The Annals of Statistics
2016, Vol. 44, No. 1, 425–453
DOI: 10.1214/15-AOS1372

ESTIMATION FOR SINGLE-INDEX AND PARTIALLY LINEAR SINGLE-INDEX INTEGRATED MODELS

BY CHAOHUA DONG[*,†,1], JITI GAO[†,1] AND DAG TJØSTHEIM[‡]

Southwestern University of Finance and Economics, Monash University† and University of Bergen‡*

Estimation mainly for two classes of popular models, single-index 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 functions in the models and a profile approach is used to derive the estimators. The findings include the dual rate of convergence of the estimators for the single-index models and a trio of convergence rates for the partially linear single-index models. A new central limit theorem is established for a plug-in estimator of the unknown link function. Meanwhile, a considerable extension to a class of partially nonlinear single-index models is discussed in Section 4. Monte Carlo simulation verifies these theoretical results. An empirical study furnishes an application of the proposed estimation procedures in practice.

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, Myklebust and Tjøstheim (2007), Gao et al. (2009a, 2009b), Phillips (2009), Wang and Phillips (2009a, 2009b, 2012), Gao (2014), Gao and Phillips (2013a, 2013b) and Phillips, Li and Gao (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; see, for example, Dong and Gao (2013, 2014).

We first consider a partially linear single-index model of the form

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

where y_t is a scalar process, $g(\cdot)$, the so-called link function, is an unknown nonlinear integrable function from $\mathbb{R}$ to $\mathbb{R}$, β_0 and θ_0 are the true but unknown d-

Received January 2015; revised August 2015.

[1]Supported in part by two Australian Research Council Discovery Grants: #DP1096374 and #DP130104229.

MSC2010 subject classifications. 62G05, 62G08, 62G20.

Key words and phrases. Integrated time series, orthogonal series expansion, single-index models, partially linear single-index models, dual convergence rates, a trio of convergence rates.

dimensional column vectors of parameters, the superscript $^\top$ signifies the transpose of a vector (or matrix, hereafter), x_t is a d-dimensional integrated 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 these into account, models such as (1.1) are often used as a reasonable compromise between fully parametric and fully nonparametric modelling. See, for example, Carroll et al. (1997), Xia, Tong and Li (1999), Xia et al. (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. 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, 2013b) have pointed out that, for multivariate $I(1)$ processes, the conventional kernel estimation method may not be workable because the limit theory may break down. This gives rise to a challenge of seeking alternative estimation methods.

When $\beta_0 = 0$, model (1.1) becomes a single-index model

$$y_t = g(\theta_0^\top x_t) + e_t, \qquad t = 1, \dots, n, \tag{1.2}$$

which has been studied extensively for the case where x_t is stationary [see, e.g., Härdle, Hall and Ichimura (1993), Xia and Li (1999) and Wu, Yu and Yu (2010)].

We shall first consider model (1.2), but this is mainly a preliminary stage on our way to the general model (1.1). Standing on its own, model (1.2) has limited applicability since it is integrable, and among other things, does not include the linear model. Coupled with the assumption that $\{x_t\}$ is a unit root process, this implies that only an order of $O(\sqrt{n})$ observations can be used in the estimations of θ_0 and g in (1.2). The function g is only capable of describing finite domain behaviour in x_t. As $\theta_0^\top x_t$ increases, $g(\theta_0^\top x_t)$ goes to zero. All of this will be made precise in the following.

When the g function is added as a component in model (1.1), one obtains a model whose behaviour is governed by the linear component with g superimposed, whereas as x_t becomes large it reduces to the linear component. The resulting model (1.1) can be likened to a smooth transition regression model [STR model, Teräsvirta, Tjøstheim and Granger (2010)], where as x_t increases the model changes smoothly to the linear model. Our model in this sense extends the smooth transition model to a situation where there is an index involved and with a nonstationary input process. We believe that this is a situation which is of interest both theoretically and practically, as witnessed for example by our empirical study where quite different index behaviour is obtained in the close domain as compared to the far out region.

It is indeed possible to generalise our model to include a nonlinear behaviour also far out, which leads to the next stage of modelling. As a linear function is a particular H-regular function while an integrable function belongs to I-regular functions, studied by Park and Phillips (1999, 2001), Wang and Phillips (2009a), in a third stage we extend model (1.1) using a known H-regular function to substitute the linear function, in order to make the model more flexible and applicable.

Following the existing identifiability condition, for example, Lin and Kulasekera (2007), we assume for models studied later that $\|\theta_0\| = 1$ and the first nonzero component of θ_0 is positive. Notice that there is no extra condition needed for β_0 to make (1.1) identifiable, as discussed in Section 2.2 below. To facilitate the theoretical development in the following sections, we assume that θ_0 is an interior point located within a compact and convex parameter space Θ, which is also a usual assumption in a parameter estimation context. To focus on the unit root case, we also assume throughout that cointegration will not happen for θ around θ_0. In other words, there exists a neighbourhood of θ_0, $\mathcal{N}(\theta_0, \delta) \subset \Theta$, such that for any $\theta \in \mathcal{N}(\theta_0, \delta)$, $\theta^\top x_t$ is always an $I(1)$ process.

The findings of this paper are summarised as follows. The rate of convergence of the estimators of θ_0 in both models (1.1) and (1.2) is a composite of two different rates in a new coordinate system where θ_0 is on one axis. $\widehat{\theta}_n$ has a rate of $n^{-1/4}$ on the θ_0-axis, and another rate as fast as $n^{-3/4}$ on all axes orthogonal to θ_0. Overall, $\widehat{\theta}_n$ possesses convergence rate $n^{-1/4}$. This is expected and comes from the integrability of $g(\cdot)$, which in turn reduces the number of effective observations to $\sqrt{n}$. Moreover, the rate of convergence of $\widehat{\beta}_n$ to β_0 is n^{-1}, consistent with that of a linear model with a unit root input. The normalisation of $\widehat{\theta}_n$, $\|\widehat{\theta}_n\|^{-1}\widehat{\theta}_n$, converges to θ_0 with a rate faster than $\widehat{\theta}_n$ in both models. A new central limit theorem for a plug-in estimator of the form $\widehat{g}_n(u)$ converging to $g(u)$, where $u \in \mathbb{R}$, is comparable with the conventional kernel estimator in the literature. These phenomena are verified with finite sample experiments below.

Theoretical results heavily depend on the level of nonstationarity of the integrated time series and the integrability of the link functions. These properties result in a slow rate of convergence for the link function involving an $I(1)$ process and fast rate of convergence for a linear model with an $I(1)$ process. These are very different from the literature where the regressors are stationary. In addition, Monte Carlo simulations generally need relative larger sample sizes than those for the cases where the regressors are stationary if the regression function is integrable, since random walk on one hand diverges at rate $\sqrt{n}$, and on the other hand it possesses recurrent property making it possible to return to the effective domain of the integrable function g.

Two papers related to this study are Chang and Park (2003) and Guerre and Moon (2006) in terms of regressor. However, Chang and Park (2003) stipulate 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 in the discussion section that their method developed for binary choice models may be applicable for the estimation of single-index models where the link function $g(x) \to \infty$ as $|x| \to \infty$. Clearly, they are quite different from the setting of this study.

The organisation of the rest of the paper is as follows. Section 2 gives estimation procedures and assumptions for models (1.1) and (1.2). Asymptotic theory is established in Section 3 for the estimator $\widehat{\theta}_n$ in model (1.2) and the estimator $(\widehat{\beta}_n, \widehat{\theta}_n)$ in model (1.1). A central limit theorem for a plug-in estimator of the form $\widehat{g}_n(u)$ is given in Section 3. An extension of model (1.1) is discussed in Section 4 and Monte Carlo simulation experiments are conducted in Section 5. Section 6 shows the implementation of the proposed estimation schedules with an empirical dataset. Appendix A presents some technical lemmas. The proof of the main results in Section 3 is given in Appendix B. A supplemental document [Dong, Gao and Tjøstheim (2015)] contains Appendices C, D and E where all the proofs of the key lemmas listed in Appendix A as well as some other lemmas are shown in Appendix C, the complete proof of the results in Section 3 is placed in Appendix D and the results in Section 4 are proven in Appendix E.

Throughout the paper, the following notation is used. $\|\cdot\|$ is Euclidean norm for vectors and element-wise norm for matrices, that is, if $A = (a_{ij})_{nm}$, $\|A\| = (\sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2)^{1/2}$; I_d is the d-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 $\dddot{f}(x)$ 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 should be understood as element-wise. Furthermore, $\phi(\cdot)$ stands for the density function of a multivariate standard normal variable; $\int f(w)\,dw$ means a multiple integral when w is a vector. Convergence in probability and convergence in distribution are signified as $\to_P$ and $\to_D$, 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 the Hermite function sequence $\{\mathscr{H}_i(x)\}$ is an orthonormal basis in $L^2(\mathbb{R})$ where by definition

$$\mathscr{H}_i(x) = \left(\sqrt{\pi}2^i i!\right)^{-1/2} H_i(x) \exp\left(-\frac{x^2}{2}\right), \qquad i \ge 0, \tag{2.1}$$

and $H_i(x)$ are Hermite polynomials orthogonal with density $\exp(-x^2)$. The orthogonality reads $\int \mathscr{H}_i(x)\mathscr{H}_j(x)\,dx = \delta_{ij}$, the Kronecker delta.

Thus, a continuous function $g(\cdot) \in L^2(\mathbb{R})$ may be expanded into an orthogonal series

$$g(x) = \sum_{i=0}^{\infty} c_i \mathscr{H}_i(x) \quad \text{and} \quad c_i = \int g(x)\mathscr{H}_i(x)\,dx. \tag{2.2}$$

Throughout, let k be a positive integer and define $g_k(x) = \sum_{i=0}^{k-1} c_i \mathscr{H}_i(x)$ as the truncation series and $\gamma_k(x) = g(x) - g_k(x) = \sum_{i=k}^{\infty} c_i \mathscr{H}_i(x)$ as the residue after truncation.

2.1. *Estimation procedure for single-index models.* By virtue of (2.2), we write model (1.2) for $t = 1, \dots, n$ as

$$y_t = Z_k^\top(\theta_0^\top x_t)c + \gamma_k(\theta_0^\top x_t) + e_t,$$

where $Z_k^\top(\cdot) = (\mathscr{H}_0(\cdot), \dots, \mathscr{H}_{k-1}(\cdot))$, $c^\top = (c_0, \dots, c_{k-1})$ and k is the truncation parameter determined later.

Let $Y = (y_1, \dots, y_n)^\top$, $Z = (Z_k(\theta_0^\top x_1), \dots, Z_k(\theta_0^\top x_n))^\top$ an $n \times k$ matrix, $\gamma = (\gamma_k(\theta_0^\top x_1), \dots, \gamma_k(\theta_0^\top x_n))^\top$ and $e = (e_1, \dots, e_n)^\top$. We have a matrix form equation $Y = Zc + \gamma + e$, and hence by the Ordinary Least Squares (OLS) method, $\widetilde{c} = \widetilde{c}(\theta_0) = (Z^\top Z)^{-1} Z^\top Y$ is an estimate for c in terms of θ_0. Nonetheless, since θ_0 is unknown, we only have a form of $\widetilde{c}$. To estimate θ_0, define for $\theta \in \Theta$, $L_n(\theta) = \frac{1}{2}\sum_{t=1}^{n}[y_t - Z_k^\top(\theta^\top x_t)\widetilde{c}(\theta)]^2$. Then we choose an optimum $\widehat{\theta}_n$ such that

$$\widehat{\theta}_n = \operatorname*{argmin}_{\theta \in \Theta} L_n(\theta), \tag{2.3}$$

as an estimator for θ_0. Once $\widehat{\theta}_n$ is available, we have a plug-in estimator $\widehat{g}_n(u) \equiv \widehat{g}_n(u; \widehat{\theta}_n) = Z_k(u)^\top \widehat{c}$ for any $u \in \mathbb{R}$ where $\widehat{c} = \widetilde{c}(\widehat{\theta}_n)$, which is purely based on the sample, and hence applicable. The estimation procedure proposed here is the profile method [see, Severini and Wong (1992), Liang et al. (2010)].

Additionally, to be in concert with the identification condition $\|\theta_0\| = 1$, we define the normalisation of $\widehat{\theta}_n$, $\widehat{\theta}_{n,\mathrm{emp}} = \|\widehat{\theta}_n\|^{-1}\widehat{\theta}_n$. An asymptotic theory for both $\widehat{\theta}_{n,\mathrm{emp}}$ and $\widehat{\theta}_n$ will be studied in Section 3 below.

2.2. *Estimation procedure for partially linear single-index models.* Usually, researchers, such as Xia, Tong and Li (1999), impose an identification condition that β_0 is perpendicular to θ_0 on the partially linear single-index models. This is because when β_0 is not perpendicular to θ_0, a new vector $\beta_0 - (\beta_0^\top \theta_0)\theta_0$ can be used in the place of β_0 and the g function will be replaced by $g(u) + (\beta_0^\top \theta_0)u$. However, in model (1.1) the lack of orthogonality between β_0 and θ_0 does not affect the identifiability of the model at all. See the verification at the end of Appendix C in the supplementary material [Dong, Gao and Tjøstheim (2015)].

Our estimation procedure in partially linear single-index models is proposed as follows. By virtue of (2.2) again, for each t rewrite (1.1) as

$$y_t - \beta_0^\top x_t = Z_k(\theta_0^\top x_t)^\top c + \gamma_k(\theta_0^\top x_t) + e_t,$$

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, γ, e remain the same as in the last subsection. We have matrix form equation: $Y - X\beta = Zc + \gamma + e$. Then the OLS gives that $\widetilde{c} = \widetilde{c}(\beta_0, \theta_0) = (Z^\top Z)^{-1} Z^\top (Y - X\beta_0)$. Due to the same reason as before, define for generic (β, θ), $L_n(\beta, \theta) = \frac{1}{2}\sum_{t=1}^n [y_t - \beta^\top x_t - Z_k^\top(\theta^\top x_t)\widetilde{c}(\beta, \theta)]^2$. The estimator of (β_0, θ_0) is given by

$$\begin{pmatrix} \widehat{\beta}_n \\ \widehat{\theta}_n \end{pmatrix} = \operatorname*{argmin}_{\beta \in \mathbb{R}^d, \theta \in \Theta} L_n(\beta, \theta). \tag{2.4}$$

Similarly, a plug-in estimator is obtained, $\widehat{g}_n(u) \equiv \widehat{g}_n(u; \widehat{\beta}_n, \widehat{\theta}_n) = Z_k^\top(u)\widehat{c}$ where $\widehat{c} = \widetilde{c}(\widehat{\beta}_n, \widehat{\theta}_n)$. Once the estimators of the parameters are available, and the normalisation $\widehat{\theta}_{n,\mathrm{emp}} = \|\widehat{\theta}_n\|^{-1}\widehat{\theta}_n$ is defined to satisfy the identification condition.

2.3. *Assumptions*. Before we establish our main theory in Section 3 below, we introduce some necessary conditions.

ASSUMTPION A. (a) Let $\{\varepsilon_j, -\infty < j < \infty\}$ be a sequence of d-dimensional independent and identically distributed (i.i.d.) continuous random variables with $E\varepsilon_1 = 0$, $E[\varepsilon_1\varepsilon_1^\top] = \Omega > 0$ and $E\|\varepsilon_1\|^p < \infty$ for some $p > 2$. The characteristic function of ε_1 is integrable, that is, $\int |E\exp(iu\varepsilon_1)|\,du < \infty$.

(b) Let $x_t = x_{t-1} + v_t$ for $t \ge 1$ and $x_0 = O_P(1)$, where $\{v_t\}$ is a linear process defined by $v_t = \sum_{j=0}^\infty \rho_j \varepsilon_{t-j}$, in which $\{\rho_j\}$ is a square matrix such that $\rho_0 = I_d$, $\sum_{j=0}^\infty \|\rho_j\| < \infty$ and $\rho = \sum_{j=0}^\infty \rho_j$ is of full rank.

(c) There is a σ-field $\mathcal{F}_t$ such that $(e_t, \mathcal{F}_t)$ is a martingale difference sequence, that is, for all t, $E(e_t|\mathcal{F}_{t-1}) = 0$ almost surely (a.s.). Also, $E(e_t^2|\mathcal{F}_{t-1}) = \sigma_e^2$ a.s. and $\mu_4 := \sup_{1\le t\le n} E(e_t^4|\mathcal{F}_{t-1}) < \infty$ 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 $(U_n(r), V_n(r)) \to_D (U(r), V(r))$ as $n \to \infty$. Here, $(U(r), V(r))$ 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, 2009b, 2012), Gao et al. (2009a, 2009b), Gao, Tjøstheim and Yin (2012), among others. The σ-field $\mathcal{F}_t$ may be taken as $\mathcal{F}_t = \sigma(\dots, \varepsilon_t, \varepsilon_{t+1}; e_1, \dots, e_t)$.

By Skorohod representation theorem [Pollard (1984), page 71] 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 $(U_n^0(r), V_n^0(r)) \to_{a.s.} (U(r), V(r))$ uniformly on $[0,1]^{d+1}$. To avoid the repetitious embedding procedure 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 $(U_n(r), V_n(r)) = (U_n^0(r), V_n^0(r))$ instead of $(U_n(r), V_n(r)) =_D (U_n^0(r), V_n^0(r))$. Since Lemma A.2 below is derived in this richer probability space, all proofs in the paper should be understood in the richer space as well. We will not repeat this again.

ASSUMTPION B. (a) $g(x)$ is differentiable on $\mathbb{R}$ and $g^{(m-\ell)}(x)x^\ell \in L^2(\mathbb{R})$ for $\ell = 0, 1, \dots, m$ with some given integer m.

(b) $k = [a \cdot n^\kappa]$ 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. Condition (a) ensures the negligibility of the truncation residuals (see the derivation at the beginning of Lemma C.1 of Appendix C of the supplementary document). Regarding condition (b), although it is stringent for κ, we may choose, for example, $\kappa \in [\frac{5}{44}, \frac{5}{41}]$ and $m = 8$ in practice. Large m and small κ are chosen such that the orthogonal series expansion for the link function converges so fast that all residues after truncation do not affect the limit theory, as can be seen in the proof of Theorem 3.1 in the supplementary material Dong, Gao and Tjøstheim (2015).

3. Asymptotic theory.

3.1. *Asymptotic theory for single-index models.* To derive an asymptotic theory for $\widehat{\theta}_n$ given by (2.3), we shall use basic ideas from Wooldridge (1994). Let $S_n(\theta) = \frac{\partial}{\partial\theta} L_n(\theta)$ and $J_n(\theta) = \frac{\partial^2}{\partial\theta\,\partial\theta^\top} L_n(\theta)$ be the score and Hessian, respectively. As usual, we have the expansion

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

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

To facilitate the development of the asymptotic theory, we consider coordinate rotation in $\mathbb{R}^d$. Let $Q = (\theta_0, Q_2)$ be an orthogonal matrix. Note that such Q does exist since $\theta_0 \neq 0$. We shall use the orthogonal matrix Q to rotate all vectors in $\mathbb{R}^d$. In particular,

$$\begin{aligned} &\alpha_0 := Q^\top\theta_0 = (\alpha_{10}, \alpha_{20}^\top)^\top \qquad \text{where } \alpha_{10} = \|\theta_0\|^2 = 1, \alpha_{20} = Q_2^\top\theta_0 = 0, \\ &z_t := Q^\top x_t = (x_{1t}, x_{2t}^\top)^\top \qquad \text{where } x_{1t} := \theta_0^\top x_t, x_{2t} := Q_2^\top x_t, \\ &\alpha := Q^\top\theta \qquad \text{for any generic } \theta. \end{aligned} \tag{3.2}$$

Accordingly, we can rewrite the single-index model as $y_t = g(\theta_0^\top Q Q^\top x_t) + e_t = g(\alpha_0^\top z_t) + e_t$. In addition, by Assumption A and the continuous mapping theorem, we have for $r \in [0, 1]$,

$$\frac{1}{\sqrt{n}} x_{1[nr]} \to_D V_1(r) = \theta_0^\top V(r) \quad \text{and} \quad \frac{1}{\sqrt{n}} x_{2[nr]} \to_D V_2(r) = Q_2^\top V(r). \tag{3.3}$$

It is noteworthy that the rotation is not necessary in practice, as shown in the simulation section, and it is also logically impossible since θ_0 is unknown. The rotation is only used as a tool to derive an asymptotic theory for the proposed estimator.

If $\widehat{\alpha}_n$ is the nonlinear least squares estimator of α_0, then $\widehat{\alpha}_n = Q^\top \widehat{\theta}_n$. Moreover, the score function $S_n(\alpha)$ and the Hessian $J_n(\alpha)$ for the parameter α can be obtained from those for θ. More precisely, $S_n(\alpha) = Q^\top S_n(\theta)$ and $J_n(\alpha) = Q^\top J_n(\theta) Q$. Premultiplying equation (3.1) by $Q^\top$, we have

$$0 = S_n(\widehat{\alpha}_n) = S_n(\alpha_0) + J_n(\alpha_n)(\widehat{\alpha}_n - \alpha_0). \tag{3.4}$$

The following theorem gives asymptotic distributions for the score $S_n(\alpha_0)$ and the Hessian $J_n(\alpha_0)$ as well as $\widehat{\alpha}_n - \alpha_0$.

THEOREM 3.1. *Denote* $D_n = \mathrm{diag}(n^{1/4}, n^{3/4} I_{d-1})$. *Under Assumptions* A *and* B, *as* $n \to \infty$

$$D_n^{-1} S_n(\alpha_0) \to_D R^{1/2} W(1) \quad \textit{and} \quad D_n^{-1} J_n(\alpha_0) D_n^{-1} \to_P R, \tag{3.5}$$

where $W(1)$ *is a* d*-dimensional vector of standard normal random variables 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*

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

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

in which V_1 *and* V_2 *given by* (3.3) *are Brownian motions of dimension* 1 *and* $d-1$, *respectively*, $L_1(r,0)$ *denotes the local time process of Brownian motion* $V_1(\cdot)$, *standing for the sojourning time of* V_1 *at zero over* $[0,r]$.

As a result, *under the same conditions*, $\widehat{\alpha}_n$ *is consistent and as* $n \to \infty$

$$D_n(\widehat{\alpha}_n - \alpha_0) \to_D R^{-1/2} W(1). \tag{3.6}$$

A standard book introducing the local time process of Brownian motion is Revuz and Yor (2005). In view of the structure of D_n, we have two limits from (3.6),

$$n^{1/4}(\widehat{\alpha}_{1n} - 1) \to_D \mathbf{MN}(0, \rho_{11}) \quad \text{and} \quad n^{3/4} \widehat{\alpha}_{2n} \to_D \mathbf{MN}(0, \rho_{22}), \tag{3.7}$$

where $\widehat{\alpha}_n = (\widehat{\alpha}_{1n}, \widehat{\alpha}_{2n}^\top)^\top$, $\mathbf{MN}(0, \Xi)$ stands for mixture normal distribution for the case where the covariance matrix Ξ is stochastic, ρ_{11} and ρ_{22} are diagonal blocks on the matrix $R^{-1} = \left(\begin{smallmatrix}\rho_{11} & \rho_{12}\\ \rho_{21} & \rho_{22}\end{smallmatrix}\right)$,

$$\rho_{11} = \left(r_{11} - r_{12} r_{22}^{-1} r_{21}\right)^{-1} \quad \text{and} \quad \rho_{22} = \left(r_{22} - r_{21} r_{11}^{-1} r_{12}\right)^{-1}. \tag{3.8}$$

Hence, $\widehat{\alpha}_n$ has two different convergence rates for its components.

Note by (3.7) that in the coordinate system Q where θ_0 is an axis, the estimator $\widehat{\theta}_n$ has dual convergence rates: the rate of convergence for the coordinate on θ_0 (i.e., $\widehat{\alpha}_{1n}$) is $n^{-1/4}$, while on all directions orthogonal to θ_0 the rate of convergence for the coordinates (i.e., $\widehat{\alpha}_{2n}$) is as fast as $n^{-3/4}$. This difference in convergence rate can be explained in the following way. Due to the unit root behaviour of $\{x_t\}$ its probability mass is spreading out in a Lebesgue type fashion. Since g is integrable, $g(\theta_0^\top x) \approx 0$ outside the effective range of g. This means that only moderate values of $\{x_t\}$ can contribute to g along θ_0, but in such directions that are orthogonal to θ_0, there is no restriction on $\{x_t\}$, so that even for far out values of $\{x_t\}$, they can contribute, and hence increase the effective sample size. Certainly, no such effect can take place in univariate models.

As defined in Section 2, $\widehat{\theta}_{n,\mathrm{emp}} = \|\widehat{\theta}_n\|^{-1}\widehat{\theta}_n$. Intuitively, $\widehat{\theta}_{n,\mathrm{emp}}$ might have a faster rate of convergence than that of $\widehat{\theta}_n$. This can be seen using the α-representation of the rotated system. Because of $\widehat{\theta}_n = Q\widehat{\alpha}_n$ and hence $\|\widehat{\theta}_n\| = \|\widehat{\alpha}_n\|$, $\widehat{\theta}_{n,\mathrm{emp}} = Q\widehat{\alpha}_{n,\mathrm{unit}}$, where $\widehat{\alpha}_{n,\mathrm{unit}} = (\widehat{\alpha}^1_{n,\mathrm{unit}}, (\widehat{\alpha}^2_{n,\mathrm{unit}})^\top)^\top = \|\widehat{\alpha}_n\|^{-1}\widehat{\alpha}_n$. The following results give the rates of convergence for $\widehat{\alpha}_{n,\mathrm{unit}}$ and then for $\widehat{\theta}_n$ and $\widehat{\theta}_{n,\mathrm{emp}}$, respectively.

COROLLARY 3.1. *Under Assumptions* A *and* B, *we have as* $n \to \infty$,

$$n^{3/2}\left(\widehat{\alpha}^1_{n,\mathrm{unit}} - 1\right) \to_D -\tfrac{1}{2}\|\xi\|^2 \quad \text{and} \quad n^{3/4}\widehat{\alpha}^2_{n,\mathrm{unit}} \to_D \xi,$$

where $\xi \sim \mathbf{MN}(0, \rho_{22})$ *is the limit given by* (3.7).

Note that, after the normalisation, the slow rate becomes as fast as $n^{-3/2}$ whereas the fast rate remains the same. Note also that $\widehat{\alpha}^1_{n,\mathrm{unit}} \to_P 1$ but $\widehat{\alpha}^1_{n,\mathrm{unit}} = \|\widehat{\alpha}_n\|^{-1}\widehat{\alpha}_{1n} \le 1$. The intuitive reason for the fast rate of $\widehat{\alpha}^1_{n,\mathrm{unit}}$ is that it takes advantage of the direction orthogonal to θ_0, where there is larger supply of information from far out x_t's as explained above. The rates are also verified with Monte Carlo simulation. The resulting rates for $\widehat{\theta}_n$ are given in Theorem 3.2 below.

THEOREM 3.2. *Under Assumptions* A *and* B, *we have as* $n \to \infty$,

$$n^{1/4}(\widehat{\theta}_n - \theta_0) \to_D \mathbf{MN}\left(0, \rho_{11}\theta_0\theta_0^\top\right), \tag{3.9}$$

$$n^{3/4}(\widehat{\theta}_{n,\mathrm{emp}} - \theta_0) \to_D \mathbf{MN}\left(0, Q_2\rho_{22}Q_2^\top\right). \tag{3.10}$$

REMARK 3.1. As can be seen, $\widehat{\theta}_n \to_P \theta_0$ at rate of $n^{-1/4}$ and $\widehat{\theta}_{n,\mathrm{emp}} \to_P \theta_0$ at rate of $n^{-3/4}$. Again roughly speaking, the normalisation scales $\widehat{\theta}_n$ to the unit ball, and hence accelerates the slow rate of $\widehat{\theta}_n$. Due to the fast convergence of $\widehat{\theta}_{n,\mathrm{emp}}$, all the following assertions regarding $\widehat{\theta}_n$ remain true if $\widehat{\theta}_n$ is replaced by $\widehat{\theta}_{n,\mathrm{emp}}$. A geometric illustration is given in Appendix C of the supplementary material [Dong, Gao and Tjøstheim (2015)] to explain the slow and fast rates. We do not wish to repeat this again.

Furthermore, by Theorem 3.2 we have $\widehat{\theta}_n \sim \mathbf{MN}(\theta_0, n^{-1/4}\rho_{11}\theta_0\theta_0^\top)$. We next show that the estimator of the covariance matrix of $\widehat{\theta}_n$ is the inverse of the Hessian matrix of the form $[J_n(\widehat{\theta}_n)]^{-1}$ or even $[\widetilde{J}_n(\widehat{\theta}_n)]^{-1}$, where $\widetilde{J}_n(\theta) = \sum_{t=1}^n \hat{\dot{g}}_n^2(\theta^\top x_t)x_t x_t^\top$ is the leading term of $J_n(\theta)$. Meanwhile, define the estimators for σ_e and $L_1(1,0)$ by

$$\text{(3.11)} \qquad \widehat{\sigma}_e^2 = \frac{1}{n}\sum_{t=1}^n \bigl[y_t - \widehat{g}_n\bigl(\widehat{\theta}_n^\top x_t\bigr)\bigr]^2 \quad \text{and} \quad \widehat{L}_{n1}(1,0) = \frac{1}{\sqrt{n}}\sum_{t=1}^n \mathscr{H}_0^2\bigl(\widehat{\theta}_n^\top x_t\bigr),$$

respectively, where $\mathscr{H}_0(\cdot)$ is the first function in the Hermite sequence.

COROLLARY 3.2. *Under Assumptions A and B, we have as $n \to \infty$,*

$$\text{(3.12)} \qquad \widehat{\sigma}_e^2 \to_P \sigma_e^2 \quad \textit{and} \quad \widehat{L}_{n1}(1,0) - L_1(1,0) \to_P 0$$

and

$$\text{(3.13)} \qquad \sqrt{n}\bigl[J_n(\widehat{\theta}_n)\bigr]^{-1} \to_P \rho_{11}\theta_0\theta_0^\top \quad \textit{and} \quad \sqrt{n}\bigl[\widetilde{J}_n(\widehat{\theta}_n)\bigr]^{-1} \to_P \rho_{11}\theta_0\theta_0^\top.$$

We then establish the following central limit theory for the plug-in estimator $\widehat{g}_n(u) = Z_k^\top(u)\widehat{c}$ defined in Section 2.1, where $u \in \mathbb{R}$.

THEOREM 3.3. *Under Assumptions A and B, as $n \to \infty$, $\sup_{u\in\mathbb{R}} |\widehat{g}_n(u) - g(u)| \to_P 0$, and*

$$\text{(3.14)} \qquad \widehat{\sigma}_e^{-1}\widehat{L}_{n1}^{1/2}(1,0)n^{1/4}\|Z_k(u)\|^{-1}\bigl(\widehat{g}_n(u) - g(u)\bigr) \to_D N(0,1).$$

REMARK 3.2. The order involved in the normality is $O_P(1)n^{1/4}k^{-1/2}$ in view of $\|Z_k(u)\|^2 = O(1)k$. This is comparable with the kernel estimate in the literature. Theorem 3.1 of Wang and Phillips [(2009a), page 721] shows that, for univariate regression $y_t = f(x_t) + u_t$, the normaliser of $\hat{f}(x) - f(x)$ is $(h\sum_{t=1}^n K_h(x_t - x))^{1/2}$ where h is a bandwidth, $K_h(\cdot) = K(\cdot/h)/h$ is a kernel function and $\hat{f}(x)$ is the kernel estimate of $f(x)$. Note that $(h\sum_{t=1}^n K_h(x_t - x))^{1/2} = O_P(1)n^{1/4}h^{1/2}$. Thinking of k^{-1} as equivalent to the bandwidth h, the normalisers in the two situations are quite comparable.

REMARK 3.3. Noting that $\widehat{g}_n(u) - g(u) = Z_k^\top(u)(\widehat{c} - c) - \gamma_k(u)$ and by the orthogonality of the basis functions, $\int(\widehat{g}_n(u) - g(u))^2\,dx = \|\widehat{c} - c\|^2 + \|\gamma_k(u)\|^2$ where $\|\gamma_k(u)\|$ is the norm of $\gamma(u)$ in the function space. Using Lemma A.3 below, $\|\widehat{c} - c\|^2 = O_P(kn^{-1/2})$ and Lemma C.1 (in the supplementary material), $\|\gamma_k(u)\|^2 = O(k^{-m})$, an optimal truncation parameter k^* may be found to be proportional to $k^* = [n^{1/2(m+1)}]$ when $\|\widehat{c} - c\|^2$ and $\|\gamma_k(u)\|^2$ have the same order going to zero. Here, m is the smoothness order of $g(u)$.

3.2. *Asymptotic theory for partially linear single-index models.* Denote $\vartheta_0 = (\beta_0^\top, \theta_0^\top)^\top$ and $\vartheta = (\beta^\top, \theta^\top)^\top$ as a generic parameter for simplicity. Let $\mathfrak{S}_n(\vartheta)$ and $\mathfrak{J}_n(\vartheta)$ be the respective score and Hessian functions of $L_n(\vartheta)$ in the minimisation problem (2.4). Let $\widehat{\vartheta}_n$ be the estimator of ϑ_0. We then have the expansion:

$$0 = \mathfrak{S}_n(\widehat{\vartheta}_n) = \mathfrak{S}_n(\vartheta_0) + \mathfrak{J}_n(\vartheta_n)(\widehat{\vartheta}_n - \vartheta_0), \tag{3.15}$$

where $\mathfrak{J}_n(\vartheta_n)$ is the Hessian matrix with the rows evaluated at a point ϑ_n between $\widehat{\vartheta}_n$ and ϑ_0.

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

$$y_t = \beta_0^\top QQ^\top x_t + g(\theta_0^\top QQ^\top x_t) + e_t = \lambda_0^\top z_t + g(\alpha_0^\top z_t) + e_t, \tag{3.16}$$

where $\lambda_0 = Q^\top\beta_0 = (\lambda_{10}, \lambda_{20}^\top)^\top$ with $\lambda_{10} = \theta_0^\top\beta_0$ a scalar, $\lambda_{20} = Q_2^\top\beta_0$ a $(d-1)$-dimensional vector, $\alpha_0 = Q^\top\theta_0$, $z_t = Q^\top x_t$ are defined the same as before. Let $\lambda = Q^\top\beta$ and $\alpha = Q^\top\theta$ for the generic vector rotation. Also, group them by $\mu_0 = (\lambda_0^\top, \alpha_0^\top)^\top$ and $\mu = (\lambda^\top, \alpha^\top)^\top$.

Let $L_n(\mu)$ be the counterpart of $L_n(\beta, \theta)$ after reparameterisation. If $\widehat{\mu}_n$, the minimiser of $L_n(\mu)$, is the estimator of μ_0, then $\widehat{\mu}_n = \operatorname{diag}(Q^\top, Q^\top)\widehat{\vartheta}_n$. Moreover, the score function $\mathfrak{S}_n(\mu)$ and the Hessian $\mathfrak{J}_n(\mu)$ for the parameter μ can be obtained from those for ϑ. Namely, $\mathfrak{S}_n(\mu) = \operatorname{diag}(Q^\top, Q^\top)\mathfrak{S}_n(\vartheta)$ and $\mathfrak{J}_n(\mu) = \operatorname{diag}(Q^\top, Q^\top)\mathfrak{J}_n(\vartheta)\operatorname{diag}(Q, Q)$. Premultiplying equation (3.15) by $\operatorname{diag}(Q^\top, Q^\top)$, we have

$$0 = \mathfrak{S}_n(\widehat{\mu}_n) = \mathfrak{S}_n(\mu_0) + \mathfrak{J}_n(\mu_n)(\widehat{\mu}_n - \mu_0), \tag{3.17}$$

from which the following theorem is derived.

THEOREM 3.4. *Under Assumptions A and B, $\widehat{\mu}_n \to_P \mu_0$. Moreover, as $n \to \infty$*

$$n(\widehat{\lambda}_n - \lambda_0) \to_D Q^\top\left(\int_0^1 V(r)V^\top(r)\,dr\right)^{-1}\int_0^1 V(r)\,dU(r), \tag{3.18}$$

$$D_n(\widehat{\alpha}_n - \alpha_0) \to_D R^{-1/2}W(1), \tag{3.19}$$

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

Theorem 3.4 shows that for the partially linear single-index model, the estimators of the parameters in the linear part have the same rates of convergence as those in the linear model, while the estimator of the index vector retains the dual rates in the system of Q. Hence, there is a trio of rates of convergence accommodated in the partially linear single-index model case. From Theorem 3.4, we may derive asymptotic distributions for both $\widehat{\beta}_n$ and $\widehat{\theta}_n$.

THEOREM 3.5. *Under Assumptions A and B, for $(\widehat{\beta}_n, \widehat{\theta}_n)$ given by (2.4) we have, as $n \to \infty$,*

$$n(\widehat{\beta}_n - \beta_0) \to_D \left(\int_0^1 V(r)V^\top(r)\,dr\right)^{-1} \int_0^1 V(r)\,dU(r), \tag{3.20}$$

$$n^{1/4}(\widehat{\theta}_n - \theta_0) \to_D \mathbf{MN}(0, \rho_{11}\theta_0\theta_0^\top). \tag{3.21}$$

Furthermore, using (3.19), for $\widehat{\theta}_n$, the results of Theorems 3.2–3.3 and Corollaries 3.1–3.2 with $\widehat{\theta}_{n,\mathrm{emp}}$ and $\widehat{g}_n(u)$ defined in the same fashion remain true.

THEOREM 3.6. *Under Assumptions A and B, the results of Theorems 3.2–3.3 and Corollaries 3.1–3.2 also remain true for $\widehat{\theta}_n$ and $\widehat{g}_n(u)$ in model (1.1).*

The proof of the main results in this section is given in Appendix B below, except that Theorem 3.1 and Corollaries 3.1–3.2 are shown in Appendix D in the supplementary material [Dong, Gao and Tjøstheim (2015)].

4. Extension to the general *H*-regular class. For integrated time series, the rate of convergence of the unknown parameters involved in a regression function heavily depends on the functional form of the regression function under consideration. The literature focuses on two classes of functions, that is, the so-called I-regular class and H-regular class. Integrable functions belong to the I-regular class, while functions like power functions and polynomial functions are H-regular. For more detail, we refer to Park and Phillips (1999, 2001).

Observe that the partially linear single-index model is a combination of the two classes. As already stated, this can be seen as a smooth transition model whose behaviour in the finite domain is a linear model perturbed by the g-function component, the influence of which is reduced for a large $\|x_t\|$. Such models have broad applications. Nonetheless, if the linear part may be relaxed to a nonlinear form, the model will be more flexible and more applicable. To do so, we combine a general H-regular function with a general I-regular function to introduce a partially nonlinear single-index model of the form:

$$y_t = f(\beta_0^\top x_t) + g(\theta_0^\top x_t) + e_t, \qquad t = 1, \dots, n, \tag{4.1}$$

where $f(\cdot)$ is parametrically known and H-regular, $g(\cdot)$ is nonparametrically unknown and integrable, and β_0 and θ_0 are the unknown parameters.

In some circumstances, one may have some idea on the trend in y_t generated by an index variable $\beta^\top x_t$, for example, linear or quadratic. Thus, model (4.1) should give an accurate description for such a relation. Certainly, the partially linear single-index model is a special case of model (4.1) with $f(u) = u$. The objective of this section is then to estimate (β_0, θ_0) and $g(\cdot)$.

Before we propose our estimation method, we give a definition for the H-regular class.

DEFINITION 4.1. We say that the function $f(x)$ is asymptotically homogeneous, or H-regular, if for all $\eta > 0$

$$f(\eta x) = \upsilon(\eta) F(x) + \xi(\eta; x), \qquad \left|\xi(\eta; x)\right| \le a(\eta) P(x), \tag{4.2}$$

where $F(x)$ and $P(x)$ are both locally integrable, and $\limsup_{\eta\to\infty} \frac{a(\eta)}{\upsilon(\eta)} = 0$.

If f is H-regular with υ and F satisfying (4.2), we call υ and F the asymptotic order and the limit homogeneous function of f, respectively. Note that any polynomial and power function with positive power are H-regular. Note also that this definition is not the exact one in the reference above, since in this section $f(\cdot)$ is required to be differentiable.

Estimation procedure: The estimation procedure follows similarly from that for model (1.1). Using expansion (2.2), for each t rewrite (4.1) as $y_t - f(\beta_0^\top x_t) = Z_k(\theta_0^\top x_t)^\top c + \gamma_k(\theta_0^\top x_t) + e_t$, where $Z_k(\cdot)$, c and $\gamma_k(\cdot)$ are defined as before. Let $\tilde{Y} = (y_1 - f(\beta_0^\top x_1), \dots, y_n - f(\beta_0^\top x_n))^\top$, and Z, γ and e remain the same as before. We then have the matrix form equation, $\tilde{Y} = Zc + \gamma + e$. Then the OLS gives $\widetilde{c} = \widetilde{c}(\beta_0, \theta_0) = (Z^\top Z)^{-1} Z^\top \tilde{Y}$.

Define, $L_n(\beta, \theta) = \frac{1}{2}\sum_{t=1}^n [y_t - f(\beta^\top x_t) - Z_k^\top(\theta^\top x_t)\widetilde{c}(\beta, \theta)]^2$. The estimator of (β_0, θ_0) is given by

$$\begin{pmatrix} \widehat{\beta}_n \\ \widehat{\theta}_n \end{pmatrix} = \operatorname*{argmin}_{\theta\in\Theta,\beta} L_n(\beta, \theta). \tag{4.3}$$

Similarly, a plug-in estimator $\widehat{g}_n(u) \equiv Z_k(u)^\top \widehat{c}$ for any real $u \in \mathbb{R}$, where $\widehat{c} = \widetilde{c}(\widehat{\beta}_n, \widehat{\theta}_n)$ is obtained once $(\widehat{\beta}_n, \widehat{\theta}_n)$ is available.

Asymptotic theory: The same notation as in Section 3.2 is used for the minimisation problem (4.3). Also, in order to derive the corresponding asymptotic theory, we need to rotate vectors. Using the orthogonal matrix $Q = (\theta_0, Q_2)$ again,

$$y_t = f\left(\beta_0^\top Q Q^\top x_t\right) + g\left(\theta_0^\top Q Q^\top x_t\right) + e_t = f\left(\lambda_0^\top z_t\right) + g\left(\alpha_0^\top z_t\right) + e_t,$$

where the notation used is the same as in (3.16), $\lambda = Q^\top \beta$ and $\alpha = Q^\top \theta$ for generic vector rotation. We also define $\mu_0 = (\lambda_0^\top, \alpha_0^\top)^\top$ and $\mu = (\lambda^\top, \alpha^\top)^\top$.

It is still true that if $\widehat{\mu}_n$ is the estimator of μ_0 given by the minimiser of $L_n(\mu)$, then $\widehat{\mu}_n = \operatorname{diag}(Q^\top, Q^\top)\widehat{\vartheta}_n$. Moreover, $\mathfrak{S}_n(\mu) = \operatorname{diag}(Q^\top, Q^\top)\mathfrak{S}_n(\vartheta)$ and $\mathfrak{J}_n(\mu) = \operatorname{diag}(Q^\top, Q^\top)\mathfrak{J}_n(\vartheta)\operatorname{diag}(Q, Q)$.

THEOREM 4.1. *Suppose that* (i) f *is* H*-regular with asymptotic order* υ *and limit homogeneous function* F; (ii) $\dot{f}$ *and* $\ddot{f}$ *are* H*-regular with asymptotic order* $\dot{\upsilon}$ *and limit homogeneous function* $\dot{F}$, *and asymptotic order* $\ddot{\upsilon}$ *and limit homogeneous function* $\ddot{F}$, *respectively*; (iii) $|F(\beta_0^\top x) - F(\beta^\top x)|$ *is not a zero function on* $\|x\| < \delta$ *for some* $\delta > 0$ *and if* $\beta \neq \beta_0$; (iv) $\upsilon(\sqrt{n})^{-1}\sqrt{k}^3 \to 0$, *where* k *is the truncation parameter satisfying Assumption* B; (v) $|\dot{\upsilon}^{-2}(u)\ddot{\upsilon}(u)\upsilon(u)|$ *is bounded in* $u \geq M_0$ *for some* $M_0 > 0$.

Under Assumptions A *and* B, *we have* $\widehat{\mu}_n \to_P \mu_0$. *Moreover, as* $n \to \infty$

$$n\dot{\upsilon}(\sqrt{n})(\widehat{\lambda}_n - \lambda_0) \to_D Q^\top \left(\int_0^1 [\dot{F}(\beta_0^\top V(r))]^2 V(r)V^\top(r)\,dr\right)^{-1} \times \int_0^1 \dot{F}(\beta_0^\top V(r))V(r)\,dU(r), \tag{4.4}$$

$$D_n(\widehat{\alpha}_n - \alpha_0) \to_D R^{-1/2}W(1), \tag{4.5}$$

where $(U(r), V(r))$, D_n, R *and* W *are the same as in Theorem* 3.1.

REMARK 4.1. It is reasonable to require that the derivatives of f are H-regular if f is H-regular, as stated in conditions (i) and (ii). Condition (iii) is simply an identification condition, while (iv) and (v) are technical requirements that can be fulfilled easily by many usual H-regular functions. Similar conditions for parameter estimation in regression models involving $I(1)$ processes can be found in Park and Phillips (2001). Particularly, $f(x) = x$ is a special case such that conditions (i)–(v) are trivially satisfied.

Similar to Theorems 3.5 and 3.6, we derive some corresponding limit distributions for $\widehat{\beta}_n$ and $\widehat{\theta}_n$ as well as a plug-in estimate $\widehat{g}_n(u)$ below.

THEOREM 4.2. *Under the same conditions as Theorem* 4.1, *we have as* $n \to \infty$

$$n\dot{\upsilon}(\sqrt{n})(\widehat{\beta}_n - \beta_0) \to_D \left(\int_0^1 [\dot{F}(\beta_0^\top V(r))]^2 V(r)V^\top(r)\,dr\right)^{-1} \times \int_0^1 \dot{F}(\beta_0^\top V(r))V(r)\,dU(r), \tag{4.6}$$

$$n^{1/4}(\widehat{\theta}_n - \theta_0) \to_D \mathbf{MN}(0, \rho_{11}\theta\theta_0^\top), \tag{4.7}$$

where the same notation is used as in Theorem 3.5.

Also, a plug-in estimate of the form: $\widehat{g}_n(u) = Z_k^\top(u)\widehat{c}$ *has the asymptotic normality as in Theorem* 3.3. *The results in Theorem* 3.2 *and Corollaries* 3.1–3.2 *remain valid.*

The proofs of Theorems 4.1–4.2 are given in Appendix E of the supplementary material [Dong, Gao and Tjøstheim (2015)].

5. Simulation experiments. This section studies the finite-sample performance of the proposed estimates. Let $d = 2$ and x_t be generated by

$$x_t = x_{t-1} + v_t \qquad \text{with } v_t = r_0 v_{t-1} + \varepsilon_t, \tag{5.1}$$

for $t = 1, \dots, n$, where $r_0 = 0.1$, $\varepsilon_t \sim iiN(0, \sigma^2 I_2)$, $x_0 = 0$ surely. Let sample size $n = 400, 600$ and 1000. The number of Monte Carlo replications is $M = 2000$. The truncation parameter is $k = [a \cdot n^\kappa]$ with $\kappa = \frac{5}{44}$ and $a = 3.65$, satisfying the conditions in Assumption B. We shall then use two examples.

EXAMPLE 5.1. Consider a single-index model $y_t = g(\theta_0^\top x_t) + e_t$, $e_t \sim N(0, 1)$, $t = 1, \dots, n$. There are two parts in the simulation, according as $\theta_0^\top = (0.6, -0.8)$ and $\theta_0^\top = (1, 0)$ that both satisfy $\|\theta_0\| = 1$.

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

$$\text{Bias} = \bar{\bar{\theta}}_n - \theta_0, \qquad \text{S.d.} = \left(\frac{1}{M} \sum_{\ell=1}^{M} (\widehat{\theta}_{n\ell} - \bar{\bar{\theta}}_n)^{\otimes 2} \right)^{\otimes 1/2}, \tag{5.2}$$

where $\otimes$ denotes an element-wise operation, and $\bar{\bar{\theta}}_n = \frac{1}{M} \sum_{\ell=1}^{M} \widehat{\theta}_{n\ell}$, in which $\widehat{\theta}_{n\ell}$ stands for the ℓth replication of the estimate.

In order to evaluate the asymptotic theory given in Theorem 3.2, we also calculate the bias and the standard deviation of $\widehat{\theta}_{n,\text{emp}} = \widehat{\theta}_n / \|\widehat{\theta}_n\|$ in the same way as in (5.2).

Part I. Set $\theta_0^\top = (0.6, 0.8)$, $\sigma = 0.6$ and $g(u) = (1 + u^2)e^{-u^2}$. We use the proposed procedure in Section 2.1 to estimate θ_0.

As can be seen from Table 1, both the biases and the standard deviations for $\widehat{\theta}_n$ decrease as the sample size increases, and $\widehat{\theta}_{1n}$ and $\widehat{\theta}_{2n}$ have similar performance. Moreover, the biases and standard deviations of $\widehat{\theta}_{n,\text{emp}}$ indicate that $\widehat{\theta}_{n,\text{emp}}$ has a rate of convergence faster than that of $\widehat{\theta}_n$, as shown in Theorem 3.2.

TABLE 1
Bias and standard deviation for single-index model

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

TABLE 2
Bias and standard deviation for single-index model

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

Part II. Put $\theta_0^\top = (1, 0)$, $\sigma = 0.6$ and $g(u) = (1+u^2)\exp(-u^2)$. As pointed out before, the rotation of the parameters is only for the derivation of the asymptotic theory. To evaluate the asymptotic theory given in Theorem 3.1, we directly take $\theta_0^\top = \alpha_0^\top = (1, 0)$ so that $\widehat{\alpha}_n = \widehat{\theta}_n$ in this experiment.

As can be seen from Table 2, both the biases and the standard deviations of $\widehat{\alpha}_{1n}$ and $\widehat{\alpha}_{2n}$ decrease as the sample size increases. Particularly, the decrease for $\widehat{\alpha}_{2n}$ is much faster than that for $\widehat{\alpha}_{1n}$. This verifies the type of rates of convergence given in Theorem 3.1 that $\widehat{\alpha}_{2n} - \alpha_{20} = O_P(n^{-3/4})$, while $\widehat{\alpha}_{1n} - \alpha_{10} = O_P(n^{-1/4})$.

Nonetheless, shown by the standard deviations, $\widehat{\alpha}^1_{n,\text{unit}}$ converges significantly faster than $\widehat{\alpha}^2_{n,\text{unit}}$. This is also implied by Corollary 3.1 that $\widehat{\alpha}^1_{n,\text{unit}} - \alpha_{10} = O_P(n^{-3/2})$ and $\widehat{\alpha}^2_{n,\text{unit}} - \alpha_{20} = O_P(n^{-3/4})$. Note also that the biases of $\widehat{\alpha}^1_{n,\text{unit}}$ are always negative (by definition, $\widehat{\alpha}^1_{n,\text{unit}} \le \alpha_{10} = 1$) for each Monte Carlo experiment. As a result, the biases of $\widehat{\alpha}^1_{n,\text{unit}}$ approach zero relatively slower than those of $\widehat{\alpha}^2_{n,\text{unit}}$. In addition, $\widehat{\alpha}^2_{n,\text{unit}}$ and $\widehat{\alpha}_{2n}$ perform quite similarly since they have the same rate of convergence.

EXAMPLE 5.2. In this example, a partially linear single-index model of the form: $y_t = \beta_0^\top x_t + g(\theta_0^\top x_t) + e_t$, $e_t \sim N(0, 1)$, $t = 1, \dots, n$, is examined with $g(u) = (1 + u^2)\exp(-u^2)$, $\beta_0^\top = (0.3, 0.5)$, $\theta_0^\top = (0.6, -0.8)$ and $\sigma = 0.8$ involved in $\varepsilon_t \sim iiN(0, \sigma^2 I_2)$.

Formulae in (5.2) are used for $\widehat{\theta}_n$, $\widehat{\theta}_{n,\text{emp}}$ and $\widehat{\beta}_n$. All simulation results with sample size $n = 400, 600, 1000$ and $\sigma = 0.8$ are reported in Table 3. As can be seen, both the biases and the standard deviations decrease as the sample size increases. Moreover, the rate of $\widehat{\theta}_{n,\text{emp}}$ approaching the true value looks faster than that of $\widehat{\theta}_n$. This is supported by Theorems 3.5 and 3.6 that $\widehat{\theta}_n - \theta_0 = O_P(n^{-1/4})$ and $\widehat{\theta}_{n,\text{emp}} - \theta_0 = O_P(n^{-3/4})$.

Meanwhile, since $\widehat{\beta}_n = (\widehat{\beta}_{1n}, \widehat{\beta}_{2n})^\top$ possesses the fastest rate of convergence of n^{-1} by Theorem 3.5, both the biases and the standard deviations of $\widehat{\beta}_n$ support the large sample behaviour. Therefore, the asymptotic theory established in Section 3 has been evaluated in these examples.

TABLE 3
Bias and standard deviation for partially linear single-index model

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

6. Empirical study. We propose to use a partially linear single-index model to fit an empirical data set before we make some comparisons with some candidate models.

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 rate, 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 trending components except r_t. To meet the theoretical assumptions, we de-trend 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 μ_i are estimated as: $\widehat{\mu}_1 = \frac{1}{198}\sum_{i=2}^{199}(c_t - c_{t-1}) = 0.1022$, $\widehat{\mu}_2 = \frac{1}{198}\sum_{i=2}^{199}(i_t - i_{t-1}) = 0.1302$, $\widehat{\mu}_3 = \frac{1}{198}\sum_{i=2}^{199}(v_t - v_{t-1}) = 0.0181$.

Define for each t, $\tilde{c}_t = c_t - \widehat{\mu}_1 t$, $\tilde{i}_t = i_t - \widehat{\mu}_2 t$ and $\tilde{v}_t = v_t - \widehat{\mu}_3 t$. They are the de-trended versions being plotted in (b) of Figure 1, correspondingly.

An ADF test is applied to each of $\tilde{c}_t$, $\tilde{i}_t$ and $\tilde{v}_t$, respectively. The ADF test fails to reject the null of possessing a 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 differences are given in Figure 2.

The model. A partially linear single-index model is proposed to fit the data $\tilde{c}_t$, $\tilde{i}_t$, $\tilde{v}_t$ as well as r_t in the following forms:

$$y_t = \beta_0^\top x_t + g(\theta_0^\top x_t) + e_t, \tag{6.1}$$

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 $x_{1t} = \tilde{i}_{t-1}$, $x_{2t} = \tilde{i}_t$, $x_{3t} = \tilde{v}_t$, $x_{4t} = \tilde{v}_{t-1}$, $x_{5t} = r_t$, and $g(\cdot)$ is an unknown integrable function, e_t is the error term. Note that we only include the first lagged information in the discussion, as they are more relevant than the other lags.

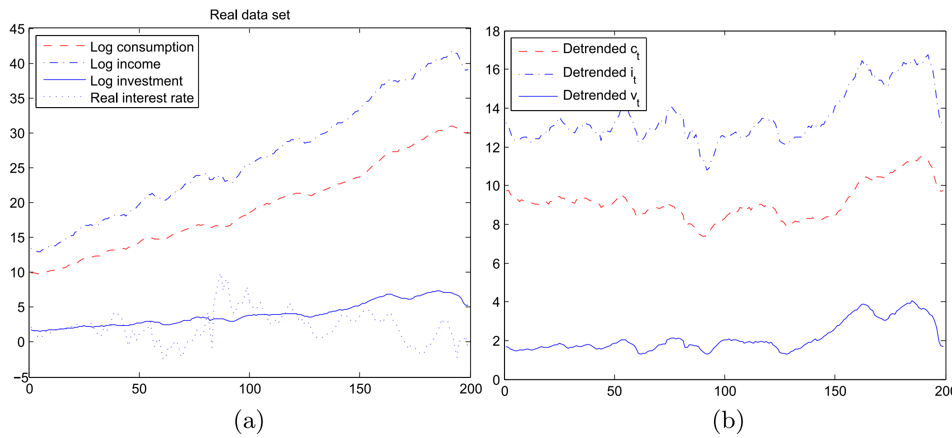


FIG. 1. *The real data and the de-trended data*. (a) *The real data*. (b) *The detrended data*.

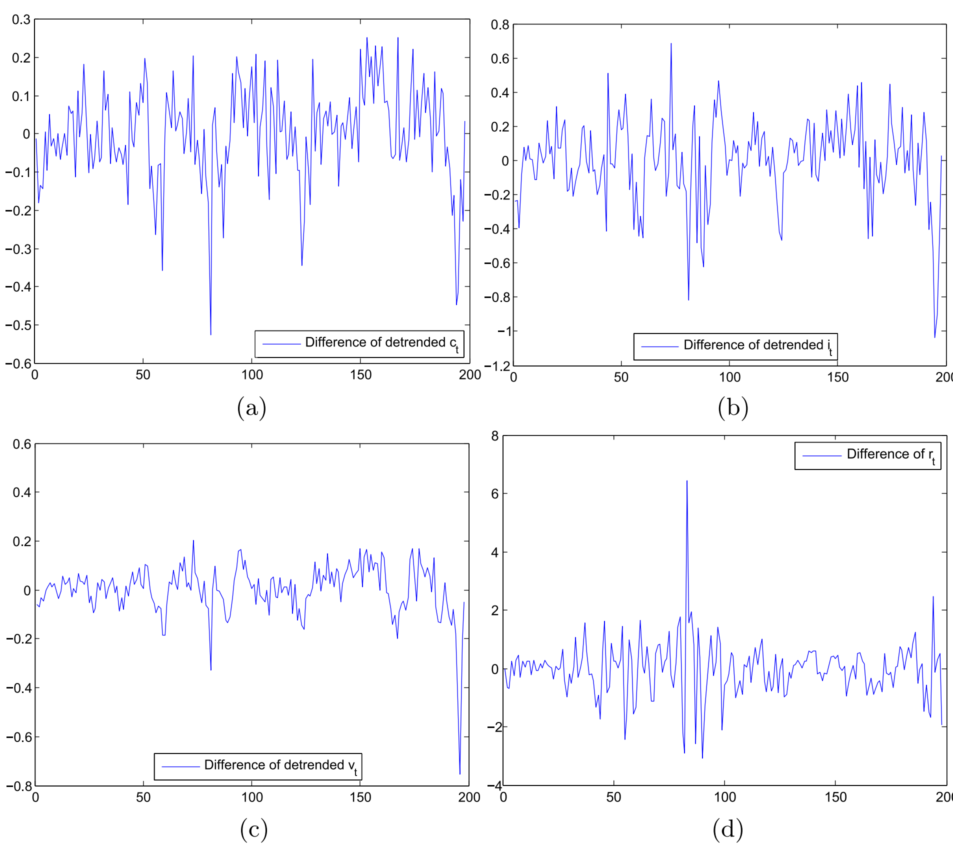
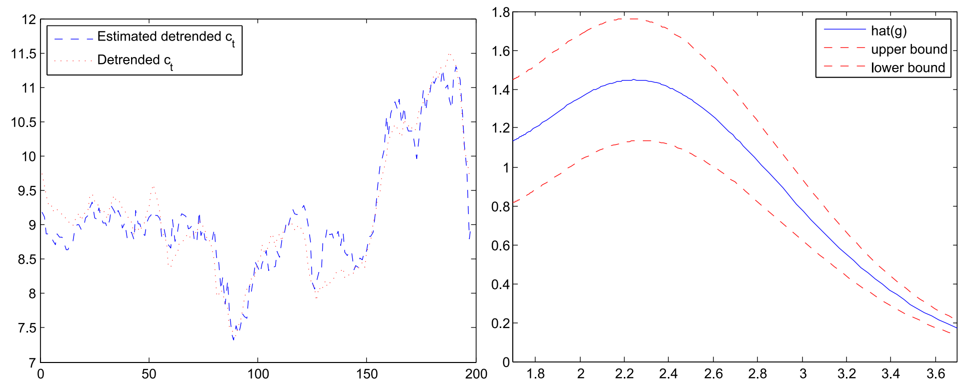


FIG. 2. *Difference of dataset*. (a) *Difference of detrended consumption*. (b) *Difference of detrended income*. (c) *Difference of detrended investment*. (d) *Difference of interest rate*.



(a) (b)

FIG. 3. *Estimated data and estimated link function*. (a) *Model* (6.1). (b) *Confidence interval curve*.

Estimation. Before implementing our proposed procedures to estimate model (6.1), one issue is to determine a suitable truncation parameter k so that the function $g(\cdot)$ can be better approximated by the first k terms in $\{\mathscr{H}_i(x)\}$. Toward this end, we propose using the Generalised Cross-Validation (GCV) method [see Gao, Tong and Wolff (2002)] as an initial step to select an optimal value k. Note that while there is no theory for such selection in the nonstationary time series case, the initial selection method works numerically in this example. Let $\widehat{k}$ denote an optimal value such that

$$\widehat{k} = \operatorname*{argmin}_{k \in K_n} \left(1 - \frac{k}{n}\right)^{-2} \widehat{\sigma}^2(k), \tag{6.2}$$

where $\widehat{\sigma}^2(k) = \frac{1}{n}\sum_{t=1}^n (y_t - \beta^\top x_t - Z_k(\theta^\top x_t)^\top \widetilde{c}(\beta, \theta))^2$, $K_n = \{1, \dots, 12\}$.

We have $\widehat{k} = 5$ by GCV, $\widehat{\beta}_n = (-0.0479, 0.5701, -1.1689, 1.8685, -0.1223)$ and $\widehat{\theta}_n = (0.2110, -0.3452, 0.0835, 2.6095, -0.2022)$. Meanwhile, we have $\widehat{c} = \widetilde{c}(\widehat{\beta}_n, \widehat{\theta}_n) = (-89.64, 112.54, -74.65, 28.94, -3.33)^\top$. This suggests

$$\begin{aligned} \widehat{g}_5(u) = {} & [-89.64 d_0^{-1} H_0(u) + 112.54 d_1^{-1} H_1(u) - 74.65 d_2^{-1} H_2(u) \\ & + 28.94 d_3^{-1} H_3(u) - 3.33 d_4^{-1} H_4(u)] e^{-u^2/2}, \end{aligned} \tag{6.3}$$

where $H_i(u)$ are the Hermite polynomials, and $d_i = (\sqrt{\pi} 2^i i!)^{1/2}$ are the norm of $H_i(u)$ in $L^2(\mathbb{R}, e^{-u^2})$ for $i = 0, 1, \dots, 4$.

In comparison, the de-trended log consumption $y_t = \tilde{c}_t$ is plotted along with the estimated de-trended log consumption by the partially linear single-index model $\widehat{y}_t = \widehat{\beta}_n^\top x_t + \widehat{g}_5(\widehat{\theta}_n^\top x_t)$ in (a) of Figure 3.

Note also by (6.3) that the estimated link function $\widehat{g}_5(u)$ is integrable on $\mathbb{R}$. According to the normality in Theorem 3.3, we draw the confidence bands for $\widehat{g}_5(u)$ at the significance level of 80% in (b) of Figure 3.

Comparison. To check whether the estimated relationship by the partially linear model is a suitable one, we shall compare model (6.1) with two natural competitors of the form

$$y_t = h(\theta_{10}^\top x_t) + e_{1t}, \tag{6.4}$$

$$y_t = \beta_{\text{linear}}^\top x_t + e_{2t}, \tag{6.5}$$

where $h(\cdot)$ is integrable and unknown, and $e_{\ell t}$ are the error terms for $\ell = 1, 2$.

To begin with, the linear model is estimated by OLS with $\widehat{\beta}_{\text{linear}} = (-0.0628, 0.7952, -1.2315, 0.9414, -0.0644)^\top$. Moreover, GCV is applied for model (6.4) with $\widehat{\sigma}^2(k) = \frac{1}{n}\sum_{t=1}^n (y_t - Z_k(\theta^\top x_t)^\top \widetilde{c}(\theta))^2$, and we have $\widehat{k} = 3$. Then $\widehat{\theta}_{1n} = (-0.0014, 0.0152, -0.0229, 0.0176, -0.0016)^\top$. Meanwhile, the estimate of $\widehat{c} = \widetilde{c}(\widehat{\theta}_{1n}) = (237.05, -61.92, 315.32)^\top$ implies

$$\begin{aligned} &\widehat{h}_3(u) \\ &\quad = [237.05 d_0^{-1} H_0(u) - 61.92 d_1^{-1} H_1(u) + 315.32 d_2^{-1} H_2(u)] e^{-u^2/2}, \end{aligned} \tag{6.6}$$

using the same notation as in (6.3).

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

(i) In-sample mean square error (MSE_{in}): As above, all unknown parameters and functions in the three models (6.1), (6.4) and (6.5) are estimated based on the whole observations (x_t, y_t), $t = 2, \ldots, 199$. Once these have been done, we shall have estimated $\widehat{y}_t^\ell$ with $\ell = 1, 2, 3$ corresponding to models (6.1), (6.4), (6.5) for $t = 2, \ldots, 199$,

$$\widehat{y}_t^1 = \widehat{\beta}_n^\top x_t + \widehat{g}_5(\widehat{\theta}_n^\top x_t), \qquad \widehat{y}_t^2 = \widehat{h}_3(\widehat{\theta}_{1n}^\top x_t) \quad \text{and} \quad \widehat{y}_t^3 = \widehat{\beta}_{\text{linear}}^\top x_t.$$

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

$$\text{MSE}_{\text{in}}(\ell) = \frac{1}{198}\sum_{t=2}^{199}(y_t - \widehat{y}_t^\ell)^2. \tag{6.7}$$

Meanwhile, to verify the choice of GCV, the MSE_{in} for model (6.1) with $k = 3, 4, 6, 7$, respectively, and for model (6.4) with $k = 1, 2, 4, 5$, respectively, are calculated as well.

(ii) Out-of-sample mean square error (MSE_{out}): Each time, one part of observations is used to estimate all unknown parameters and functions in the three models; then the next value of the dependent variable is forecasted using the estimated models. More precisely, letting $j = 1, 2, \ldots, 10$, we use the observations $\{(y_t, x_t) : 2 \le t \le 178 + 2j\}$ to estimate the unknown parameters and functions [with fixed $\widehat{k} = 5$ for model (6.1) and $\widehat{k} = 3$ for model (6.4)] in the three models, then the next y_{179+2j} is forecasted by the three models using the estimated

TABLE 4
The MSEs for models (6.1), (6.4), (6.5)

$k=$	3	4	5	6	7	
			Partially linear single-index model (6.1)			
$\mathrm{MSE_{in}}$	0.0968	0.1018	0.0946	0.1460	0.1559	
$\mathrm{MSE_{out}}$	0.2418	0.1761	0.1232	0.2146	0.1786	
			Single-index model (6.4)			Linear model (6.5)
$\mathrm{MSE_{in}}$	0.3011	0.1641	0.1544	0.7709	1.4076	0.1666
$\mathrm{MSE_{out}}$	0.6962	0.2733	0.2607	1.9060	3.0838	0.2598

parameters,

$$\widehat{y}^1_{179+2j} = \widehat{\beta}_j^\top x_{179+2j} + \widehat{g}_5^j(\widehat{\theta}_j^\top x_{179+2j}),$$

$$\widehat{y}^2_{179+2j} = \widehat{h}_3^j(\widehat{\theta}_{1j}^\top x_{179+2j}) \quad \text{and} \quad \widehat{y}^3_{179+2j} = \widehat{\beta}_{j,\mathrm{linear}}^\top x_{179+2j}.$$

The $\mathrm{MSE_{out}}$ are evaluated, for $\ell = 1, 2, 3$, by

$$\mathrm{MSE_{out}}(\ell) = \frac{1}{10}\sum_{j=1}^{10}(y_t - \widehat{y}^\ell_{179+2j})^2. \tag{6.8}$$

In addition, to assess the choice of GCV, the $\mathrm{MSE_{out}}$ for model (6.1) with $k = 3, 4, 6, 7$, respectively, and for model (6.4) with $k = 1, 2, 4, 5$, respectively, are computed as well. All $\mathrm{MSE_{in}}$ and $\mathrm{MSE_{out}}$ are given in Table 4.

In summary, among the three models, the partially linear single-index model (6.1) performs much better than the other two, in the sense that both its $\mathrm{MSE_{in}}$ and $\mathrm{MSE_{out}}$ are the smallest within the models. Particularly, model (6.1) outperforms models (6.4) and (6.5) over all choices of the truncation parameter regardless of whether or not it is chosen by GCV method. This is possibly because model (6.1) is the combination of a linear trend and a local adjustment by the link function such that it is more flexible than the other two.

Note also that, with $\widehat{k} = 5$ model (6.1) has the best performance. Therefore, model (6.1) with $\widehat{k} = 5$ is the most favourable one to explain the empirical relationship between the consumption and the income, investment and real interest rate for the US data from the period of 1960 to 2009.

7. Conclusions. The estimation procedures for both single-index and partially single-index models in the presence of nonstationarity and integrability have been proposed. New asymptotic properties for the proposed estimates have been established. The rate of convergence of the estimators of the index vector θ_0 consists of two different components in a new coordinate system for both the single-index and the partially linear single-index models, while the estimator of the coefficient vector β_0 has the super n-rate. The normality of the plug-in estimate of the link function involved in each model has been established. To satisfy the identification condition, the normalisation of the estimator of θ_0 in each case has been proposed and interestingly it possesses a fast rate of convergence. Motivated by more applicability, the partially linear single-index model is extended by using a general H-regular function to replace the linear function. New results have been obtained. Meanwhile, Monte Carlo simulations have supported the key theoretical properties. 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.

APPENDIX A: LEMMAS

Three lemmas are given in this appendix while their proofs are shown in Appendix C of the supplementary material [Dong, Gao and Tjøstheim (2015)].

LEMMA A.1. *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 density functions are bounded uniformly in* (x, w) *as well as* t *and* (t, s), *respectively*.

(2) *For large* t *and* $t - s$, *we have* $\psi_t(x, w^\top) = \phi(w) f_t(x)(1 + o(1))$ *and* $\psi_{ts}(x, w^\top) = \phi(w) f_{ts}(x)(1 + o(1))$ *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* $\frac{1}{\sqrt{t-s}}(x_{1t} - x_{1s})$.

LEMMA A.2. (1) *Under Assumptions* A *and* B, *we have as* $n \to \infty$, $\|\frac{1}{\sqrt{n}} Z^\top Z - L_1(1, 0) I_k\| = o_P(1)$ *in a richer probability space*.

(2) *Let* $\widehat{Z}$ *be the matrix* Z *defined in Section* 2 *with* θ *being replaced by* $\widehat{\theta}_n$. *Under Assumptions* A *and* B, *we have* $\frac{1}{\sqrt{n}}\|Z^\top Z - \widehat{Z}^\top \widehat{Z}\| = o_P(1)$.

LEMMA A.3. *Under Assumptions* A *and* B, *we have* $\|\widetilde{c}(\theta_0) - c\|^2 = O_P(1)\frac{k}{\sqrt{n}}$ *as* $n \to \infty$ *where* $\widetilde{c}(\theta_0)$ *is defined in Section* 2.1.

APPENDIX B: PROOFS OF THE MAIN RESULTS

The full proof of Theorem 3.2 and the outlines of the proofs of Theorems 3.3 and 3.4 are given below. In the meantime, all detailed proofs for the theorems and corollaries in Section 3 and that in Section 4 are given in Appendices D and E, respectively, of the supplementary material [Dong, Gao and Tjøstheim (2015)].

PROOF OF THEOREM 3.2. Noting that $\sqrt[4]{n}D_n^{-1} \to \mathrm{diag}(1, \mathbf{0}_{d-1})$ as $n \to \infty$ where $\mathbf{0}_{d-1}$ is a zero matrix of $(d-1) \times (d-1)$, by the continuous mapping theorem and Theorem 3.1 we have

$$
\begin{aligned}
&\sqrt[4]{n}(\widehat{\theta}_n - \theta_0) \\
&\quad = \sqrt[4]{n}(D_n Q^\top)^{-1} D_n Q^\top (\widehat{\theta}_n - \theta_0) = Q\sqrt[4]{n}D_n^{-1} D_n(\widehat{\alpha}_n - \alpha_0) \\
&\quad \to_D Q\,\mathrm{diag}(1, \mathbf{0}_{d-1}) R^{-1/2} W(1) = \mathbf{MN}(0, \rho_{11}\theta_0\theta_0^\top).
\end{aligned} \tag{B.1}
$$

In addition, it follows from $\widehat{\theta}_{n,\mathrm{emp}} = Q\widehat{\alpha}_{n,\mathrm{unit}}$, $\theta_0 = Q\alpha_0$ and Corollary 3.1 that

$$
\begin{aligned}
n^{3/4}(\widehat{\theta}_{n,\mathrm{emp}} - \theta_0) &= Qn^{3/4}\begin{pmatrix} \widehat{\alpha}^1_{n,\mathrm{unit}} - 1 \\ \widehat{\alpha}^2_{n,\mathrm{unit}} \end{pmatrix} = (\theta_0 Q_2)\begin{pmatrix} 0 \\ n^{3/4}\widehat{\alpha}^2_{n,\mathrm{unit}} \end{pmatrix} + o_P(1) \\
&= Q_2 n^{3/4}\widehat{\alpha}^2_{n,\mathrm{unit}} + o_P(1) \to_D \mathbf{MN}(0, Q_2\rho_{22}Q_2^\top). \qquad \square
\end{aligned}
$$

OUTLINE OF THE PROOF OF THEOREM 3.3. The uniform consistency of $\widehat{g}_n(u)$ follows from Lemma A.3 directly. Indeed, for large n and by the consistency of $\widehat{\theta}_n$ and the continuity of $\widetilde{c}(\theta)$ in θ, we have $\|\widehat{c}-c\|^2 = \|\widetilde{c}(\widehat{\theta}_n)-c\|^2 = O_P(1)\frac{k}{\sqrt{n}}$.

$$
\begin{aligned}
\sup_{u\in\mathbb{R}}\bigl|\widehat{g}_n(u) - g(u)\bigr| &\le \sup_{u\in\mathbb{R}}\bigl|Z_k^\top(u)[\widehat{c} - c]\bigr| + \sup_{u\in\mathbb{R}}\bigl|\gamma_k(u)\bigr| \le \sup_{u\in\mathbb{R}}\bigl\|Z_k(u)\bigr\|\|\widehat{c} - c\| \\
&\quad + \sup_{u\in\mathbb{R}}\bigl|\gamma_k(u)\bigr| = O_P(1)\sqrt{k}n^{-1/4+\kappa/2} + o(1)k^{-(m-2)/2-1/12} \\
&= o_P(1),
\end{aligned}
$$

where the facts that $\sup_{u\in\mathbb{R}}\|Z_k(u)\| \le C\sqrt{k}$ and $\sup_{u\in\mathbb{R}}|\gamma_k(u)| \le Ck^{-(m-2)/2-1/12}$ with some constant $C > 0$ are given in Lemma C.1 in the supplementary material of the paper.

For the normality, in view of the consistency of $\widehat{\sigma}_e$ and $\widehat{L}_{n1}(1, 0)$, we show the result with the replacement of σ_e and $L_1(1, 0)$. Meanwhile, in order to correspond to the plug-in of $\widehat{\theta}_n$, denote by $\widehat{Z}$ the matrix Z defined in Section 2 with replacement of θ_0 by $\widehat{\theta}_n$. Noting that $\widetilde{c} = \widetilde{c}(\theta_0) = (Z^\top Z)^{-1}Z^\top Y$ and $Y = Zc + \gamma + e$ given in Section 2.1,

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

It follows from Lemma A.2 that

$$
\begin{aligned}
&\widehat{g}_n(u) - g(u) \\
&\quad = Z_k^\top(u)\widehat{c} - g(u) = Z_k^\top(u)(\widehat{c} - c) - \gamma_k(u) \\
&\quad = Z_k^\top(u)(\widehat{Z}^\top\widehat{Z})^{-1}\widehat{Z}^\top(\gamma + e) + Z_k^\top(u)(\widehat{Z}^\top\widehat{Z})^{-1}\widehat{Z}^\top(Z - \widehat{Z})c - \gamma_k(u)
\end{aligned}
$$

$$= \frac{1}{\sqrt{n}} L_1^{-1}(1,0) Z_k^\top(u) \widehat{Z}^\top e\bigl(1+o_P(1)\bigr) + \frac{1}{\sqrt{n}} L_1^{-1}(1,0) Z_k^\top(u) \widehat{Z}^\top \gamma$$
$$+ \frac{1}{\sqrt{n}} L_1^{-1}(1,0) Z_k^\top(u) \widehat{Z}^\top (Z - \widehat{Z})c - \gamma_k(u)$$
$$= \frac{1}{\sqrt{n} L_1(1,0)} Z_k^\top(u) Z^\top e + \frac{1}{\sqrt{n} L_1(1,0)} Z_k^\top(u)(\widehat{Z} - Z)^\top e$$
$$+ \frac{1}{\sqrt{n} L_1(1,0)} Z_k^\top(u) \widehat{Z}^\top \gamma$$
$$+ \frac{1}{\sqrt{n} L_1(1,0)} Z_k^\top(u) \widehat{Z}^\top (Z - \widehat{Z})c - \gamma_k(u).$$

To fulfill the normality, we need to show

$$(1) \quad \sigma_e^{-1} L_1(1,0)^{-1/2} \frac{1}{\sqrt[4]{n}} \|Z_k(u)\|^{-1} Z_k^\top(u) Z^\top e \to_D N(0,1),$$

$$(2) \quad \frac{1}{\sqrt[4]{n}\|Z_k(u)\|} Z_k^\top(u) \widehat{Z}^\top \gamma = o_P(1),$$

$$(3) \quad \frac{1}{\sqrt[4]{n}\|Z_k(u)\|} Z_k^\top(u) \widehat{Z}^\top (Z - \widehat{Z})c = o_P(1),$$

$$(4) \quad \sqrt[4]{n}\|Z_k(u)\|^{-1} \gamma_k(u) = o(1),$$

$$(5) \quad \frac{1}{\sqrt[4]{n}} \|Z_k(u)\|^{-1} Z_k^\top(u)(\widehat{Z} - Z)^\top e = o_P(1).$$

To begin with (1), observe that

$$n^{-1/4}\sigma_e^{-1} L_1^{-1/2}(1,0)\|Z_k(u)\|^{-1} Z_k^\top(u) Z^\top e$$
$$= n^{-1/4}\sigma_e^{-1} L_1^{-1/2}(1,0)\|Z_k(u)\|^{-1} \sum_{t=1}^n Z_k^\top(u) Z_k\bigl(\theta_0^\top x_t\bigr) e_t,$$

which is a martingale array in view of Assumption A. We shall use Corollary 3.1 of Hall and Heyde (1980) to show the normality of (1).

The conditional variance process is, by Lemma A.2,

$$\frac{1}{\sqrt{n}} \sigma_e^{-2} L_1^{-1}(1,0)\|Z_k(u)\|^{-2} \sum_{t=1}^n \bigl(Z_k^\top(u) Z_k\bigl(\theta_0^\top x_t\bigr)\bigr)^2 E\bigl(e_t^2 | \mathcal{F}_{n,t-1}\bigr)$$
$$= \frac{1}{\sqrt{n}} L_1^{-1}(1,0)\|Z_k(u)\|^{-2} \sum_{t=1}^n \bigl(Z_k^\top(u) Z_k(x_{1t})\bigr)^2$$
$$= \frac{1}{\sqrt{n}} L_1^{-1}(1,0)\|Z_k(u)\|^{-2} Z_k^\top(u) \left(\sum_{t=1}^n Z_k(x_{1t})^\top Z_k(x_{1t})\right) Z_k(u)$$

$$= L_1^{-1}(1,0)\|Z_k(u)\|^{-2}Z_k^\top(u)\Big(\frac{1}{\sqrt{n}}Z^\top Z\Big)Z_k(u)$$

$$= \|Z_k(u)\|^{-2}Z_k^\top(u)Z_k(u)\big(1+o_P(1)\big) = 1+o_P(1).$$

Moreover, to make the conditional Lindeberg's condition fulfilled it suffices to show

$$\|Z_k(u)\|^{-4}\frac{1}{n}\sum_{t=1}^n E\big[\big(Z_k^\top(u)Z_k(x_{1t})e_t\big)^4|\mathcal{F}_{n,t-1}\big]$$

$$\le C\|Z_k(u)\|^{-4}\frac{1}{n}\sum_{t=1}^n\|Z_k(u)\|^4\|Z_k(x_{1t})\|^4 = C\frac{1}{n}\sum_{t=1}^n\|Z_k(x_{1t})\|^4 = o_P(1)$$

by a routine calculation using the density of $t^{-1/2}x_{1t}$ in Lemma A.1. This finishes the normality for (1). For the sake of brevity, the proof for (2)–(5) is relegated to the supplementary material. The outline then is completed. $\square$

OUTLINE OF THE PROOF OF THEOREM 3.4. Denote for any $\vartheta = (\beta,\theta)$,

$$\mathfrak{S}_n(\vartheta) = \begin{pmatrix}\mathfrak{S}_{n,1}(\vartheta)\\ \mathfrak{S}_{n,2}(\vartheta)\end{pmatrix} = \begin{pmatrix}\dfrac{\partial L_n(\vartheta)}{\partial\beta}\\ \dfrac{\partial L_n(\vartheta)}{\partial\theta}\end{pmatrix},$$

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

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

Denote $\widetilde{D}_n = \operatorname{diag}(nI_d, D_n)$. Thus, (3.17) may be equivalently written as

$$\widetilde{D}_n^{-1}\mathfrak{S}_n(\mu_0) + \widetilde{D}_n^{-1}\mathfrak{J}_n(\mu_n)\widetilde{D}_n^{-1}\widetilde{D}_n(\widehat{\mu}_n-\mu_0) = 0. \tag{B.2}$$

It follows from (B.2) that

$$\begin{aligned} &n^{-1}\mathfrak{S}_{n,1}(\mu_0) + n^{-2}\mathfrak{J}_{n,11}(\mu_n)n(\widehat{\lambda}_n-\lambda_0)\\ &\qquad + n^{-1}\mathfrak{J}_{n,12}(\mu_n)D_n^{-1}D_n(\widehat{\alpha}_n-\alpha_0) = 0,\end{aligned} \tag{B.3}$$

$$\begin{aligned} &D_n^{-1}\mathfrak{S}_{n,2}(\mu_0) + D_n^{-1}\mathfrak{J}_{n,21}(\mu_n)n^{-1}n(\widehat{\lambda}_n-\lambda_0)\\ &\qquad + D_n^{-1}\mathfrak{J}_{n,22}(\mu_n)D_n^{-1}D_n(\widehat{\alpha}_n-\alpha_0) = 0.\end{aligned} \tag{B.4}$$

The results of (3.18) and (3.19) will be derived from (B.3) and (B.4), respectively. These are shown in the following two steps.

Step 1: We first prove (3.19) from (B.4). First of all, note 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,

$$D_n^{-1}\mathfrak{S}_{n,2}(\mu_0) \to_D R^{1/2}W(1) \quad \text{and} \quad D_n^{-1}\mathfrak{J}_{n,22}(\mu_0)D_n^{-1} \to_P R, \tag{B.5}$$

where R and W are defined in Theorem 3.1.

To prove (3.19), it therefore suffices to show that

$$D_n(\widehat{\alpha}_n - \alpha_0) = \left[D_n^{-1}\mathfrak{J}_{n,22}(\mu_0)D_n^{-1}\right]^{-1} D_n^{-1}\mathfrak{S}_{n,2}(\mu_0) + o_P(1). \tag{B.6}$$

Once again, Theorem 10.1 of Wooldridge (1994) is used to prove (B.6). It is shown in detail in the supplemental material that $n(\widehat{\lambda}_n - \lambda_0) = Q^\top n(\widehat{\beta}_n - \beta_0) = O_P(1)$ and $D_n^{-1}\mathfrak{J}_{n,21}(\mu_0)n^{-1} = o_P(1)$. Define for some $\delta > 0$, $\widetilde{C}_n = n^{-\delta}\widetilde{D}_n = \operatorname{diag}(n^{1-\delta}I_d, C_n)$ such that $\widetilde{C}_n\widetilde{D}_n^{-1} \to 0$ as $n \to \infty$, where $C_n = n^{-\delta}D_n$ used in the proof of Theorem 3.1. It follows from (B.4) that

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

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

$$\sup_{\{\mu:\|\widetilde{C}_n(\mu-\mu_0)\|<1\}} \|n^{-1+\delta}C_n^{-1}[\mathfrak{J}_{n,21}(\mu) - \mathfrak{J}_{n,21}(\mu_0)]\| = o_P(1), \tag{B.7}$$

$$\sup_{\{\mu:\|\widetilde{C}_n(\mu-\mu_0)\|<1\}} \|C_n^{-1}[\mathfrak{J}_{n,22}(\mu) - \mathfrak{J}_{n,22}(\mu_0)]C_n^{-1}\| = o_P(1). \tag{B.8}$$

With the choice of $\delta \in (0, 1/24)$, both (B.7) and (B.8) are proved in the supplemental material of the paper, and hence the requirement (B.6) is verified if we choose $\delta \in (0, 1/24)$.

Furthermore, (B.5) shows that condition (iv) in Wooldridge's theorem holds. Thus, the limit distribution (3.19) now follows directly.

Step 2: We now turn to prove (3.18) from (B.3). Since $\mathfrak{J}_{n,12}(\mu_n) = \mathfrak{J}_{n,21}(\mu_n)^\top$, $D_n(\widehat{\alpha}_n - \alpha_0) = O_P(1)$ by Step 1, $D_n^{-1}\mathfrak{J}_{n,21}(\mu_0)n^{-1} = o_P(1)$ (as shown in the supplementary material), and $\mathfrak{J}_{n,11}(\mu_n)$ is independent of μ_n, we have

$$n(\widehat{\lambda}_n - \lambda_0) = \left(n^{-2}\mathfrak{J}_{n,11}(\mu_n)\right)^{-1}n^{-1}\mathfrak{S}_{n,1}(\mu_0) + o_P(1)$$

by (B.7). Note that

$$\begin{aligned}
\frac{1}{n}\mathfrak{S}_{n,1}(\mu_0) &= \frac{1}{n}Q^\top\frac{\partial L_n(\vartheta_0)}{\partial\beta} = \frac{1}{n}Q^\top\sum_{t=1}^n\left(y_t - \beta_0^\top x_t - \widehat{g}_n(\theta_0^\top x_t)\right)x_t \\
&= \frac{1}{n}\sum_{t=1}^n e_tQ^\top x_t + o_P(1) \to_D Q^\top\int_0^1 V(r)\,dU(r),
\end{aligned}$$

as shown in Appendix D of the supplementary material when $n \to \infty$. Note also that

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

almost surely using Theorem 3.1 of Park and Phillips (2001), from which (3.18) follows. The outline of the proof is completed. □

PROOF OF THEOREM 3.5. The result of (3.20) follows directly from (3.18). In view of the proof of Theorem 3.2 as well as (3.19), equation (3.21) holds. □

PROOF OF THEOREM 3.6. In view of (3.19) and the proofs of Theorems 3.2–3.3, it holds. □

Acknowledgements. The authors are grateful to the Editor, Professor Runze Li, an Associate Editor and three referees for their valuable and constructive comments and suggestions that substantially improve earlier versions of the paper. Thanks also go to many seminar and conference participants for their useful comments when this work was presented. The authors acknowledge the financial support from the Norwegian Research Council.

SUPPLEMENTARY MATERIAL

Additional technical details (DOI: 10.1214/15-AOS1372SUPP; .pdf). The proofs and technical details that are omitted in the paper are provided in the supplement that accompanies the paper.

REFERENCES

CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *J. Amer. Statist. Assoc.* **92** 477–489. MR1467842

CHANG, Y. and PARK, J. Y. (2003). Index models with integrated time series. *J. Econometrics* **114** 73–106. MR1962373

DONG, C. and GAO, J. (2013). Orthogonal expansion of lévy process functionals: Theory and practice. Available at http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2013/03-13.php.

DONG, C. and GAO, J. (2014). Specification testing in structure nonparametric cointegration. Available at http://www.buseco.monash.edu.au/ebs/pubs/wpapers/2014/wp02-14.pdf.

DONG, C., GAO, J. and TJØSTHEIM, D. (2015). Supplement to "Estimation for single-index and partially linear single-index integrated models." DOI:10.1214/15-AOS1372SUPP.

GAO, J. (2014). Identification, estimation, and specification in a class of semilinear time series models. In *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics* 421–443. Oxford Univ. Press, Oxford. MR3306932

GAO, J. and PHILLIPS, P. C. B. (2013a). Functional coefficient nonstationary regression with non- and semi-parametric cointegration. Available at http://ideas.repec.org/p/msh/ebswps/2013-16.html.

GAO, J. and PHILLIPS, P. C. B. (2013b). Semiparametric estimation in triangular system equations with nonstationarity. *J. Econometrics* **176** 59–79. MR3067025

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. *J. Multivariate Anal.* **83** 324–359. MR1945957

GAO, J., KING, M., LU, Z. and TJØSTHEIM, D. (2009a). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory* **25** 1869–1892. MR2557585

GAO, J., KING, M., LU, Z. and TJØSTHEIM, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. *Ann. Statist.* **37** 3893–3928. MR2572447

GUERRE, E. and MOON, H. R. (2006). A study of a semiparametric binary choice model with integrated covariates. *Econometric Theory* **22** 721–742. MR2283033

HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press, New York. MR0624435

HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21** 157–178. MR1212171

KARLSEN, H. A., MYKLEBUST, T. and TJØSTHEIM, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Ann. Statist.* **35** 252–299. MR2332276

KARLSEN, H. A. and TJØSTHEIM, D. (2001). Nonparametric estimation in null recurrent time series. *Ann. Statist.* **29** 372–416. MR1863963

LIANG, H., LIU, X., LI, R. and TSAI, C.-L. (2010). Estimation and testing for partially linear single-index models. *Ann. Statist.* **38** 3811–3836. MR2766869

LIN, W. and KULASEKERA, K. B. (2007). Identifiability of single-index models and additive-index models. *Biometrika* **94** 496–501. MR2380574

MA, Y. and ZHU, L. (2013). Doubly robust and efficient estimators for heteroscedastic partially linear single-index models allowing high dimensional covariates. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 305–322. MR3021389

PARK, J. Y. and PHILLIPS, P. C. B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory* **15** 269–298. MR1704225

PARK, J. Y. and PHILLIPS, P. C. B. (2000). Nonstationary binary choice. *Econometrica* **68** 1249–1280. MR1779149

PARK, J. Y. and PHILLIPS, P. C. B. (2001). Nonlinear regressions with integrated time series. *Econometrica* **69** 117–161. MR1806536

PHILLIPS, P. C. B. (2009). Local limit theory and spurious nonparametric regression. *Econometric Theory* **25** 1466–1497. MR2557571

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, New York. MR0762984

REVUZ, D. and YOR, M. (2005). *Continuous Martingales and Brownian Motion*. *A Series of Comprehensive Studies in Mathematics* **293**. Springer, Berlin.

SEVERINI, T. A. and WONG, W. H. (1992). Profile likelihood and conditionally parametric models. *Ann. Statist.* **20** 1768–1802. MR1193312

TERÄSVIRTA, T., TJØSTHEIM, D. and GRANGER, C. W. J. (2010). *Modelling Nonlinear Economic Time Series*. Oxford Univ. Press, Oxford. MR3185399

WANG, Q. and PHILLIPS, P. C. B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* **25** 710–738. MR2507529

WANG, Q. and PHILLIPS, P. C. B. (2009b). Structural nonparametric cointegrating regression. *Econometrica* **77** 1901–1948. MR2573873

WANG, Q. and PHILLIPS, P. C. B. (2012). A specification test for nonlinear nonstationary models. *Ann. Statist.* **40** 727–758. MR2933664

WANG, J.-L., XUE, L., ZHU, L. and CHONG, Y. S. (2010). Estimation for a partial-linear single-index model. *Ann. Statist.* **38** 246–274. MR2589322

WOOLDRIDGE, J. M. (1994). Estimation and inference for dependent processes. In *Handbook of Econometrics, Vol. IV. Handbooks in Econom.* **2** 2639–2738. North-Holland, Amsterdam. MR1315980

WU, T. Z., YU, K. and YU, Y. (2010). Single-index quantile regression. *J. Multivariate Anal.* **101** 1607–1621. MR2610735

XIA, Y. and LI, W. K. (1999). On single-index coefficient regression models. *J. Amer. Statist. Assoc.* **94** 1275–1285. MR1731489

XIA, Y., TONG, H. and LI, W. K. (1999). On extended partially linear single-index models. *Biometrika* **86** 831–842. MR1741980

XIA, Y., TONG, H., LI, W. K. and ZHU, L.-X. (2002). An adaptive estimation of dimension reduction space. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **64** 363–410. MR1924297

YU, Y. and RUPPERT, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Amer. Statist. Assoc.* **97** 1042–1054. MR1951258

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. MR2278341

C. DONG
SCHOOL OF ECONOMICS
SOUTHWESTERN UNIVERSITY
OF FINANCE AND ECONOMICS
CHENGDU
CHINA
AND
DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS
MONASH UNIVERSITY
MELBOURNE
AUSTRALIA
E-MAIL: dchaohua@hotmail.com

J. GAO
DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS
MONASH UNIVERSITY
MELBOURNE
AUSTRALIA
E-MAIL: jiti.gao@monash.edu

D. TJØSTHEIM
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF BERGEN
5020 BERGEN
NORWAY
E-MAIL: dag.tjostheim@math.uib.no