Ann. Inst. Statist. Math.
Vol. 49, No. 3, 493–517 (1997)

STATISTICAL INFERENCE IN SINGLE-INDEX AND PARTIALLY NONLINEAR MODELS

JITI GAO[1]* AND HUA LIANG[2]**

[1]*Department of Statistics, University of Auckland, Auckland, New Zealand*
[2]*Institute of Systems Science, Academia Sinica, Beijing, China*

(Received May 22, 1995; revised August 20, 1996)

Abstract. A finite series approximation technique is introduced. We first apply this approximation technique to a semiparametric single-index model to construct a nonlinear least squares (LS) estimator for an unknown parameter and then discuss the confidence region for this parameter based on the asymptotic distribution of the nonlinear LS estimator. Meanwhile, a computational algorithm and a small sample study for this nonlinear LS estimator are developed. Additionally, we apply the finite series approximation technique to a partially nonlinear model and obtain some new results.

Key words and phrases: Asymptotic normality, semiparametric single-index regression model, finite series approximation, partially nonlinear model.

1. Introduction

Consider the model given by

$$Y_i = g(x_i^\tau \beta_0) + e_i, \qquad i = 1, 2, \dots, \tag{1.1}$$

where $x_i = (x_{i1}, \dots, x_{ip})^\tau \in X$ ($p \geq 1$) (X is a convex subset of R^p) are known and nonrandom design points, $\beta_0 = (\beta_{01}, \dots, \beta_{0p})^\tau$ is an unknown true parameter vector over Θ_1 (a compact parameter space of R^p), $g(\cdot)$ is an unknown function over $T = \{x^\tau \beta_0 : x \in X\}$, and the e_i are *i.i.d.* random errors with $Ee_i = 0$ and $Ee_i^2 = \sigma^2 < \infty$.

The model defined in (1.1) belongs to a class of semiparametric single-index regression models, which was discussed by some authors. See, for example, Stoker (1986) and Ichimura (1987) proposed the least squares approach to estimate β_0 up to scale that uses kernel estimation in the optimizing conditions and Härdle

* Now at the School of Mathematical Sciences, the Queensland University of Technology, Gardens Point, GPO Box 2434, Brisbane, Qld 4001, Australia.

** Now at the Institute of Statistics and Econometrics, Humboldt-Universität Zu Berlin, Berlin, Germany.

and Stoker (1989) and Powell *et al.* (1989) gave the solution to the problem of estimating coefficients of index models, through the estimations of the average derivatives and the density-weighted average derivative vector of a general regression function. Other related work is that of Ruud (1986), Han (1987), Li (1989), Andrews (1991), Stoker (1993), Samarov (1993), and Whang and Andrews (1993). More recently, Härdle *et al.* (1993) investigated the asymptotic normality for an estimator of β_0, which is under the case where the x_i are *i.i.d.* random samples and $g(\cdot)$ is estimated by nonparametric kernel function.

In this paper, based on the finite series approximation technique, we first establish the asymptotic normality for the nonlinear LS estimator of β_0 and then apply the finite series approximation technique to a partially nonlinear model $Y_i = f(x_i, \beta_0) + g(t_i) + e_i$, and obtain some asymptotic results. Additionally, a computational algorithm and a small sample study for the nonlinear LS estimator are proposed to support the asymptotic theory.

The organization of this paper is as follows: Section 2 states the assumptions and the main results. Section 3 discusses the partially nonlinear model. The proofs of Theorems are given in the Appendix.

2. Statement of the main results

2.1 *Asymptotic normality*

Consider the model given by (1.1). The objective is to estimate the parameter-of-interest β_0. Here g is regarded as an infinite-dimensional nuisance parameter in the problem of estimating β_0. The approach taken here is to approximate $g(\cdot)$ by a finite series sum $\sum_{k=1}^q z_k(\cdot)\gamma_{0k}$, where $\{z_k(\cdot); k = 1, 2, \ldots, q\}$ is a prespecified family of functions from $T \subset R$ to R, $\gamma_0 = (\gamma_{01}, \ldots, \gamma_{0q})^\tau$ is an unknown parameter vector, and $q = q_n \geq 1$ is the number of summands and is taken to be nonrandom.

To construct the least squares estimator of β_0, we need to introduce the following assumption.

ASSUMPTION 2.1. For $1 \leq q = q_n < n - p$ ($q_n \to \infty$ as $n \to \infty$) and $\{z_k(\cdot), k = 1, 2, \ldots, q\}$ given above, there exists an unknown parameter vector $\gamma_0 = (\gamma_{01}, \ldots, \gamma_{0q})^\tau \in \Theta_2$ (a compact subset of R^q) such that for n large enough

$$q_n^{1/2} \max_{1 \leq i \leq n} \left| \sum_{k=1}^q z_k(x_i^\tau \beta_0)\gamma_{0k} - g(x_i^\tau \beta_0) \right| = o(n^{-1/2}). \tag{2.1}$$

In order to construct some estimators, we consider the approximate version of the model (1.1) below

$$Y_i = F(x_i, \theta_0) + \tilde{e}_i, \qquad 1 \leq i \leq n \tag{2.2}$$

where $\tilde{e}_i = e_i + \delta_i$, $\delta_i = g(x_i^\tau \beta_0) - F(x_i, \theta_0)$, $F(x_i, \theta_0) = Z(x_i^\tau \beta_0)^\tau \gamma_0$, $\theta_0 = (\beta_0^\tau, \gamma_0^\tau)^\tau \in \Theta = \Theta_1 \times \Theta_2$, and $Z(\cdot) = Z_q(\cdot) = (z_1(\cdot), \ldots, z_q(\cdot))^\tau$.

Now, based on (2.2) we can define the nonlinear least squares estimator $\hat{\theta}_n = \hat{\theta}_{nq} = (\hat{\beta}_n^\tau, \hat{\gamma}_n^\tau)^\tau$ (its existence and measurability were proved in Jennrich (1969)) of $\theta_0 = (\beta_0^\tau, \gamma_0^\tau)^\tau$ by

$$S_n(\hat{\theta}_n) = \sum_{i=1}^{n} (Y_i - F(x_i, \hat{\theta}_n))^2 = \min!, \tag{2.3}$$

where $S_n(\theta_0) = \sum_{i=1}^{n} (Y_i - F(x_i, \theta_0))^2$.

The corresponding LS estimator $\hat{g}_n$ of g is

$$\hat{g}_n(\cdot) = Z(\cdot)^\tau \hat{\gamma}_n. \tag{2.4}$$

In order to state the main result of this section, we introduce the following assumption.

ASSUMPTION 2.2.

(i) Assume that $g'(\cdot)$ exists on T and there exist some functions $h_j(\cdot)$ over T such that for all $1 \le i \le n$ and $1 \le j \le p$

$$g_{ij} = h_j(x_i^\tau \beta_0) + u_{ij}, \tag{2.5}$$

where $g_{ij} = g'(x_i^\tau \beta_0) x_{ij}$ and u_{ij} are real sequences satisfying

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} u_i u_i^\tau = B, \tag{2.6}$$

and

$$\limsup_{n\to\infty} \frac{1}{n^{1/2}\log n} \max_{1\le m\le n} \left\| \sum_{i=1}^{m} u_{j_i} \right\| < \infty \tag{2.7}$$

for any permutation $(j_1, j_2, \ldots, j_n)$ of the integers $(1, 2, \ldots, n)$, where $u_i = (u_{i1}, \ldots, u_{ip})^\tau$, B is a positive definite matrix with order $p \times p$, and $\|\cdot\|$ denotes the Euclidean norm;

(ii) For $1 \le q = q_n < n - p$ ($q_n \to \infty$ as $n \to \infty$) and $\{z_k(\cdot), k = 1, 2, \ldots\}$ given above, there exist $\gamma_j = (\gamma_{j1}, \ldots, \gamma_{jq})^\tau \in \Theta_2$ such that for n large enough and $1 \le j \le p$

$$\max_{1\le i\le n} \left| \sum_{k=1}^{q} z_k(x_i^\tau \beta_0)\gamma_{jk} - h_j(x_i^\tau \beta_0) \right| = o(n^{-1/4}). \tag{2.8}$$

Remark 2.1.

(i) There is one key reason to explain why we have proposed the Assumption 2.2. From (1.1) we know that $m_i(\beta_0) = EY_i = g(x_i^\tau \beta_0)$ and the first order derivatives of $m_i(\beta_0)$ with respect to β_0 are $g_{ij} = m'_{ij}(\beta_0) = g'(x_i^\tau \beta_0) x_{ij}$. So, the scaled coefficients g_{ij} contain the information about β_0 and therefore the relationship between $\{g_{ij}\}$ and $\{x_i^\tau \beta_0\}$ affects the estimability of β_0. Thus, how to determine this relationship is very important in constructing an asymptotically efficient estimate for β_0. Assumption 2.2(i) determines this relationship with the errors u_i satisfying some mild conditions. In the case of $p = 1$, the above equations (2.6) and (2.7) can be interpreted as the restriction of the asymptotic finiteness of $\frac{1}{n}\sum_{i=1}^n u_i^2$ and the asymptotic negligibility of $\frac{1}{n}\sum_{i=1}^n u_i$. Some similar discussions about the relationship between the fixed design points $\{x_i\}$ and $\{t_i\}$ in the partly linear model $y_i = x_i^\tau \beta_0 + g(t_i) + e_i$ were given by Rice (1986), Speckman (1988), and Gao (1992). Rice (1986) assumed that there is a function $h(\cdot)$ such that $x_i = h(t_i) + \epsilon_i$ with the ϵ_i satisfying some conditions more complicated than (2.6) and (2.7); Speckman (1988) placed a random prior on $\{\epsilon_i\}$ and interpreted ϵ_i as independent random errors, and Gao (1992) improved the conditions of Rice (1986) and proposed some reasonable assumptions which are similar to (2.6) and (2.7). Based on the points of Rice (1986), Speckman (1988), and Gao (1992), we also propose Assumption 3.2(ii) below for the partially nonlinear model. As a matter of fact, the above $\{u_i\}$ behaves like zero mean *i.i.d.* random variable, and $\{h_j(t_i)\}$ is the regression of $\{g_{ij}\}$ on $t_i = x_i^\tau \beta_0$. Specifically, if the design point $\{x_i\}$ is the *i.i.d.* random sample, and let $h_j(t_i) = E(g_{ij} \mid t_i) = g'(t_i) \times E(x_{ij} \mid t_i)$ and $u_{ij} = g_{ij} - h_j(t_i)$ with $Eu_i u_i^\tau > 0$. Then (2.5) holds automatically, and (2.6) and (2.7) hold with probability one by the law of strong large numbers and the law of the iterated logarithm respectively. When the x_i are *i.i.d.* random samples, Hardle *et al.* (1993) proposed some sufficient assumptions for constructing an asymptotically normal estimator for β_0 in the single-index model.

(ii) As a matter of fact, it is not difficult to show that $\{h_j(\cdot)\}$ is determined uniquely under Assumption 2.2. This is necessary to ensure the uniqueness of B in (2.6).

Remark 2.2. Assumptions 2.1 and 2.2(ii) are some smoothness conditions. In almost all cases, they hold if g and h_j are sufficiently smooth. See, for example, g and h_j can be approximated by trigonometric series used by Eubank and Speckman (1990) and Eastwood and Gallant (1991). More generally, they hold whenever T is compact (see Nurnberger (1989)). The details about the polynomials and splines can be found in Nurnberger (1989). See, for example, Theorems 2.8, 3.30, Corollary 2.11, and Theorem 4.27. The required smoothness of the regression functions g and h_j given here is not just an artifact of the proof, but is a property of some estimators.

ASSUMPTION 2.3.

(i) For any $b > 0$ and some $c > 0$

$$\limsup_{n\to\infty} \frac{\{\sum_{i=1}^n \sup_{\|\theta-\theta_0\|\geq b}(F_i(\theta) - F_i(\theta_0))^2\}^{(1+c)/2}}{\inf_{\|\theta-\theta_0\|\geq b}\sum_{i=1}^n (F_i(\theta) - F_i(\theta_0))^2} < \infty,$$

where $F_i(\theta) = F(x_i, \theta)$, $\theta \in \Theta$;

(ii) $\{F_i(\theta)\}$ is Lipschitz function on Θ and

$$\sup_{\theta_1\neq\theta_2\in\Theta} \frac{|F_i(\theta_1) - F_i(\theta_2)|}{\|\theta_1 - \theta_2\|} \leq M_0 \sup_{\|\theta-\theta_0\|\geq \bar{b}} |F_i(\theta_1) - F_i(\theta_2)|$$

for some $\bar{b}$ and for all i, where M_0 is independent of i.

Remark 2.3.

(i) Some conditions for establishing the asymptotic normality of nonlinear least squares estimates have been discussed by some authors. See, for example, Jennrich (1969), Wu (1981), Gallant (1987), and Seber and Wild (1989). In this section, we adopt the assumptions of Wu (1981).

(ii) Assumption 2.3 is similar to Assumption A of Wu (1981), which is satisfied for the case where T is compact and $\{z_k(t), k \geq 1\}$ is of the form: $1, t, t^2, \ldots, t^m$ or $\cos(kt)$ or $\sin(kt)$ used by Eubank and Speckman (1990).

ASSUMPTION 2.4.

(i) Z_0 has of full column rank q for n large enough, and $\liminf_{n\to\infty} \cdot \lambda_{\min}(Z_0^\tau Z_0/n) > 0$, where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of matrix A and $Z_0 = Z_{\beta_0} = (Z(x_1^\tau\beta_0), \ldots, Z(x_n^\tau\beta_0))^\tau$;

(ii) $\lim_{n\to\infty}\max_{1\leq i\leq n} p_{ii} = 0$ and $\lim_{n\to\infty}\max_{1\leq i\neq j\leq n} p_{ij}(\log n)^2 = 0$, where $\{p_{ij}\}$ denotes the i-th row element and j-th rank element of the $n\times n$ projection matrix $P_0 = Z_0(Z_0^\tau Z_0)^{-1}Z_0^\tau$;

(iii) for any $j \geq 1$, $\limsup_{n\to\infty}\frac{1}{n}\max_{j\geq 1}\sum_{i=1}^n d_i^2 x_{ij}^2 < \infty$, where $d_i^2 = \max(1, g(x_i^\tau\beta_0)^2)$;

(iv) for any given $\delta > 0$, $\limsup_{n\to\infty}\frac{q}{n}\sup_{\|\beta-\beta_0\|\leq\delta}\max_{1\leq k\leq q} \cdot \sum_{i=1}^n z_k(x_i^\tau\beta)^2 < \infty$;

(v) both $\frac{q}{n}\max_{1\leq j\leq p, 1\leq k,l\leq q}|\sum_{i=1}^n z_k'(x_i^\tau\beta)z_l(x_i^\tau\beta_0)x_{ij}|$ and $\frac{q}{n}\max_{1\leq k\neq l\leq q}\cdot |\sum_{i=1}^n z_k(x_i^\tau\beta)z_l(x_i^\tau\beta_0)|$ converge to zero uniformly as $\|\beta-\beta_0\|\to 0$ and $n\to\infty$;

(vi) for any $k \geq 1$, $z_k^{(l)}(\cdot)$ and $g'(\cdot)$ exist over T, both $\max_{1\leq i\leq n}|Z'(x_i^\tau\beta)^\tau\gamma - g'(x_i^\tau\beta)|$ and $\max_{1\leq i\leq n}|g'(x_i^\tau\beta) - g'(x_i^\tau\beta_0)|$ tend to 0 uniformly as $\|\theta-\theta_0\|\to 0$ and $n\to\infty$, where $Z^{(l)}(\cdot) = (z_1^{(l)}(\cdot), \ldots, z_q^{(l)}(\cdot))^\tau$ and $l = 1, 2$;

(vii) there exists an absolute constant $d > 0$ such that

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^n \sup_{\|\theta-\theta_0\|\leq d}\lambda_{\max}(F_i''(\theta))^2 < \infty;$$

(viii) if

$$\sum_{i=1}^\infty \sup_{\|\theta-\theta_0\|\leq d}\lambda_{\min}(F_i''(\theta))^2 = \infty,$$

then there exists an M independent of i such that for all i

$$\sup_{\theta_1\neq\theta_2\in\Theta_d}\frac{|\lambda_{\max}(F_i''(\theta_1) - F_i''(\theta_2))|}{\|\theta_1-\theta_2\|} \leq M\sup_{\theta\in\Theta_d}|\lambda_{\max}(F_i''(\theta))| < \infty,$$

where $\Theta_d = \{\theta\in\Theta; \|\theta-\theta_0\|\leq d\}$, $F_i''(\theta) = \{a_{ijk}(\theta)\}_{1\leq j,k<p+q}$, $a_{ijk}(\theta) = Z''(x_i^\tau\beta)^\tau\gamma\cdot x_{ij}x_{ik}$ $(1\leq j,k\leq p)$, $a_{i,j+p,k+p}(\theta) = 0$ $(1\leq j,k\leq q)$ and $a_{i,j,k+p}(\theta) = a_{i,k+p,j}(\theta) = z_k'(x_i^\tau\beta)x_{ij}$ $(1\leq j\leq p, 1\leq k\leq q)$.

Remark 2.4.

(i) Assumption 2.4(i) is sufficient for constructing a $n^{1/2}$ consistent estimator of β_0. Assumptions 2.4(i) and the second part of 2.4(v) hold in many cases. See, for example, Eastwood and Gallant (1991) proposed that $Z_0^\tau Z_0 = \mathrm{diag}(n, n/2, \ldots, n/2)$ and Eubank and Hart (1992) assumed that $Z_0^\tau Z_0 = n \cdot I_q$.

(ii) Assumption 2.4(ii) restricts the growth rate of q_n. More precisely, it restricts the growth rate of the number of linearly independent column of Z_0. If $Z_0^\tau Z_0$ is nonsingular, the first part of 2.4(ii) implies $q_n/n \to 0$ as $n \to \infty$, since $\max_{1 \le i \le n} p_{ii} \ge \frac{1}{n} \mathrm{tr}(Z_o(Z_o^\tau Z_o)^{-1} Z_o) = q_n/n$.

(iii) Assumptions 2.4(iii)(iv) hold for most design point $\{x_i\}$ and most series functions. For example, Assumption 2.4(iv) holds for trigonmetric polynomial and FFF functions. See Eubank and Speckman (1990), Eubank and Hart (1992), Eastwood and Gallant (1991), and Andrews (1991). For the case where the $\{x_i\}$ is *i.i.d.* random sample, Assumptions 2.4 (iii)(iv) hold with probability one by using the law of strong large numbers.

(iv) Assumptions 2.4(iv) and 2.4(v) also hold in probability when the $\{x_i\}$ is *i.i.d.* random sample, which can be verified easily in the case where $\{z_k(\cdot), k \ge 1\}$ is of the form: $\{\cos(kt); \sin(kt)\}$. See Lemma A.2 of Gao (1996*b*) for more details.

(v) Like Assumption 2.1, Assumption 2.4(vi) restricts the smoothness of $g(\cdot)$ and $z_k(\cdot)$, which are satisfied when g is approximated by a class of smoothing spline functions (see Nurnberger (1989)). Generally, it is not difficult to verify above Assumptions 2.4(i)–(v) for the case where T is compact and $\{z_k(\cdot), k \ge 1\}$ is of the form: $\{1, t, t^2, \ldots, t^m$; or $\cos(kt)$; or $\sin(kt)\}$.

(vi) As p is finite integer, Assumptions 2.4(vii)(viii) can be replaced by some assumptions imposed on $\{a_{ijk}(\theta)\}$, which are similar to Assumptions B(iv)(v) of Wu (1981). Since $q = q_n$ proposed in this paper depends on n and $q = q_n \to \infty$ as $n \to \infty$, we impose the conditions on the eigenvalue of the matrix $\{F_i''(\theta)\}$ not on $\{a_{ijk}(\theta)\}$. Assumptions 2.4(vii)(viii) are general conditions for constructing the asymptotic distribution of $\hat{\beta}_n$ for the case where p is finite or $p = p_n \to \infty$ as $n \to \infty$, which are needed for the application of Corollary A of Wu (1981) (the law of uniformly strong large numbers). Recently, Pollard (1984, 1990) proposed some new results about the uniform strong laws under some suitable conditions. See, for example, Theorem 8.3 of Pollard (1990) is one of the key results.

Now, we give the asymptotic results of this section.

THEOREM 2.1. *Assume that Assumptions* 2.1–2.3 *hold. Let* $D_n(\theta, \theta_0) = \sum_{i=1}^n (F_i(\theta) - F_i(\theta_0))^2 \to \infty$ *for all* $\theta \ne \theta_0$ *and* $n \to \infty$. *Then as* $n \to \infty$

$$\hat{\theta}_n - \theta_0 \to 0 \quad a.s. \tag{2.9}$$

THEOREM 2.2. *Assume that Assumptions* 2.1, 2.2 *and* 2.4 *hold. Let* θ_0 *be in the interior of* Θ *and* $\hat{\theta}_n$ *be a strongly consistent least squares estimator of* θ_0. *Then as* $n \to \infty$

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \to_D N(0, \sigma^2 B^{-1}), \tag{2.10}$$

where B *is defined by* (2.6).

Remark 2.5.

(i) Theorems 2.1 and 2.2 only consider the case where the e_i are *i.i.d.* random errors with $Ee_i = 0$ and $Ee_i^2 = \sigma^2 < \infty$. As a matter of fact, Theorems 2.1 and 2.2 can be modified to the case where $Ee_i = 0$ and $Ee_i^2 = f(x_i^\tau \beta_0)$ with unknown function $f(\cdot)$. First, based on the finite series approximation $\hat{g}_n$ to g and a probability weighting function $\{W_{ni}(\cdot)\}$, we can construct an estimator $\hat{f}_n(\cdot) = \sum_{j=1}^n W_{nj}(\cdot)(Y_j - \hat{g}_n(x_j^\tau \beta_0))^2$ for $f(\cdot)$ and then define a weighted LS estimator $\bar{\beta}_n$ for β_0 by minimizing $\sum_{i=1}^n \frac{(Y_i - \hat{g}_n(x_i^\tau \beta_0))^2}{\hat{f}_n(x_i^\tau \beta_0)}$ over $\beta_0 \in R^p$. Some additional conditions are required to obtain the asymptotic normality of $\bar{\beta}_n$.

(ii) The proposed estimators have the following advantages: the first is that Theorem 2.2 and its proof can be generalized to the case where $p = p_n$ depends on n. In this case, we only need to modify the conditions of Theorem 2.2 slightly and new conditions and corresponding conclusion are similar to those of Theorem 2.3 below. See Remark 2.4(vi), Subsection 2.3 below and Remark A.1 in the Appendix. The second advantage is that through using the finite series approximation, we can apply some previous results including computational algorithms for classical partially linear models (see Seber and Wild (1989), p. 654) to the new partially linear model (2.13) below, and obtain some new results for $\hat{\beta}_n$. See Subsection 2.2 below for more details. From Section 4 below, we know that the small sample results support the new estimates proposed in this paper.

If $\sigma^2 = Ee_1^2$ is unknown, then we can define an estimator for σ^2 by

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{g}_n(x_i^\tau \hat{\beta}_n))^2 = \frac{1}{n}\sum_{i=1}^n (Y_i - Z(x_i^\tau \hat{\beta}_n)^\tau \hat{\gamma}_n)^2. \tag{2.11}$$

And, the strong consistency and asymptotic normality of $\hat{\sigma}_n^2$ can be obtained by using Theorems 2.1 and 2.2. Here we omit the details.

2.2 *Confidence region for parameter vector β_0*

Based on the finite series approximate technique, the semiparametric single-index regression model

$$Y_i = g(x_i^\tau \beta_0) + e_i \tag{2.12}$$

has the following approximation version

$$Y_i = Z(x_i^\tau \beta_0)^\tau \gamma_0 + \tilde{e}_i, \tag{2.13}$$

which generalizes the classical partially linear models (see Seber and Wild (1989), p. 654) to the case where the numbers p and q may depend on n.

The confidence region for the parameter subsets in the classical partially linear models has been discussed by some authors. See, for example, Hamilton (1986), Gallant (1987), and Seber and Wild (1989).

In this section, the objective is to construct the confidence region for the parameter-of-interest β_0. Here γ_0 is regarded as a nuisance parameter.

The approach to finding confidence region for the parameter vector β_0 is based on the large sample distribution of the nonlinear LS estimator $\hat{\beta}_n$. By Theorem 2.2 we have as $n \to \infty$

$$F_{1n} = n(\hat{\beta} - \beta_0)^\tau B(\hat{\beta}_n - \beta_0)\sigma^{-2} \to {}_D\chi_p^2. \tag{2.14}$$

Since B depends on g (unknown link function), we need to modify F_{1n}. By (2.3) we can define the estimate of $g'(\cdot)$ by

$$\hat{g}_n'(\cdot) = Z'(\cdot)^\tau \hat{\gamma}_n, \tag{2.15}$$

where $Z'(\cdot) = (z_1'(\cdot), \ldots, z_q'(\cdot))^\tau$ is defined as in Assumption 2.4(vi).

Now, F_{1n} can be replaced by

$$\hat{F}_{1n} = n(\hat{\beta}_n - \beta_0)^\tau \hat{B}(\hat{\beta}_n - \beta_0)\sigma^{-2}, \tag{2.16}$$

where $\hat{B} = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \hat{u}_i\hat{u}_i^\tau$, $\hat{u}_i = Z'(x_i^\tau\hat{\beta}_n)^\tau\hat{\gamma}_n x_i - H(x_i^\tau\hat{\beta}_n)$, $\hat{\beta}_n = (\hat{\beta}_{n1}, \ldots, \hat{\beta}_{np})^\tau$, and $H(\cdot) = (h_1(\cdot), \ldots, h_p(\cdot))^\tau$.

If σ^2 is unknown and $h_j(\cdot)$ in (2.5) is unknown, then σ^2 and $h_j(\cdot)$ can be replaced by the above consistent estimates $\hat{\sigma}_n^2$ and $\hat{h}_j(\cdot) = Z(\cdot)^\tau\hat{\gamma}_{nj}$, where $\{\hat{\gamma}_{nj}\}$ minimizes $\sum_{i=1}^n \tilde{u}_{ij}^2$ and $\tilde{u}_{ij} = Z'(x_i^\tau\hat{\beta}_n)^\tau\hat{\gamma}_n x_{ij} - Z(x_i^\tau\hat{\beta}_n)^\tau\hat{\gamma}_{nj}$. We now have another test statistic

$$\tilde{F}_{1n} = n(\hat{\beta}_n - \beta_0)^\tau \tilde{B}(\hat{\beta}_n)(\hat{\beta}_n - \beta_0)\hat{\sigma}_n^{-2}, \tag{2.17}$$

where $\tilde{B}(\hat{\beta}_n) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \tilde{u}_i\tilde{u}_i^\tau$, $\tilde{u}_i = Z'(x_i^\tau\hat{\beta}_n)^\tau\hat{\gamma}_n x_i - \hat{H}(x_i^\tau\hat{\beta}_n)$, and $\hat{H}(\cdot) = (\hat{h}_1(\cdot), \ldots, \hat{h}_p(\cdot))^\tau$.

By (2.10) and (2.15)–(2.17), it is not difficult to show that as $n \to \infty$

$$\tilde{F}_{1n} \to {}_D\chi_p^2. \tag{2.18}$$

Thus, based on (2.18) we can obtain the approximate $(1-\alpha)$ confidence region for β_0,

$$\{\beta_0 \in R^p : (\beta_0 - \hat{\beta}_n)^\tau \tilde{B}(\hat{\beta}_n)(\beta_0 - \hat{\beta}_n) \le \frac{1}{n}\hat{\sigma}_n^2\chi_{1-\alpha}^2(p)\}. \tag{2.19}$$

Writing $\{\tilde{b}_{jj}(\hat{\beta}_n)\}$ for the j-th diagonal element of $\tilde{B}(\hat{\beta}_n)$; we also have the approximate $(1-\alpha)$ confidence region for $\{\beta_{0j}\}$,

$$\{\beta_{0j} \in R : |\beta_{0j} - \hat{\beta}_{nj}| \le n^{-1/2}\hat{\sigma}_n\tilde{b}_{jj}(\hat{\beta}_n)^{1/2}\Phi_{1-\alpha/2}\}. \tag{2.20}$$

Remark 2.6. Here we only discuss the construction of the confidence region for β_0, which is based on the large sample distribution of $\hat{\beta}_n$. In fact, the approaches based on the likelihood ratio and the score statistics can also be given. Here we omit the details which are similar to the discussion of Hamilton (1986).

Remark 2.7. The computation of $\hat{F}_{1n}$ (or $\tilde{F}_{1n}$) is very important in small sample situation, which depends on the computations of $\hat{\beta}_n$ and $\hat{\gamma}_n$. The computational procedures of the nonlinear least squares estimators $\hat{\beta}_n$ and $\hat{\gamma}_n$ are detailed in Section 4 below.

2.3 *Hypothesis test for testing $g = g_0$*

Another important problem considered in this section is testing a null hypothesis $H_0 : g(\cdot) = g_0(\cdot)$. The hypothesis test for testing $g = g_0$ has been discussed by some authors. See, for example, Eubank and Speckman (1990), Eubank and Hart (1992), Gozalo (1993), and Gao (1996*a*). Recently, Azzalini and Bowman (1993) investigated the test statistic for testing $H_0 : g$ is linear and alternative hypothesis $H_0' : g$ is a smooth function, and constructed the test statistic which displays the same form as the statistic of Durbin-Watson and Munson and Jernigan (1989). Like Azzalini and Bowman (1993), we can construct a test statistic for testing null hypothesis test $H_0 : g = g_0$. But here $q = q_n \to \infty$ as $n \to \infty$, so we need to modify the F-statistic used by Azzalini and Bowman (1993). Because of the same reason, some test statistics proposed by Gallant (1987) and Seber and Wild (1989) for testing non-linear regression functions should also be modified here.

In the following, we will construct a test statistic for testing the null hypothesis $H_0 : g = g_0$.

Based on Assumption 2.1, the null hypothesis $H_0 : g = g_0$ is equivalent to $H_0'' : \gamma_0 = \tilde{\gamma}_0$ (known vector). This suggests using a test statistic of the form

$$F_{2n} = (2q_n)^{-1/2}\sigma^{-2}((\hat{\gamma}_n - \tilde{\gamma}_0)^\tau W(\hat{\theta}_n)^\tau W(\hat{\theta}_n)(\hat{\gamma}_n - \tilde{\gamma}_0) - q\sigma^2), \tag{2.21}$$

where $W(\hat{\theta}_n) = Z_{\hat{\beta}_n} = (Z(x_1^\tau \hat{\beta}_n), \ldots, Z(x_n^\tau \hat{\beta}_n))^\tau$ and $\sigma^2 = Ee_1^2$ is known. If $\sigma^2 = Ee_1^2$ is unknown, it can be replaced by the above consistent estimate $\hat{\sigma}_n^2$ without affecting the following Theorem 2.3.

For investigating the asymptotic behaviour of the test statistic F_{2n}, we introduce the following Assumption 2.5.

ASSUMPTION 2.5.

(i) There exists a sequence of real numbers j_n ($j_n \to \infty$ as $n \to \infty$) such that for $n \to \infty$

$$j_n q_n^{-1} \max_{1 \le i \le n} \sum_{j=1, \ne i}^{n} p_{ij}^2 \to 0;$$

(ii) $q_n^{-1} \max_{1 \le i \le n} k_i^2 \to 0$ as $n \to \infty$, where k_i denote the eigenvalues of the matrix P_0;

(iii) $\lim_{n\to\infty} \frac{1}{nq_n} \max_{1 \le j \le p} \sum_{k=1}^{n} (\sum_{i=1 \ne k}^{n} p_{ik} v_{ij})^2 = 0$, where $t_i = x_i^\tau \beta_0$ and $v_{ij} = Z'(t_i)^\tau \gamma_0 x_{ij}$.

Remark 2.8.

(i) Like Assumption 2.4(ii), Assumptions 2.5(i)(ii) restrict the growth rate of q_n. Assumption 2.5(i)(ii) are needed for the application of Theorem 5.1 of De Jong (1987).

(ii) Assumption 2.5(iii) imposes some orthogonality restriction on $\{v_{ij}\}$ and $\{z_k(t_i)\}$, which holds in many cases. See, for example, Eastwood and Gallant (1991) gave $p_{ij} = \frac{2}{n} \sum_{k=1}^{q} c_k z_k(t_i) z_k(t_j)$ ($c_1 = \frac{1}{2}, c_k = 1, k \ge 2$). If $\{x_{ij} z_l'(t_i)\}$ and $\{z_k(t_i)\}$ are orthogonal for all j, k, and l (i.e. $\sum_{i=1}^{n} x_{ij} z_l'(t_i) z_k(t_i) = 0$), then Assumption 2.5(iii) holds immediately. Recently, Eubank and Hart (1992) also used some orthogonality conditions. For the case where $\{x_i\}$ is *i.i.d.* random variable, Assumption 2.5(iii) holds with probability one by applying the law of strong large numbers.

Now, we give one of the main results below.

THEOREM 2.3. *Assume that the conditions of Theorem 2.2 and that Assumption 2.5 hold. Let $Ee_1^4 < \infty$. Then under the null hypothesis $H_0 : g = g_0$*

$$F_{2n} \to_D N(0,1), \qquad as \qquad n \to \infty. \tag{2.22}$$

Furthermore, if $H_1 : g \neq g_0$ holds, then $F_{2n} \to_p \infty$ as $n \to \infty$.

Remark 2.9. Theorem 2.3 states that F_{2n} has an asymptotic standard normal distribution under null hypothesis H_0. In general, H_0 should be rejected if F_{2n} exceeds some approximate upper-tail critical value, F_0, of the standard normal distribution.

The proofs of Theorems 2.1 through 2.3 are given in the Appendix.

Remark 2.10. As there is a close connection between hypothesis test and data-driven smoothing parameter. Eubank and Hart (1992) and Eubank *et al.* (1993) proposed some new approaches to construct test statistics for testing $H_1 : g$ is linear or $H_1' : g$ is a smooth function.

In this section, we will give some similar discussions. Consider the hypothesis test problem

$$\begin{aligned} &H_0 : g(t) = t, \\ &H_0' : g(t) = t + g_1(t), \qquad t \in T \end{aligned} \tag{2.23}$$

where g_1 is an unknown smooth function.

The following discussion is based on g_1 being approximated by a class of orthogonal series functions $\{p_k(\cdot), k = 1, 2, \ldots, n-p\}$.

For the following discussion, we introduce the orthogonality conditions below.

ASSUMPTION 2.6. Let $\{p_k(\cdot), k = 1, 2, \ldots, n-p\}$ be the functions on T that satisfy the orthogonality conditions

$$\begin{aligned} &\sum_{i=1}^n p_k(x_i^\tau \beta_0) p_l(x_i^\tau \beta_0) = n\delta_{kl}, \qquad k, l = 1, 2, \ldots, n-p, \\ &\sum_{i=1}^n p_k(x_i^\tau \beta_0) x_{ij} = 0, \qquad j = 1, 2, \ldots, p; \qquad k = 1, 2, \ldots, n-p. \end{aligned} \tag{2.24}$$

To test H_0, we consider fitting the alternative "model"

$$Y_i = x_i^\tau \beta_0 + \sum_{k=1}^q p_k(x_i^\tau \beta_0) a_k + e_i, \qquad i = 1, 2, \ldots, n. \tag{2.25}$$

Let $X = (x_1, \ldots, x_n)^\tau$, $Y = (Y_1, \ldots, Y_n)^\tau$, and define

$$\bar{\beta}_n = (\bar{\beta}_{n1}, \ldots, \bar{\beta}_{np})^\tau = (X^\tau X)^{-1} X^\tau Y. \tag{2.26}$$

Here it needs to assume that X is of full column rank. Based on $\{t_i, Y_i; 1 \le i \le n\}$ $(t_i = x_i^\tau \beta_0)$ we define the sample "Fourier" coefficients

$$\tilde{a}_{kn} = a_{kn}(\beta_0) = \frac{1}{n} \sum_{i=1}^{n} p_k(x_i^\tau \beta_0) Y_i. \tag{2.27}$$

Since $\{a_{kn}(\beta_0)\}$ depends on the unknown parameter β_0, we need to estimate $a_{kn}(\beta_0)$. A natural estimate of $\{a_{kn}(\beta_0)\}$ is

$$\bar{a}_{kn} = a_{kn}(\bar{\beta}_n) = \frac{1}{n} \sum_{i=1}^{n} p_k(x_i^\tau \bar{\beta}_n) Y_i. \tag{2.28}$$

Assume that σ^2 and β_0 are known, an estimate of the risk with (2.25) is

$$\begin{aligned} &\frac{1}{n} \sum_{i=1}^{n} (Y_i - x_i^\tau \beta_0 - \sum_{k=1}^{q} p_k(x_i^\tau \beta_0) \tilde{a}_{kn})^2 + \frac{2\sigma^2}{n}(p+q) \\ &= \frac{1}{n} \sum_{i=1}^{n} (Y_i - x_i^\tau \beta_0)^2 - \sum_{k=1}^{q} \tilde{a}_{kn}^2 + \frac{2\sigma^2}{n}(p+q). \end{aligned} \tag{2.29}$$

An indication that H_0 is false is provided if (2.29) is minimized by any value of q other than zero.

Minimizing (2.29) is equivalent to maximizing

$$\sum_{k=1}^{q} \tilde{a}_{kn}^2 - \frac{2\sigma^2}{n} q. \tag{2.30}$$

Similar to (2.6) of Eubank and Hart (1992), we propose using as test statistic the maximizer $\bar{q}$ of

$$\begin{aligned} R(q) &= 0 & &\text{if } q = 0, \quad \text{or} \\ &= \sum_{k=1}^{q} \bar{a}_{kn}^2 - \frac{\bar{\sigma}^2 c \alpha_q}{n} & &\text{if } q = 1, 2, \ldots, n-p, \end{aligned} \tag{2.31}$$

where $\bar{\sigma}^2$ is any consistent estimator of σ^2 and c_α is chosen so that $P(\bar{q} = 0) = 1 - \alpha$ under H_0.

Hence, a natural estimate of $g(t)$ is

$$\bar{g}_n(t) = t + \sum_{k=1}^{\bar{q}} p_k(t) \bar{a}_{kn}. \tag{2.32}$$

When the null hypothesis is not rejected $\bar{g}_n(t)$ is simply the least squares estimate of the null model. Otherwise, $\bar{g}_n(t)$ is a nonparametric estimate of the regression function g. Under some suitable conditions, similar results can be given. The proofs are very tedious and therefore are omitted.

2.4 *Optimal smoothing parameter selection*

To discuss the smoothing parameter h selection, we consider fitting the alternative "model"

$$Y_i = \sum_{k=1}^{h} z_k(x_i^\tau \beta_0)\gamma_{0k} + e_i, \quad i = 1, 2, \ldots, n, \tag{2.33}$$

where $h \in H_n = \{1, 2, \ldots, q_n\}$.

Various methods have been suggested for objectively choosing the smoothing parameter h. See, for example, Li (1986, 1987), Eastwood and Gallant (1991), and Eubank *et al.* (1993). Recently, Eubank and Hart (1992) proposed a new procedure of selecting h, which is connected with a null hypothesis test problem.

In this section, we will use C_p and GCV and present some theoretical results to support their applications.

For applying the results of Li (1987), let $m_i(h) = \sum_{k=1}^{h} z_k(x_i^\tau \beta_0)\gamma_{0k}$. Now (2.33) can be written as the "model"

$$Y = M_n(h) + e, \tag{2.34}$$

where $M_n(h) = (m_1(h), \ldots, m_n(h))^\tau$, $e = (e_1, \ldots, e_n)^\tau$, and $Y = (Y_1, \ldots, Y_n)^\tau$.

If β_0 is the known true parameter-of-interest, then we can define the pseudo-LS estimator $M_n = (h, \beta_0)$ of $M_n(h)$ by

$$\tilde{M}_n = \tilde{M}_n(h, \beta_0) = Z_0(h)(Z_0(h)^\tau Z_0(h))^{-1} Z_0(h)^\tau Y = J_0(h)Y, \tag{2.35}$$

where $Z_0(h) = (Z_h(x_1^\tau \beta_0), \ldots, Z_h(x_n^\tau \beta_0))^\tau$ and $Z_h(\cdot) = (z_1(\cdot), \ldots, z_h(\cdot))^\tau$.

Obviously, a natural estimate for $\tilde{M}_n(h, \beta_0)$ is

$$\hat{M}_n(h) = \tilde{M}_n(h, \hat{\beta}_n(h)) = Z_n(h)(Z_n(h)^\tau Z_n(h))^{+} Z_n(h)^\tau Y = J_n(h)Y, \tag{2.36}$$

where $\hat{\beta}_n(h)$ is defined by (2.3), $Z_n(h) = (Z_h(x_1^\tau \hat{\beta}_n(h)), \ldots, Z_h(x_n^\tau \hat{\beta}_n(h)))^\tau$, and $(\cdot)^+$ denotes the Moore-Penrose inverse.

The following two well-known procedures of selecting h will be studied in detail in this section.

(i) Mallows C_p (Mallows (1973)): select h, denoted by $\hat{h}_p$, that achieves

$$\min_{h \in H_n} n^{-1}\|Y - \hat{M}_n(h)\|^2 + 2\sigma^2 n^{-1} h. \tag{2.37}$$

(ii) Generalized Cross-Validation (Craven and Wahba (1979)): select h, denoted by $\hat{h}_G$, that achieves

$$\min_{h \in H_n} n(n-h)^{-2}\|Y - \hat{M}_n(h)\|^2. \tag{2.38}$$

The primary goal of this section is to demonstrate that under reasonable conditions, these procedures are asymptotically optimal in the sense that

$$\frac{\hat{L}_n(\hat{h})}{\inf_{h \in H_n} \hat{L}_n(h)} \to 1 \quad \text{in probab.,} \tag{2.39}$$

where $\hat{L}_n(h) = n^{-1}\|M_n - \hat{M}_n(h)\|^2$, $\hat{M}_n(h) = J_n(h)Y$, and $M_n = M_n(q)$.

Another way to define those procedures which are asymptotically optimal is in the sense that

$$\frac{\tilde{L}_n(\tilde{h})}{\inf_{h\in H_n}\tilde{L}_n(h)} \to 1 \quad \text{in probab.,} \tag{2.40}$$

where $\tilde{M}_n(h) = J_0(h)Y$, $J_0(h) = Z_0(h)(Z_0(h)^\tau Z_0(h))^{-1}Z_0(h)^\tau$, and $\tilde{L}_n(h) = n^{-1}\|M_n - \tilde{M}_n(h)\|^2$.

In this section, we only consider selecting $\tilde{h}$ in the sense that (2.40) holds. In order to give the results of this subsection, we introduce the following assumption.

ASSUMPTION 2.7.

(i) $Ee_1^8 < \infty$;

(ii) $\inf_{h\in H_n} n\tilde{R}_n(h) \to \infty$, where $\tilde{R}_n(h) = E(\tilde{L}_n(h))$;

(iii) $\inf_{h\in H_n}\tilde{L}_n(h) \to 0$ as $n \to \infty$;

(iv) $\sup_{h\in H_n}\frac{\lambda_{\max}(I_n(h;\beta,\beta_0))}{\tilde{R}_n(h)} \to 0$ uniformly as $\|\beta - \beta_0\| \to 0$ and $n \to \infty$, where $Z_\beta(h) = (Z_h(x_1^\tau\beta),\ldots,Z_h(x_n^\tau\beta))^\tau$, $J_\beta(h) = Z_\beta(h)(Z_\beta(h)^\tau Z_\beta(h))^{-1}Z_\beta(h)^\tau$, and $I_n(h;\beta,\beta_0) = (J_\beta(h) - J_0(h))^\tau(J_\beta(h) - J_0(h))$;

(v) there exists an absolute constant b such that for any $a > 0$

$$\sup_{x\in R} P(x - a \le e_1 \le x + a) \le ba.$$

Remark 2.11. We can show by using the similar reason as the proof of Lemma A.3 in the Appendix that $\lambda_{\max}(I_n(h;\beta,\beta_0)) \to 0$ and we also observe from Assumption 2.7(ii) that $\tilde{R}_n(h)$ is typically of order n^{-1+t} for some $0 < t < 1$. Assumption 2.7(iv) restricts that $\lambda_{\max}(I_n(h;\beta,\beta_0))$ needs to have order higher than n^{-1+t} for some $0 < t < 1$. Assumption 2.7(iv) is easy to verify when both $Z_\beta^\tau Z_\beta$ and $Z_0^\tau Z_0$ are diagonal matrices (see Remark 2.4(i)). Assumption 2.7(v) is satisfied if e_1 has a bounded density.

Now, we give the following results for the case of $g(t) \neq t$.

THEOREM 2.4.

(i) *Assume that the conditions of Theorem* 2.2 *and Assumption* 2.7(i)(ii)(iv) *hold. Then $\hat{h}_p$ defined by* (2.37) *is asymptotically optimal when $g(t) \neq t$.*

(ii) *If σ^2 in* (2.37) *is unknown and σ^2 is replaced by a consistent estimator $\hat{\sigma}^2$, then $\hat{h}_p$ is asymptotically optimal.*

THEOREM 2.5. *Assume that the conditions of Theorem* 2.2 *and that Assumption* 2.7 *hold. Then $\hat{h}_G$ defined by* (2.38) *is asymptotically optimal when $g(t) \neq t$.*

Remark 2.12.

(i) As a matter of fact, there do exist examples, which C_p is consistent and asymptotically optimal while *GCV* is consistent but not asymptotically optimal (see Li (1986)).

(ii) Other useful data-driven techniques for selecting optimal smoothing parameter such as Cross-validation and Stein estimates can also be discussed similarly in this subsection.

Remark 2.13. An important problem is that what happens to the conclusion of Theorem 2.2 when q is data-driven. In fact, it can be shown that the conclusion of Theorem 2.2 still holds when q_n in the definition of $\hat{\theta}_n$ is replaced by $\hat{h}_p$ (or $\hat{h}_G$). Here we omit the details for they are lengthy, which will be given in another paper.

Remark 2.14. The following useful criteria for selecting the optimal smoothing parameter q (nonrandom selection) are based on the order of the average squared error (ASE) and mean average squared error (MASE).

Now, we give the final results of this section.

THEOREM 2.6. *Under the conditions of Theorem* 2.2, *we have*

$$\hat{D}_n(h) = \frac{1}{n}\sum_{i=1}^{n}(\hat{g}_n(t_i) - g(t_i))^2 = \frac{h}{n}\sigma^2 + o_p(n^{-1}h) + o_p(n^{-1}), \tag{2.41}$$

where $\hat{g}_n(\cdot)$ *is defined by* (2.4) *and* $t_i = x_i^\tau \beta_0$.

THEOREM 2.7. *Under the conditions of Theorem* 2.2, *we have*

$$D_n(h) = \frac{h}{n}\sigma^2 + o(n^{-1}h). \tag{2.42}$$

where $D_n(h) = E\hat{D}_n(h) = \frac{1}{n}\sum_{i=1}^{n} E(\hat{g}_n(t_i) - g(t_i))^2$.

Remark 2.15.

(i) If β_0 is unknown true parameter in (2.41) and (2.42), it can be replaced by the above LS estimate $\hat{\beta}_n$ without affecting the results of Theorems 2.6 and 2.7.

(ii) It is clear from the results of Stone (1982) that $h = h_n = O(n^{1/3})$ in (2.41) and (2.42) is optimal, so $\hat{D}_n(h)$ and $D_n(h)$ can achieve the optimal rate of convergence $O(n^{-2/3})$.

The proofs of Theorems 2.4–2.7 will be given in the Appendix.

3. Discussion of a certain related partially nonlinear model

Consider the model given by

$$Y_i = f(x_i, \beta_0) + g(t_i) + e_i, \qquad i = 1, 2, \ldots, \tag{3.1}$$

where $x_i = (x_{i1}, \ldots, x_{ib})^\tau$ $(b \geq 1)$ and $t_i = (t_{i1}, \ldots, t_{id})^\tau$ $(d \geq 1)$ are known and nonrandom design points, $\beta_0 = (\beta_{01}, \ldots, \beta_{0p})^\tau$ $(p \geq 1)$ is an unknown parameter vector over Θ_1 (a compact subset of R^p), $f_i(\beta_0) = f(x_i, \beta_0)$ are some known function, $g(\cdot)$ is an unknown function over $T \subset R$, and the e_i are *i.i.d.* random errors with $Ee_i = 0$ and $Ee_i^2 = \sigma^2 < \infty$.

The model defined in (3.1) includes some classical models: such as nonlinear models $Y_i = f(x_i, \beta_0) + e_i$ ($g = 0$ in (3.1)); nonparametric regression models $Y_i = g(t_i) + e_i$ ($f = 0$ in (3.1)); and semiparametric regression models $Y_i = x_i^\tau \beta_0 + g(t_i) + e_i$ ($f(x_i, \beta_0) = x_i^\tau \beta_0$ in (3.1)). As can be seen, there is a bewildering array of interesting special cases of the partially nonlinear models. Some of them have been thoroughly studied and specific estimation procedures for them have been developed. See, for example, Gallant (1987), Seber and Wild (1989), Hardle (1990), Hastie and Tibshirani (1990), Wahba (1990), Robinson (1988), and Gao *et al.* (1994).

In this section, we will apply the above finite series approximation technique to $g(\cdot)$ to construct a nonlinear least squares estimator of β_0. Meanwhile, the asymptotic normality of a test statistic for testing $H_0 : g(\cdot) = 0$ is also obtained. Based on $g(\cdot)$ satisfying Assumption 2.1, the LS estimator $\bar{\theta}_n = ((\bar{\beta}_n)^\tau, (\bar{\gamma}_n)^\tau)^\tau$ of $\theta_0 = (\beta_0^\tau, \gamma_0^\tau)^\tau$ can be defined as the solution of

$$S_n(\theta_0) = \sum_{i=1}^{n} (Y_i - m_i(\theta_0))^2 = \min!, \tag{3.2}$$

where $m_i(\theta_0) = f_i(\beta_0) + Z(t_i)^\tau \gamma_0$, and $Z(\cdot) = (z_1(\cdot), \ldots, z_q(\cdot))^\tau$ with $z_k(\cdot)$ defined over $T \subset R$ and γ_0 is defined as in Assumption 2.1.

For constructing the main results of this section, we introduce the following assumptions. Most of them are similar to Assumptions 2.1 to 2.4.

Let $f_i'(\beta) = (\frac{\partial f_i(\beta)}{\partial \beta_j})_{1 \le j \le p}$, $f_i''(\beta) = (\frac{\partial^2 f_i(\beta)}{\partial \beta_j \partial \beta_k})_{1 \le j,k \le p}$, and $u_j(x_i, \beta_0) = |\frac{\partial f_i(\beta)}{\partial \beta_j}|_{\beta = \beta_0}$.

ASSUMPTION 3.1. Assumption 2.1 holds with $\{x_i^\tau \beta_0\}$ replaced by $\{t_i\}$.

ASSUMPTION 3.2.

(i) $f_i'(\beta)$ and $f_i''(\beta)$ exist for β near β_0;

(ii) there exist some bounded functions $s_j(\cdot)$ over T such that for all $1 \le i \le n$ and $1 \le j \le p$

$$u_j(x_i, \beta_0) = s_j(t_i) + \epsilon_{ij}, \tag{3.3}$$

where ϵ_{ij} are real sequences satisfying

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \epsilon_i \epsilon_i^\tau = B_1 \tag{3.4}$$

and

$$\limsup_{n \to \infty} \frac{1}{n^{1/2} \log n} \max_{1 \le m \le n} \| \sum_{i=1}^{m} \epsilon_{j_i} \| < \infty \tag{3.5}$$

for any permutation $(j_1, j_2, \ldots, j_n)$ of the integers $(1, 2, \ldots, n)$, where $\epsilon_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^\tau$ and B_1 is a positive definite matrix with order $p \times p$;

(iii) Assumption 2.2 (ii) holds with $\{h_j(x_i^\tau \beta_0)\}$ replaced by $\{s_j(t_i)\}$.

ASSUMPTION 3.3. $m_i(\theta)$ satisfies Assumption 2.3.

ASSUMPTION 3.4.

(i) Z is of full column rank q_n for n large enough, and $\liminf_{n\to\infty} \cdot \lambda_{\min}(Z^\tau Z/n) > 0$, where $Z = (Z(t_1), \ldots, Z(t_n))^\tau$;

(ii) Assumption 2.4(ii) holds with P_0 replaced by $P = Z(Z^\tau Z)^{-1} Z^\tau$;

(iii) Assumption 2.4(iii) holds with $\{d_i x_{ij}\}$ replaced by $\{u_j(x_i, \beta_0)\}$;

(iv) Assumption 2.4(iv) holds with $\{z_k(x_i^\tau \beta_0)\}$ replaced by $\{z_k(t_i)\}$;

(v) $\max_{1\le i\le n} \|f_i(\beta) - f_i(\beta_0)\| \to 0$ uniformly as $\|\beta - \beta_0\| \to 0$ and $n \to \infty$;

(vi) there exists an absolute constant $\tilde{d} > 0$ such that for all (j,k)

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sup_{\|\beta-\beta_0\|\le \tilde{d}} \left(\frac{\partial^2 f_i(\beta)}{\partial\beta_j \partial\beta_k} \right)^2 < \infty;$$

(vii) if, for a pair (j,k)

$$\sum_{i=1}^{\infty} \sup_{\|\beta-\beta_0\|\le \tilde{d}} \left(\frac{\partial^2 f_i(\beta)}{\partial\beta_j \partial\beta_k} \right)^2 = \infty,$$

then there exists an M independent of i such that for all i

$$\sup_{s\ne t, s,t\in\Theta_{\tilde{d}}} \frac{\left| \frac{\partial^2 f_i(s)}{\partial s_j \partial s_k} - \frac{\partial^2 f_i(t)}{\partial t_j \partial t_k} \right|}{\|s-t\|} \le \sup_{\beta\in\Theta_{\tilde{d}}} \left| \frac{\partial^2 f_i(\beta)}{\partial\beta_j \partial\beta_k} \right|,$$

where $\Theta_h = \{\beta \in \Theta_1, \|\beta - \beta_0\| \le h\}$.

ASSUMPTION 3.5. Assumption 2.5 holds with P_0 replaced by P and Assumption 2.5(iii) holds with $\{v_{ij}\}$ replaced by $\{u_j(x_i, \beta_0)\}$.

Remark 3.1. The discussions of Assumptions 3.1 through 3.5 follow similarly from Remarks 2.1 through 2.9 given in Section 2.

Now, we state the main results of this section.

THEOREM 3.1. *Assume that Assumptions* 3.1 *and* 3.3 *hold. Let* $M_n(\theta, \theta_0) = \sum_{i=1}^n (m_i(\theta) - m_i(\theta_0))^2 \to \infty$ *for all* $\theta \ne \theta_0$ *and* $n \to \infty$. *Then as* $n \to \infty$

$$\bar{\theta}_n - \theta_0 \to 0 \quad a.s. \tag{3.6}$$

THEOREM 3.2. *Assume that Assumptions* 3.1, 3.2 *and* 3.4 *hold. Let* θ_0 *be in the interior of* Θ *and* $\bar{\theta}_n$ *be a strongly consistent least squares estimator of* θ_0. *Then as* $n \to \infty$

$$\sqrt{n}(\bar{\beta}_n - \beta_0) \to_D N(0, \sigma^2 B_1^{-1}). \tag{3.7}$$

Remark 3.2. Theorem 3.2 only gives the asymptotic normality of the LS estimator of β_0 for the case where p is finite integer. The case of $p = p_n \to \infty$ as $n \to \infty$ can also be discussed. For this case, Assumptions 3.4(vi) and (vii) should be replaced by those which are similar to Assumptions 2.4(vii)–(viii) and the proof is similar to that of Theorem 3.3 below.

Remark 3.3. As in (2.4) and (2.11), we define the estimates of $g(\cdot)$ and $\sigma^2 = Ee_1^2$ by

$$\bar{g}_n(\cdot) = Z(\cdot)^\tau \bar{\gamma}_n \tag{3.8}$$

and

$$\bar{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(Y_i - f(x_i, \bar{\beta}_n) - \bar{g}_n(t_i))^2. \tag{3.9}$$

Also, we may obtain easily some asymptotic properties of $\bar{\sigma}_n^2$ and $\bar{g}_n(\cdot)$.

Another important problem considered in this section is testing for a nonlinear model or partially nonlinear model, that is, testing $H_0 : g(\cdot) = 0$ or $H_0' : g(\cdot)$ is an unknown smooth function. Based on Assumption 3.1, the null hypothesis $H_0 : g(\cdot) = 0$ is equivalent to $H_0'' : \gamma_0 = 0$. This suggests using a test statistic of the form

$$F_{3n} = (2q_n)^{-1/2}\sigma^{-2}(\bar{\gamma}_n^\tau(Z^\tau Z)\bar{\gamma}_n - q\sigma^2), \tag{3.10}$$

where $\sigma^2 = Ee_1^2$ is known and $\bar{\gamma}_n$ is defined by (3.2). Also, if σ^2 is unknown, it can be replaced by $\bar{\sigma}_n^2$ without affecting the conclusion of Theorem 3.3 below.

THEOREM 3.3. *Assume that the conditions of Theorem 3.2 and that Assumption 3.5 hold. Let $Ee_1^4 < \infty$. Then under the null hypothesis $H_0 : g(\cdot) = 0$*

$$F_{3n} \to_D N(0, 1), \quad \text{as} \quad n \to \infty. \tag{3.11}$$

Furthermore, we have $F_{3n} \to_p \infty$ under $H_1 : g \neq 0$.

Remark 3.4. The discussions similar to Remarks 2.1 through 2.17 are omitted here. The proofs of Theorems 3.1–3.3 are given in the Appendix.

4. Computational aspects

In this subsection, we only give some computational procedures in outline for the nonlinear LS estimators $\hat{\beta}_n$ and $\hat{\gamma}_n$. Consider the model given by (let $g(t) = e^t$ and $p = 1$ in (1.1))

$$Y_i = e^{x_i\beta_0} + e_i, \quad i = 1, 2, \ldots, n, \tag{4.1}$$

where $x_i = 2\pi(i-1)/n$ and e_i are *i.i.d.* $N(0, 10)$ random variables.

The approximation function is the family of trigonometric functions $\{\cos(jt), \sin(jt); j = 1, 2, \ldots, k_n = [n^{1/5}]\}$ used by Eastwood and Gallant (1991), where the $[x]$ denotes the largest integer part of x satisfying $[x] \le x$.

Now, the approximation function of $g(t)$ is defined by

$$\tilde{g}_n(t) = \sum_{j=1}^{q} z_j(t)\gamma_j = \sum_{j=1}^{k}(\cos(jt)\gamma_{j1} + \sin(jt)\gamma_{j2}), \tag{4.2}$$

where $q_n = 2k_n = 2[n^{1/5}]$ and $\gamma_0 = (\gamma_1, \ldots, \gamma_q)^\tau = (\gamma_{11}, \gamma_{12}, \ldots, \gamma_{k1}, \gamma_{k2})^\tau$.

Then, it is easy to see that Assumptions 2.3 and 2.4 hold.

Next, by (2.3) we define the nonlinear LS estimator $\hat{\theta}_n = (\hat{\beta}_n^\tau, \hat{\gamma}_n^\tau)^\tau$ of $\theta_0 = (\beta_0^\tau, \gamma_0^\tau)^\tau$ by

$$S(\beta, \gamma) = \sum_{i=1}^{n}[Y_i - \sum_{j=1}^{k}(\gamma_{j1}\cos(jx_i\beta) + \gamma_{j2}\sin(jx_i\beta))]^2 = \min! \tag{4.3}$$

In the following, we give some iterative algorithms in outline for the above nonlinear LS estimator.

For an initial value $\beta^{(0)}$, we can obtain from Eastwood and Gallant (1991) that the LS estimators of $(\gamma_{j1}, \gamma_{j2})$ are

$$\gamma_{j1}^{(1)} = \gamma_{j1}(\beta^{(0)}) = \frac{2}{n}\sum_{i=1}^{n} Y_i \cos(jx_i\beta^{(0)}), \tag{4.4}$$

$$\gamma_{j2}^{(1)} = \gamma_{j2}(\beta^{(0)}) = \frac{2}{n}\sum_{i=1}^{n} Y_i \sin(jx_i\beta^{(0)}). \tag{4.5}$$

Now, we can define an estimator $\beta^{(1)}$ of β_0 by minimizing

$$S(\beta_0, \gamma^{(1)}) = \sum_{i=1}^{n}\{Y_i - \sum_{j=1}^{k}[\gamma_{j1}^{(1)}\cos(jx_i\beta_0) + \gamma_{j2}^{(1)}\sin(jx_i\beta_0)]\}^2. \tag{4.6}$$

Iteratively, based on $(\beta^{(p)}, p = 1, 2, \ldots)$ we can find $(\gamma_{j1}^{(p+1)}, \gamma_{j2}^{(p+1)})$ which minimizes $S(\beta^{(p)}, \gamma)$.

Ruhe and Wedin (1980) proved that the above iterative algorithms are convergent and have asymptotical convergence rates. The details can be obtained from Ruhe and Wedin (1980).

The iterative procedure below is known as the Gauss-Newton method for nonlinear least squares.

For an initial vector $(\tilde{\beta}^{(1)}, \tilde{\gamma}^{(1)})$, the following iterative procedure is easy to use in practice,

$$\begin{aligned} \tilde{\beta}^{(p+1)} &= \tilde{\beta}^{(p)} + (V(p)^\tau(I - P(p))V(p))^{-1}V(p)^\tau(I - P(p))e, \\ \tilde{\gamma}^{(p+1)} &= \tilde{\gamma}^{(p)} + n^{-1}\operatorname{diag}(1, 2, \ldots, 2)W(\tilde{\theta}^{(p)})^\tau \\ &\quad \cdot (I - V(p)(V(p)^\tau(I - P(p))V(p))^{-1}V(p)^\tau(I - P(p)))e, \end{aligned} \tag{4.7}$$

where $V(p) = (v_1(\tilde{\theta}^{(p)}), \ldots, v_n(\tilde{\theta}^{(p)}))^\tau$, $v_i(\tilde{\theta}^{(p)}) = Z'(x_i\tilde{\beta}^{(p)})^\tau\tilde{\gamma}^{(p)} \cdot x_i$, $P(p) = n^{-1} \cdot W(\tilde{\theta}^{(p)}) \operatorname{diag}(1, 2, \ldots, 2) W(\tilde{\theta}^{(p)})^\tau$, $W(\tilde{\theta}^{(p)}) = (w_1(\tilde{\theta}^{(p)}), \ldots, w_n(\tilde{\theta}^{(p)}))^\tau$, and $w_i(\tilde{\theta}^{(p)}) = Z(x_i\tilde{\beta}^{(p)})$.

At some step p_0, if $\|\tilde{\theta}^{(p_0+1)} - \tilde{\theta}^{(p_0)}\| < 0.001$, then the above iterative procedure is terminated and $\tilde{\theta}^{(p_0+1)} = (\tilde{\beta}^{(p_0+1)^\tau}, \tilde{\gamma}^{(p_0+1)^\tau})^\tau$ is taken to be the desired estimate of $\hat{\theta}_n = (\hat{\beta}_n^\tau, \hat{\gamma}_n^\tau)^\tau$ defined by (4.3).

Thus, based on the above algorithms, we can discuss the finite sample properties of some estimates. The measures for the estimates $\hat{\beta}_n$ and $\hat{\sigma}_n^2$ are taken to be

$$\|\hat{\beta}_n - \beta_0\| \quad \text{and} \quad |\hat{\sigma}_n^2 - \sigma_0^2|. \tag{4.8}$$

The simulation results for the cases where $n = 100$ and $n = 200$ can be given based on (4.1) through (4.7). The details obtained using the Splus commands (see Chapter 10 of Chambers and Hastie (1992)) are given in the following Tables 1 and 2.

Table 1. The Estimates and the Biases for $n = 100$.

β_0	$\hat{\beta}_n$	$\lvert\hat{\beta}_n - \beta_0\rvert$	$\hat{\sigma}_n^2$	$\lvert\hat{\sigma}_n^2 - 10\rvert$
0.1	0.0892	0.0108	10.0232	0.0232
0.2	0.2007	0.0007	10.0093	0.0093
0.3	0.2907	0.0093	10.0103	0.0103
0.4	0.4017	0.0017	10.0035	0.0035
0.5	0.5009	0.0009	10.0016	0.0016
0.6	0.6008	0.0008	9.9989	0.0011
0.7	0.6991	0.0009	9.9990	0.0010
0.8	0.7992	0.0008	10.0025	0.0025
0.9	0.9011	0.0011	10.0016	0.0016
1.0	0.9984	0.0016	10.0045	0.0045

Table 2. The Estimates and the Biases for $n = 200$.

β_0	$\hat{\beta}_n$	$\lvert\hat{\beta}_n - \beta_0\rvert$	$\hat{\sigma}_n^2$	$\lvert\hat{\sigma}_n^2 - 10\rvert$
0.1	0.0896	0.0104	10.0103	0.0103
0.2	0.2004	0.0004	10.0051	0.0051
0.3	0.2909	0.0091	10.0062	0.0062
0.4	0.4016	0.0016	10.0026	0.0026
0.5	0.5008	0.0008	10.0011	0.0011
0.6	0.6007	0.0007	9.9990	0.0010
0.7	0.6993	0.0007	10.0009	0.0009
0.8	0.7994	0.0006	10.0024	0.0024
0.9	0.9009	0.0009	9.9987	0.0013
1.0	0.9989	0.0011	10.0032	0.0032

Remark 4.1. The small sample results above support the new estimators given in (2.3), (2.4), and (2.11) and the above Theorems 2.1 and 2.2 above.

Acknowledgements

The first author wishes to give his thanks to Professor George Seber for providing him with some constructive suggestions and references. Thanks also go to the Associate Editor and the referees for some constructive comments which have greatly improved this paper.

Appendix

For convenience and simplicity, let C $(0 < C < \infty)$ denote some constant which may have different values at each appearance throughout this paper. In the following, we only prove Theorems 2.1 through 2.7 and the proofs of Theorems 3.1 through 3.3 follow similarly from those of Theorems 2.1 to 2.3.

In order to complete the proofs of Theorems 2.1–2.7, we need to introduce the following lemmas and their proofs can be found in Gao (1995).

LEMMA A.1. *Assume that Assumptions* 2.2 *and* 2.4(ii)(iii)(iv) *hold. Then*

$$\lim_{n\to\infty} \frac{1}{n} V(\theta_0)^\tau (I - P_0) V(\theta_0) = B, \tag{A.1}$$

where $V(\theta) = (v_1(\theta), \ldots, v_n(\theta))^\tau$, $v_i(\theta) = Z'(x_i^\tau \beta)^\tau \gamma x_i$, $v_{ij} = Z'(x_i^\tau \beta)^\tau \gamma x_{ij}$, $1 \le i \le n$, $1 \le j \le p$, $P_0 = P_{\beta_0} = W(\theta_0)(W(\theta_0)^\tau W(\theta_0))^{-1} W(\theta_0)^\tau$, $W(\theta) = (w_1(\theta), \ldots, w_n(\theta))^\tau$, $w_i(\theta) = Z(x_i^\tau \beta)$, $1 \le i \le n$, *and* B *is defined as in Theorem* 2.2.

LEMMA A.2. *Under the conditions of Theorem* 2.2, *we have*

$$\limsup_{n\to\infty} \lambda_{\max}(n A_{\theta_0}^{-1}) < \infty, \tag{A.2}$$

where

$$\begin{aligned}
A_{\theta_0}^{-1} &= (A_{\theta_0}^{ij})_{1\le i,j\le 2}, \\
A_{\theta_0}^{11} &= (V(\theta_0)^\tau (I - P_0) V(\theta_0))^{-1}, \\
A_{\theta_0}^{12} &= -A_{\theta_0}^{11} V(\theta_0)^\tau W(\theta_0)(W(\theta_0)^\tau W(\theta_0))^{-1}, \\
A_{\theta_0}^{21} &= (W(\theta_0)^\tau W(\theta_0))^{-1} W(\theta_0)^\tau V(\theta_0) A_{\theta_0}^{11},
\end{aligned}$$

and

$$\begin{aligned}
A_{\theta_0}^{22} &= (W(\theta_0)^\tau W(\theta_0))^{-1} + (W(\theta_0)^\tau W(\theta_0))^{-1} \\
&\quad \cdot W(\theta_0)^\tau V(\theta_0) A_{\theta_0}^{11} V(\theta_0)^\tau W(\theta_0)(W(\theta_0)^\tau W(\theta_0))^{-1}.
\end{aligned}$$

LEMMA A.3. *Assume that the conditions of Theorem* 2.2 *hold. Let* $A_\theta = A(\theta)^\tau A(\theta)$ *and* $A(\theta) = (V(\theta), W(\theta))^\tau$. *Then* $\lambda_{\max}(A_\theta A_{\theta_0}^{-1} - I) \to 0$ *uniformly as* $\|\theta - \theta_0\| \to 0$ *and* $n \to \infty$.

A.1 *Proof of Theorems* 2.1 *and* 2.2

The proof of Theorem 2.1 can also be found in Gao (1995). The proof of Theorem 2.2 is very lengthy and here we only give the proof in outline.

First, note that the following

$$S_n'(\theta) = 2\sum_{i=1}^{n}(F_i(\theta) - g(x_i^\tau \beta_0) - e_i)F_i'(\theta) \tag{A.3}$$

and

$$\begin{aligned} S_n''(\theta) = 2\Bigg(&\sum_{i=1}^{n} F_i'(\theta)F_i'(\theta)^\tau - \sum_{i=1}^{n} F_i''(\theta)e_i \\ &+ \sum_{i=1}^{n} d_i(\theta)F_i''(\theta) - \sum_{i=1}^{n} r_i(\theta_0)F_i''(\theta)\Bigg), \end{aligned} \tag{A.4}$$

where $F_i'(\theta) = (\frac{\partial}{\partial\theta_j}F_i(\theta))$, $F_i''(\theta) = (\frac{\partial^2}{\partial\theta_j\partial\theta_k}F_i(\theta))$, and $d_i(\theta) = F_i(\theta) - F_i(\theta_0)$.

From the mean-value theorem, there exists a $p_n \in [0,1]$ such that

$$\begin{aligned} S_n'(\theta_0) &= -2\sum_{i=1}^{n} F_i'(\theta_o)e_i - 2\sum_{i=1}^{n} F_i'(\theta_0)r_i(\theta_0) \\ &= S_n(\hat\theta_n) + S_n''(\theta_n^*)(\theta_0 - \hat\theta_n), \end{aligned} \tag{A.5}$$

where $\theta_n^* = (1-p_n)\theta_0 + p_n\hat\theta_n$ is measurable from Lemma 3 of Jennrich (1969). Since $\hat\theta_n$ is in the interior of Θ eventually, $S_n'(\hat\theta_n) = 0$. Now, (A.5) can be rewritten as

$$\begin{aligned} \frac{1}{2}S_n'(\theta_0) &= -\sum_{i=1}^{n} F_i'(\theta_0)e_i - \sum_{i=1}^{n} F_i'(\theta_0)r_i(\theta_0) \\ &= K_n\left(\sum_{i=1}^{n} F_i'(\theta_0)F_i'(\theta_0)^\tau\right)(\theta_0 - \hat\theta_n), \end{aligned} \tag{A.6}$$

where

$$\begin{aligned} K_n = &\sum_{i=1}^{n} F_i'(\theta_n^*)F_i(\theta_n^*)^\tau \left(\sum_{i=1}^{n} F_i'(\theta_0)F_i'(\theta_0)^\tau\right)^{-1} \\ &- \sum_{i=1}^{n} F_i''(\theta_n^*)e_i\left(\sum_{i=1}^{n} F_i'(\theta_o)F_i'(\theta_0)^\tau\right)^{-1} \\ &+ \sum_{i=1}^{n}(d_i(\theta_n^*) - r_i(\theta_0))F_i''(\theta_n^*)\left(\sum_{i=1}^{n} F_i'(\theta_0)F_i'(\theta_0)^\tau\right)^{-1} \\ = &A_{\theta_n^*}A_{\theta_0}^{-1} - \sum_{i=1}^{n} F_i''(\theta_n^*)e_i \cdot A_{\theta_0}^{-1} + \sum_{i=1}^{n}(d_i(\theta_n^*) - r_i(\theta_0))F_i''(\theta_n^*)\cdot A_{\theta_0}^{-1} \\ = &A_{\theta_n^*}A_{\theta_0}^{-1} \quad B(\theta_n^*)A_{\theta_0}^{-1} \mid C(\theta_n^*)A_{\theta_0}^{-1}, \end{aligned} \tag{A.7}$$

where A_θ and $A_{\theta_0}^{-1}$ are defined as before.

Note that (A.7) and Lemmas A.2 and A.3, in order to show that $\lambda_{\max}(K_n - I_{p+q}) \to 0$ a.s., it suffices to show that

$$n^{-1}\lambda_{\max}(B(\theta_n^*)) \to 0 \quad \text{a.s.} \tag{A.8}$$

and

$$n^{-1}\lambda_{\max}(C(\theta_n^*)) \to 0 \quad \text{a.s.} \tag{A.9}$$

uniformly as $\|\theta_n^* - \theta_0\| \to 0$ a.s. and $n \to \infty$.

Now, observe that $F_i''(\theta) = (a_{ijk}(\theta))_{1\le j,k\le p+q}$ defined by Assumption 2.4(viii). For any nonzero vector $c = (c_1, \ldots, c_p; c_{p+1}, \ldots, c_{p+q})^\tau$ satisfying $c^\tau c = 1$, we have

$$c^\tau B(\theta)c = \sum_{i=1}^n (c^\tau F_i''(\theta)c)e_i = \sum_{i=1}^n A_i(\theta)e_i. \tag{A.10}$$

Using the same reason as the proof of (4.7) of Wu (1981) and applying Corollary A of Wu (1981) and Assumptions 2.4(vii)(viii), we obtain that

$$n^{-1}\sup_{\|c\|=1}(c^\tau B(\theta_n^*)c) = o(1) \quad \text{a.s.} \tag{A.11}$$

uniformly as $\|\theta_n^* - \theta_0\| \to 0$ a.s. and $n \to \infty$.

Next, applying the Cauchy-Schwarz inequality and using Assumptions 2.1, 2.4(vi), and 2.4(vii) we get that

$$n^{-1}\sup_{\|c\|=1}(c^\tau C(\theta_n^*)c) = o(1) \quad \text{a.s.} \tag{A.12}$$

uniformly as $\|\theta_n^* - \theta_0\| \to 0$ a.s. and $n \to \infty$.

On the other hand, by Assumptions 2.1, 2.4(iii), and 2.4(iv) we have that for all j

$$\begin{aligned}\left|\sum_{i=1}^n v_{ij}(\theta_0)r_i(\theta_0)\right|^2 &= \left|\sum_{i=1}^n k_i(\theta_0)x_{ij}r_i(\theta_0)\right|^2 \\ &\le \left(\sum_{i=1}^n k_i^2(\theta_0)x_{ij}^2\right)\cdot\sum_{i=1}^n r_i^2(\theta_0) = O(n)\cdot o(1) = o(n)\end{aligned} \tag{A.13}$$

and

$$\sum_{k=1}^q\left|\sum_{i=1}^n z_k(x_i^\tau\beta_0)r_i(\theta_0)\right|^2 \le \sum_{k=1}^q\left(\sum_{i=1}^n z_k(x_i^\tau\beta_0)^2\right)\cdot\sum_{i=1}^n r_i^2(\theta_0) = o(n). \tag{A.14}$$

Thus

$$\left\|\sum_{i=1}^n F_i'(\theta_0)r_i(\theta_0)\right\| = o(n^{1/2}), \tag{A.15}$$

where $F_i'(\theta_0) = (v_{i1}(\theta_0), \ldots, v_{ip}(\theta_0); z_1(x_i^\tau\beta_0), \ldots, z_q(x_i^\tau\beta_0))^\tau$.

Therefore, by (A.6)–(A.15) we get

$$\left(\sum_{i=1}^{n} F_i'(\theta_0)F_i'(\theta_0)^\tau\right)(\hat{\theta}_n - \theta_0) = \sum_{i=1}^{n} F_i'(\theta_0)e_i + o_p(n^{1/2}). \tag{A.16}$$

Thus, by Lemma A.1 and Lemma 3 of Wu (1981) we get that

$$\begin{aligned}\sqrt{n}(\hat{\beta}_n - \beta_0) &= \sqrt{n}(A_{\theta_0}^{11}, A_{\theta_0}^{12})\begin{pmatrix} V(\theta_0)^\tau e \\ W(\theta_0)^\tau e\end{pmatrix} + o_p(1) \\ &= \sqrt{n}(V(\theta_0)^\tau(I - P_0)V(\theta_0))^{-1}V(\theta_0)^\tau(I - P_0)e + o_p(1) \\ &\to {}_D N(0, B^{-1}\sigma^2),\end{aligned} \tag{A.17}$$

where $e = (e_1, \ldots, e_n)^\tau$. This completes the proof of Theorem 2.2.

Remark A.1. The above (A.17) establishes the asymptotic normality of $\hat{\beta}_n$ for the case where p is finite integer. As $p = p_n \to \infty$ as $n \to \infty$, we can study the asymptotic distribution of a functional of $\hat{\beta}_n$,

$$F(\hat{\beta}_n) = (2p)^{-1/2}\sigma^{-2}[\hat{\beta}_n^\tau(V(\hat{\beta}_n)^\tau(I - P(\hat{\beta}_n))V(\hat{\beta}_n))\hat{\beta}_n - \sigma^2 p], \tag{A.18}$$

where $P(\hat{\beta}_n) = Z_{\hat{\beta}_n}(Z_{\hat{\beta}_n}^\tau Z_{\hat{\beta}_n})^+ Z_{\hat{\beta}_n}^\tau$. The proof is similar to that of Theorem 2.3 below.

A.2 *Proof of Theorem* 2.3

It follows from (A.16) that

$$\begin{aligned}\hat{\gamma}_n - \gamma_0 &= (A_{\theta_0}^{21}, A_{\theta_0}^{22})\begin{pmatrix} V(\theta_0)^\tau e \\ W(\theta_0)^\tau e\end{pmatrix} + o_p(n^{-1/2}) \\ &= (W(\theta_0)^\tau W(\theta_0))^{-1}W(\theta_0)^\tau(I - V(\theta_0)(\hat{V}(\theta_0)^\tau\hat{V}(\theta_0))^{-1}\hat{V}(\theta_0)^\tau)e \\ &\quad + o_p(n^{-1/2}) \\ &= (W(\theta_0)^\tau W(\theta_0))^{-1/2}(A_{1n} + A_{2n}) + o_p(n^{-1/2}),\end{aligned} \tag{A.19}$$

where $\hat{V}(\theta_0) = (I - P_0)V(\theta_0)$, $A_{1n} = (W(\theta_0)^\tau W(\theta_0))^{-1/2}W(\theta_0)^\tau e$, and $A_{2n} = -(W(\theta_0)^\tau W(\theta_0))^{-1/2}W(\theta_0)^\tau V(\theta_0)(\hat{V}(\theta_0)^\tau\hat{V}(\theta_0))^{-1}\hat{V}(\theta_0)^\tau e$.

By using the conditions of Theorem 2.3 and applying Theorem 5.1 of De Jong (1987), we can obtain the proof of Theorem 2.3. See Gao (1995) for more details.

A.3 *The proof of Theorems* 2.4–2.7

Proofs of Theorems 2.4–2.7 are omitted here for they are very lengthy. See Gao (1995) for details.

REFERENCES

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

Azzalini, A. and Bowman, A. (1993) On the use of nonparametric regression for checking linear relationships, *J. Roy. Statist. Soc. Ser. B*, **55**, 549–557.

Chambers, J. M. and Hastie, T. J. (1992). *Statistical Models in S*, Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, California.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation, *Numer. Math.*, **31**, 377–403.

De Jong, P. (1987). A central limit theorem for generalized quadratic forms, *Probab. Theory Related Fields*, **75**, 261–277.

Eastwood, B. and Gallant, R. (1991). Adaptive truncation rules for seminonparametric estimates that achieve asymptotic normality, *Econometric Theory*, **7**, 307–340.

Eubank, R. and Hart, D. (1992). Testing goodness-of-fit in regression via order selection criterion, *Ann. Statist.*, **20**, 1412–1425.

Eubank, R., Hart, D. and Lariccia, V. (1993). Testing goodness-of-fit via nonparametric function estimation techniques, *Comm. Statist. Theory Methods*, **22**, 3327–3354.

Eubank, R. and Speckman, P. (1990). Curve fitting by polynomial-trigonometric regression, *Biometrika*, **77**, 1–9.

Gallant, R. (1987). *Nonlinear Statistical Models*, Wiley, New York.

Gao, J. T. (1992). *A large sample theory in semiparametric regression models*, Ph.D. thesis, Graduate School at the University of Science and Technology of China, Anhui, P. R. China.

Gao, J. T. (1995). Statistical inference in semiparametric single-index regression models, Tech. Report, No. 9510, Department of Statistics, the University of Auckland, Auckland, New Zealand.

Gao, J. T. (1996*a*). Parametric and nonparametric tests in semiparametric regression models, *Comm. Statist. Theory Methods* **26**, 787–800.

Gao, J. T. (1996*b*). Semiparametric approach to nonlinear regression models, Tech. Report, No. 9612, School of Mathematics, the Queensland University of Technology, Australia.

Gao, J. T., Hong, S. Y., Liang, H. and Shi, P. D. (1994). Survey of Chinese work on semiparametric regression models, *Chin. J. Appl. Probab. Statist.*, **10**, 96–104.

Gozalo, P. L. (1993). A consistent model specification nonparametric estimates of regression models, *Econometric Theory*, **9**, 451–477.

Hamilton, D. C. (1986). Confidence regions for parameter subsets in nonlinear regression, *Biometrika*, **73**, 57–64.

Han, A. K. (1987). Nonparametric analysis of a generalized regression model: the maximum rank correlation estimator, *J. Econometrics*, **35**, 303–316.

Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, Boston.

Härdle, W. and Stoker, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives, *J. Amer. Statist. Assoc.*, **84**, 986–995.

Härdle, W., Hall, P. and Ichimura, H. (1993). Optimal smoothing in single-index models, *Ann. Statist.*, **21**, 157–178.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*, Chapman and Hall.

Ichimura, H. (1987). *Estimation of Single-Index Models*, Ph.D. thesis, Department of Economics, MIT, Massachusetts.

Jennrich, R. I. (1969). Asymptotic properties of nonlinear least squares estimators, *Ann. Math. Statist.*, **40**, 633–643.

Li, K. C. (1986). Asymptotic optimality of C_p and generalized cross-validation in ridge regression with application to spline smoothing, *Ann. Statist.*, **14**, 1101–1112.

Li, K. C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: discrete index set, *Ann. Statist.*, **15**, 958–975.

Li, K. C. (1989). Regression analysis under link violation, *Ann. Statist.*, **17**, 1009–1052.

Mallows, C. L. (1973). Some comments on C_p, *Technometrics*, **15**, 661–675.

Munson, P. J. and Jernigan, R. W. (1989). A cubic spline extension of the Durbin-Watson test, *Biometrika*, **76**, 39–47.

Nurnberger, G. (1989). *Approximation by Spline Functions*, Springer, New York.

Pollard, D. (1984). *Convergence of Stochastic Processes*, Springer, New York.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*, CBMS/NSF Regional Conference Series in Applied Mathematics, SIAM/IMS, Philadelphia.

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

Rice, J. (1986). Convergence rates for partially splined models, *Statist. Probab. Lett.*, **4**, 203–208.

Robinson, P. (1988). Root-N-consistent semiparametric regression, *Econometrica*, **56**, 931–954.

Ruhe, A. and Wedin, P. A. (1980). Algorithms for separable nonlinear least squares problems, *SIAM Review*, **22**, 318–337.

Ruud, P. A. (1986). Consistent estimation of limited dependent variable models despite misspecification of distribution, *J. Econometrics*, **32**, 157–187.

Samarov, A. (1993). Exploring regression structure using nonparametric functional estimation, *J. Amer. Statist. Assoc.*, **423**, 836–847.

Seber, G. A. F. and Wild, C. J. (1989). *Nonlinear Regression*, Wiley, New York.

Speckman, P. (1988). Kernel smoothing in partial linear models, *J. Roy. Statist. Soc. Ser. B*, **50**, 413–436.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients, *Econometrica*, **54**, 1461–1481.

Stoker, T. M. (1993). Smoothing bias in density derivative estimation, *J. Amer. Statist. Assoc.*, **423**, 855–863.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression, *Ann. Statist.*, **10**, 1040–1053.

Wahba, G. (1990). *Spline Models for Observational Data*, SIAM, Philadelphia.

Whang, Y. J. and Andrews, D. W. K. (1993). Tests of specification for parametric and semiparametric models, *J. Econometrics*, **57**, 277–318.

Wu, C. F. (1981). Asymptotic theory of nonlinear least squares estimate, *Ann. Statist.*, **9**, 501–513.