that

$$E\|T_{32}\| = \frac{1}{\sqrt{n}} E \left\| \sum_{t=1}^n \dot{g}(x_{1t}) Z_k(x_{1t})^\top Z^\top \gamma \xi_{n,t} \right\| = \frac{1}{\sqrt{n}} E \left\| \sum_{t=1}^n \dot{g}(x_{1t}) \sum_{s=1}^n \mathcal{U}_k(x_{1t}, x_{1s}) \gamma_k(x_{1s}) \xi_{n,t} \right\|$$

$$\begin{aligned}
&\leq \frac{1}{\sqrt{n}} E \sum_{t=1}^n |\dot{g}(x_{1t})| \sum_{s=1}^n \|Z_k(x_{1t})\| \|Z_k(x_{1s})\| |\gamma_k(x_{1s})| \|\xi_{n,t}\| \\
&= \frac{1}{\sqrt{n}} E \sum_{t=1}^n |\dot{g}(x_{1t})| \|Z_k(x_{1t})\|^2 |\gamma_k(x_{1t})| \|\xi_{n,t}\| + \frac{1}{\sqrt{n}} E \sum_{t=2}^n |\dot{g}(x_{1t})| \sum_{s=1}^{t-1} \|Z_k(x_{1t})\| \|Z_k(x_{1s})\| |\gamma_k(x_{1s})| \|\xi_{n,t}\| \\
&\quad + \frac{1}{\sqrt{n}} E \sum_{t=1}^{n-1} |\dot{g}(x_{1t})| \sum_{s=t+1}^n \|Z_k(x_{1t})\| \|Z_k(x_{1s})\| |\gamma_k(x_{1s})| \|\xi_{n,t}\| \\
&= \frac{1}{\sqrt[4]{n^3}} E \sum_{t=1}^n |\dot{g}(x_{1t})| \|Z_k(x_{1t})\|^2 |\gamma_k(x_{1t})| \|x_{1t}\| + \frac{1}{n\sqrt[4]{n}} E \sum_{t=1}^n |\dot{g}(x_{1t})| \|Z_k(x_{1t})\|^2 |\gamma_k(x_{1t})| \|x_{2t}\| \\
&\quad + \frac{1}{\sqrt[4]{n^3}} E \sum_{t=2}^n |\dot{g}(x_{1t})| \sum_{s=1}^{t-1} \|Z_k(x_{1t})\| \|Z_k(x_{1s})\| |\gamma_k(x_{1s})| \|x_{1t}\| \\
&\quad + \frac{1}{n\sqrt[4]{n}} E \sum_{t=2}^n |\dot{g}(x_{1t})| \sum_{s=1}^{t-1} \|Z_k(x_{1t})\| \|Z_k(x_{1s})\| |\gamma_k(x_{1s})| \|x_{2t}\| \\
&\quad + \frac{1}{\sqrt[4]{n^3}} E \sum_{t=1}^{n-1} |\dot{g}(x_{1t})| \sum_{s=t+1}^n \|Z_k(x_{1t})\| \|Z_k(x_{1s})\| |\gamma_k(x_{1s})| \|x_{1t}\| \\
&\quad + \frac{1}{n\sqrt[4]{n}} E \sum_{t=1}^{n-1} |\dot{g}(x_{1t})| \sum_{s=t+1}^n \|Z_k(x_{1t})\| \|Z_k(x_{1s})\| |\gamma_k(x_{1s})| \|x_{2t}\| \\
&\leq O(1) \frac{k}{\sqrt[4]{n^3}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\dot{g}(x) \gamma_k(x)| |x| dx \\
&\quad + O(1) \frac{k}{n\sqrt[4]{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\dot{g}(x)| |\gamma_k(x)| \|\sqrt{t}w\| \phi(w) (1 + o(1)) dx dw \\
&\quad + O(1) \frac{1}{\sqrt[4]{n^3}} \sum_{t=2}^n \sum_{s=1}^{t-1} \frac{1}{\sqrt{t-s}} \frac{1}{\sqrt{t}} \iint |\dot{g}(x)| \|Z_k(x)\| \|Z_k(y)\| |\gamma_k(y)| |x| dx dy \\
&\quad + O(1) \frac{1}{n\sqrt[4]{n}} \sum_{t=2}^n \sum_{s=1}^{t-1} \frac{1}{\sqrt{t-s}} \frac{1}{\sqrt{t}} \int \cdots \int |\dot{g}(x)| \|Z_k(x)\| \|Z_k(y)\| \\
&\quad \times |\gamma_k(y)| \|\sqrt{t-s}w + \sqrt{sv}\| \phi(w) \phi(v) (1 + o(1)) dx dy dw dv \\
&\quad + \frac{1}{\sqrt[4]{n^3}} \sum_{t=1}^{n-1} \sum_{s=t+1}^n \frac{1}{\sqrt{s-t}} \frac{1}{\sqrt{t}} \iint |\dot{g}(y)| \|Z_k(x)\| \|Z_k(y)\| |\gamma_k(x)| |y| dx dy \\
&\quad + \frac{1}{n\sqrt[4]{n}} \sum_{t=1}^{n-1} \sum_{s=t+1}^n \frac{1}{\sqrt{s-t}} \frac{1}{\sqrt{t}} \\
&\quad \times \iiint |\dot{g}(y)| \|Z_k(x)\| \|Z_k(y)\| |\gamma_k(x)| \|\sqrt{t}w\| \phi(w) (1 + o(1)) dx dy dw \\
&\leq O(1) \frac{k}{\sqrt[4]{n}} \left( \int |x \dot{g}(x)|^2 dx \int \gamma_k^2(x) dx \right)^{1/2} + O(1) \frac{k}{\sqrt[4]{n}} \left( \int |\dot{g}(x)|^2 dx \int \gamma_k^2(x) dx \right)^{1/2} \\
&\quad + O(1) \sqrt[4]{n} \left( \int |x \dot{g}(x)|^2 dx \int \|Z_k(x)\|^2 dx \int \|Z_k(y)\|^2 dy \int |\gamma_k(y)|^2 dy \right)^{1/2}
\end{aligned}$$

$$\begin{aligned}
& + O(1)\sqrt[4]{n} \left( \int |y\dot{g}(y)|^2 dy \int \|Z_k(y)\|^2 dy \int \|Z_k(x)\|^2 dx \int |\gamma_k(x)|^2 dx \right)^{1/2} \\
& = o(1) \frac{k}{\sqrt[4]{n}} k^{-m/2} + o(1) \sqrt[4]{n} k^{-(m-2)/2} = o(1)
\end{aligned}$$

using Lemma A.1 and Assumption B. Moreover,

$$\begin{aligned}
T_4 &= \sum_{t=1}^n [y_t - \hat{g}_n(x_{1t})] [\dot{Z}_k(x_{1t})^\top (\hat{c} - c) - \dot{\gamma}_k(x_{1t})] \xi_{n,t} \\
&= \sum_{t=1}^n [y_t - \hat{g}_n(x_{1t})] \dot{Z}_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} - \sum_{t=1}^n [y_t - \hat{g}_n(x_{1t})] \dot{\gamma}_k(x_{1t}) \xi_{n,t} \\
&= \sum_{t=1}^n e_t \dot{Z}_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} + \sum_{t=1}^n \gamma_k(x_{1t}) \dot{Z}_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} \\
&\quad - \sum_{t=1}^n Z_k(x_{1t})^\top (\hat{c} - c) \dot{Z}_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} - \sum_{t=1}^n e_t \dot{\gamma}_k(x_{1t}) \xi_{n,t} \\
&\quad - \sum_{t=1}^n \gamma_k(x_{1t}) \dot{\gamma}_k(x_{1t}) \xi_{n,t} + \sum_{t=1}^n Z_k(x_{1t})^\top (\hat{c} - c) \dot{\gamma}_k(x_{1t}) \xi_{n,t} \\
&:= T_{41} + T_{42} - T_{43} - T_{44} - T_{45} + T_{46}, \quad \text{say.}
\end{aligned}$$

We shall tackle them one by one. Notice that

$$\begin{aligned}
\|T_{41}\|^2 &= \left\| \sum_{t=1}^n e_t \dot{Z}_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} \right\|^2 = \left\| \left( \sum_{t=1}^n e_t \xi_{n,t}^\top \dot{Z}_k(x_{1t}) \right)^\top (\hat{c} - c) \right\|^2 \\
&\leq \left\| \sum_{t=1}^n e_t \xi_{n,t} \dot{Z}_k(x_{1t})^\top \right\|^2 \|\hat{c} - c\|^2 = O_P(1) \frac{k}{\sqrt{n}} \left\| \sum_{t=1}^n e_t \xi_{n,t} \dot{Z}_k(x_{1t})^\top \right\|^2
\end{aligned}$$

by Theorem 2.1. However, by martingale difference structure for  $e_t$ ,

$$\begin{aligned}
& \frac{k}{\sqrt{n}} E \left\| \sum_{t=1}^n e_t \xi_{n,t} \dot{Z}_k(x_{1t})^\top \right\|^2 = \sigma_e^2 \frac{k}{\sqrt{n}} E \sum_{t=1}^n \left\| \xi_{n,t} \dot{Z}_k(x_{1t})^\top \right\|^2 \\
&= \sigma_e^2 \frac{k}{\sqrt{n}} E \sum_{t=1}^n \left( \frac{1}{\sqrt{n}} x_{1t}^2 + \frac{1}{\sqrt{n}^3} \|x_{2t}\|^2 \right) \|\dot{Z}_k(x_{1t})\|^2 \\
&\leq O(1) \frac{k}{n} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int x^2 \|\dot{Z}_k(x)\|^2 dx + O(1) \frac{k}{n^2} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint t \|w\|^2 \|\dot{Z}_k(x_{1t})\|^2 \phi(w) dx dw (1 + o(1)) \\
&= O(1) \frac{k}{\sqrt{n}} \int x^2 \|\dot{Z}_k(x)\|^2 dx = O(1) \frac{k}{\sqrt{n}} k^3 = o(1),
\end{aligned}$$

by Assumption B and (4) of Lemma A.1, indicating that  $T_{41} = o_P(1)$ .

$$\|T_{42}\| = \left\| \sum_{t=1}^n \gamma_k(x_{1t}) \dot{Z}_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} \right\| \leq \sum_{t=1}^n |\gamma_k(x_{1t}) \dot{Z}_k(x_{1t})^\top (\hat{c} - c)| \|\xi_{n,t}\|$$



$$\leq \sum_{t=1}^n |\gamma_k(x_{1t})| \|\dot{Z}_k(x_{1t})\| \|\widehat{c} - c\| \|\xi_{n,t}\| = O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n |\gamma_k(x_{1t})| \|\dot{Z}_k(x_{1t})\| \|\xi_{n,t}\|,$$

and

$$\begin{aligned} & \frac{\sqrt{k}}{\sqrt[4]{n}} E \sum_{t=1}^n |\gamma_k(x_{1t})| \|\dot{Z}_k(x_{1t})\| \|\xi_{n,t}\| \\ & \leq \frac{\sqrt{k}}{\sqrt{n}} E \sum_{t=1}^n |\gamma_k(x_{1t})| \|\dot{Z}_k(x_{1t})\| |x_{1t}| + \frac{\sqrt{k}}{n} E \sum_{t=1}^n |\gamma_k(x_{1t})| \|\dot{Z}_k(x_{1t})\| |x_{2t}| \\ & \leq O(1) \frac{\sqrt{k}}{\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\gamma_k(x)| \|\dot{Z}_k(x)\| |x| dx \\ & \quad + O(1) \frac{\sqrt{k}}{n} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\gamma_k(x)| \|\dot{Z}_k(x)\| \|\sqrt{t}w\| \phi(w) dx dw (1 + o(1)) \\ & = O(1) \sqrt{k} \int |\gamma_k(x)| \|\dot{Z}_k(x)\| |x| dx + O(1) \sqrt{k} \int |\gamma_k(x)| \|\dot{Z}_k(x)\| dx \\ & \leq O(1) \sqrt{k} \left( \int |x \gamma_k(x)|^2 dx \int \|\dot{Z}_k(x)\|^2 dx \right)^{1/2} \\ & = o(1) \sqrt{k} k^{-(m-1)/2} = o(1) k^{-(m-4)/2} = o(1). \end{aligned}$$

Also, using Lemma A.2 again,

$$\begin{aligned} T_{43} &= \sum_{t=1}^n Z_k(x_{1t})^\top (\widehat{c} - c) \dot{Z}_k(x_{1t})^\top (\widehat{c} - c) \xi_{n,t} \\ &= \sum_{t=1}^n Z_k(x_{1t})^\top (Z^\top Z)^{-1} Z^\top (e + \gamma) \dot{Z}_k(x_{1t})^\top (Z^\top Z)^{-1} Z^\top (e + \gamma) \xi_{n,t} \\ &= \frac{1 + o_P(1)}{nL_1^2(1, 0)} \sum_{t=1}^n Z_k(x_{1t})^\top Z^\top (e + \gamma) \dot{Z}_k(x_{1t})^\top Z^\top (e + \gamma) \xi_{n,t} \\ &= \frac{1 + o_P(1)}{nL_1^2(1, 0)} \sum_{t=1}^n Z_k(x_{1t})^\top Z^\top e \dot{Z}_k(x_{1t})^\top Z^\top e \xi_{n,t} \\ & \quad + \frac{1 + o_P(1)}{nL_1^2(1, 0)} \sum_{t=1}^n Z_k(x_{1t})^\top Z^\top \gamma \dot{Z}_k(x_{1t})^\top Z^\top \gamma \xi_{n,t} \\ & \quad + \frac{1 + o_P(1)}{nL_1^2(1, 0)} \sum_{t=1}^n Z_k(x_{1t})^\top Z^\top e \dot{Z}_k(x_{1t})^\top Z^\top \gamma \xi_{n,t} \\ & \quad + \frac{1 + o_P(1)}{nL_1^2(1, 0)} \sum_{t=1}^n Z_k(x_{1t})^\top Z^\top \gamma \dot{Z}_k(x_{1t})^\top Z^\top e \xi_{n,t} \\ & := \frac{1 + o_P(1)}{L_1^2(1, 0)} \sum_{j=1}^4 T_{43}(j), \quad \text{say.} \end{aligned}$$

We shall show  $T_{43}(j) = o_P(1)$ , for  $j = 1, \dots, 4$ . Denoting  $\widetilde{\mathcal{U}}_k(x, y) = \sum_{i=0}^{k-1} \mathcal{H}_i(x) \mathcal{H}_i(y)$ ,

we have

$$\begin{aligned}
E\|T_{43}(1)\|^2 &= \frac{1}{n^2} E \left\| \sum_{t=1}^n Z_k(x_{1t})^\top Z^\top e \dot{Z}_k(x_{1t})^\top Z^\top e \xi_{n,t} \right\|^2 \\
&= \frac{1}{n^2} E \left\| \sum_{t=1}^n \sum_{s=1}^n \mathcal{U}_k(x_{1t}, x_{1s}) e_s \sum_{\ell=1}^n \widetilde{\mathcal{U}}_k(x_{1t}, x_{1\ell}) e_\ell \xi_{n,t} \right\|^2 \\
&= \frac{1}{n^2} E \left\| \sum_{s=1}^n \sum_{\ell=1}^n e_s e_\ell \sum_{t=1}^n \mathcal{U}_k(x_{1t}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t}, x_{1\ell}) \xi_{n,t} \right\|^2 \\
&= \mu_4 \frac{1}{n^2} E \sum_{s=1}^n \left\| \sum_{t=1}^n \mathcal{U}_k(x_{1t}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t}, x_{1s}) \xi_{n,t} \right\|^2 \\
&\quad + 2\sigma_e^4 \frac{1}{n^2} E \sum_{s=2}^n \sum_{\ell=1}^{s-1} \left\| \sum_{t=1}^n \mathcal{U}_k(x_{1t}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t}, x_{1\ell}) \xi_{n,t} \right\|^2 \\
&\leq \mu_4 \frac{1}{n^2} E \sum_{s=1}^n \left( \sum_{t=1}^n |\mathcal{U}_k(x_{1t}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t}, x_{1s})| |\xi_{n,t}| \right)^2 \\
&\quad + 2\sigma_e^4 \frac{1}{n^2} E \sum_{s=2}^n \sum_{\ell=1}^{s-1} \left( \sum_{t=1}^n |\mathcal{U}_k(x_{1t}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t}, x_{1\ell})| |\xi_{n,t}| \right)^2 \\
&= \mu_4 \frac{1}{n^2} E \sum_{s=1}^n \sum_{t=1}^n |\mathcal{U}_k(x_{1t}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t}, x_{1s})|^2 |\xi_{n,t}|^2 \\
&\quad + 2\mu_4 \frac{1}{n^2} E \sum_{s=1}^n \sum_{t=2}^n \sum_{t'=1}^{t-1} |\mathcal{U}_k(x_{1t}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t}, x_{1s})| |\xi_{n,t}| |\mathcal{U}_k(x_{1t'}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t'}, x_{1s})| |\xi_{n,t'}| \\
&\quad + 2\sigma_e^4 \frac{1}{n^2} E \sum_{s=2}^n \sum_{\ell=1}^{s-1} \sum_{t=1}^n |\mathcal{U}_k(x_{1t}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t}, x_{1\ell})|^2 |\xi_{n,t}|^2 \\
&\quad + 4\sigma_e^4 \frac{1}{n^2} E \sum_{s=2}^n \sum_{\ell=1}^{s-1} \sum_{t=2}^n \sum_{t'=1}^{t-1} |\mathcal{U}_k(x_{1t}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t}, x_{1\ell})| |\xi_{n,t}| |\mathcal{U}_k(x_{1t'}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t'}, x_{1\ell})| |\xi_{n,t'}| \\
&\leq O(1) \frac{1}{n^2 \sqrt{n}} E \sum_{s=1}^n \sum_{t=1}^n \|Z_k(x_{1t})\|^2 \|\dot{Z}_k(x_{1t})\|^2 \|Z_k(x_{1s})\|^4 |x_{1t}|^2 \\
&\quad + O(1) \frac{1}{n^3 \sqrt{n}} E \sum_{s=1}^n \sum_{t=1}^n \|Z_k(x_{1t})\|^2 \|\dot{Z}_k(x_{1t})\|^2 \|Z_k(x_{1s})\|^4 |x_{2t}|^2 \\
&\quad + O(1) \frac{1}{n^2} E \sum_{s=1}^n \sum_{t=2}^n \sum_{t'=1}^{t-1} \|Z_k(x_{1t})\| \|\dot{Z}_k(x_{1t})\| \|Z_k(x_{1t'})\| \|\dot{Z}_k(x_{1t'})\| \|Z_k(x_{1s})\|^4 \\
&\quad \quad \times \left( \frac{1}{\sqrt[4]{n}} |x_{1t}| + \frac{1}{\sqrt[4]{n^3}} \|x_{2t}\| \right) \left( \frac{1}{\sqrt[4]{n}} |x_{1t'}| + \frac{1}{\sqrt[4]{n^3}} \|x_{2t'}\| \right) \\
&\quad + O(1) \frac{1}{n^2 \sqrt{n}} E \sum_{s=2}^n \sum_{\ell=1}^{s-1} \sum_{t=1}^n \|Z_k(x_{1t})\|^2 \|\dot{Z}_k(x_{1t})\|^2 \|Z_k(x_{1s})\|^2 \|Z_k(x_{1\ell})\|^2 |x_{1t}|^2 \\
&\quad + O(1) \frac{1}{n^3 \sqrt{n}} E \sum_{s=2}^n \sum_{\ell=1}^{s-1} \sum_{t=1}^n \|Z_k(x_{1t})\|^2 \|\dot{Z}_k(x_{1t})\|^2 \|Z_k(x_{1s})\|^2 \|Z_k(x_{1\ell})\|^2 |x_{2t}|^2
\end{aligned}$$

$$\begin{aligned}
& + 4\sigma_e^4 \frac{1}{n^2} E \sum_{s=2}^n \sum_{\ell=1}^{s-1} \sum_{t=2}^n \sum_{t'=1}^{t-1} |\mathcal{U}_k(x_{1t}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t}, x_{1\ell})| \|\xi_{n,t}\| |\mathcal{U}_k(x_{1t'}, x_{1s}) \widetilde{\mathcal{U}}_k(x_{1t'}, x_{1\ell})| \|\xi_{n,t'}\| \\
\leq & O(1) \frac{k^4}{n^2 \sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int \|Z_k(x)\|^2 |x|^2 dx \\
& + O(1) \frac{k^3}{n^2 \sqrt{n}} \sum_{s=2}^n \sum_{t=1}^{s-1} \frac{1}{\sqrt{s-t}} \frac{1}{\sqrt{t}} \iint \|Z_k(y)\|^2 \|Z_k(x)\|^2 |y|^2 dx dy \\
& + O(1) \frac{k^4}{n^3 \sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|Z_k(x)\|^2 \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
& + O(1) \frac{k^3}{n^3 \sqrt{n}} \sum_{s=2}^n \sum_{t=1}^{s-1} \frac{1}{\sqrt{s-t}} \frac{1}{\sqrt{t}} \iiint \|Z_k(y)\|^2 \|Z_k(x)\|^2 \|\sqrt{t}w\|^2 \phi(w) dx dy dw \\
& + O(1) \frac{k^3}{n^2 \sqrt{n}} E \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{t'=1}^{t-1} \|Z_k(x_{1t})\|^2 \|Z_k(x_{1t'})\|^2 \|Z_k(x_{1s})\|^2 |x_{1t}| |x_{1t'}| \\
& + O(1) \frac{k^3}{n^3 \sqrt{n}} E \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{t'=1}^{t-1} \|Z_k(x_{1t})\|^2 \|Z_k(x_{1t'})\|^2 \|Z_k(x_{1s})\|^2 \|x_{2t}\| \|x_{2t'}\| \\
& + O(1) \frac{k^2}{n^2 \sqrt{n}} \sum_{s=3}^n \sum_{\ell=2}^{s-1} \sum_{t=1}^{\ell-1} \frac{1}{\sqrt{s-\ell}} \frac{1}{\sqrt{\ell-t}} \frac{1}{\sqrt{t}} \iiint \|Z_k(z)\|^2 \|Z_k(x)\|^2 \|Z_k(y)\|^2 |z|^2 dx dy dz \\
& + O(1) \frac{k^2}{n^3 \sqrt{n}} \sum_{s=3}^n \sum_{\ell=2}^{s-1} \sum_{t=1}^{\ell-1} \frac{1}{\sqrt{s-\ell}} \frac{1}{\sqrt{\ell-t}} \frac{1}{\sqrt{t}} \\
& \times \int \cdots \int \|Z_k(z)\|^2 \|Z_k(x)\|^2 \|Z_k(y)\|^2 \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dy dz dw \\
& + O(1) \frac{1}{n^2} E \sum_{s=4}^n \sum_{\ell=3}^{s-1} \sum_{t=2}^{\ell-1} \sum_{t'=1}^{t-1} \|\dot{Z}_k(x_{1t})\| \|\mathcal{U}_k(x_{1t}, x_{1s})\| \|\mathcal{U}_k(x_{1t'}, x_{1s})\| \|\dot{Z}_k(x_{1t'})\| \\
& \quad \times \|Z_k(x_{1\ell})\|^2 \left( \frac{1}{\sqrt[4]{n}} |x_{1t}| + \frac{1}{\sqrt[4]{n^3}} \|x_{2t}\| \right) \left( \frac{1}{\sqrt[4]{n}} |x_{1t'}| + \frac{1}{\sqrt[4]{n^3}} \|x_{2t'}\| \right) \\
\leq & O(1) \frac{k^4}{n^2} k^2 + O(1) \frac{k^3}{n \sqrt{n}} k^3 + O(1) \frac{k^4}{n^2 \sqrt{n}} k + O(1) \frac{k^3}{n \sqrt{n}} k^2 \\
& + O(1) \frac{k^3}{n^2 \sqrt{n}} \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{t'=1}^{t-1} \frac{1}{\sqrt{s-t}} \frac{1}{\sqrt{t-t'}} \frac{1}{\sqrt{t'}} \iiint \|Z_k(y)\|^2 \|Z_k(z)\|^2 \|Z_k(x)\|^2 |y| |z| dx dy dz \\
& + O(1) \frac{k^3}{n^3 \sqrt{n}} \sum_{s=3}^n \sum_{t=2}^{s-1} \sum_{t'=1}^{t-1} \frac{1}{\sqrt{s-t}} \frac{1}{\sqrt{t-t'}} \frac{1}{\sqrt{t'}} \int \cdots \int \|Z_k(y)\|^2 \|Z_k(z)\|^2 \\
& \times \|Z_k(x)\|^2 \|\sqrt{t-t'}w + \sqrt{t'}v\| \|\sqrt{t'}v\| \phi(w) \phi(v) (1 + o(1)) dx \cdots dv \\
& + O(1) \frac{k^2}{n} k^4 + O(1) \frac{1}{n^2 \sqrt{n}} \sum_{s=4}^n \sum_{\ell=3}^{s-1} \sum_{t=2}^{\ell-1} \sum_{t'=1}^{t-1} \frac{1}{\sqrt{s-\ell}} \frac{1}{\sqrt{\ell-t}} \frac{1}{\sqrt{t-t'}} \frac{1}{\sqrt{t'}} \int \cdots \int_{x \neq z \neq x'} \\
& \times \|\dot{Z}_k(z)\| \|\mathcal{U}_k(z, x)\| \|\mathcal{U}_k(x', x)\| \|\dot{Z}_k(x')\| \|Z_k(y)\|^2 |z| |x'| dx dy dz dx'
\end{aligned}$$

$$\begin{aligned}
& + O(1) \frac{1}{n^3 \sqrt{n}} \sum_{s=4}^n \sum_{\ell=3}^{s-1} \sum_{t=2}^{\ell-1} \sum_{t'=1}^{t-1} \frac{1}{\sqrt{s-\ell}} \frac{1}{\sqrt{\ell-t}} \frac{1}{\sqrt{t-t'}} \frac{1}{\sqrt{t'}} \int \cdots \int_{x \neq z \neq x'} \\
& \times \|\dot{Z}_k(z)\| \|\mathcal{U}_k(z, x)\| \|\mathcal{U}_k(x', x)\| \|\dot{Z}_k(x')\| \|Z_k(y)\|^2 \\
& \times \|\sqrt{t-t'}w + \sqrt{t'}v\| \|\sqrt{t'}v\| \|\phi(w)\phi(v)(1+o(1))\| dx dy dz dx' dw dv \\
& = o(1) + O(1) \frac{k^2}{n} k^5 + O(1) \frac{k^2}{n} k^3 + O(1) \frac{k^4}{\sqrt{n}} \iiint_{x \neq z \neq x'} |\mathcal{U}_k(z, x)| |\mathcal{U}_k(x', x)| dx dz dx' \rightarrow 0,
\end{aligned}$$

by virtue of Assumption B, Lemma A.6, where several relations are used, such as  $|\widetilde{\mathcal{U}}_k(x, y)| \leq \|\dot{Z}_k(x)\| \|Z_k(y)\|$  and  $\int |x| \|Z(x)\|^2 dx \leq O(1) \int |x|^2 \|Z(x)\|^2 dx = O(1)k^2$  as well as  $\|x \dot{Z}_k(x)\|^2 \leq O(1)k^3$  uniformly in  $x$ . We also omit some items in the above derivation which are obviously not larger than the existing items. We mean we ignore the cross-product terms in

$$\begin{aligned}
& \left( \frac{1}{\sqrt[4]{n}} |x_{1t}| + \frac{1}{\sqrt[4]{n^3}} \|x_{2t}\| \right) \\
& \left( \frac{1}{\sqrt[4]{n}} |x_{1t'}| + \frac{1}{\sqrt[4]{n^3}} \|x_{2t'}\| \right).
\end{aligned}$$

Next, consider  $T_{43}(2)$ . Observe that

$$\begin{aligned}
E\|T_{43}(2)\| & \leq \frac{1}{n} E \sum_{t=1}^n \|Z_k(x_{1t})^\top Z^\top \gamma\| \|\dot{Z}_k(x_{1t})^\top Z^\top \gamma\| \|\xi_{n,t}\| \\
& \leq \frac{1}{n} E \sum_{t=1}^n \sum_{s=1}^n \sum_{\ell=1}^n |\mathcal{U}_k(x_{1t}, x_{1s}) \gamma_k(x_{1s})| |\widetilde{\mathcal{U}}_k(x_{1t}, x_{1\ell}) \gamma_k(x_{1\ell})| \|\xi_{n,t}\| \\
& = \frac{1}{n} E \sum_{t=1}^n |\mathcal{U}_k(x_{1t}, x_{1t}) \gamma_k(x_{1t})| |\widetilde{\mathcal{U}}_k(x_{1t}, x_{1t}) \gamma_k(x_{1t})| \|\xi_{n,t}\| \\
& \quad + 6 \frac{1}{n} E \sum_{t=2}^n \sum_{s=1}^{t-1} |\mathcal{U}_k(x_{1t}, x_{1s}) \gamma_k(x_{1s})| |\widetilde{\mathcal{U}}_k(x_{1t}, x_{1s}) \gamma_k(x_{1s})| \|\xi_{n,t}\| \\
& \quad + 4 \frac{1}{n} E \sum_{t=3}^n \sum_{s=2}^{t-1} \sum_{\ell=1}^{s-1} |\mathcal{U}_k(x_{1t}, x_{1s}) \gamma_k(x_{1s})| |\widetilde{\mathcal{U}}_k(x_{1t}, x_{1\ell}) \gamma_k(x_{1\ell})| \|\xi_{n,t}\| \\
& \leq \frac{1}{n} E \sum_{t=1}^n \|Z_k(x_{1t})\|^3 \|\dot{Z}_k(x_{1t})\| |\gamma_k(x_{1t})|^2 \left( \frac{1}{\sqrt[4]{n}} |x_{1t}| + \frac{1}{\sqrt[4]{n^3}} \|x_{2t}\| \right) \\
& \quad + 6 \frac{1}{n} E \sum_{t=2}^n \sum_{s=1}^{t-1} \|Z_k(x_{1t})\| \|Z_k(x_{1s})\|^2 \|\dot{Z}_k(x_{1t})\| |\gamma_k(x_{1s})|^2 \left( \frac{1}{\sqrt[4]{n}} |x_{1t}| + \frac{1}{\sqrt[4]{n^3}} \|x_{2t}\| \right) \\
& \quad + 4 \frac{1}{n} E \sum_{t=3}^n \sum_{s=2}^{t-1} \sum_{\ell=1}^{s-1} \|Z_k(x_{1t})\| \|Z_k(x_{1s})\| \|\dot{Z}_k(x_{1t})\| \|Z_k(x_{1\ell})\| |\gamma_k(x_{1s}) \gamma_k(x_{1\ell})| \\
& \quad \times \left( \frac{1}{\sqrt[4]{n}} |x_{1t}| + \frac{1}{\sqrt[4]{n^3}} \|x_{2t}\| \right) \\
& \leq O(1) \frac{k^{5/2}}{n \sqrt[4]{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\gamma_k(x)|^2 |x| dx \\
& \quad + O(1) \frac{k^{5/2}}{n \sqrt[4]{n^3}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\gamma_k(x)|^2 \|\sqrt{t}w\| \|\phi(w)\| dw dx (1+o(1))
\end{aligned}$$

$$\begin{aligned}
& + O(1) \frac{k}{n\sqrt[4]{n}} \sum_{t=2}^n \sum_{s=1}^{t-1} \frac{1}{\sqrt{t-s}} \frac{1}{\sqrt{s}} \iint \|Z_k(x)\| \|\dot{Z}_k(x)\| |\gamma_k(y)|^2 |x| dx dy \\
& + O(1) \frac{k}{n\sqrt[4]{n^3}} \sum_{t=2}^n \sum_{s=1}^{t-1} \frac{1}{\sqrt{t-s}} \frac{1}{\sqrt{s}} \\
& \times \iiint \|Z_k(x)\| \|\dot{Z}_k(x)\| |\gamma_k(y)|^2 \|\sqrt{t}w\| \phi(w) dw dx dy (1 + o(1)) \\
& + O(1) \frac{1}{n\sqrt[4]{n}} \sum_{t=3}^n \sum_{s=2}^{t-1} \sum_{\ell=1}^{s-1} \frac{1}{\sqrt{t-s}} \frac{1}{\sqrt{s-\ell}} \frac{1}{\sqrt{\ell}} \\
& \times \iiint |x| \|Z_k(x)\| \|Z_k(y)\| \|\dot{Z}_k(x)\| \|Z_k(z)\| |\gamma_k(y)\gamma_k(z)| dx dy dz \\
& + O(1) \frac{1}{n\sqrt[4]{n^3}} \sum_{t=3}^n \sum_{s=2}^{t-1} \sum_{\ell=1}^{s-1} \frac{1}{\sqrt{t-s}} \frac{1}{\sqrt{s-\ell}} \frac{1}{\sqrt{\ell}} \int \cdots \int \|Z_k(x)\| \|Z_k(y)\| \|\dot{Z}_k(x)\| \|Z_k(z)\| \\
& \times |\gamma_k(y)\gamma_k(z)| \|\sqrt{t-s}w + \sqrt{s-\ell}w' + \sqrt{\ell}w''\| \phi(w)\phi(w')\phi(w'') dx dy dz dw dw' dw'' \\
& = \frac{k^{5/2}}{\sqrt[4]{n^3}} k^{-(m-1)} + \frac{k^{5/2}}{\sqrt[4]{n^3}} k^{-m} + \frac{k}{\sqrt[4]{n}} k^{-m} \left( \int \|Z_k(x)\|^2 x^2 dx \int \|\dot{Z}_k(x)\|^2 dx \right)^{1/2} \\
& + o(1) \frac{k}{\sqrt[4]{n}} k^{-m} \left( \int \|Z_k(x)\|^2 dx \int \|\dot{Z}_k(x)\|^2 dx \right)^{1/2} \\
& + O(1) \sqrt[4]{n} \left( \int \|Z_k(x)\|^2 x^2 dx \int \|\dot{Z}_k(x)\|^2 dx \right)^{1/2} \int \|Z_k(z)\|^2 dz \int |\gamma_k(z)|^2 dz \\
& + O(1) \sqrt[4]{n} \left( \int \|Z_k(x)\|^2 dx \int \|\dot{Z}_k(x)\|^2 dx \right)^{1/2} \int \|Z_k(z)\|^2 dz \int |\gamma_k(z)|^2 dz \\
& = o(1) + \frac{k}{\sqrt[4]{n}} k^{-m} k^2 + o(1) \frac{k}{\sqrt[4]{n}} k^{-m} k^{3/2} + O(1) \sqrt[4]{n} k^3 k^{-m} + O(1) \sqrt[4]{n} k^{5/2} k^{-m} = o(1),
\end{aligned}$$

by Cauchy-Schwarz inequality, Lemma A.1 and Assumption B.

Note that  $T_{43}(3) = o_P(1)$  and  $T_{43}(4) = o_P(1)$  can be shown similarly. Next, using martingale difference structure again we have

$$\begin{aligned}
E\|T_{44}\|^2 & = E \left\| \sum_{t=1}^n e_t \dot{\gamma}_k(x_{1t}) \xi_{n,t} \right\|^2 = \sigma_e^2 E \sum_{t=1}^n |\dot{\gamma}_k(x_{1t})|^2 \|\xi_{n,t}\|^2 \\
& = \sigma_e^2 \frac{1}{\sqrt{n}} E \sum_{t=1}^n |\dot{\gamma}_k(x_{1t})|^2 |x_{1t}|^2 + \sigma_e^2 \frac{1}{\sqrt{n^3}} E \sum_{t=1}^n |\dot{\gamma}_k(x_{1t})|^2 \|x_{2t}\|^2 \\
& \leq O(1) \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\dot{\gamma}_k(x)|^2 |x|^2 dx + O(1) \frac{1}{\sqrt{n^3}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\dot{\gamma}_k(x)|^2 \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dw \\
& = O(1) \int |\dot{\gamma}_k(x)|^2 |x|^2 dx = o(1) k^{-(m-2)} = o(1),
\end{aligned}$$

in view of (7) in Lemma A.1. Moreover,

$$E\|T_{45}\| = E \left\| \sum_{t=1}^n \gamma_k(x_{1t}) \dot{\gamma}_k(x_{1t}) \xi_{n,t} \right\| \leq E \sum_{t=1}^n |\gamma_k(x_{1t}) \dot{\gamma}_k(x_{1t})| \|\xi_{n,t}\|$$

$$\begin{aligned}
&\leq \frac{1}{\sqrt[4]{n}} E \sum_{t=1}^n |\gamma_k(x_{1t}) \dot{\gamma}_k(x_{1t})| |x_{1t}| + \frac{1}{\sqrt[4]{n^3}} E \sum_{t=1}^n |\gamma_k(x_{1t}) \dot{\gamma}_k(x_{1t})| |x_{2t}| \\
&\leq \frac{O(1)}{\sqrt[4]{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\gamma_k(x) \dot{\gamma}_k(x)| |x| dx + \frac{O(1)}{\sqrt[4]{n^3}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\gamma_k(x) \dot{\gamma}_k(x)| |\sqrt{tw}| \phi(w) dx dw \\
&\leq O(1) \sqrt[4]{n} \left( \int |x \gamma_k(x)|^2 dx \int |\dot{\gamma}_k(x)|^2 dx \right)^{1/2} = o(1) \sqrt[4]{n} k^{-(m-1)} = o(1) n^{1/4 - (m-1)\kappa} = o(1),
\end{aligned}$$

due to (7) of Lemma A.1 and Assumption B. Furthermore,

$$\begin{aligned}
\|T_{46}\| &= \left\| \sum_{t=1}^n Z_k(x_{1t})^\top (\hat{c} - c) \dot{\gamma}_k(x_{1t}) \xi_{n,t} \right\| \leq \sum_{t=1}^n |Z_k(x_{1t})^\top (\hat{c} - c) \dot{\gamma}_k(x_{1t})| |\xi_{n,t}| \\
&\leq \sum_{t=1}^n \|Z_k(x_{1t})\| \|\hat{c} - c\| |\dot{\gamma}_k(x_{1t})| |\xi_{n,t}| = O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n \|Z_k(x_{1t})\| |\dot{\gamma}_k(x_{1t})| |\xi_{n,t}|
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\sqrt{k}}{\sqrt[4]{n}} E \sum_{t=1}^n \|Z_k(x_{1t})\| |\dot{\gamma}_k(x_{1t})| |\xi_{n,t}| \\
&\leq \frac{\sqrt{k}}{\sqrt{n}} E \sum_{t=1}^n \|Z_k(x_{1t})\| |\dot{\gamma}_k(x_{1t})| |x_{1t}| + \frac{\sqrt{k}}{n} E \sum_{t=1}^n \|Z_k(x_{1t})\| |\dot{\gamma}_k(x_{1t})| |x_{2t}| \\
&\leq O(1) \frac{\sqrt{k}}{\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int \|Z_k(x)\| |\dot{\gamma}_k(x)| |x| dx \\
&\quad + O(1) \frac{\sqrt{k}}{n} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|Z_k(x)\| |\dot{\gamma}_k(x)| |\sqrt{tw}| \phi(w) (1 + o(1)) dx dw \\
&\leq O(1) \sqrt{k} \left( \int \|Z_k(x)\|^2 dx \int |x \dot{\gamma}_k(x)|^2 dx \right)^{1/2} \\
&= o(1) \sqrt{k} (kk^{-(m-2)})^{1/2} = o(1) k^{-(m-4)/2} = o(1).
\end{aligned}$$

The proof for  $T_2, T_3, T_4$  is complete.

(b) The hessian. We are about to show  $I_j = o_P(1)$  for  $j = 2, 3, 4$ . Noting that  $\hat{g}_n(x_{1t}) - \dot{g}(x_{1t}) = \dot{Z}_k(x_{1t})^\top (\hat{c} - c) - \dot{\gamma}_k(x_{1t})$ , we may write

$$\begin{aligned}
I_2 &= \sum_{t=1}^n \dot{g}(x_{1t}) [\dot{Z}_k(x_{1t})^\top (\hat{c} - c) - \dot{\gamma}_k(x_{1t})] \xi_{n,t} \xi_{n,t}^\top, \\
I_3 &= \sum_{t=1}^n [\dot{Z}_k(x_{1t})^\top (\hat{c} - c) - \dot{\gamma}_k(x_{1t})]^2 \xi_{n,t} \xi_{n,t}^\top.
\end{aligned}$$

We start on  $I_3$  since if we can show  $I_3 = o_P(1)$ , we may have  $I_2 = o_P(1)$  by Cauchy-Schwarz inequality and  $I_1 = O_P(1)$ . Observe that

$$\|I_3\| = \left\| \sum_{t=1}^n [\dot{Z}_k(x_{1t})^\top (\hat{c} - c) - \dot{\gamma}_k(x_{1t})]^2 \xi_{n,t} \xi_{n,t}^\top \right\| \leq \sum_{t=1}^n [\dot{Z}_k(x_{1t})^\top (\hat{c} - c) - \dot{\gamma}_k(x_{1t})]^2 \|\xi_{n,t} \xi_{n,t}^\top\|$$

$$\begin{aligned}
&\leq 2 \sum_{t=1}^n [|\dot{Z}_k(x_{1t})^\top(\hat{c} - c)|^2 + |\dot{\gamma}_k(x_{1t})|^2] \|\xi_{n,t} \xi_{n,t}^\top\| \leq 2 \sum_{t=1}^n [\|\dot{Z}_k(x_{1t})\|^2 \|\hat{c} - c\|^2 + |\dot{\gamma}_k(x_{1t})|^2] \|\xi_{n,t} \xi_{n,t}^\top\| \\
&= O_P(1) \frac{k}{\sqrt{n}} \sum_{t=1}^n \|\dot{Z}_k(x_{1t})\|^2 \|\xi_{n,t} \xi_{n,t}^\top\| + 2 \sum_{t=1}^n |\dot{\gamma}_k(x_{1t})|^2 \|\xi_{n,t} \xi_{n,t}^\top\|
\end{aligned}$$

and noting that  $\|\xi_{n,t} \xi_{n,t}^\top\| = \|\xi_{n,t}\|^2 = \frac{1}{\sqrt{n}} x_{1t}^2 + \frac{1}{\sqrt{n}^3} \|x_{2t}\|^2$ ,

$$\begin{aligned}
&\frac{k}{\sqrt{n}} E \sum_{t=1}^n \|\dot{Z}_k(x_{1t})\|^2 \|\xi_{n,t} \xi_{n,t}^\top\| = \frac{k}{\sqrt{n}} E \sum_{t=1}^n \|\dot{Z}_k(x_{1t})\|^2 \left( \frac{1}{\sqrt{n}} x_{1t}^2 + \frac{1}{\sqrt{n}^3} \|x_{2t}\|^2 \right) \\
&\leq O(1) \frac{k}{n} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int \|\dot{Z}_k(x)\|^2 x^2 dx + O(1) \frac{k}{n^2} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|\dot{Z}_k(x)\|^2 \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
&= O(1) \frac{k}{\sqrt{n}} k^3 + O(1) \frac{k}{\sqrt{n}} k^2 = O(1) \frac{1}{\sqrt{n}} k^4 = o(1)
\end{aligned}$$

using (4) of Lemma A.1 and Assumption B, and furthermore,

$$\begin{aligned}
&E \sum_{t=1}^n |\dot{\gamma}_k(x_{1t})|^2 \|\xi_{n,t} \xi_{n,t}^\top\| = E \sum_{t=1}^n |\dot{\gamma}_k(x_{1t})|^2 \left( \frac{1}{\sqrt{n}} x_{1t}^2 + \frac{1}{\sqrt{n}^3} \|x_{2t}\|^2 \right) \\
&\leq O(1) \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\dot{\gamma}_k(x)|^2 x^2 dx \\
&\quad + O(1) \frac{1}{n\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\dot{\gamma}_k(x)|^2 \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
&= O(1) \int |\dot{\gamma}_k(x)|^2 x^2 dx + O(1) \int |\dot{\gamma}_k(x)|^2 dx = o(1) k^{-(m-2)} = o(1),
\end{aligned}$$

once again by virtue of (7) of Lemma A.1 and Assumption B, proving  $I_3 = o_P(1)$ .

Since  $y_t - \hat{g}_n(x_{1t}) = e_t + \gamma_k(x_{1t}) - Z_k(x_{1t})^\top(\hat{c} - c)$ , we have

$$\begin{aligned}
I_4 &= \sum_{t=1}^n [y_t - \hat{g}_n(x_{1t})] \ddot{g}_n(x_{1t}) \xi_{n,t} \xi_{n,t}^\top \\
&= \sum_{t=1}^n [y_t - \hat{g}_n(x_{1t})] \ddot{g}(x_{1t}) \xi_{n,t} \xi_{n,t}^\top + \sum_{t=1}^n [y_t - \hat{g}_n(x_{1t})] [\ddot{Z}_k(x_{1t})^\top(\hat{c} - c) - \ddot{\gamma}_k(x_{1t})] \xi_{n,t} \xi_{n,t}^\top \\
&= \sum_{t=1}^n e_t \ddot{g}(x_{1t}) \xi_{n,t} \xi_{n,t}^\top + \sum_{t=1}^n \gamma_k(x_{1t}) \ddot{g}(x_{1t}) \xi_{n,t} \xi_{n,t}^\top - \sum_{t=1}^n Z_k(x_{1t})^\top(\hat{c} - c) \ddot{g}(x_{1t}) \xi_{n,t} \xi_{n,t}^\top \\
&\quad + \sum_{t=1}^n [y_t - \hat{g}_n(x_{1t})] [\ddot{Z}_k(x_{1t})^\top(\hat{c} - c) - \ddot{\gamma}_k(x_{1t})] \xi_{n,t} \xi_{n,t}^\top \\
&:= I_{41} + I_{42} - I_{43} + I'_4, \quad \text{say.}
\end{aligned}$$

As discussed in the calculation of  $T_{31}$ , we may take advantage of the martingale difference structure in the following computation.

$$E \|I_{41}\|^2 = E \left\| \sum_{t=1}^n e_t \ddot{g}(x_{1t}) \xi_{n,t} \xi_{n,t}^\top \right\|^2 = \sigma_e^2 E \sum_{t=1}^n |\ddot{g}(x_{1t})|^2 \|\xi_{n,t} \xi_{n,t}^\top\|^2$$

$$\begin{aligned}
&= \sigma_e^2 E \sum_{t=1}^n |\ddot{g}(x_{1t})| \left( \frac{1}{n} x_{1t}^4 + 2 \frac{1}{n^2} |x_{1t}|^2 \|x_{2t}\|^2 + \frac{1}{n^3} \|x_{2t}\|^4 \right) \\
&\leq O(1) \frac{1}{n} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\ddot{g}(x)| x^4 dx + O(1) \frac{1}{n^2} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\ddot{g}(x)| |x|^2 \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
&\quad + O(1) \frac{1}{n^3} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\ddot{g}(x)| \|\sqrt{t}w\|^4 \phi(w) (1 + o(1)) dx dw \\
&= O(1) \frac{1}{\sqrt{n}} = o(1),
\end{aligned}$$

and

$$\begin{aligned}
E \|I_{42}\| &= E \left\| \sum_{t=1}^n \gamma_k(x_{1t}) \ddot{g}(x_{1t}) \xi_{n,t} \xi_{n,t}^\top \right\| \leq E \sum_{t=1}^n |\gamma_k(x_{1t}) \ddot{g}(x_{1t})| \|\xi_{n,t} \xi_{n,t}^\top\| \\
&= E \sum_{t=1}^n |\gamma_k(x_{1t}) \ddot{g}(x_{1t})| \left( \frac{1}{\sqrt{n}} x_{1t}^2 + \frac{1}{\sqrt{n}^3} \|x_{2t}\|^2 \right) \\
&\leq O(1) \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\gamma_k(x) \ddot{g}(x)| x^2 dx \\
&\quad + O(1) \frac{1}{n\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\gamma_k(x) \ddot{g}(x)| \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
&= o(1) k^{-(m-1)/2-1/12} \int |\ddot{g}(x)| x^2 dx = o(1).
\end{aligned}$$

Moreover, notice that

$$\begin{aligned}
\|I_{43}\| &= \left\| \sum_{t=1}^n Z_k(x_{1t})^\top (\hat{c} - c) \ddot{g}(x_{1t}) \xi_{n,t} \xi_{n,t}^\top \right\| \leq \sum_{t=1}^n |Z_k(x_{1t})^\top (\hat{c} - c) \ddot{g}(x_{1t})| \|\xi_{n,t} \xi_{n,t}^\top\| \\
&\leq \|\hat{c} - c\| \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{g}(x_{1t})\| \|\xi_{n,t}\|^2 = O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{g}(x_{1t})\| \|\xi_{n,t}\|^2
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\sqrt{k}}{\sqrt[4]{n}} E \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{g}(x_{1t})\| \|\xi_{n,t}\|^2 = \frac{\sqrt{k}}{\sqrt[4]{n}} E \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{g}(x_{1t})\| \left( \frac{1}{\sqrt{n}} x_{1t}^2 + \frac{1}{\sqrt{n}^3} \|x_{2t}\|^2 \right) \\
&\leq O(1) \frac{\sqrt{k}}{\sqrt[4]{n} \sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int \|Z_k(x)\| \|\ddot{g}(x)\| x^2 dx \\
&\quad + O(1) \frac{\sqrt{k}}{n\sqrt{n} \sqrt[4]{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|Z_k(x)\| \|\ddot{g}(x)\| \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
&\leq O(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \left( \int \|Z_k(x)\|^2 dx \int |\ddot{g}(x)|^2 x^4 dx \right)^{1/2} = O(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \sqrt{k} = O(1) \frac{1}{\sqrt[4]{n}} k = o(1).
\end{aligned}$$

Now, turn to  $I'_4$ . Once again, noting  $y_t - \hat{g}_n(x_{1t}) = e_t + \gamma_k(x_{1t}) - Z_k(x_{1t})^\top (\hat{c} - c)$ ,

$$I'_4 = \sum_{t=1}^n [y_t - \hat{g}_n(x_{1t})] [\ddot{Z}_k(x_{1t})^\top (\hat{c} - c) - \ddot{\gamma}_k(x_{1t})] \xi_{n,t} \xi_{n,t}^\top$$



$$\begin{aligned}
&= \sum_{t=1}^n e_t \ddot{Z}_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} \xi_{n,t}^\top - \sum_{t=1}^n e_t \ddot{\gamma}_k(x_{1t}) \xi_{n,t} \xi_{n,t}^\top \\
&\quad + \sum_{t=1}^n \gamma_k(x_{1t}) \ddot{Z}_k(x_{1t})^\top (\hat{c} - c) - \sum_{t=1}^n \gamma_k(x_{1t}) \ddot{\gamma}_k(x_{1t}) \xi_{n,t} \xi_{n,t}^\top \\
&\quad - \sum_{t=1}^n Z_k(x_{1t})^\top (\hat{c} - c) \ddot{Z}_k(x_{1t})^\top (\hat{c} - c) + \sum_{t=1}^n Z_k(x_{1t})^\top (\hat{c} - c) \ddot{\gamma}_k(x_{1t}) \xi_{n,t} \xi_{n,t}^\top \\
&:= I'_{41} - I'_{42} + I'_{43} - I'_{44} - I'_{45} + I'_{46}, \quad \text{say.}
\end{aligned}$$

We shall deal with them one by one. Temporarily, denote  $(a_{ij}) = \xi_{n,t} \xi_{n,t}^\top$  to facilitate the following calculation.

$$\begin{aligned}
\|I'_{41}\|^2 &= \left\| \sum_{t=1}^n e_t \ddot{Z}_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} \xi_{n,t}^\top \right\|^2 = \sum_i \sum_j \left( \sum_{t=1}^n a_{ij} e_t \ddot{Z}_k(x_{1t})^\top (\hat{c} - c) \right)^2 \\
&= \sum_i \sum_j \left[ \left( \sum_{t=1}^n a_{ij} e_t \ddot{Z}_k(x_{1t}) \right)^\top (\hat{c} - c) \right]^2 \leq \|\hat{c} - c\|^2 \sum_i \sum_j \left\| \sum_{t=1}^n a_{ij} e_t \ddot{Z}_k(x_{1t}) \right\|^2 \\
&= O_P(1) \frac{k}{\sqrt{n}} \sum_i \sum_j \left\| \sum_{t=1}^n a_{ij} e_t \ddot{Z}_k(x_{1t}) \right\|^2
\end{aligned}$$

and by martingale difference structure we have

$$\begin{aligned}
&\frac{k}{\sqrt{n}} E \sum_i \sum_j \left\| \sum_{t=1}^n a_{ij} e_t \ddot{Z}_k(x_{1t}) \right\|^2 = \sigma_e^2 \frac{k}{\sqrt{n}} E \sum_i \sum_j \sum_{t=1}^n a_{ij}^2 \|\ddot{Z}_k(x_{1t})\|^2 \\
&= \sigma_e^2 \frac{k}{\sqrt{n}} E \sum_{t=1}^n \|\ddot{Z}_k(x_{1t})\|^2 \|\xi_{n,t} \xi_{n,t}^\top\|^2 = \sigma_e^2 \frac{k}{\sqrt{n}} E \sum_{t=1}^n \|\ddot{Z}_k(x_{1t})\|^2 \|\xi_{n,t}\|^4 \\
&= \sigma_e^2 \frac{k}{\sqrt{n}} E \sum_{t=1}^n \|\ddot{Z}_k(x_{1t})\|^2 \left( \frac{1}{n} x_{1t}^4 + 2 \frac{1}{n^2} |x_{1t}|^2 \|x_{2t}\|^2 + \frac{1}{n^3} \|x_{2t}\|^4 \right) \\
&\leq O(1) \frac{k}{n\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int \|\ddot{Z}_k(x)\|^2 x^4 dx \\
&\quad + O(1) \frac{k}{n^2\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|\ddot{Z}_k(x)\|^2 |x|^2 \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
&\quad + O(1) \frac{k}{n^3\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|\ddot{Z}_k(x)\|^2 \|\sqrt{t}w\|^4 \phi(w) (1 + o(1)) dx dw \\
&= O(1) \frac{k}{n} \int \|\ddot{Z}_k(x)\|^2 x^4 dx + O(1) \frac{k}{n} \int \|\ddot{Z}_k(x)\|^2 |x|^2 dx + O(1) \frac{k}{n} \int \|\ddot{Z}_k(x)\|_2^2 dx \\
&= O(1) \frac{k}{n} k^5 = O(1) n^{-1+6\kappa} = o(1),
\end{aligned}$$

using (4) of Lemma A.1 and Assumption B. Similarly,

$$E\|I'_{42}\|^2 = E \left\| \sum_{t=1}^n e_t \ddot{\gamma}_k(x_{1t}) \xi_{n,t} \xi_{n,t}^\top \right\|^2 = E \sum_i \sum_j \left( \sum_{t=1}^n a_{ij} e_t \ddot{\gamma}_k(x_{1t}) \right)^2$$

$$\begin{aligned}
&= \sigma_e^2 E \sum_i \sum_j \sum_{t=1}^n (a_{ij} \ddot{\gamma}_k(x_{1t}))^2 = \sigma_e^2 E \sum_{t=1}^n \ddot{\gamma}_k^2(x_{1t}) \|\xi_{n,t} \xi_{n,t}^\top\|_2^2 \\
&= \sigma_e^2 E \sum_{t=1}^n \ddot{\gamma}_k^2(x_{1t}) \left( \frac{1}{n} x_{1t}^4 + 2 \frac{1}{n^2} |x_{1t}|^2 \|x_{2t}\|^2 + \frac{1}{n^3} \|x_{2t}\|^4 \right) \\
&\leq O(1) \frac{1}{n} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int \ddot{\gamma}_k^2(x) x^4 dx \\
&\quad + O(1) \frac{1}{n^2} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \ddot{\gamma}_k^2(x) |x|^2 \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
&\quad + O(1) \frac{1}{n^3} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \ddot{\gamma}_k^2(x) \|\sqrt{t}w\|^4 \phi(w) (1 + o(1)) dx dw \\
&= O(1) \frac{1}{\sqrt{n}} \int \ddot{\gamma}_k^2(x) x^4 dx = o(1) \frac{1}{\sqrt{n}} k^{-(m-4)} = o(1).
\end{aligned}$$

Similar derivation gives

$$\begin{aligned}
\|I'_{43}\| &= \left\| \sum_{t=1}^n \gamma_k(x_{1t}) \ddot{Z}_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} \xi_{n,t}^\top \right\| \leq \|\hat{c} - c\| \sum_{t=1}^n |\gamma_k(x_{1t})| \|\ddot{Z}_k(x_{1t})\| \|\xi_{n,t} \xi_{n,t}^\top\| \\
&= O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n |\gamma_k(x_{1t})| \|\ddot{Z}_k(x_{1t})\| \|\xi_{n,t}\|^2
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\sqrt{k}}{\sqrt[4]{n}} E \sum_{t=1}^n |\gamma_k(x_{1t})| \|\ddot{Z}_k(x_{1t})\| \|\xi_{n,t}\|^2 = \frac{\sqrt{k}}{\sqrt[4]{n}} E \sum_{t=1}^n |\gamma_k(x_{1t})| \|\ddot{Z}_k(x_{1t})\| \left( \frac{1}{\sqrt{n}} x_{1t}^2 + \frac{1}{\sqrt{n}^3} \|x_{2t}\|^2 \right) \\
&\leq O(1) \frac{\sqrt{k}}{\sqrt[4]{n} \sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\gamma_k(x)| \|\ddot{Z}_k(x)\| x^2 dx \\
&\quad + O(1) \frac{\sqrt{k}}{\sqrt[4]{n} \sqrt{n}^3} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\gamma_k(x)| \|\ddot{Z}_k(x)\| \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
&\leq O(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \left( \int |\gamma_k(x)|^2 dx \int \|\ddot{Z}_k(x)\|^2 x^4 dx \right)^{1/2} = o(1) \frac{\sqrt{k}}{\sqrt[4]{n}} k^{-(m-5)/2} = o(1).
\end{aligned}$$

Meanwhile,

$$\begin{aligned}
E \|I'_{44}\| &= E \left\| \sum_{t=1}^n \gamma_k(x_{1t}) \dot{\gamma}_k(x_{1t}) \xi_{n,t} \xi_{n,t}^\top \right\| \leq E \sum_{t=1}^n |\gamma_k(x_{1t}) \dot{\gamma}_k(x_{1t})| \|\xi_{n,t}\|^2 \\
&= E \sum_{t=1}^n |\gamma_k(x_{1t}) \dot{\gamma}_k(x_{1t})| \left( \frac{1}{\sqrt{n}} x_{1t}^2 + \frac{1}{\sqrt{n}^3} \|x_{2t}\|^2 \right) \\
&\leq O(1) \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int |\gamma_k(x) \dot{\gamma}_k(x)| x^2 dx \\
&\quad + O(1) \frac{1}{\sqrt{n}^3} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint |\gamma_k(x) \dot{\gamma}_k(x)| \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw
\end{aligned}$$

$$\begin{aligned}
&\leq O(1) \left( \int |\gamma_k(x)|^2 x^2 dx \int |\ddot{\gamma}_k(x)|^2 x^2 dx \right)^{1/2} \\
&= o(1) (k^{-(m-1)} k^{-(m-3)})^{1/2} = o(1) k^{-(m-2)} = o(1).
\end{aligned}$$

Moreover,

$$\begin{aligned}
\|I'_{45}\| &= \left\| \sum_{t=1}^n Z_k(x_{1t})^\top (\hat{c} - c) \ddot{Z}_k(x_{1t})^\top (\hat{c} - c) \xi_{n,t} \xi_{n,t}^\top \right\| \\
&\leq \sum_{t=1}^n \|Z_k(x_{1t})^\top (\hat{c} - c) \ddot{Z}_k(x_{1t})^\top (\hat{c} - c)\| \|\xi_{n,t} \xi_{n,t}^\top\| \\
&\leq \|\hat{c} - c\|^2 \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{Z}_k(x_{1t})\| \|\xi_{n,t}\|^2 = O_P(1) \frac{k}{\sqrt{n}} \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{Z}_k(x_{1t})\| \|\xi_{n,t}\|^2,
\end{aligned}$$

and

$$\begin{aligned}
&\frac{k}{\sqrt{n}} E \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{Z}_k(x_{1t})\| \|\xi_{n,t}\|^2 = \frac{k}{\sqrt{n}} E \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{Z}_k(x_{1t})\| \left( \frac{1}{\sqrt{n}} x_{1t}^2 + \frac{1}{\sqrt{n}^3} \|x_{2t}\|^2 \right) \\
&\leq O(1) \frac{k}{n} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int \|Z_k(x)\| \|\ddot{Z}_k(x)\| x^2 dx \\
&\quad + O(1) \frac{1}{n^2} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|Z_k(x)\| \|\ddot{Z}_k(x)\| \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
&= O(1) \frac{k}{\sqrt{n}} \int \|Z_k(x)\| \|\ddot{Z}_k(x)\| x^2 dx \leq O(1) \frac{k}{\sqrt{n}} \left( \int \|Z_k(x)\|^2 x^2 dx \int \|\ddot{Z}_k(x)\|^2 x^2 dx \right)^{1/2} \\
&= O(1) \frac{k}{\sqrt{n}} (k^2 k^4)^{1/2} = O(1) \frac{1}{\sqrt{n}} k^4 = O(1) n^{-1/2+4\kappa} = o(1).
\end{aligned}$$

Finally,

$$\begin{aligned}
\|I'_{46}\| &\leq \sum_{t=1}^n \|Z_k(x_{1t})^\top (\hat{c} - c) \dot{\gamma}_k(x_{1t})\| \|\xi_{n,t} \xi_{n,t}^\top\| \leq \|\hat{c} - c\| \sum_{t=1}^n \|Z_k(x_{1t})\| \|\dot{\gamma}_k(x_{1t})\| \|\xi_{n,t} \xi_{n,t}^\top\| \\
&= O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n \|Z_k(x_{1t})\| \|\dot{\gamma}_k(x_{1t})\| \|\xi_{n,t}\|^2,
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\sqrt{k}}{\sqrt[4]{n}} E \sum_{t=1}^n \|Z_k(x_{1t})\| \|\dot{\gamma}_k(x_{1t})\| \|\xi_{n,t}\|^2 = \frac{\sqrt{k}}{\sqrt[4]{n}} E \sum_{t=1}^n \|Z_k(x_{1t})\| \|\dot{\gamma}_k(x_{1t})\| \left( \frac{1}{\sqrt{n}} x_{1t}^2 + \frac{1}{\sqrt{n}^3} \|x_{2t}\|^2 \right) \\
&\leq O(1) \frac{\sqrt{k}}{\sqrt[4]{n}^3} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int \|Z_k(x)\| \|\dot{\gamma}_k(x)\| x^2 dx \\
&\quad + \frac{\sqrt{k}}{\sqrt[4]{n} \sqrt[4]{n}^3} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|Z_k(x)\| \|\dot{\gamma}_k(x)\| \|\sqrt{t}w\|^2 \phi(w) (1 + o(1)) dx dw \\
&= O(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \int \|Z_k(x)\| \|\dot{\gamma}_k(x)\| x^2 dx \leq O(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \left( \int \|Z_k(x)\|^2 x^2 dx \int |\dot{\gamma}_k(x)|^2 x^2 dx \right)^{1/2}
\end{aligned}$$

$$=o(1)\frac{\sqrt{k}}{\sqrt[4]{n}}\left(k^2k^{-(m-3)}\right)^{1/2}=o(1)\frac{\sqrt{k}}{\sqrt[4]{n}}k^{-(m-5)/2}=o(1)$$

proving  $I_4 = o_P(1)$  and hence the whole proof is complete.  $\square$

*Proof of Theorem 3.2.* Motivated by (3.3) and the assertion of Theorem 3.1, in order to prove the theorem it is sufficient to show that

$$D_n(\hat{\alpha}_n - \alpha_0) = -[D_n^{-1}J_n(\alpha_0)D_n^{-1}]^{-1}D_n^{-1}S_n(\alpha_0) + o_P(1). \quad (\text{B.2})$$

Theorem 10.1 of Wooldridge (1994) provides a sufficient condition which we may utilize in this proof. Our task thereby is to check the four requirements in the aforementioned theorem, the first two of which, however, are trivially satisfied because of the assumption on  $\theta_0$ :  $\theta_0 \in \text{int}(\Theta)$  and Assumption B on the smoothness of  $g(x)$  function. To verify the third requirement, rewrite (3.3) as

$$S_n(\alpha_0) + J_n(\alpha_0)(\hat{\alpha}_n - \alpha_0) + [J_n(\alpha_n) - J_n(\alpha_0)](\hat{\alpha}_n - \alpha_0) = 0, \quad (\text{B.3})$$

where  $S_n(\alpha_0)$  and  $J_n(\alpha_0)$  are the score and hessian at  $\alpha_0$ , respectively, and  $J_n(\alpha_n)$  is the hessian matrix with rows evaluated at mean values that lie on the line segment joining  $\hat{\alpha}_n$  and  $\alpha_0$ . Define  $C_n = n^{-\delta}D_n$  for some  $\delta > 0$  such that  $C_nD_n^{-1} = o(1)$  as  $n \rightarrow \infty$ . It follows from (B.3) that

$$\begin{aligned} 0 &= D_n^{-1}S_n(\alpha_0) + D_n^{-1}J_n(\alpha_0)D_n^{-1}D_n(\hat{\alpha}_n - \alpha_0) + D_n^{-1}[J_n(\alpha_n) - J_n(\alpha_0)]D_n^{-1}D_n(\hat{\alpha}_n - \alpha_0) \\ &= D_n^{-1}S_n(\alpha_0) + D_n^{-1}J_n(\alpha_0)D_n^{-1}D_n(\hat{\alpha}_n - \alpha_0) + n^{-2\delta}C_n^{-1}[J_n(\alpha_n) - J_n(\alpha_0)]C_n^{-1}D_n(\hat{\alpha}_n - \alpha_0). \end{aligned} \quad (\text{B.4})$$

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

$$\sup_{\{\alpha: \|C_n(\alpha - \alpha_0)\| < 1\}} \|C_n^{-1}[J_n(\alpha) - J_n(\alpha_0)]C_n^{-1}\| = o_P(1). \quad (\text{B.5})$$

Towards this end, we partition the hessian matrix conformably with  $\alpha^\top = (\alpha^1, (\alpha^2)^\top)$  as

$$J_n(\alpha) = \begin{pmatrix} J_{n,11}(\alpha) & J_{n,12}(\alpha) \\ J_{n,21}(\alpha) & J_{n,22}(\alpha) \end{pmatrix} \quad (\text{B.6})$$

where

$$\begin{aligned} J_{n,11}(\alpha) &= \sum_{t=1}^n [\hat{g}_n(\eta_t)]^2 x_{1t}^2 - \sum_{t=1}^n [y_t - \hat{g}_n(\eta_t)] \ddot{g}_n(\eta_t) x_{1t}^2 \\ J_{n,12}(\alpha) &= \sum_{t=1}^n [\dot{g}_n(\eta_t)]^2 x_{1t} x_{2t}^\top - \sum_{t=1}^n [y_t - \hat{g}_n(\eta_t)] \ddot{g}_n(\eta_t) x_{1t} x_{2t}^\top \end{aligned}$$

$$\begin{aligned}
J_{n,21}(\alpha) &= J_{n,12}(\alpha)^\top \\
J_{n,22}(\alpha) &= \sum_{t=1}^n [\dot{\hat{g}}_n(\eta_t)]^2 x_{2t} x_{2t}^\top - \sum_{t=1}^n [y_t - \hat{g}_n(\eta_t)] \ddot{\hat{g}}_n(\eta_t) x_{2t} x_{2t}^\top
\end{aligned}$$

in which  $\eta_t = \alpha^\top Q^\top x_t = \alpha^1 x_{1t} + (\alpha^2)^\top x_{2t}$  for brevity.

Thus, for (B.5) to hold, it suffices to show that

$$\begin{aligned}
n^{-1/2+2\delta} |J_{n,11}(\alpha) - J_{n,11}(\alpha_0)| &= o_P(1) \\
n^{-1+2\delta} \|J_{n,21}(\alpha) - J_{n,21}(\alpha_0)\| &= o_P(1) \\
n^{-3/2+2\delta} \|J_{n,22}(\alpha) - J_{n,22}(\alpha_0)\| &= o_P(1)
\end{aligned} \tag{B.7}$$

uniformly in  $\alpha^1$  and  $\alpha^2$  satisfying

$$|\alpha^1 - \alpha_0^1| < n^{-1/4+\delta} \quad \text{and} \quad \|\alpha^2 - \alpha_0^2\| < n^{-3/4+\delta} \tag{B.8}$$

for some  $\delta > 0$  where  $\alpha_0^1 = 1$  and  $\alpha_0^2 = (0, \dots, 0)$ . Using the mean value theorem we have, for  $i, j \in \{1, 2\}$ ,

$$\begin{aligned}
&J_{n,ij}(\alpha) - J_{n,ij}(\alpha_0) \\
&= \sum_{t=1}^n \{[\dot{\hat{g}}_n(\eta_t)]^2 - [\dot{\hat{g}}_n(x_{1t})]^2\} x_{it} x_{jt}^\top - \sum_{t=1}^n \{[y_t - \hat{g}_n(\eta_t)] \ddot{\hat{g}}_n(\eta_t) - [y_t - \hat{g}_n(x_{1t})] \ddot{\hat{g}}_n(x_{1t})\} x_{it} x_{jt}^\top \\
&= 2 \sum_{t=1}^n \dot{\hat{g}}_n(\eta_t^*) \ddot{\hat{g}}_n(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) - \sum_{t=1}^n e_t [\ddot{\hat{g}}_n(\eta_t) - \ddot{\hat{g}}_n(x_{1t})] x_{it} x_{jt}^\top \\
&\quad - \sum_{t=1}^n \{[g(x_{1t}) - \hat{g}_n(\eta_t)] \ddot{\hat{g}}_n(\eta_t) - [g(x_{1t}) - \hat{g}_n(x_{1t})] \ddot{\hat{g}}_n(x_{1t})\} x_{it} x_{jt}^\top \\
&= 2 \sum_{t=1}^n \dot{\hat{g}}_n(\eta_t^*) \ddot{\hat{g}}_n(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) - \sum_{t=1}^n e_t [\ddot{\hat{g}}_n(\eta_t) - \ddot{\hat{g}}_n(x_{1t})] x_{it} x_{jt}^\top \\
&\quad - \sum_{t=1}^n [g(x_{1t}) - \hat{g}_n(x_{1t})] [\ddot{\hat{g}}_n(\eta_t) - \ddot{\hat{g}}_n(x_{1t})] x_{it} x_{jt}^\top + \sum_{t=1}^n [\hat{g}_n(\eta_t) - \hat{g}_n(x_{1t})] \ddot{\hat{g}}_n(\eta_t) x_{it} x_{jt}^\top \\
&= 2 \sum_{t=1}^n \dot{\hat{g}}_n(\eta_t^*) \ddot{\hat{g}}_n(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) - \sum_{t=1}^n e_t \ddot{\hat{g}}_n(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\
&\quad - \sum_{t=1}^n [g(x_{1t}) - \hat{g}_n(x_{1t})] \ddot{\hat{g}}_n(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) + \sum_{t=1}^n \dot{\hat{g}}_n(\eta_t^*) \ddot{\hat{g}}_n(\eta_t) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\
&= \sum_{\ell=1}^4 \Xi_{n,\ell}^{i,j}, \quad \text{say,}
\end{aligned}$$

where  $\eta_t^* := \alpha_*^\top Q^\top x_t$  and  $\alpha_*$  is on the segment connecting  $\alpha$  and  $\alpha_0$ . Definitely, each  $\alpha_*$  may not be equal to each other but they are all close to  $\alpha_0$ .

Define for any function  $f : \mathbb{R} \mapsto \mathbb{R}$

$$f_{\text{sup}}^\epsilon(x) = \sup_{|a - \alpha_0^1| < \epsilon} \sup_{|b| < \epsilon} |f(ax + b)| \tag{B.9}$$

for  $\epsilon > 0$  given.

If  $f(x)$  is bounded and integrable function, so is  $f_{\text{sup}}^\epsilon(x)$ , as shown in Lemma A1 of Park and Phillips (2000, p. 1266). Since  $|\alpha^1 - \alpha_0^1| < n^{-1/4+\delta}$  and  $\|\alpha^2\| < n^{-3/4+\delta}$ , as well as  $\sup_{1 \leq t \leq n} \|x_{2t}\|/\sqrt{n} = O_P(1)$ , we have for any  $\epsilon > 0$ ,

$$|f(\eta_t)| = |f(\alpha^1 x_{1t} + (\alpha^2)' x_{2t})| < f_{\text{sup}}^\epsilon(x_{1t}) + o_P(1)$$

for large  $n$ , uniformly in  $t$ .

On the other hand, suppose further that  $f(x)$  is continuous. Because of integrability, we may assume without loss of generality that for sufficient large  $|x|$ , say  $|x| > M$ ,  $|f(x)|$  is monotone. Hence,  $f_{\text{sup}}^\epsilon(x) = |f(x - \epsilon)|$  for  $x > M + \epsilon$  and  $f_{\text{sup}}^\epsilon(x) = |f(x + \epsilon)|$  for  $x < -M - \epsilon$ , which gives  $\int_{|x| > M + \epsilon} f_{\text{sup}}^\epsilon(x) dx = \int_{|x| > M + \epsilon} |f(x \pm \epsilon)| dx = \int_{|x| > M} |f(x)| dx$ . Meanwhile, on the interval  $[-M - \epsilon, M + \epsilon]$  the function  $f_{\text{sup}}^\epsilon(x)$  can be approximated by  $|f(x)|$  as accurate as we wish due to continuity as long as  $\epsilon$  is sufficient small. Therefore, we conclude

$$\int f_{\text{sup}}^\epsilon(x) dx = \int |f(x)| dx (1 + o(1)). \quad (\text{B.10})$$

Therefore, ignoring some unimportant constant,

$$\begin{aligned} \Xi_{n,1}^{i,j} &= \sum_{t=1}^n \dot{\hat{g}}_n(\eta_t^*) \ddot{\hat{g}}_n(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) = \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top \widehat{c} \ddot{Z}_k(\eta_t^*)^\top \widehat{c} x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &= \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top (c + \widehat{c} - c) \ddot{Z}_k(\eta_t^*)^\top (c + \widehat{c} - c) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &= \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top c \ddot{Z}_k(\eta_t^*)^\top c x_{it} x_{jt}^\top (\eta_t - x_{1t}) + \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top c \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &\quad + \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top (\widehat{c} - c) \ddot{Z}_k(\eta_t^*)^\top c x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &\quad + \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top (\widehat{c} - c) \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &= \sum_{t=1}^n [\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*)][\ddot{g}(\eta_t^*) - \ddot{\gamma}_k(\eta_t^*)] x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &\quad + \sum_{t=1}^n [\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*)] \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &\quad + \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top (\widehat{c} - c) [\ddot{g}(\eta_t^*) - \ddot{\gamma}_k(\eta_t^*)] x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &\quad + \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top (\widehat{c} - c) \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &= \sum_{\ell=1}^4 T_{1\ell}^{i,j}, \quad \text{say.} \end{aligned}$$

It follows that

$$\begin{aligned}
T_{11}^{i,j} &= \sum_{t=1}^n [\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*)][\ddot{g}(\eta_t^*) - \ddot{\gamma}_k(\eta_t^*)]x_{it}x_{jt}^\top(\eta_t - x_{1t}) \\
&= \sum_{t=1}^n \dot{g}(\eta_t^*)\ddot{g}(\eta_t^*)x_{it}x_{jt}^\top(\eta_t - x_{1t}) - \sum_{t=1}^n \dot{\gamma}_k(\eta_t^*)\ddot{\gamma}_k(\eta_t^*)x_{it}x_{jt}^\top(\eta_t - x_{1t}) \\
&\quad - \sum_{t=1}^n \dot{g}(\eta_t^*)\ddot{\gamma}_k(\eta_t^*)x_{it}x_{jt}^\top(\eta_t - x_{1t}) + \sum_{t=1}^n \dot{\gamma}_k(\eta_t^*)\ddot{g}(\eta_t^*)x_{it}x_{jt}^\top(\eta_t - x_{1t}).
\end{aligned}$$

By virtue of Lemma A.5 and (3) of Lemma A.1, the leading term in  $T_{11}^{i,j}$  is the first term  $\sum_{t=1}^n \dot{g}(\eta_t^*)\ddot{g}(\eta_t^*)x_{it}x_{jt}^\top(\eta_t - x_{1t})$  because  $\dot{\gamma}_k(x)$  and  $\ddot{\gamma}_k(x)$  are infinitesimal uniformly in  $x$ , and the integrals  $\int x^2\dot{\gamma}_k^2(x)dx$  and  $\int x^4\ddot{\gamma}_k^2(x)dx$  are infinitesimal as well. Whence, using notation  $f_{\text{sup}}$  by suppressing  $\epsilon$ , the first term in  $T_{11}^{i,j}$  is bounded in absolute value by

$$\begin{aligned}
&\sum_{t=1}^n (\dot{g}\ddot{g})_{\text{sup}}(x_{1t})\|x_{it}\|\|x_{jt}\|[\|\alpha^1 - \alpha_0^1\|\|x_{1t}\| + \|\alpha^2\|\|x_{2t}\|] \\
&\leq n^{-1/4+\delta} \sum_{t=1}^n (x\dot{g}\ddot{g})_{\text{sup}}(x_{1t})\|x_{it}\|\|x_{jt}\| + n^{-3/4+\delta} \sum_{t=1}^n (\dot{g}\ddot{g})_{\text{sup}}(x_{1t})\|x_{it}\|\|x_{jt}\|\|x_{2t}\|
\end{aligned}$$

where we have used the fact that  $f_{\text{sup}}(x)|x| \leq (xf)_{\text{sup}}(x)$ . In view of Lemma A.5,  $T_{11}^{i,j} = O_P(n^{1/4+\delta})$  for  $i = j = 1$ ,  $T_{11}^{i,j} = O_P(n^{3/4+\delta})$  for  $i = 1, j = 2$  or  $i = 2, j = 1$  and  $T_{11}^{i,j} = O_P(n^{5/4+\delta})$  for  $i = j = 2$ .

Next, consider  $T_{12}^{i,j}$ . It follows from Theorem 2.1 and (5) of Lemma A.1 that

$$\begin{aligned}
\|T_{12}^{i,j}\| &= \left\| \sum_{t=1}^n [\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*)]\ddot{Z}_k(\eta_t^*)^\top(\hat{c} - c)x_{it}x_{jt}^\top(\eta_t - x_{1t}) \right\| \\
&\leq \sum_{t=1}^n |\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*)|\|\ddot{Z}_k(\eta_t^*)^\top(\hat{c} - c)\|\|x_{it}\|\|x_{jt}\|[\|\alpha^1 - \alpha_0^1\|\|x_{1t}\| + \|\alpha^2\|\|x_{2t}\|] \\
&\leq n^{-1/4+\delta} \sum_{t=1}^n |\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*)|\|\ddot{Z}_k(\eta_t^*)^\top(\hat{c} - c)\|\|x_{it}\|\|x_{jt}\|\|x_{1t}\| \\
&\quad + n^{-3/4+\delta} \sum_{t=1}^n |\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*)|\|\ddot{Z}_k(\eta_t^*)^\top(\hat{c} - c)\|\|x_{it}\|\|x_{jt}\|\|x_{2t}\| \\
&\leq O_P(1)k^2n^{-1/2+\delta} \sum_{t=1}^n |\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*)|\|x_{it}\|\|x_{jt}\|\|x_{1t}\| \\
&\quad + O_P(1)k^2n^{-1+\delta} \sum_{t=1}^n |\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*)|\|x_{it}\|\|x_{jt}\|\|x_{2t}\|.
\end{aligned}$$

Due to the integrability of  $(x^3\dot{g})_{\text{sup}}$  and using the same argument as in  $T_{11}^{i,j}$ , we have  $T_{12}^{i,j} = O_P(1)k^2n^\delta$  for  $i = j = 1$ ,  $T_{12}^{i,j} = O_P(1)k^2n^{1/2+\delta}$  for  $i = 1, j = 2$  or  $i = 2, j = 1$  and  $T_{12}^{i,j} = O_P(1)k^2n^{1+\delta}$  for  $i = j = 2$ . They are all less than their counterparts in  $T_{11}^{i,j}$  that is

because  $\ddot{Z}_k(\eta_t^*)^\top(\widehat{c} - c)$  is a residue. This happens once again for  $T_{13}^{i,j}$ , so we neglect  $T_{13}^{i,j}$  as well. Moreover, for  $T_{14}^{i,j}$  there are two residues in the expression.

$$\begin{aligned} \|T_{14}^{i,j}\| &= \left\| \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top(\widehat{c} - c)\ddot{Z}_k(\eta_t^*)^\top(\widehat{c} - c)x_{it}x_{jt}^\top(\eta_t - x_{1t}) \right\| \\ &\leq \|\widehat{c} - c\|^2 \sum_{t=1}^n \|\dot{Z}_k(\eta_t^*)\| \|\ddot{Z}_k(\eta_t^*)\| \|x_{it}\| \|x_{jt}\| |\eta_t - x_{1t}| \\ &\leq O_P(1) \frac{k}{\sqrt{n}} n^{-1/4+\delta} \sum_{t=1}^n (x \|\dot{Z}_k\| \|\ddot{Z}_k\|)_{\text{sup}}(x_{1t}) \|x_{it}\| \|x_{jt}\| \\ &\quad + O_P(1) \frac{k}{\sqrt{n}} n^{-3/4+\delta} \sum_{t=1}^n (\|\dot{Z}_k\| \|\ddot{Z}_k\|)_{\text{sup}}(x_{1t}) \|x_{it}\| \|x_{jt}\| \|x_{2t}\|. \end{aligned}$$

Notice that  $(x \|\dot{Z}_k\| \|\ddot{Z}_k\|)_{\text{sup}}(x_{1t}) |x_{1t}|^p \leq (x^{p+1} \|\dot{Z}_k\| \|\ddot{Z}_k\|)_{\text{sup}}(x_{1t})$  for any  $p > 0$ . In view of Lemma A.5, we need consider the following integral,

$$\begin{aligned} &\int (x^p \|\dot{Z}_k\| \|\ddot{Z}_k\|)_{\text{sup}}(u) du = \int |u|^p \|\dot{Z}_k(u)\| \|\ddot{Z}_k(u)\| du + o(1) \\ &\leq \left( \int |u|^{2p_1} \|\dot{Z}_k(u)\|^2 du \int |u|^{2p_2} \|\ddot{Z}_k(u)\|^2 du \right)^{1/2} + o(1) \\ &= \begin{cases} O(1)k^3 & \text{when } p_1 = 1, p_2 = 0, \\ O(1)k^{7/2} & \text{when } p_1 = 0, p_2 = 2, \\ O(1)k^4 & \text{when } p_1 = 1, p_2 = 2, \end{cases} \end{aligned}$$

by (4) of Lemma A.1, where  $p$  is decomposed as  $p = p_1 + p_2$ . This entails  $T_{14}^{i,j} = O_P(1)k^5 n^{-1/4+\delta}$ ,  $O_P(1)k^{9/2} n^{1/4+\delta}$ , and  $O_P(1)k^4 n^{3/4+\delta}$  according as  $i = j = 1$ ,  $i = 1, j = 2$  or  $i = 2, j = 1$ , and  $i = j = 2$ .

After discussion for  $T_{1\ell}^{i,j}$ ,  $\ell = 1, \dots, 4$ , we have

$$\Xi_{n,1}^{i,j} = \begin{cases} O_P(1)n^{\max(5\kappa-1/4, 1/4)+\delta}, & \text{when } i = j = 1, \\ O_P(1)n^{\max(4.5\kappa+1/4, 3/4)+\delta}, & \text{when } i = 1, j = 2, \\ O_P(1)n^{5/4+\delta}, & \text{when } i = j = 2. \end{cases} \quad (\text{B.11})$$

Now we move on to  $\Xi_{n,2}^{i,j}$ .

$$\begin{aligned} \Xi_{n,2}^{i,j} &= \sum_{t=1}^n e_t \ddot{g}_n(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) = \sum_{t=1}^n e_t \ddot{Z}_k(\eta_t^*)^\top \widehat{c} x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &= \sum_{t=1}^n e_t [\ddot{Z}_k(\eta_t^*)^\top c + \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c)] x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\ &= \sum_{t=1}^n e_t [\ddot{g}(\eta_t^*) - \ddot{\gamma}_k(\eta_t^*) + \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c)] x_{it} x_{jt}^\top (\eta_t - x_{1t}) \end{aligned}$$



$$\begin{aligned}
&= \sum_{t=1}^n e_t \ddot{g}(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) - \sum_{t=1}^n e_t \ddot{\gamma}_k(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\
&\quad + \sum_{t=1}^n e_t \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\
&:= T_{21}^{i,j} - T_{22}^{i,j} + T_{23}^{i,j}, \quad \text{say.}
\end{aligned}$$

Because  $(e_t, \mathcal{F}_t)$  is a martingale difference sequence, as discussed in  $T_{31}$  of Lemma B.1, we may evaluate

$$\begin{aligned}
E \|T_{21}^{i,j}\|^2 &= E \left\| \sum_{t=1}^n e_t \ddot{g}(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \right\|^2 = \sigma_e^2 E \sum_{t=1}^n |\ddot{g}(\eta_t^*)|^2 \|x_{it} x_{jt}^\top\|^2 |\eta_t - x_{1t}|^2 \\
&\leq O(1) n^{-1/2+2\delta} E \sum_{t=1}^n (x \ddot{g})_{\text{sup}}^2(x_{1t}) \|x_{it}\|^2 \|x_{jt}\|^2 \\
&\quad + O(1) n^{-3/2+2\delta} E \sum_{t=1}^n (\ddot{g})_{\text{sup}}^2(x_{1t}) \|x_{it}\|^2 \|x_{jt}\|^2 \|x_{2t}\|^2.
\end{aligned}$$

Thus, similar to Lemma A.5,  $E \|T_{21}^{1,1}\|^2 \leq O(1) n^{2\delta}$ , implying  $T_{21}^{1,1} = O_P(1) n^\delta$ ;  $E \|T_{21}^{1,2}\|^2 \leq O(1) n^{1+2\delta}$ , implying  $T_{21}^{1,2} = O_P(1) n^{1/2+\delta}$ ; and  $E \|T_{21}^{2,2}\|^2 \leq O(1) n^{2+2\delta}$ , implying  $T_{21}^{2,2} = O_P(1) n^{1+\delta}$ .

Regarding of  $T_{22}^{i,j}$ , notice that  $\ddot{\gamma}_k(x)$  entails the integrability of  $(x^3 \ddot{\gamma}_k(x))^2$  as well as the attenuation of this integration when  $k$  increases. This implies  $T_{22}^{i,j}$  is dominated by  $T_{21}^{i,j}$ . Moreover, for  $T_{23}^{i,j}$ ,

$$\begin{aligned}
\|T_{23}^{i,j}\|^2 &= \left\| \sum_{t=1}^n e_t \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \right\|^2 = \left\| \left( \sum_{t=1}^n e_t (\eta_t - x_{1t}) x_{jt} x_{it}^\top \ddot{Z}_k(\eta_t^*) \right)^\top (\widehat{c} - c) \right\|^2 \\
&\leq \left\| \sum_{t=1}^n e_t (\eta_t - x_{1t}) x_{jt} x_{it}^\top \ddot{Z}_k(\eta_t^*) \right\|^2 \|\widehat{c} - c\|^2 = O_P(1) \frac{k}{\sqrt{n}} \left\| \sum_{t=1}^n e_t (\eta_t - x_{1t}) x_{jt} x_{it}^\top \ddot{Z}_k(\eta_t^*) \right\|^2.
\end{aligned}$$

By virtue of the martingale difference property for  $e_t$ ,

$$\begin{aligned}
&\frac{k}{\sqrt{n}} E \left\| \sum_{t=1}^n e_t (\eta_t - x_{1t}) x_{jt} x_{it}^\top \ddot{Z}_k(\eta_t^*) \right\|^2 = \sigma_e^2 \frac{k}{\sqrt{n}} E \sum_{t=1}^n (\eta_t - x_{1t})^2 \|x_{jt} x_{it}^\top \ddot{Z}_k(\eta_t^*)\|^2 \\
&\leq O(1) \frac{k}{\sqrt{n}} n^{-1/2+2\delta} E \sum_{t=1}^n x_{1t}^2 \|x_{it}\|^2 \|x_{jt}\|^2 \|\ddot{Z}_k(\eta_t^*)\|^2 \\
&\quad + O(1) \frac{k}{\sqrt{n}} n^{-3/2+2\delta} E \sum_{t=1}^n \|x_{2t}\|^2 \|x_{it}\|^2 \|x_{jt}\|^2 \|\ddot{Z}_k(\eta_t^*)\|^2.
\end{aligned}$$

We may calculate the expectations for different  $i$  and  $j$ , similar to Lemma A.5. The case where  $i = j = 1$  is given as an exemplar. Using (B.10) and (4) of Lemma A.1,

$$\frac{k}{\sqrt{n}} n^{-1/2+2\delta} E \sum_{t=1}^n x_{1t}^6 \|\ddot{Z}_k(\eta_t^*)\|^2 + \frac{k}{\sqrt{n}} n^{-3/2+2\delta} E \sum_{t=1}^n \|x_{2t}\|^2 x_{1t}^4 \|\ddot{Z}_k(\eta_t^*)\|^2$$

$$\begin{aligned}
&\leq \frac{k}{\sqrt{n}} n^{-1/2+2\delta} E \sum_{t=1}^n (x_{1t}^6 \|\ddot{Z}_k(x)\|^2)_{\sup}(x_{1t}) + \frac{k}{\sqrt{n}} n^{-3/2+2\delta} E \sum_{t=1}^n \|x_{2t}\|^2 (x_{1t}^4 \|\ddot{Z}_k(x)\|^2)_{\sup}(x_{1t}) \\
&\leq O(1) \frac{k}{\sqrt{n}} n^{-1/2+2\delta} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int (x^6 \|\ddot{Z}_k(x)\|^2)_{\sup}(u) du \\
&\quad + O(1) \frac{k}{\sqrt{n}} n^{-3/2+2\delta} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|\sqrt{t}w\|^2 \phi(w) (x^4 \|\ddot{Z}_k(x)\|^2)_{\sup}(u) dudw \\
&= O(1) \frac{k}{\sqrt{n}} n^{2\delta} \int (x^6 \|\ddot{Z}_k(x)\|^2)_{\sup}(u) du + O(1) \frac{k}{\sqrt{n}} n^{2\delta} \int (x^4 \|\ddot{Z}_k(x)\|^2)_{\sup}(u) du \\
&= O(1) \frac{k}{\sqrt{n}} n^{2\delta} k^7 + O(1) \frac{k}{\sqrt{n}} n^{2\delta} k^6 \\
&= O(1) n^{-1/2+2\delta} k^8 = O(1) n^{-1/2+2\delta+8\kappa},
\end{aligned}$$

implying  $T_{23}^{1,1} = O_P(1) n^{-1/4+\delta+4\kappa}$ .

When  $i = 1$  and  $j = 2$ , the expectations are computed to be  $O(1) n^{1/2+2\delta+7\kappa}$  which implies  $T_{23}^{1,2} = O_P(1) n^{1/4+\delta+7\kappa/2}$ . When  $i = j = 2$ , the expectations are computed to be  $O(1) n^{3/2+2\delta+6\kappa}$  which implies  $T_{23}^{2,2} = O_P(1) n^{3/4+\delta+3\kappa}$ . Therefore,

$$\Xi_{n,2}^{i,j} = \begin{cases} O_P(1) n^{4\kappa-1/4+\delta} & \text{when } i = j = 1, \\ O_P(1) n^{3.5\kappa+1/4+\delta} & \text{when } i = 1, j = 2, \\ O_P(1) n^{3\kappa+3/4+\delta} & \text{when } i = j = 2. \end{cases} \quad (\text{B.12})$$

Next, noting that  $\widehat{g}_n(u) = g(u) - \gamma_k(u) + Z_k(u)^\top (\widehat{c} - c)$ ,

$$\begin{aligned}
\Xi_{n,3}^{i,j} &= \sum_{t=1}^n [g(x_{1t}) - \widehat{g}_n(x_{1t})] \ddot{\widehat{g}}_n(\eta_t^*) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\
&= \sum_{t=1}^n [Z_k(x_{1t})^\top (c - \widehat{c}) + \gamma_k(x_{1t})] [\ddot{g}(\eta_t^*) - \ddot{\gamma}_k(\eta_t^*) + \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c)] x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\
&= \sum_{t=1}^n Z_k(x_{1t})^\top (c - \widehat{c}) [\ddot{g}(\eta_t^*) - \ddot{\gamma}_k(\eta_t^*) + \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c)] x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\
&\quad + \sum_{t=1}^n \gamma_k(x_{1t}) [\ddot{g}(\eta_t^*) - \ddot{\gamma}_k(\eta_t^*) + \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c)] x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\
&:= T_{31}^{i,j} + T_{32}^{i,j}, \quad \text{say.}
\end{aligned}$$

It follows from Lemma 2.1 and Lemma A.1 that

$$\begin{aligned}
\|T_{31}^{i,j}\| &= \left\| \sum_{t=1}^n Z_k(x_{1t})^\top (c - \widehat{c}) [\ddot{g}(\eta_t^*) - \ddot{\gamma}_k(\eta_t^*) + \ddot{Z}_k(\eta_t^*)^\top (\widehat{c} - c)] x_{it} x_{jt}^\top (\eta_t - x_{1t}) \right\| \\
&\leq \|c - \widehat{c}\| \sum_{t=1}^n \|Z_k(x_{1t})\| [|\ddot{g}(\eta_t^*)| + |\ddot{\gamma}_k(\eta_t^*)| + \|\ddot{Z}_k(\eta_t^*)\|] \|\widehat{c} - c\| \|x_{it}\| \|x_{jt}\| |\eta_t - x_{1t}|
\end{aligned}$$

$$\begin{aligned}
&= O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{g}(\eta_t^*)\| \|x_{it}\| \|x_{jt}\| |\eta_t - x_{1t}| \\
&\quad + O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{\gamma}_k(\eta_t^*)\| \|x_{it}\| \|x_{jt}\| |\eta_t - x_{1t}| \\
&\quad + O_P(1) \frac{k}{\sqrt{n}} \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{Z}_k(\eta_t^*)\| \|x_{it}\| \|x_{jt}\| |\eta_t - x_{1t}| \\
&\leq O_P(1) \frac{k}{\sqrt[4]{n}} \sum_{t=1}^n \|\ddot{g}(\eta_t^*)\| \|x_{it}\| \|x_{jt}\| [n^{-1/4+\delta} |x_{1t}| + n^{-3/4+\delta} \|x_{2t}\|] \\
&\quad + O_P(1) \frac{k}{\sqrt{n}} \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{Z}_k(\eta_t^*)\| \|x_{it}\| \|x_{jt}\| [n^{-1/4+\delta} |x_{1t}| + n^{-3/4+\delta} \|x_{2t}\|],
\end{aligned}$$

where we have ignored the second term because  $\ddot{\gamma}_k$  is the residue of  $\ddot{g}$  such that  $x^3 \ddot{\gamma}_k$  is integrable and attenuates to zero with the increase of  $k$ .

We shall evaluate the following expectations for  $i = j = 1$  as an exemplar. The other cases can be done almost the same. For the first term in  $T_{31}^{i,j}$  when  $i = j = 1$

$$\begin{aligned}
&\frac{k}{\sqrt[4]{n}} E \sum_{t=1}^n \|\ddot{g}(\eta_t^*)\| x_{1t}^2 [n^{-1/4+\delta} |x_{1t}| + n^{-3/4+\delta} \|x_{2t}\|] \\
&\leq \frac{k}{\sqrt[4]{n}} n^{-1/4+\delta} E \sum_{t=1}^n (x^3 \ddot{g})_{\sup}(x_{1t}) + \frac{k}{\sqrt[4]{n}} n^{-3/4+\delta} E \sum_{t=1}^n (x^2 \ddot{g})_{\sup}(x_{1t}) \|x_{2t}\| \\
&\leq O(1) \frac{k}{\sqrt[4]{n}} n^{-1/4+\delta} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int (|x|^3 \ddot{g})_{\sup}(u) du \\
&\quad + O(1) \frac{k}{\sqrt[4]{n}} n^{-3/4+\delta} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint (|x|^2 \ddot{g})_{\sup}(u) \|\sqrt{t}w\| \phi(w) dudw \\
&= O(1) \frac{k}{\sqrt[4]{n}} n^{-1/4+\delta} \sqrt{n} \int (|x|^3 \ddot{g})_{\sup}(u) du + O(1) \frac{k}{\sqrt[4]{n}} n^{-3/4+\delta} n \int (|x|^2 \ddot{g})_{\sup}(u) du \\
&= O(1) k n^\delta = O(1) n^{\kappa+\delta},
\end{aligned}$$

and by (4) of Lemma A.1 and (B.10), for the second term

$$\begin{aligned}
&\frac{k}{\sqrt{n}} E \sum_{t=1}^n \|Z_k(x_{1t})\| \|\ddot{Z}_k(\eta_t^*)\| x_{1t}^2 [n^{-1/4+\delta} |x_{1t}| + n^{-3/4+\delta} \|x_{2t}\|] \\
&\leq \frac{k}{\sqrt{n}} n^{-1/4+\delta} E \sum_{t=1}^n (\|Z_k(x)\| \|x^3\| \|\ddot{Z}_k(x)\|)_{\sup}(x_{1t}) \\
&\quad + \frac{k}{\sqrt{n}} n^{-3/4+\delta} E \sum_{t=1}^n (\|Z_k(x)\| \|x^2\| \|\ddot{Z}_k(x)\|)_{\sup}(x_{1t}) \|x_{2t}\| \\
&\leq O(1) \frac{k}{\sqrt{n}} n^{-1/4+\delta} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int (\|Z_k(x)\| \|x^3\| \|\ddot{Z}_k(x)\|)_{\sup}(u) du \\
&\quad + O(1) \frac{k}{\sqrt{n}} n^{-3/4+\delta} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint (\|Z_k(x)\| \|x^2\| \|\ddot{Z}_k(x)\|)_{\sup}(u) \|\sqrt{t}w\| \phi(w) dudw
\end{aligned}$$

$$\begin{aligned}
&= O(1) \frac{k}{\sqrt{n}} n^{-1/4+\delta} \sqrt{n} \int \|Z_k(u)\| |u|^3 \|\ddot{Z}_k(u)\| du (1 + o(1)) \\
&\quad + O(1) \frac{k}{\sqrt{n}} n^{-3/4+\delta} n \int \|Z_k(u)\| u^2 \|\ddot{Z}_k(u)\| du (1 + o(1)) \\
&\leq O(1) k n^{-1/4+\delta} \left( \int \|Z_k(u)\|^2 du \int |u|^6 \|\ddot{Z}_k(u)\|^2 du \right)^{1/2} \\
&\quad + O(1) k n^{-1/4+\delta} \left( \int \|Z_k(u)\|^2 du \int |u|^4 \|\ddot{Z}_k(u)\|^2 du \right)^{1/2} \\
&= O(1) k n^{-1/4+\delta} \sqrt{k k^7} = O(1) k^5 n^{-1/4+\delta} = n^{-1/4+5\kappa+\delta},
\end{aligned}$$

we have  $T_{3,1}^{1,1} = O_P(1) n^{3\kappa+\delta}$  as  $2\kappa < 1/4$ . Similarly, when  $i = 1, j = 2$ ,  $T_{3,1}^{1,2} = O_P(1) n^{1/4+9\kappa/2+\delta}$ , and when  $i = j = 2$ ,  $T_{3,1}^{2,2} = O_P(1) n^{3/4+4\kappa+\delta}$ .

Regarding of  $T_{3,2}^{i,j}$ , notice that

$$\begin{aligned}
\|T_{3,2}^{i,j}\| &= \left\| \sum_{t=1}^n \gamma_k(x_{1t}) [\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*) + \ddot{Z}_k(\eta_t^*)^\top (\hat{c} - c)] x_{it} x_{jt}^\top (\eta_t - x_{1t}) \right\| \\
&\leq \sum_{t=1}^n |\gamma_k(x_{1t}) \dot{g}(\eta_t^*)| \|x_{it}\| \|x_{jt}\| |\eta_t - x_{1t}| \\
&\quad + \|\hat{c} - c\| \sum_{t=1}^n |\gamma_k(x_{1t})| \|\ddot{Z}_k(\eta_t^*)\| \|x_{it}\| \|x_{jt}\| |\eta_t - x_{1t}| \\
&\leq o(1) k^{-(m-1)/2-1/12} \sum_{t=1}^n |\dot{g}(\eta_t^*)| \|x_{it}\| \|x_{jt}\| [n^{-1/4+\delta} |x_{1t}| + n^{-3/4+\delta} \|x_{2t}\|] \\
&\quad + O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n |\gamma_k(x_{1t})| \|\ddot{Z}_k(\eta_t^*)\| \|x_{it}\| \|x_{jt}\| [n^{-1/4+\delta} |x_{1t}| + n^{-3/4+\delta} \|x_{2t}\|],
\end{aligned}$$

by (1) of Lemma A.1, where the second term involving  $\dot{\gamma}_k$  has been omitted due to the aforementioned reason.

To find out the order of  $T_{32}^{1,1}$ , using Lemma A.5 and the function in (B.9), the first term in the above is of order  $n^{1/4+\delta-(m-1)\kappa/2-\kappa/12}$ , while the second term is of order  $n^\delta k^{-m/2+4}$  utilising  $\int \gamma_k(u) \|\ddot{Z}_k(u)\| u^3 du \leq k^{(7-m)/2}$  estimated by Cauchy-Schwarz inequality and (4) of Lemma A.5, implying  $T_{32}^{1,1} = O_P(1) n^\delta k^{-m/2+4} = O_P(1) n^{(8-m)\kappa/2+\delta}$ . Similarly,  $T_{32}^{1,2} = O_P(1) n^{1/2+(7-m)\kappa/2+\delta}$  and  $T_{32}^{2,2} = O_P(1) n^{1+\delta+(6-m)\kappa/2}$ . We then have

$$\Xi_{n,3}^{i,j} = \begin{cases} O_P(1) n^{3\kappa+\delta} & \text{when } i = j = 1, \\ O_P(1) n^{4.5\kappa+1/4+\delta} & \text{when } i = 1, j = 2, \\ O_P(1) n^{4\kappa+3/4+\delta} & \text{when } i = j = 2. \end{cases} \quad (\text{B.13})$$

Finally, noting again that  $\hat{g}_n(u) = Z_k(u)^\top \hat{c} = g(u) - \gamma_k(u) + Z_k(u)^\top (\hat{c} - c)$ ,

$$\Xi_{n,4}^{i,j} = \sum_{t=1}^n \hat{g}_n(\eta_t^*) \ddot{g}_n(\eta_t) x_{it} x_{jt}^\top (\eta_t - x_{1t})$$

$$\begin{aligned}
&= \sum_{t=1}^n [\dot{g}(\eta_t^*) - \dot{\gamma}_k(\eta_t^*) + \dot{Z}_k(\eta_t^*)^\top (\hat{c} - c)] [\ddot{g}(\eta_t) - \ddot{\gamma}_k(\eta_t) + \ddot{Z}_k(\eta_t)^\top (\hat{c} - c)] x_{it} x_{jt}^\top (\eta_t - x_{1t}) \\
&= \sum_{t=1}^n \dot{g}(\eta_t^*) \ddot{g}(\eta_t) x_{it} x_{jt}^\top (\eta_t - x_{1t}) + \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top (\hat{c} - c) \ddot{Z}_k(\eta_t)^\top (\hat{c} - c) x_{it} x_{jt}^\top (\eta_t - x_{1t}) + \dots \\
&= T_{41}^{i,j} + T_{42}^{i,j} + \dots
\end{aligned}$$

where the reminder terms (totally 7) are of smaller order than the first two terms due to Theorem 2.1 and (5) of Lemma A.1 that  $|\dot{Z}_k(u)^\top (\hat{c} - c)| \leq O_P(1) \frac{\sqrt{k}}{\sqrt{n}} \|\dot{Z}_k(u)\| = O_P(1) \frac{k\sqrt{k}}{\sqrt{n}} = o_P(1)$  uniformly in  $u$ , as well as the properties for  $\dot{\gamma}_k(u)$  and  $\ddot{\gamma}_k(u)$  listed (2)-(3) of Lemma A.1. That means we only need to evaluate the first two terms.

Using the function form in (B.9) for  $\dot{g}(\eta_t^*)$  and  $\ddot{g}(\eta_t)$ , Lemma A.5 gives  $T_{41}^{1,1} = O_P(1)n^{1/4+\delta}$ ,  $T_{41}^{1,2} = O_P(1)n^{3/4+\delta}$ , and  $T_{41}^{2,2} = O_P(1)n^{5/4+\delta}$ . On the other hand,

$$\begin{aligned}
\|T_{42}^{i,j}\| &= \left\| \sum_{t=1}^n \dot{Z}_k(\eta_t^*)^\top (\hat{c} - c) \dot{Z}_k(\eta_t)^\top (\hat{c} - c) x_{it} x_{jt}^\top (\eta_t - x_{1t}) \right\| \\
&\leq \|\hat{c} - c\|^2 \sum_{t=1}^n \|\dot{Z}_k(\eta_t^*)\| \|\dot{Z}_k(\eta_t)\| \|x_{it}\| \|x_{jt}\| \|\eta_t - x_{1t}\| \\
&\leq O_P(1) \frac{k}{\sqrt{n}} n^{-1/4+\delta} \sum_{t=1}^n \|\dot{Z}_k(\eta_t^*)\| \|\dot{Z}_k(\eta_t)\| \|x_{it}\| \|x_{jt}\| \|x_{1t}\| \\
&\quad + O_P(1) \frac{k}{\sqrt{n}} n^{-3/4+\delta} \sum_{t=1}^n \|\dot{Z}_k(\eta_t^*)\| \|\dot{Z}_k(\eta_t)\| \|x_{it}\| \|x_{jt}\| \|x_{2t}\|.
\end{aligned}$$

Whence, for  $i = j = 1$ , similar calculation as in Lemma A.5 yields  $T_{42}^{1,1} = O_P(1) \frac{k}{\sqrt{n}} n^{1/4+\delta} \int |u|^3 \|\dot{Z}_k(u)\| \|\dot{Z}_k(u)\| du = O_P(1) n^{-1/4+5\kappa+\delta}$  by virtue of Cauchy-Schwarz inequality and (4) of Lemma A.1; for  $i = 1, j = 2$ ,  $T_{42}^{1,2} = O_P(1) n^{1/4+9\kappa/2+\delta}$ ; for  $i = j = 2$ ,  $T_{42}^{2,2} = O_P(1) n^{3/4+4\kappa+\delta}$ .

Therefore, we have

$$\Xi_{n,4}^{i,j} = \begin{cases} O_P(1) n^{\max(5\kappa-1/4, 1/4)+\delta} & \text{when } i = j = 1, \\ O_P(1) n^{\max(4.5\kappa+1/4, 3/4)+\delta} & \text{when } i = 1, j = 2, \\ O_P(1) n^{5/4+\delta} & \text{when } i = j = 2. \end{cases} \quad (\text{B.14})$$

In view of (B.11), (B.12), (B.13) and (B.14),  $|J_{n,11}(\alpha) - J_{n,11}(\alpha_0)|$  is of order  $O_P(1) n^{3\kappa+\delta}$ ,  $\|J_{n,12}(\alpha) - J_{n,12}(\alpha_0)\|$  is of order  $O_P(1) n^{\max(4.5\kappa+1/4, 3/4)+\delta}$ , and  $\|J_{n,22}(\alpha) - J_{n,22}(\alpha_0)\|$  is of order  $O_P(1) n^{5/4+\delta}$ . To fulfill (B.7), we may choose  $\delta$ :  $0 < \delta < 1/24$ , because  $\kappa < 1/8$  stipulated in Assumption B.

We have shown in Theorem 3.1 the convergence

$$D_n^{-1} S_n(\alpha_0) \rightarrow_D (R^{1/2} W(1)), \quad \text{and} \quad D_n^{-1} J_n(\alpha_0) D_n^{-1} \rightarrow_P R, \quad (\text{B.15})$$

where  $R$  is positive definite with probability one. This indicates the condition (iv) in Wooldridge's theorem holds. Thus, there exists a sequence of estimator  $\widehat{\alpha}_n$  of  $\alpha_0$  such that  $D_n(\widehat{\alpha}_n - \alpha_0) = O_P(1)$  and hence the limit distribution now follows from (B.2) and (B.15) directly.  $\square$

*Proof of Theorem 3.3.* Noting that  $\sqrt[4]{n}D_n^{-1} \rightarrow \text{diag}(1, \mathbf{0}_{d-1})$  as  $n \rightarrow \infty$  where  $\mathbf{0}_{d-1}$  is a zero matrix of  $(d-1) \times (d-1)$ , by the continuous mapping theorem we have

$$\begin{aligned} \sqrt[4]{n}(\widehat{\theta}_n - \theta_0) &= \sqrt[4]{n}(D_n Q^\top)^{-1} D_n Q^\top (\widehat{\theta}_n - \theta_0) = Q \sqrt[4]{n} D_n^{-1} D_n Q^\top (\widehat{\theta}_n - \theta_0) \\ &\rightarrow_D Q \text{diag}(1, \mathbf{0}_{d-1}) R^{-1/2} W(1) = \mathbf{MN}(0, \rho_{11} \theta_0 \theta_0^\top). \end{aligned} \quad (\text{B.16})$$

$\square$

*Proof of Theorem 3.4.* Note that

$$\sqrt{n}[J_n(\widehat{\theta}_n)]^{-1} = \sqrt{n}Q[J_n(\widehat{\alpha}_n)]^{-1}Q^\top = \sqrt{n}QD_n^{-1}[D_n^{-1}J_n(\widehat{\alpha}_n)D_n^{-1}]^{-1}D_n^{-1}Q^\top,$$

and by Theorems 3.1 and 3.2,  $D_n^{-1}J_n(\widehat{\alpha}_n)D_n^{-1} = D_n^{-1}J_n(\alpha_0)D_n^{-1} + o_P(1) \rightarrow_P R$ . Therefore, noting  $\sqrt[4]{n}QD_n^{-1} \rightarrow (\theta_0, \mathbf{0})$ , we have

$$\sqrt{n}[J_n(\widehat{\alpha}_n)]^{-1} = \sqrt{n}QD_n^{-1}[D_n^{-1}J_n(\widehat{\alpha}_n)D_n^{-1}]^{-1}D_n^{-1}Q^\top \rightarrow_P (\theta_0, \mathbf{0})R^{-1}(\theta_0, \mathbf{0})^\top = \rho_{11}\theta_0\theta_0^\top,$$

where  $\rho_{11}$  is the element of  $R^{-1}$  at the first row and first column given by (3.7).

The convergence of  $\sqrt{n}[\widetilde{J}_n(\widehat{\theta}_n)]^{-1}$  is manifest from the proof of Theorem 3.1 and the above derivation.  $\square$

*Proof of Theorem 3.5.* Recall that  $\widehat{\theta}_n = Q\widehat{\alpha}_n$  and  $\theta_0 = Q\alpha_0$  with  $\alpha_0 = (\alpha_0^1, (\alpha_0^2)^\top)^\top = (1, 0, \dots, 0)^\top$ . Notice that  $\|\widehat{\theta}_n\| = \|\widehat{\alpha}_n\|$  by the orthogonality of  $Q$ . In the meantime, it follows from Theorem 3.2 that  $\widehat{\alpha}_{1n} \rightarrow_P 1$ , implying  $\widehat{\alpha}_{1n} > 0$  with probability approaching 1.

Using Taylor expansion  $(1+x)^{-1/2} = 1 - \frac{1}{2}x + O(x^2)$  and  $(1+x)^{-2} = 1 - 2x + O(x^2)$  for small  $x$ , we have

$$\begin{aligned} \widehat{\alpha}_{n,\text{unit}}^1 &= \frac{1}{\|\widehat{\alpha}_n\|} \widehat{\alpha}_{1n} = \frac{1}{(\widehat{\alpha}_{1n}^2 + \|\widehat{\alpha}_{2n}\|^2)^{1/2}} \widehat{\alpha}_{1n} = \left(1 + \frac{\|\widehat{\alpha}_{2n}\|^2}{\widehat{\alpha}_{1n}^2}\right)^{-1/2} \\ &= 1 - \frac{1}{2} \frac{n^{-3/2} \|\sqrt[4]{n}^3 \widehat{\alpha}_{2n}\|^2}{(1 + \widehat{\alpha}_{1n} - 1)^2} + O_P(n^{-3}) = 1 - \frac{1}{2} n^{-3/2} \|\sqrt[4]{n}^3 \widehat{\alpha}_{2n}\|^2 + O_P(n^{-7/4}), \end{aligned}$$

by virtue of Theorem 3.2, implying

$$n^{3/2} (\widehat{\alpha}_{n,\text{unit}}^1 - \alpha_0^1) = -\frac{1}{2} \|\sqrt[4]{n}^3 \widehat{\alpha}_{2n}\|^2 + o_P(1) \rightarrow_D -\frac{1}{2} \|\xi\|^2,$$

due to the continuous mapping theorem. Meanwhile,

$$\widehat{\alpha}_{n,\text{unit}}^2 = \frac{1}{(\widehat{\alpha}_{1n}^2 + \|\widehat{\alpha}_{2n}\|^2)^{1/2}} \widehat{\alpha}_{2n} = \widehat{\alpha}_{2n} (1 + (\widehat{\alpha}_{1n}^2 - 1) + \|\widehat{\alpha}_{2n}\|)^{-1/2}$$

$$= \widehat{\alpha}_{2n} - \frac{1}{2} \widehat{\alpha}_{2n} O_P[(\widehat{\alpha}_{1n}^2 - 1)^2 + \|\widehat{\alpha}_{2n}\|^2] = \widehat{\alpha}_{2n}(1 + o_P(1)).$$

Thus,  $\widehat{\alpha}_{n,\text{unit}}^2$  has the same asymptotics as  $\widehat{\alpha}_{2n}$ . Consequently,

$$\begin{aligned} n^{3/4} \left( \widehat{\theta}_{n,\text{emp}} - \theta_0 \right) &= Q n^{3/4} \left( \frac{1}{\|\widehat{\alpha}_n\|} \widehat{\alpha}_n - \alpha_0 \right) = Q n^{3/4} \begin{pmatrix} \frac{1}{\|\widehat{\alpha}_n\|} \widehat{\alpha}_{1n} - 1 \\ \frac{1}{\|\widehat{\alpha}_n\|} \widehat{\alpha}_{2n} \end{pmatrix} \\ &= (\theta_0 \ Q_2) \begin{pmatrix} 0 \\ n^{3/4} \widehat{\alpha}_{2n} \end{pmatrix} + o_P(1) = Q_2 n^{3/4} \widehat{\alpha}_{2n} + o_P(1) \rightarrow_D \mathbf{MN}(0, Q_2 \rho_{22} Q_2^\top). \end{aligned}$$

□

Before proving Theorem 3.6, we first establish the consistency of the estimator  $(\widehat{\beta}_n, \widehat{\theta}_n)$  for the partially linear single-index models. Towards this end, define for any  $\theta \in \Theta$ ,

$$\widetilde{\beta}_n(\theta) = \underset{\beta}{\operatorname{argmin}} L_n(\beta, \theta) = \frac{1}{2} \sum_{t=1}^n [y_t - \beta^\top x_t - \widehat{g}_n(\theta^\top x_t)]^2. \quad (\text{B.17})$$

**Lemma B.2.** (1) The estimator  $\widetilde{\beta}_n(\theta)$  given by (B.17) is consistent uniformly for  $\theta \in \Theta$ , that is,  $\widetilde{\beta}_n(\theta) \rightarrow_P \beta_0$  uniformly in  $\theta$ . In addition, for any fixed  $\theta \in \Theta$  such that  $\|\theta\| \neq 0$ ,

$$\begin{aligned} n(\widetilde{\beta}_n(\theta) - \beta_0) &\rightarrow_D \left( \int_0^1 V(r) V(r)^\top dr \right)^{-1} \left[ \int_0^1 V(r) dU(r) \right. \\ &\quad \left. + Q \begin{pmatrix} 0 \\ \tau \int_0^1 V_2(r) dL_1(r, 0) \end{pmatrix} - \widetilde{Q} \begin{pmatrix} 0 \\ \frac{1}{\|\theta\|} \tau \int_0^1 \widetilde{V}_2(r) d\widetilde{L}_1(r, 0) \end{pmatrix} \right], \end{aligned} \quad (\text{B.18})$$

where  $Q = (\theta_0, Q_2)$  an orthogonal matrix defined as before,  $\widetilde{Q} = (\theta/\|\theta\|, \widetilde{Q}_2)$  an orthogonal matrix,  $\tau = \int g(x) dx$ ,  $\widetilde{V}_2(r)$  and  $\widetilde{L}_1(r, 0)$  are defined similar to  $V_2(r)$  and  $L_1(r, 0)$  depending on  $\theta$  other than  $\theta_0$ . Particularly,

$$n(\widetilde{\beta}_n(\theta_0) - \beta_0) \rightarrow_D \left( \int_0^1 V(r) V(r)^\top dr \right)^{-1} \int_0^1 V(r) dU(r). \quad (\text{B.19})$$

(2)  $(\widehat{\beta}_n, \widehat{\theta}_n)$  given by (2.8) is consistent,  $((\widehat{\beta}_n - \beta_0)^\top, (\widehat{\theta}_n - \theta_0)^\top)^\top \rightarrow_P 0$  as  $n \rightarrow \infty$ .

*Remark B.1.* The first assertion of Lemma B.2 gives a relationship of the estimators  $\widehat{\beta}_n$  and  $\widehat{\theta}_n$ . Basically,  $\beta_0$  can be estimated without knowing any knowledge of  $\theta$ . This is because in the partially linear single-index model (1.1),  $\beta_0^\top x_t = O_P(\sqrt{n})$  and  $g(\theta_0^\top x_t)$  is a bounded variable, so that the linear part dominates the nonlinear part. Therefore, in the estimator  $\widehat{\beta}_n$  the effect of the nonlinear part is negligible.

*Proof of Lemma B.2.* (1) Define  $E_n(\beta; \beta_0) = L_n(\beta, \theta) - L_n(\beta_0, \theta)$ . To prove the consistency for  $\tilde{\beta}_n$ , by Jennrich (1969) it suffices to show that there is a sequence  $\nu_n$  such that  $\frac{1}{\nu_n} E_n(\beta; \beta_0) \rightarrow_P E(\beta; \beta_0)$  uniformly in  $\beta$  as  $n \rightarrow \infty$ , where  $E(\cdot; \beta_0)$  is continuous and has unique minimum  $\beta_0$  a.s.

Let  $\nu_n = n^2$ . Note that

$$\begin{aligned} \frac{1}{n^2} E_n(\beta; \beta_0) &= \frac{1}{2n^2} \sum_{t=1}^n [y_t - \beta^\top x_t - \hat{g}_n(\theta^\top x_t)]^2 - \frac{1}{2n^2} \sum_{t=1}^n [y_t - \beta_0^\top x_t - \hat{g}_n(\theta^\top x_t)]^2 \\ &= \frac{1}{2n^2} \sum_{t=1}^n [y_t - \beta_0^\top x_t - \hat{g}_n(\theta^\top x_t) + (\beta_0 - \beta)^\top x_t]^2 - \frac{1}{2n^2} \sum_{t=1}^n [y_t - \beta_0^\top x_t - \hat{g}_n(\theta^\top x_t)]^2 \\ &= \frac{1}{2n^2} \sum_{t=1}^n [(\beta_0 - \beta)^\top x_t]^2 + \frac{1}{n^2} \sum_{t=1}^n (\beta_0 - \beta)^\top x_t [y_t - \beta_0^\top x_t - \hat{g}_n(\theta^\top x_t)] \\ &= \frac{1}{2n^2} \sum_{t=1}^n [(\beta_0 - \beta)^\top x_t]^2 + \frac{1}{n^2} \sum_{t=1}^n (\beta_0 - \beta)^\top x_t [e_t + g(\theta_0^\top x_t) - \hat{g}_n(\theta^\top x_t)] \\ &:= A_1 + A_2, \quad \text{say.} \end{aligned}$$

Note further that

$$\begin{aligned} A_1 &= \frac{1}{2n^2} \sum_{t=1}^n [(\beta_0 - \beta)^\top x_t]^2 = \frac{1}{2n} \sum_{t=1}^n [(\beta_0 - \beta)^\top V_n(t/n)]^2 \\ &= \frac{1}{2} \int_0^1 [(\beta_0 - \beta)^\top V_n(r)]^2 dr + o_P(1) \rightarrow_P \frac{1}{2} \int_0^1 [(\beta_0 - \beta)^\top V(r)]^2 dr, \end{aligned}$$

uniformly in  $\beta$  in any neighborhood of  $\beta_0$  by the continuous mapping theorem since  $V_n(r) \rightarrow V(r)$  almost surely by the Skorohod representation theorem. Meanwhile,

$$\begin{aligned} A_2 &= \frac{1}{n^2} \sum_{t=1}^n (\beta_0 - \beta)^\top x_t [e_t + g(\theta_0^\top x_t) - \hat{g}_n(\theta^\top x_t)] \\ &= \frac{1}{n^2} \sum_{t=1}^n (\beta_0 - \beta)^\top x_t e_t + \frac{1}{n^2} \sum_{t=1}^n (\beta_0 - \beta)^\top x_t g(\theta_0^\top x_t) - \frac{1}{n^2} \sum_{t=1}^n (\beta_0 - \beta)^\top x_t Z_k(\theta^\top x_t)^\top \tilde{c} \\ &= o_P(1), \end{aligned}$$

as every term is  $o_P(1)$  uniformly in  $\beta$  and  $\theta$ . Precisely,  $x_t/\sqrt{n} = O_P(1)$  uniformly in  $t$ ,  $\sum_{t=1}^n |e_t| = O_P(n)$  by Assumption A,  $g(u)$  is bounded on  $\mathbb{R}$ ,  $\|Z_k(x)\|^2 \leq O(1)k$  uniformly in  $x$  and  $\|\tilde{c}\| \leq \|c\| + \|\tilde{c} - c\| \leq \|g\|_{L^2} + o_P(1)$  by Theorem 2.1.

Thus,  $E(\beta; \beta_0) = \frac{1}{2} \int_0^1 [(\beta_0 - \beta)^\top V(r)]^2 dr$ , which clearly is continuous in  $\beta$  and has unique minimum  $\beta_0$  a.s. This gives the consistency. Notice that during the derivation of the consistency, we do not make use of any knowledge of  $\theta$ , since  $\theta$  is included in  $Z_k(\cdot)$  and  $\tilde{c}$ . However,  $\|Z_k(x)\|^2 \leq O(1)k$  uniformly in  $x$  and  $\|\tilde{c}\|$  is uniformly bounded. Hence, the consistency is independent of  $\theta$ .



For any fixed  $\theta \in \Theta$ , by definition  $\tilde{\beta}_n(\theta)$  minimises  $L_n(\beta, \theta)$  with respect to  $\beta$ , implying  $\frac{\partial L_n(\beta, \theta)}{\partial \beta} = 0$  at  $\tilde{\beta}_n(\theta)$ . Since its second partial derivative  $\sum_{t=1}^n x_t x_t^\top$  is positive definite almost surely, it follows that  $\tilde{\beta}_n(\theta)$  is the unique minimum and

$$\begin{aligned}\tilde{\beta}_n(\theta) &= \left( \sum_{t=1}^n x_t x_t^\top \right)^{-1} \sum_{t=1}^n [y_t - \hat{g}_n(\theta^\top x_t)] x_t \\ &= \left( \sum_{t=1}^n x_t x_t^\top \right)^{-1} \sum_{t=1}^n [e_t + \beta_0^\top x_t + g(\theta_0^\top x_t) - \hat{g}_n(\theta^\top x_t)] x_t \\ &= \beta_0 + \left( \sum_{t=1}^n x_t x_t^\top \right)^{-1} \sum_{t=1}^n [e_t + g(\theta_0^\top x_t) - \hat{g}_n(\theta^\top x_t)] x_t,\end{aligned}$$

which gives

$$\begin{aligned}n(\tilde{\beta}_n(\theta) - \beta_0) &= \left( \frac{1}{n^2} \sum_{t=1}^n x_t x_t^\top \right)^{-1} \frac{1}{n} \sum_{t=1}^n [e_t + g(\theta_0^\top x_t) - \hat{g}_n(\theta^\top x_t)] x_t \\ &= \left( \frac{1}{n^2} \sum_{t=1}^n x_t x_t^\top \right)^{-1} \left[ \frac{1}{n} \sum_{t=1}^n x_t e_t + \frac{1}{n} \sum_{t=1}^n g(\theta_0^\top x_t) x_t - \frac{1}{n} \sum_{t=1}^n \hat{g}_n(\theta^\top x_t) x_t \right].\end{aligned}$$

Note that  $\frac{1}{n^2} \sum_{t=1}^n x_t x_t^\top \rightarrow \int_0^1 V(r) V(r)^\top dr$  a.s. by Theorem 3.1 in Park and Phillips (2001). Also, since for any real vector  $a$ ,  $a^\top x_t / \sqrt{n} \rightarrow a^\top V(r)$  a.s. we have  $\frac{1}{n} \sum_{t=1}^n a^\top x_t e_t \rightarrow_D \int_0^1 a^\top V(r) dU(r)$  by Theorem 3.1 in Park and Phillips (2001) again, implying that  $\frac{1}{n} \sum_{t=1}^n x_t e_t \rightarrow_D \int_0^1 V(r) dU(r)$ . Meanwhile, by Theorem 1 of Phillips (2009, p. 1475)

$$\begin{aligned}\frac{1}{n} \sum_{t=1}^n g(\theta_0^\top x_t) x_t &= Q \frac{1}{n} \sum_{t=1}^n g(x_{1t}) Q^\top x_t = Q \frac{1}{n} \sum_{t=1}^n g(x_{1t}) \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} \\ &\rightarrow_P Q \begin{pmatrix} 0 \\ \int g(x) dx \int_0^1 V_2(r) dL_1(r, 0) \end{pmatrix}.\end{aligned}$$

Additionally,

$$\begin{aligned}\frac{1}{n} \sum_{t=1}^n \hat{g}_n(\theta^\top x_t) x_t &= \frac{1}{n} \sum_{t=1}^n [g(\theta^\top x_t) + Z_k(\theta^\top x_t)^\top (\tilde{c} - c) - \gamma_k(\theta^\top x_t)] x_t \\ &= \frac{1}{n} \sum_{t=1}^n g(\theta^\top x_t) x_t + \frac{1}{n} \sum_{t=1}^n Z_k(\theta^\top x_t)^\top (\tilde{c} - c) x_t - \frac{1}{n} \sum_{t=1}^n \gamma_k(\theta^\top x_t) x_t.\end{aligned}$$

Again, by Theorem 1 of Phillips (2009, p. 1475),

$$\frac{1}{n} \sum_{t=1}^n g(\theta^\top x_t) x_t = \tilde{Q} \frac{1}{n} \sum_{t=1}^n g(\|\theta\| \tilde{x}_{1t}) \tilde{Q}^\top x_t = \tilde{Q} \frac{1}{n} \sum_{t=1}^n g(\|\theta\| \tilde{x}_{1t}) \begin{pmatrix} \tilde{x}_{1t} \\ \tilde{x}_{2t} \end{pmatrix}$$

$$\rightarrow_P \tilde{Q} \begin{pmatrix} 0 \\ \int g(\|\theta\|x) dx \int_0^1 \tilde{V}_2(r) d\tilde{L}_1(r, 0) \end{pmatrix},$$

where  $\tilde{x}_{1t} = \theta^\top x_t / \|\theta\|$ ,  $\tilde{x}_{2t} = \tilde{Q}_2^\top x_t$ . On the other hand,

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n |Z_k(\theta^\top x_t)^\top (\tilde{c} - c)| \|x_t\| \leq \|\tilde{c} - c\| \frac{1}{n} \sum_{t=1}^n \|Z_k(\theta^\top x_t)\| \|x_t\| \\ & \leq O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \frac{1}{n} \sum_{t=1}^n \|Z_k(\theta^\top x_t)\| \|\tilde{Q}x_t\| \leq O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \frac{1}{n} \sum_{t=1}^n \|Z_k(\|\theta\|\tilde{x}_{1t})\| (|\tilde{x}_{1t}| + \|\tilde{x}_{2t}\|). \end{aligned}$$

Similar to  $x_{1t}$  and  $x_{2t}$ ,  $\frac{1}{\sqrt{t}}\tilde{x}_{1t}$  and  $\frac{1}{\sqrt{t}}\tilde{x}_{2t}$  have joint density  $\tilde{\psi}_t(x, w)$ , and for  $t$  large,  $\tilde{\psi}_t(x, w) = \tilde{f}_t(x)\phi(w)(1 + o(1))$  where  $\tilde{f}_t(x)$  is the marginal density of  $\frac{1}{\sqrt{t}}\tilde{x}_{1t}$  which is uniformly bounded in  $t$  and  $x$ , and  $\phi(w)$  is the density of standard normal variable of dimension  $d - 1$ . We only consider in what follows for large  $t$ , while for small  $t$  part we always can control it to be  $o_P(1)$ , as what we did before. Hence,

$$\begin{aligned} & \frac{\sqrt{k}}{\sqrt[4]{n}} \frac{1}{n} E \sum_{t=m_n}^n \|Z_k(\|\theta\|\tilde{x}_{1t})\| (|\tilde{x}_{1t}| + \|\tilde{x}_{2t}\|) \\ & \leq \frac{\sqrt{k}}{\sqrt[4]{n}} \frac{1}{n} \sum_{t=1}^n \frac{1}{\sqrt{t}} \int \|Z_k(\|\theta\|x)\| |x| dx + \frac{\sqrt{k}}{\sqrt[4]{n}} \frac{1}{n} \sum_{t=1}^n \frac{1}{\sqrt{t}} \iint \|Z_k(\|\theta\|x)\| \|\sqrt{t}w\| \phi(w) dx dw \\ & = O(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \frac{1}{\sqrt{n}} \int \|Z_k(x)\| |x| dx + O(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \int \|Z_k(x)\| dx \\ & = O(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \frac{1}{\sqrt{n}} k^{17/12} + \frac{\sqrt{k}}{\sqrt[4]{n}} k^{11/12} = o(1) \end{aligned}$$

by (6) of Lemma A.1 and Assumption B. Similarly, by Lemma A.1 again,  $\frac{1}{n} \sum_{t=1}^n \gamma_k(\theta^\top x_t) x_t = o_P(1)$ . This shows that (B.18) holds and hence (B.19) is valid trivially.

(2) By definition,  $\hat{\beta}_n$  and  $\hat{\theta}_n$  given by (2.8) solve that  $\frac{\partial L_n(\beta, \theta)}{\partial \beta} = 0$  and  $\frac{\partial L_n(\beta, \theta)}{\partial \theta} = 0$ , which gives  $\hat{\beta}_n = \tilde{\beta}_n(\hat{\theta}_n)$ , since  $\frac{\partial^2 L_n(\beta, \theta)}{\partial \beta \partial \beta^\top} = \sum_{t=1}^n x_t x_t^\top$  is positive definite almost surely, implying the uniqueness of the minimizer  $\tilde{\beta}_n$ . Because of the consistency of  $\tilde{\beta}_n(\theta)$  and  $\tilde{\beta}_n(\theta) \rightarrow_P \beta_0$  independent of  $\theta$ ,  $\hat{\beta}_n$  given by (2.8) is consistent as well.

On the other hand, in view of (B.18), we have  $\hat{\beta}_n - \beta_0 = O_P(n^{-1})$ . Note that  $\hat{\theta}_n$  is the minimizer of

$$L_n(\hat{\beta}_n, \theta) = \frac{1}{2} \sum_{t=1}^n [y_t - \hat{\beta}_n^\top x_t - \hat{g}_n(\theta^\top x_t)]^2 = \frac{1}{2} \sum_{t=1}^n [e_t + (\beta_0 - \hat{\beta}_n)^\top x_t + g(\theta_0^\top x_t) - \hat{g}_n(\theta^\top x_t)]^2.$$

Since  $(\beta_0 - \hat{\beta}_n)^\top x_t = O_P(n^{-1/2})$  uniformly in  $t$  and independent of  $\theta$ , the problem is equivalent to a single-index model estimation. Therefore,  $\hat{\theta}_n$  is consistent by virtue of Theorem 3.3.  $\square$

*Proof of Theorem 3.6.* Denote for any  $\vartheta = (\beta, \theta)$ ,

$$\mathfrak{S}_n(\vartheta) = \begin{pmatrix} \mathfrak{S}_{n,1}(\vartheta) \\ \mathfrak{S}_{n,2}(\vartheta) \end{pmatrix} = \begin{pmatrix} \frac{\partial L_n(\vartheta)}{\partial \beta} \\ \frac{\partial L_n(\vartheta)}{\partial \theta} \end{pmatrix}, \quad \mathfrak{J}_n(\vartheta) = \begin{pmatrix} \mathfrak{J}_{n,11}(\vartheta) & \mathfrak{J}_{n,12}(\vartheta) \\ \mathfrak{J}_{n,21}(\vartheta) & \mathfrak{J}_{n,22}(\vartheta) \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 L_n(\vartheta)}{\partial \beta \partial \beta^\top} & \frac{\partial^2 L_n(\vartheta)}{\partial \beta \partial \theta^\top} \\ \frac{\partial^2 L_n(\vartheta)}{\partial \theta \partial \beta^\top} & \frac{\partial^2 L_n(\vartheta)}{\partial \theta \partial \theta^\top} \end{pmatrix}.$$

Also, for any  $\mu = (\lambda, \alpha)$ ,  $\mathfrak{S}_n(\mu)$  and  $\mathfrak{J}_n(\mu)$  are defined in the same way but with the parameters rotated.

Denote  $\tilde{D}_n = \text{diag}(nI_d, D_n) = \text{diag}(nI_d, n^{1/4}, n^{3/4}I_{d-1})$ , in which  $D_n$  is the same as in Theorem 3.1. Thus, the relationship (3.15):

$$0 = \mathfrak{S}_n(\hat{\mu}_n) = \mathfrak{S}_n(\mu_0) + \mathfrak{J}_n(\mu_n)(\hat{\mu}_n - \mu_0) \quad (\text{B.20})$$

may be equivalently written as

$$\tilde{D}_n^{-1} \mathfrak{S}_n(\mu_0) + \tilde{D}_n^{-1} \mathfrak{J}_n(\mu_n) \tilde{D}_n^{-1} \tilde{D}_n (\hat{\mu}_n - \mu_0) = 0. \quad (\text{B.21})$$

It follows from (B.21) that

$$n^{-1} \mathfrak{S}_{n,1}(\mu_0) + n^{-2} \mathfrak{J}_{n,11}(\mu_n) n (\hat{\lambda}_n - \lambda_0) + n^{-1} \mathfrak{J}_{n,12}(\mu_n) D_n^{-1} D_n (\hat{\alpha}_n - \alpha_0) = 0 \quad (\text{B.22})$$

$$D_n^{-1} \mathfrak{S}_{n,2}(\mu_0) + D_n^{-1} \mathfrak{J}_{n,21}(\mu_n) n^{-1} n (\hat{\lambda}_n - \lambda_0) + D_n^{-1} \mathfrak{J}_{n,22}(\mu_n) D_n^{-1} D_n (\hat{\alpha}_n - \alpha_0) = 0. \quad (\text{B.23})$$

The results of (3.16) and (3.17) will be derived from (B.22) and (B.23), respectively, which are shown in the following two steps.

**Step I** We first prove (3.17) from (B.23). To begin, noting that  $\mathfrak{S}_{n,2}(\mu_0)$  and  $\mathfrak{J}_{n,22}(\mu_0)$  are exactly the  $S_n(\alpha_0)$  and  $J_n(\alpha_0)$  in Theorem 3.1, respectively, since  $y_t - \beta_0^\top x_t$  in model (1.1) plays the same role as  $y_t$  in model (1.2). Therefore,

$$D_n^{-1} \mathfrak{S}_{n,2}(\mu_0) \rightarrow_D R^{1/2} W(1), \quad \text{and} \quad D_n^{-1} \mathfrak{J}_{n,22}(\mu_0) D_n^{-1} \rightarrow_P R, \quad (\text{B.24})$$

where  $R$  and  $W$  are defined in Theorem 3.1.

To prove (3.17), it therefore suffices to show that

$$D_n (\hat{\alpha}_n - \alpha_0) = [D_n^{-1} \mathfrak{J}_{n,22}(\mu_0) D_n^{-1}]^{-1} D_n^{-1} \mathfrak{S}_{n,2}(\mu_0) + o_P(1). \quad (\text{B.25})$$

We shall follow Theorem 10.1 of Wooldridge (1994) to prove (B.25). Observe that  $n(\hat{\lambda}_n - \lambda_0) = Q^\top n(\hat{\beta}_n - \beta_0) = o_P(1)$  due to Lemma B.2. Additionally,  $D_n^{-1} \mathfrak{J}_{n,21}(\mu_0) n^{-1} = o_P(1)$ . In fact,

$$\begin{aligned} D_n^{-1} \mathfrak{J}_{n,21}(\mu_0) n^{-1} &= n^{-1} D_n^{-1} Q^\top \mathfrak{J}_{n,21}(\vartheta_0) Q = n^{-1} D_n^{-1} Q^\top \sum_{t=1}^n \dot{g}_n(\theta_0^\top x_t) x_t x_t^\top Q \\ &= n^{-1} D_n^{-1} Q^\top \sum_{t=1}^n \dot{g}(\theta_0^\top x_t) x_t x_t^\top Q + n^{-1} D_n^{-1} Q^\top \sum_{t=1}^n [\dot{g}_n(\theta_0^\top x_t) - \dot{g}(\theta_0^\top x_t)] x_t x_t^\top Q. \end{aligned}$$

Here, by virtue of Theorem 1 of Phillips (2009),

$$n^{-1}D_n^{-1}Q^\top \sum_{t=1}^n \dot{g}(\theta_0^\top x_t) x_t x_t^\top Q = n^{-1} \sum_{t=1}^n \dot{g}(x_{1t}) \begin{pmatrix} n^{-1/4} x_{1t}^2 & n^{-1/4} x_{1t} x_{2t}^\top \\ n^{-3/4} x_{1t} x_{2t} & n^{-3/4} x_{2t} x_{2t}^\top \end{pmatrix} \rightarrow_P 0.$$

At the meantime, as  $\hat{g}_n(u) - \dot{g}(u) = \dot{Z}_k(u)^\top (\tilde{c} - c) - \dot{\gamma}_k(u)$  for any  $u$ ,

$$\begin{aligned} & n^{-1}D_n^{-1}Q^\top \sum_{t=1}^n [\hat{g}_n(\theta_0^\top x_t) - \dot{g}(\theta_0^\top x_t)] x_t x_t^\top Q \\ &= n^{-1}D_n^{-1}Q^\top \sum_{t=1}^n [\dot{Z}_k(\theta_0^\top x_t)^\top (\tilde{c} - c) - \dot{\gamma}_k(\theta_0^\top x_t)] x_t x_t^\top Q \\ &= n^{-1} \sum_{t=1}^n [\dot{Z}_k(x_{1t})^\top (\tilde{c} - c) - \dot{\gamma}_k(x_{1t})] \begin{pmatrix} n^{-1/4} x_{1t}^2 & n^{-1/4} x_{1t} x_{2t}^\top \\ n^{-3/4} x_{1t} x_{2t} & n^{-3/4} x_{2t} x_{2t}^\top \end{pmatrix}, \end{aligned}$$

where every block is  $o_P(1)$ . Indeed, in the block of (1,1),

$$\begin{aligned} & n^{-5/4} \sum_{t=1}^n [\dot{Z}_k(x_{1t})^\top (\tilde{c} - c) - \dot{\gamma}_k(x_{1t})] x_{1t}^2 \\ &= n^{-5/4} \sum_{t=1}^n [\dot{Z}_k(x_{1t})^\top (\tilde{c} - c) x_{1t}^2 - n^{-5/4} \sum_{t=1}^n \dot{\gamma}_k(x_{1t}) x_{1t}^2] \\ &= n^{-5/4} \left( \sum_{t=1}^n x_{1t}^2 \dot{Z}_k(x_{1t}) \right)^\top (\tilde{c} - c) + o_P(1) \end{aligned}$$

because  $n^{-5/4} E \sum_{t=1}^n |\dot{\gamma}_k(x_{1t})| x_{1t}^2 \leq O(1) n^{-3/4} \int |\dot{\gamma}_k(x)| x^2 dx = o(1) n^{-3/4} k^{-m/2+29/12} = o(1)$

by virtue of the density of  $\frac{1}{\sqrt{t}} x_{1t}$  and (7) of Lemma A.1, while

$$\begin{aligned} & n^{-5/4} \left\| \left( \sum_{t=1}^n x_{1t}^2 \dot{Z}_k(x_{1t}) \right)^\top (\tilde{c} - c) \right\| \leq n^{-5/4} \sum_{t=1}^n x_{1t}^2 \|\dot{Z}_k(x_{1t})\| \|\tilde{c} - c\| \\ &= O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} n^{-5/4} \sum_{t=1}^n x_{1t}^2 \|\dot{Z}_k(x_{1t})\| = O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} n^{-3/4} \int x^2 \|\dot{Z}_k(x)\| dx \\ &= O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} n^{-3/4} k^{29/12} = o_P(1) \end{aligned}$$

again by virtue of the density of  $\frac{1}{\sqrt{t}} x_{1t}$  and (6) of Lemma A.1 and Theorem 2.1 for  $\tilde{c} - c$ . In the block of (2,2),

$$\begin{aligned} & n^{-7/4} \left\| \sum_{t=1}^n [\dot{Z}_k(x_{1t})^\top (\tilde{c} - c) - \dot{\gamma}_k(x_{1t})] x_{2t} x_{2t}^\top \right\| \\ & \leq n^{-7/4} \sum_{t=1}^n \|\dot{Z}_k(x_{1t})\| \|\tilde{c} - c\| \|x_{2t} x_{2t}^\top\| + n^{-7/4} \sum_{t=1}^n |\dot{\gamma}_k(x_{1t})| \|x_{2t} x_{2t}^\top\| \end{aligned}$$

$$\begin{aligned}
&= O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} n^{-7/4} \sum_{t=1}^n \|\dot{Z}_k(x_{1t})\| \|\tilde{c} - c\| \|x_{2t} x_{2t}^\top\| + n^{-7/4} \sum_{t=1}^n |\dot{\gamma}_k(x_{1t})| \|x_{2t} x_{2t}^\top\| \\
&= O_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} n^{-7/4} n^{3/2} \int \|\dot{Z}_k(x)\| dx + n^{-7/4} n^{3/2} \int |\dot{\gamma}_k(x)| dx \\
&= O_P(1) \frac{\sqrt{k}}{\sqrt{n}} k^{17/12} + n^{-1/4} k^{-m/2+17/12} = o_P(1).
\end{aligned}$$

The other two blocks can be shown similarly.

Define  $\tilde{C}_n = n^{-\delta} \tilde{D}_n = \text{diag}(n^{1-\delta} I_d, n^{-\delta} D_n) = \text{diag}(n^{1-\delta} I_d, C_n)$  for some  $\delta > 0$  such that  $\tilde{C}_n \tilde{D}_n^{-1} \rightarrow 0$  as  $n \rightarrow \infty$ , where  $C_n = n^{-\delta} D_n$  which we use in the proof of Theorem 3.2. It follows from (B.23) that

$$\begin{aligned}
0 &= D_n^{-1} \mathfrak{S}_{n,2}(\mu_0) + D_n^{-1} [\mathfrak{J}_{n,21}(\mu_n) - \mathfrak{J}_{n,21}(\mu_0)] n^{-1} n(\hat{\lambda}_n - \lambda_0) + o_P(1) \\
&\quad + D_n^{-1} \mathfrak{J}_{n,22}(\mu_0) D_n^{-1} D_n(\hat{\alpha}_n - \alpha_0) + D_n^{-1} [\mathfrak{J}_{n,22}(\mu_n) - \mathfrak{J}_{n,22}(\mu_0)] D_n^{-1} D_n(\hat{\alpha}_n - \alpha_0) \\
&= D_n^{-1} \mathfrak{S}_{n,2}(\mu_0) + n^{-2\delta} C_n^{-1} [\mathfrak{J}_{n,21}(\mu_n) - \mathfrak{J}_{n,21}(\mu_0)] n^{-1+\delta} n(\hat{\lambda}_n - \lambda_0) + o_P(1) \\
&\quad + D_n^{-1} \mathfrak{J}_{n,22}(\mu_0) D_n^{-1} D_n(\hat{\alpha}_n - \alpha_0) + n^{-2\delta} C_n^{-1} [\mathfrak{J}_{n,22}(\mu_n) - \mathfrak{J}_{n,22}(\mu_0)] C_n^{-1} D_n(\hat{\alpha}_n - \alpha_0).
\end{aligned}$$

The requirements (i)-(ii) in Theorem 10.1 of Wooldridge (1994) are trivially fulfilled and the requirement (iii) will be satisfied if we can show

$$\sup_{\{\mu: \|\tilde{C}_n(\mu - \mu_0)\| < 1\}} \|n^{-1+\delta} C_n^{-1} [\mathfrak{J}_{n,21}(\mu) - \mathfrak{J}_{n,21}(\mu_0)]\| = o_P(1) \quad (\text{B.26})$$

$$\sup_{\{\mu: \|\tilde{C}_n(\mu - \mu_0)\| < 1\}} \|C_n^{-1} [\mathfrak{J}_{n,22}(\mu) - \mathfrak{J}_{n,22}(\mu_0)] C_n^{-1}\| = o_P(1). \quad (\text{B.27})$$

Denote  $\eta_t = \alpha^1 x_{1t} + (\alpha^2)^\top x_{2t}$  for brevity. Note that once  $\alpha$  takes value  $\alpha_0$ ,  $\eta_t$  would become  $x_{1t}$ . Thus,

$$\begin{aligned}
\mathfrak{J}_{n,22}(\mu) &= \sum_{t=1}^n \hat{g}_n^2(\eta_t) Q^\top x_t x_t^\top Q - \sum_{t=1}^n [y_t - (\lambda^1 x_{1t} + (\lambda^2)^\top x_{2t}) - \hat{g}_n(\eta_t)] \ddot{g}_n(\eta_t) Q^\top x_t x_t^\top Q \\
&= \sum_{t=1}^n \hat{g}_n^2(\eta_t) Q^\top x_t x_t^\top Q - \sum_{t=1}^n [y_t - (\lambda_0^1 x_{1t} + (\lambda_0^2)^\top x_{2t}) - \hat{g}_n(\eta_t)] \ddot{g}_n(\eta_t) Q^\top x_t x_t^\top Q \\
&\quad - \sum_{t=1}^n [(\lambda_0^1 - \lambda^1) x_{1t} + (\lambda_0^2 - \lambda^2)^\top x_{2t}] \ddot{g}_n(\eta_t) Q^\top x_t x_t^\top Q \\
&:= J_n(\alpha) - \mathfrak{J}'_{n,22}(\mu),
\end{aligned}$$

where  $\mathfrak{J}'_{n,22}(\mu) = \sum_{t=1}^n [(\lambda_0^1 - \lambda^1) x_{1t} + (\lambda_0^2 - \lambda^2)^\top x_{2t}] \ddot{g}_n(\eta_t) Q^\top x_t x_t^\top Q$ , and  $J_n(\alpha)$  exactly is the same one in the proof of Theorem 3.2 (because  $y_t - (\lambda_0^1 x_{1t} + (\lambda_0^2)^\top x_{2t}) = y_t - \beta_0^\top x_t$  plays the same role as  $y_t$  in the single-index model) for which we have shown that

$$\sup_{\{\alpha: \|C_n(\alpha - \alpha_0)\| < 1\}} \|C_n^{-1} [J_n(\alpha) - J_n(\alpha_0)] C_n^{-1}\| = o_P(1) \quad (\text{B.28})$$

if we select  $\delta \in (0, 1/24)$ .

Accordingly, we only need to consider  $\mathfrak{J}_{n,21}(\mu)$  and  $\mathfrak{J}'_{n,22}(\mu)$ . In other words, to find some  $\delta > 0$  such that (B.26) and (B.27) are valid, it is sufficient to find a  $\delta \in (0, 1/24)$  such that

$$\sup_{\{\mu: \|\tilde{C}_n(\mu - \mu_0)\| < 1\}} \left\| \frac{1}{n^{1-\delta}} C_n^{-1} (\mathfrak{J}_{n,21}(\mu) - \mathfrak{J}_{n,21}(\mu_0)) \right\| = o_P(1), \quad (\text{B.29})$$

$$\sup_{\{\mu: \|\tilde{C}_n(\mu - \mu_0)\| < 1\}} \left\| C_n^{-1} (\mathfrak{J}'_{n,22}(\mu) - \mathfrak{J}'_{n,22}(\mu_0)) C_n^{-1} \right\| = o_P(1). \quad (\text{B.30})$$

To do so, we have to rewrite both  $\mathfrak{J}_{n,21}(\mu)$  and  $\mathfrak{J}'_{n,22}(\mu)$  conformably with  $C_n$  as

$$\mathfrak{J}_{n,21}(\mu) = \sum_{t=1}^n \hat{g}_n(\eta_t) \begin{pmatrix} x_{1t}^2 & x_{1t}x_{2t}^\top \\ x_{1t}x_{2t} & x_{2t}x_{2t}^\top \end{pmatrix},$$

$$\mathfrak{J}'_{n,22}(\mu) = \sum_{t=1}^n [(\lambda_0^1 - \lambda^1)x_{1t} + (\lambda_0^2 - \lambda^2)^\top x_{2t}] \ddot{g}_n(\eta_t) \begin{pmatrix} x_{1t}^2 & x_{1t}x_{2t}^\top \\ x_{1t}x_{2t} & x_{2t}x_{2t}^\top \end{pmatrix}.$$

Thus, it suffices to demonstrate

$$\frac{1}{n^{5/4-2\delta}} \sum_{t=1}^n [\hat{g}_n(\eta_t) - \hat{g}_n(x_{1t})] x_{1t}^2 = o_P(1), \quad (\text{B.31})$$

$$\frac{1}{n^{5/4-2\delta}} \sum_{t=1}^n [\hat{g}_n(\eta_t) - \hat{g}_n(x_{1t})] x_{1t}x_{2t} = o_P(1), \quad (\text{B.32})$$

$$\frac{1}{n^{7/4-2\delta}} \sum_{t=1}^n [\hat{g}_n(\eta_t) - \hat{g}_n(x_{1t})] x_{2t}x_{2t}^\top = o_P(1), \quad (\text{B.33})$$

and noting that  $\mathfrak{J}'_{n,22}(\mu_0) \equiv 0$ ,

$$\frac{1}{n^{1/2-2\delta}} \sum_{t=1}^n [(\lambda_0^1 - \lambda^1)x_{1t} + (\lambda_0^2 - \lambda^2)^\top x_{2t}] \ddot{g}_n(\eta_t) x_{1t}^2 = o_P(1), \quad (\text{B.34})$$

$$\frac{1}{n^{1-2\delta}} \sum_{t=1}^n [(\lambda_0^1 - \lambda^1)x_{1t} + (\lambda_0^2 - \lambda^2)^\top x_{2t}] \ddot{g}_n(\eta_t) x_{1t}x_{2t} = o_P(1), \quad (\text{B.35})$$

$$\frac{1}{n^{3/2-2\delta}} \sum_{t=1}^n [(\lambda_0^1 - \lambda^1)x_{1t} + (\lambda_0^2 - \lambda^2)^\top x_{2t}] \ddot{g}_n(\eta_t) x_{2t}x_{2t}^\top = o_P(1), \quad (\text{B.36})$$

uniformly in  $\lambda^1, \lambda^2, \alpha^1$  and  $\alpha^2$  satisfying

$$\begin{aligned} |\lambda^1 - \lambda_0^1| &< n^{-1+\delta} \quad \text{and} \quad \|\lambda^2 - \lambda_0^2\| < n^{-1+\delta}, \\ |\alpha^1 - \alpha_0^1| &< n^{-1/4+\delta} \quad \text{and} \quad \|\alpha^2 - \alpha_0^2\| < n^{-3/4+\delta}. \end{aligned} \quad (\text{B.37})$$

In what follows, we shall use the function  $f_{\text{sup}}^\epsilon(x)$  defined by (B.9) for any integrable function and its property in (B.10) to prove (B.31)-(B.36) one by one. By virtue of the mean value theorem we have

$$\frac{1}{n^{5/4-2\delta}} \left| \sum_{t=1}^n [\hat{g}_n(\eta_t) - \hat{g}_n(x_{1t})] x_{1t}^2 \right| \leq \frac{1}{n^{5/4-2\delta}} \sum_{t=1}^n |\ddot{g}_n(\eta_t^*)| [|\alpha^1 - 1| |x_{1t}| + \|\alpha^2\| \|x_{2t}\|] x_{1t}^2$$

$$\begin{aligned}
&\leq \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{\hat{g}}_n(\eta_t^*)| |x_{1t}|^3 + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{\hat{g}}_n(\eta_t^*)| |x_{1t}|^2 \|x_{2t}\| \\
&\leq \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{g}(\eta_t^*)| |x_{1t}|^3 + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{\hat{g}}_n(\eta_t^*) - \ddot{g}(\eta_t^*)| |x_{1t}|^3 \\
&\quad + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{g}(\eta_t^*)| |x_{1t}|^2 \|x_{2t}\| + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{\hat{g}}_n(\eta_t^*) - \ddot{g}(\eta_t^*)| |x_{1t}|^2 \|x_{2t}\| \\
&\leq \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n (x_{1t}^3 \ddot{g})_{\text{sup}}(x_{1t}) + \frac{1}{n^{3/2-3\delta}} \|c - \tilde{c}\| \sum_{t=1}^n \|\ddot{Z}_k(\eta_t^*)\| |x_{1t}|^3 + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{\gamma}_k(\eta_t^*)| |x_{1t}|^3 \\
&\quad + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n (x_{1t}^2 \ddot{g})_{\text{sup}}(x_{1t}) \|x_{2t}\| + \frac{1}{n^{2-3\delta}} \|c - \tilde{c}\| \sum_{t=1}^n \sum_{t=1}^n \|\ddot{Z}_k(\eta_t^*)\| |x_{1t}|^2 \|x_{2t}\| \\
&\quad + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{\gamma}_k(\eta_t^*)| |x_{1t}|^2 \|x_{2t}\| \\
&= O_P(1) \frac{1}{n^{1-3\delta}} + O_P(1) \frac{1}{n^{1-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \int \|\ddot{Z}_k(x)\| |x|^3 dx + O_P(1) \frac{1}{n^{1-3\delta}} \int |\ddot{\gamma}_k(x)| |x|^3 dx \\
&\quad + O_P(1) \frac{1}{n^{1-3\delta}} + O_P(1) \frac{1}{n^{1-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \int \|\ddot{Z}_k(x)\| |x|^2 dx + O_P(1) \frac{1}{n^{1-3\delta}} \int |\ddot{\gamma}_k(x)| |x|^2 dx \\
&= O_P(1) \frac{1}{n^{1-3\delta}} + O_P(1) \frac{1}{n^{1-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} k^{41/12} + O_P(1) \frac{1}{n^{1-3\delta}} k^{-m/2+41/12} \leq O_P(1) \frac{1}{n^{3/4-3\delta}},
\end{aligned}$$

due to  $k^4 < \sqrt{n}$  by Assumption B, where we have used Theorem 2.1, Lemma A.5 and (6) and (7) in Lemma A.1. So (B.31) is true as long as  $\delta < 1/4$ . For (B.32), note that

$$\begin{aligned}
&\frac{1}{n^{5/4-2\delta}} \left\| \sum_{t=1}^n [\dot{\hat{g}}_n(\eta_t) - \dot{\hat{g}}_n(x_{1t})] x_{1t} x_{2t} \right\| \leq \frac{1}{n^{5/4-2\delta}} \sum_{t=1}^n |\dot{\hat{g}}_n(\eta_t) - \dot{\hat{g}}_n(x_{1t})| |x_{1t}| \|x_{2t}\| \\
&\leq \frac{1}{n^{5/4-2\delta}} \sum_{t=1}^n |\ddot{\hat{g}}_n(\eta_t^*)| [|\alpha^1 - 1| |x_{1t}| + |\alpha^2| \|x_{2t}\|] |x_{1t}| \|x_{2t}\| \\
&\leq \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{\hat{g}}_n(\eta_t^*)| |x_{1t}|^2 \|x_{2t}\| + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{\hat{g}}_n(\eta_t^*)| |x_{1t}| \|x_{2t}\|^2 \\
&= \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{g}(\eta_t^*)| |x_{1t}|^2 \|x_{2t}\| + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{\hat{g}}_n(\eta_t^*) - \ddot{g}(\eta_t^*)| |x_{1t}|^2 \|x_{2t}\| \\
&\quad + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{g}(\eta_t^*)| |x_{1t}| \|x_{2t}\|^2 + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{\hat{g}}_n(\eta_t^*) - \ddot{g}(\eta_t^*)| |x_{1t}| \|x_{2t}\|^2 \\
&\leq \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n (x_{1t}^2 \ddot{g})_{\text{sup}}(x_{1t}) \|x_{2t}\| + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{Z}_k(\eta_t^*)^\top (\tilde{c} - c)| |x_{1t}|^2 \|x_{2t}\| \\
&\quad + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{\gamma}_k(\eta_t^*)| |x_{1t}|^2 \|x_{2t}\| + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n (x_{1t} \ddot{g})_{\text{sup}}(x_{1t}) \|x_{2t}\|^2 \\
&\quad + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{Z}_k(\eta_t^*)^\top (\tilde{c} - c)| |x_{1t}| \|x_{2t}\|^2 + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{\gamma}_k(\eta_t^*)| |x_{1t}| \|x_{2t}\|^2
\end{aligned}$$

$$\begin{aligned}
&\leq O_P(1) \frac{1}{n^{1/2-3\delta}} + \frac{1}{n^{3/2-3\delta}} \|\tilde{c} - c\| \sum_{t=1}^n (x_{1t}^2 \|\ddot{Z}_k(x)\|)_{\sup}(x_{1t}) \|x_{2t}\| \\
&\quad + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n (x_{1t}^2 \ddot{\gamma}_k(x))_{\sup}(x_{1t}) \|x_{2t}\| + O_P(1) \frac{1}{n^{1/2-3\delta}} \\
&\quad + \frac{1}{n^{2-3\delta}} \|\tilde{c} - c\| \sum_{t=1}^n (x_{1t} \|\ddot{Z}_k(x)\|)_{\sup}(x_{1t}) \|x_{2t}\|^2 + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n (x_{1t} \ddot{\gamma}_k(x))_{\sup}(x_{1t}) \|x_{2t}\|^2 \\
&= O_P(1) \frac{1}{n^{1/2-3\delta}} + O_P(1) \frac{1}{n^{1/2-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \int x^2 \|\ddot{Z}_k(x)\| dx + O_P(1) \frac{1}{n^{1/2-3\delta}} \int x^2 |\ddot{\gamma}_k(x)| dx \\
&\quad + O_P(1) \frac{1}{n^{1/2-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \int |x| \|\ddot{Z}_k(x)\| dx + O_P(1) \frac{1}{n^{1/2-3\delta}} \int |x| \ddot{\gamma}_k(x) dx \\
&= O_P(1) \frac{1}{n^{1/2-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} k^{35/12} \leq O_P(1) \frac{1}{n^{3/4-3\delta}} k^{3.5} \leq O_P(1) \frac{1}{n^{1/4-3\delta}},
\end{aligned}$$

implying that (B.32) holds if  $\delta < 1/12$ , where we have used Theorem 2.1, Lemma A.5 and (6) and (7) in Lemma A.1 as well as Assumption B.

As for (B.33), notice that there is an extra  $\sqrt{n}$  in the normalizer than (B.32) which will be used to address one extra  $x_{2t}$ . Thus, under the same condition as (B.32), (B.33) is satisfied.

Now, we turn to (B.34). Note that

$$\begin{aligned}
&\frac{1}{n^{1/2-2\delta}} \left| \sum_{t=1}^n [(\lambda_0^1 - \lambda^1)x_{1t} + (\lambda_0^2 - \lambda^2)^\top x_{2t}] \ddot{g}_n(\eta_t) x_{1t}^2 \right| \\
&\leq \frac{1}{n^{1/2-2\delta}} \sum_{t=1}^n [|\lambda_0^1 - \lambda^1| |x_{1t}| + |\lambda_0^2 - \lambda^2| \|x_{2t}\|] |\ddot{g}_n(\eta_t)| x_{1t}^2 \\
&\leq \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{g}_n(\eta_t)| |x_{1t}|^3 + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{g}_n(\eta_t)| |x_{1t}|^2 \|x_{2t}\| \\
&\leq \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{g}(\eta_t)| |x_{1t}|^3 + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{g}_n(\eta_t) - \ddot{g}(\eta_t)| |x_{1t}|^3 \\
&\quad + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{g}(\eta_t)| |x_{1t}|^2 \|x_{2t}\| + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n |\ddot{g}_n(\eta_t) - \ddot{g}(\eta_t)| |x_{1t}|^2 \|x_{2t}\| \\
&\leq \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n (x_{1t}^3 \ddot{g}(x))_{\sup}(x_{1t}) + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n [|\ddot{Z}_k(\eta_t)^\top (\tilde{c} - c)| + |\ddot{\gamma}_k(\eta_t)|] |x_{1t}|^3 \\
&\quad + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n (x_{1t}^2 \ddot{g}(x))_{\sup}(x_{1t}) \|x_{2t}\| + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n [|\ddot{Z}_k(\eta_t)^\top (\tilde{c} - c)| + |\ddot{\gamma}_k(\eta_t)|] |x_{1t}|^2 \|x_{2t}\| \\
&\leq O_P(1) \frac{1}{n^{1-3\delta}} + O_P(1) \frac{1}{n^{3/2-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n (x_{1t}^3 \|\ddot{Z}_k(x)\|)_{\sup}(x_{1t}) + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n (x_{1t}^3 \ddot{\gamma}_k(x))_{\sup}(x_{1t}) \\
&\quad + O_P(1) \frac{1}{n^{1/2-3\delta}} + O_P(1) \frac{1}{n^{3/2-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n (x_{1t}^2 \|\ddot{Z}_k(x)\|)_{\sup}(x_{1t}) \|x_{2t}\| \\
&\quad + \frac{1}{n^{3/2-3\delta}} \sum_{t=1}^n (x_{1t}^2 \ddot{\gamma}_k(x))_{\sup}(x_{1t}) \|x_{2t}\|
\end{aligned}$$



$$\begin{aligned}
&= O_P(1) \frac{1}{n^{1/2-3\delta}} + O_P(1) \frac{1}{n^{1-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \int |x|^3 \|\ddot{Z}_k(x)\| dx + O_P(1) \frac{1}{n^{1-3\delta}} \int |x^3 \dot{\gamma}_k(x)| dx \\
&\quad + O_P(1) \frac{1}{n^{1/2-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \int |x|^2 \|\ddot{Z}_k(x)\| dx + O_P(1) \frac{1}{n^{1/2-3\delta}} \int |x^2 \ddot{\gamma}_k(x)| dx \\
&= O_P(1) \frac{1}{n^{1/2-3\delta}} + O_P(1) \frac{1}{n^{5/4-3\delta}} \sqrt{k} k^{41/12} + O_P(1) \frac{1}{n^{1-3\delta}} k^{-m/2+41/12} \\
&\quad + O_P(1) \frac{1}{n^{3/4-3\delta}} \sqrt{k} k^{35/12} + O_P(1) \frac{1}{n^{1/2-3\delta}} k^{-m/2+35/12} \\
&= O_P(1)
\end{aligned}$$

as long as  $\delta < 1/12$  since  $k^4 < \sqrt{n}$  by Assumption B, where again we have used Theorem 2.1, Lemma A.5 and (6) and (7) in Lemma A.1. Regarding (B.35), notice that

$$\begin{aligned}
&\frac{1}{n^{1-2\delta}} \left\| \sum_{t=1}^n [(\lambda_0^1 - \lambda^1) x_{1t} + (\lambda_0^2 - \lambda^2)^\top x_{2t}] \ddot{g}_n(\eta_t) x_{1t} x_{2t} \right\| \\
&\leq \frac{1}{n^{1-2\delta}} \sum_{t=1}^n |(\lambda_0^1 - \lambda^1) x_{1t} + (\lambda_0^2 - \lambda^2)^\top x_{2t}| \|\ddot{g}_n(\eta_t) x_{1t}\| \|x_{2t}\| \\
&\leq \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{g}_n(\eta_t) x_{1t}^2| \|x_{2t}\| + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{g}_n(\eta_t) x_{1t}| \|x_{2t}\|^2 \\
&\leq \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{g}(\eta_t) x_{1t}^2| \|x_{2t}\| + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{g}_n(\eta_t) - \ddot{g}(\eta_t)| x_{1t}^2 \|x_{2t}\| \\
&\quad + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{g}(\eta_t) x_{1t}| \|x_{2t}\|^2 + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n |\ddot{g}_n(\eta_t) - \ddot{g}(\eta_t)| |x_{1t}| \|x_{2t}\|^2 \\
&\leq \frac{1}{n^{2-3\delta}} \sum_{t=1}^n (x_{1t}^2 \ddot{g}(x))_{\sup}(x_{1t}) \|x_{2t}\| + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n [|\ddot{Z}_k(\eta_t)^\top (\tilde{c} - c)| + |\ddot{\gamma}_k(\eta_t)|] x_{1t}^2 \|x_{2t}\| \\
&\quad + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n (x_{1t} \ddot{g}(x))_{\sup}(x_{1t}) \|x_{2t}\|^2 + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n [|\ddot{Z}_k(\eta_t)^\top (\tilde{c} - c)| + |\ddot{\gamma}_k(\eta_t)|] |x_{1t}| \|x_{2t}\|^2 \\
&\leq O_P(1) \frac{1}{n^{1-3\delta}} + O_P(1) \frac{1}{n^{2-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n (x_{1t}^2 \|\ddot{Z}_k(x)\|)_{\sup}(x_{1t}) \|x_{2t}\| + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n (x_{1t}^2 \ddot{\gamma}_k(x))_{\sup}(x_{1t}) \|x_{2t}\| \\
&\quad + O_P(1) \frac{1}{n^{1/2-3\delta}} + O_P(1) \frac{1}{n^{2-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \sum_{t=1}^n (x_{1t} \|\ddot{Z}_k(x)\|)_{\sup}(x_{1t}) \|x_{2t}\|^2 \\
&\quad + \frac{1}{n^{2-3\delta}} \sum_{t=1}^n (x_{1t} \ddot{\gamma}_k(x))_{\sup}(x_{1t}) \|x_{2t}\|^2 \\
&= O_P(1) \frac{1}{n^{1/2-3\delta}} + O_P(1) \frac{1}{n^{1-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \int x^2 \|\ddot{Z}_k(x)\| dx + O_P(1) \frac{1}{n^{1-3\delta}} \int x^2 |\ddot{\gamma}_k(x)| dx \\
&\quad + O_P(1) \frac{1}{n^{1/2-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} \int |x| \|\ddot{Z}_k(x)\| dx + O_P(1) \frac{1}{n^{1/2-3\delta}} \int |x| \ddot{\gamma}_k(x) dx \\
&= O_P(1) \frac{1}{n^{1/2-3\delta}} + O_P(1) \frac{1}{n^{1-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} k^{35/12} + O_P(1) \frac{1}{n^{1-3\delta}} k^{-m/2+35/12}
\end{aligned}$$

$$+ O_P(1) \frac{1}{n^{1/2-3\delta}} \frac{\sqrt{k}}{\sqrt[4]{n}} k^{29/12} + O_P(1) \frac{1}{n^{1/2-3\delta}} k^{-m/2+29/12}$$

$$= o_P(1)$$

if  $\delta < 1/12$ , where again we have used Theorem 2.1, Lemma A.5 and (6) and (7) in Lemma A.1 as well as Assumption B.

Similarly, (B.36) will be satisfied if  $\delta < 1/12$ , as there is an extra  $\sqrt{n}$  in the normalizer comparing with (B.35). Taking (B.28) into account, the requirement (B.25) is verified if we choose  $\delta \in (0, 1/24)$ .

Furthermore, (B.24) indicates the condition (iv) in Wooldridge's theorem holds. Thus, the limit distribution (3.17) now follows directly.

**Step 2** We now turn to prove (3.16) from (B.22). Because  $\mathfrak{J}_{n,12}(\mu_n) = \mathfrak{J}_{n,21}(\mu_n)^\top$ ,  $D_n(\hat{\alpha}_n - \alpha_0) = o_P(1)$ ,  $\mathfrak{J}_{n,11}(\mu_n)$  is independent of  $\mu_n$  and we have shown  $D_n^{-1} \mathfrak{J}_{n,21}(\mu_0) n^{-1} = o_P(1)$ , in view of (B.26) it follows that  $n(\hat{\lambda}_n - \lambda_0) = (n^{-2} \mathfrak{J}_{n,11}(\mu_n))^{-1} n^{-1} \mathfrak{S}_{n,1}(\mu_0) + o_P(1)$ . Note that

$$\begin{aligned} \frac{1}{n} \mathfrak{S}_{n,1}(\mu_0) &= \frac{1}{n} Q^\top \frac{\partial L_n(\vartheta_0)}{\partial \beta} = \frac{1}{n} Q^\top \sum_{t=1}^n (y_t - \beta_0^\top x_t - \hat{g}_n(\theta_0^\top x_t)) x_t \\ &= \frac{1}{n} \sum_{t=1}^n (e_t + g(\theta_0^\top x_t) - \hat{g}_n(\theta_0^\top x_t)) Q^\top x_t \\ &= \frac{1}{n} \sum_{t=1}^n e_t Q^\top x_t + \frac{1}{n} \sum_{t=1}^n (g(x_{1t}) - \hat{g}_n(x_{1t})) Q^\top x_t \\ &= \frac{1}{n} \sum_{t=1}^n e_t Q^\top x_t + \frac{1}{n} \sum_{t=1}^n (Z_k(x_{1t})^\top (c - \tilde{c}) + \gamma_k(x_{1t})) Q^\top x_t. \end{aligned}$$

However, as shown in Lemma B.2,

$$\frac{1}{n} \sum_{t=1}^n e_t Q^\top x_t \rightarrow_D Q^\top \int_0^1 V(r) dU(r)$$

as  $n \rightarrow \infty$ . In addition,

$$\begin{aligned} &\left\| \frac{1}{n} \sum_{t=1}^n (Z_k(x_{1t})^\top (c - \tilde{c}) + \gamma_k(x_{1t})) Q^\top x_t \right\| \\ &\leq \frac{1}{n} \sum_{t=1}^n |Z_k(x_{1t})^\top (c - \tilde{c})| (|x_{1t}| + \|x_{2t}\|) + \frac{1}{n} \sum_{t=1}^n |\gamma_k(x_{1t})| (|x_{1t}| + \|x_{2t}\|) \\ &\leq o_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \frac{1}{n} \sum_{t=1}^n \|Z_k(x_{1t})\| (|x_{1t}| + \|x_{2t}\|) + o_P(1) \frac{1}{n} \sqrt{n} \int |x \gamma_k(x)| dx + o_P(1) \int |\gamma_k(x)| dx \\ &= o_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \frac{1}{n} \sqrt{n} \int \|Z_k(x)\| |x| dx + o_P(1) \frac{\sqrt{k}}{\sqrt[4]{n}} \int \|Z_k(x)\| dx + o_P(1) = o_P(1), \end{aligned}$$

by virtue of the joint density of  $(\frac{1}{\sqrt{t}} x_{1t}, \frac{1}{\sqrt{t}} x_{2t})$  and Theorem 2.1 as well as Lemma A.1.

Notice also that

$$\frac{1}{n^2} \mathfrak{J}_{n,11}(\mu_n) = \frac{1}{n^2} Q^\top \sum_{t=1}^n x_t x_t^\top Q \rightarrow Q^\top \int V(r) V(r)^\top dr Q$$

almost surely using Theorem 3.1 of Park and Phillips (2001), from which (3.16) follows. The proof is complete.  $\square$

*Proof of Theorem 3.7.* The result of (3.18) follows directly from (3.16). In view of the proof of Theorem 3.3 as well as (3.17), (3.19) holds.  $\square$

*Proof of Theorem 3.8.* In view of (3.17) and the proofs of Theorems 3.4-3.5, it holds.  $\square$

## References

- Cai, B. and Gao, J. (2013). Hermite series estimation in nonlinear cointegrating models. Working paper available at <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2013>.
- Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92:477–489.
- Chang, Y. and Park, J. Y. (2003). Index models with integrated time series. *Journal of Econometrics*, 114:73–106.
- Dong, C. and Gao, J. (2012). Specification testing driven by orthogonal series in nonstationary time series models. Working paper at <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2012/20-12.php>.
- Dong, C. and Gao, J. (2013). Orthogonal expansion of lévy process functionals: theory and practice. Working paper available at <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2013/03-13.php>.
- Gao, J. (2012). Identification, estimation and specification in a class of semi-linear time series models. Working paper available at <http://mpira.ub.uni-muenchen.de/39256/>.
- Gao, J., King, M., Lu, Z., and Tjøstheim, D. (2009a). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory*, 25:1869–1892.
- Gao, J., King, M., Lu, Z., and Tjøstheim, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. *Annals of Statistics*, 37(68):3893–3928.
- Gao, J. and Phillips, P. C. B. (2013a). Functional coefficient nonstationary regression with non- and semi-parametric cointegration. <http://ideas.repec.org/p/msh/ebswps/2013-16.html>, 13:1–56.
- Gao, J. and Phillips, P. C. B. (2013b). Semiparametric estimation in triangular system equations with nonstationarity. *Journal of Econometrics*, 176:59–79.
- Gao, J., Tjøstheim, D., and Yin, J. (2012). Parametric model specification testing between parametric and nonparametric co-integrating regression models. Working paper available at <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2012/wp18-12.pdf>.

- Gao, J., Tong, H., and Wolff, R. (2002). Model specification tests in nonparametric stochastic regression models. *Journal of Multivariate Analysis*, 83:324–359.
- Gel'fand, I. M. and Shilov, G. E. (1964). *Generalized Functions*. Academic Press, New York.
- Guerre, E. and Moon, H. R. (2006). A study of a semiparametric binary choice model with integrated covariates. *Econometrics Theory*, 22:721–742.
- Härdle, W., Hall, P., and Ichimura, H. (1993). Optimal smoothing in single-index models. *Annals of Statistics*, 21:157–178.
- Jennrich, R. I. (1969). Asymptotic properties of nonlinear least square estimation. *Annals of Mathematical Statistics*, 40:633–643.
- Karlsen, H. A., Mykelbust, T., and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics*, 35:252–299.
- Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Annals of Statistics*, 29:372–416.
- Liang, H., Liu, X., Li, R., and Tsai, C.-L. (2010). Estimation and Testing for partially linear single-index models. *Annals of Statistics*, 38(6):3811–3836.
- Lin, W. and Kulasekera, K. B. (2007). Identifiability of single-index models and additive-index models. *Biometrika*, 94:496–501.
- Ma, Y. and Zhu, L. (2013). Doubly robust and efficient estimators for heteroscedastic partially single-index models allowing high dimension covariates. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75(2):305–322.
- Park, J. Y. and Phillips, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, 15:269–298.
- Park, J. Y. and Phillips, P. C. B. (2000). Nonstationary binary choice. *Econometrica*, 68:1249–1280.
- Park, J. Y. and Phillips, P. C. B. (2001). Nonlinear regression with integrated time series. *Econometrica*, 69(1):117–161.
- Phillips, P. C. B. (2009). Local time theory and spurious nonparametric regression. *Econometric Theory*, 25:1466–1497.
- Phillips, P. C. B., Li, D., and Gao, J. (2013). Estimating smooth structure change in cointegration models. Working paper available at <http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2013/22-13.php>.
- Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer-verlag, New York.
- Robinson, P. M. (1988). Root-N-consistent semiparametric regression. *Econometrica*, 56:931–954.
- Severini, T. A. and Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *Annals of Statistics*, 20:1768–1802.

- Szego, G. (1975). *Orthogonal Polynomials*. Colloquium publications XXIII. American Mathematical Association, Providence, Rhode Island.
- Wang, J.-L., Xue, L., Zhu, L., and Chong, Y. S. (2010). Estimation for a partial-linear single-index model. *Annals of Statistics*, 38:246–274.
- Wang, Q., Lin, Y. X., and Gulati, C. M. (2003). Asymptotics for general fractionally integrated processes with applications to unit root tests. *Econometric Theory*, 19:143–164.
- Wang, Q. and Phillips, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, 25:710–738.
- Wang, Q. and Phillips, P. C. B. (2009b). Structure nonparametric cointegrating regression. *Econometrica*, 77:1901–1948.
- Wang, Q. and Phillips, P. C. B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, 40(2):727–758.
- Wooldridge, J. M. (1994). *Handbook of Econometrics*, volume IV, chapter Estimation and Inference for Dependent Processes, pages 2639–2738. Amsterdam: North Holland.
- Xia, Y., Tong, H., and Li, W. K. (1999). On extended partially linear single-index models. *Boimetrica*, 86:831–842.
- Xia, Y., Tong, H., Li, W. K., and Zhu, L.-X. (2002). An adaptive estimation of dimension reduction. *Journal of the Royal Statistical Society B*, 64:363–410.
- Yu, Y. and Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *Journal of the American Statistical Association*, 97:1042–1054.
- Zhu, L. and Xue, L. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. R. Stat. Soc. Ser B Stat. Methodol.*, 68:549–570.