Statistica Sinica **21** (2011), 749-778

TESTING IN NONPARAMETRIC VARYING COEFFICIENT ADDITIVE MODELS

Byeong U. Park[1], Jun H. Hwang[2] and Min S. Park[1]

[1]*Seoul National University and* [2]*Hyundai Capital*

Abstract: We consider the problem of testing for a general parametric form against a nonparametric alternative for a coefficient function in a varying coefficient multivariate regression model. We propose a test statistic and derive its asymptotic null and alternative distributions. We analyze the asymptotic power of the test in shrinking neighborhoods of the null hypothesis and show that the test is asymptotically optimal. These results are derived under the fairly general condition of absolute regularity (β-mixing) for the predictor variables. We give numerical results that support the theory. We also illustrate usefulness of the method through an application to a body fat dataset where we build a simple, yet accurate, model that predicts individual body fat values well.

Key words and phrases: Backfitting, local polynomial fitting, marginal integration, varying coefficient models, wild bootstrap.

1. Introduction

Many useful nonparametric models have been suggested and studied to relax the parametric assumptions usually imposed on traditional regression models. One of them is the varying coefficient model, originally suggested by Hastie and Tibshirani (1993). This model inherits simplicity and easy interpretation of the classical linear model, yet is intrinsically nonparametric so that it is flexible enough to accommodate various complicated relationships between the response and predictor variables. The model has been studied by Chen and Tsay (1993), Cai, Fan, and Li (2000), Cai, Fan, and Yao (2000), and Yang et al. (2006), among others.

For the observations $(\mathbf{X}^i, \mathbf{T}^i, Y^i)$, $i = 1, \ldots, n$, with Y^i being responses and $\mathbf{X}^i \equiv (X_1^i, \ldots, X_d^i)^\top$, $\mathbf{T}^i \equiv (T_1^i, \ldots, T_d^i)^\top$ being d-dimensional predictors, the varying coefficient model is given by

$$Y^i = \sum_{j=1}^{d} f_j(X_j^i) T_j^i + \sigma(\mathbf{X}^i, \mathbf{T}^i)\epsilon^i, \tag{1.1}$$

where ϵ^i are independent and identically distributed white noises, independent of the predictors $(\mathbf{X}^i, \mathbf{T}^i)$. In this model we assume

$$\sum_{j=1}^{d} m_j(X_j^i)T_j^i = 0 \ \text{ a.s.} \quad \Rightarrow \ m_j \equiv 0 \ \text{ for } \ 1 \le j \le d. \tag{1.2}$$

One needs the condition (1.2) for the identifiability of f_j. It is a sufficient condition for avoiding concurvity as termed by Hastie and Tibshirani (1990). Concurvity in additive models is known as an analog of collinearity in linear models.

Yang et al. (2006) proposed and studied a testing procedure for the hypothesis that a particular coefficient function f_s is constant. The method is certainly useful when some of the coefficient functions are in fact constant and the true model actually partially linear. However, a more interesting and more general problem is to test whether a coefficient function has a certain parametric form, since in many cases partial linear models fail to fit the data sufficiently well, as in the data example we consider in Section 5.3. In those cases, a testing procedure for such hypothesis may be used to build a simple, yet accurate, parametric or semiparametric model that predicts sufficiently well the values of the response. We illustrate this in the data example.

Here we consider the problem of testing for the null hypothesis that a coefficient function f_s belongs to a general parametric family, against a nonparametric alternative. We propose a testing procedure based on the distance between a nonparametric estimator of f_s and the parametric model of the null hypothesis. We use the local polynomial marginal integration method to define a nonparametric estimator of f_s; this was studied by Yang et al. (2006) for the varying coefficient model (1.1). One might use other techniques of additive fitting, such as the ordinary backfitting of Friedman and Stuetzle (1981) or the smooth backfitting of Mammen, Linton, and Nielsen (1999). However, there has been no estimation theory developed for these methods when applied to fitting the varying coefficient model (1.1).

The problem with the general parametric hypothesis is quite different from the one with the constant null, the latter being studied by Yang et al. (2006). We find that a smoothing bias incurred by nonconstant coefficient functions in the null hypothesis produces some non-negligible terms in the expansion of the test statistic. Due to the additional terms, the procedure of Yang et al. (2006) to obtain critical values for their test is inappropriate for our problem. This leads us to propose a new method of obtaining critical values. Another point is that we derive the asymptotic distribution of the test statistic not only under the null hypothesis, but also under the alternative. Using these results we show that the proposed test achieves an optimal rate in the sense of the minimum distance between the null and alternative hypotheses that is necessary to attain a specified level of power. The theory is developed under fairly general condition of absolute regularity (β-mixing) for the predictors, to accommodate dependent data. Thus, it allows endogenous as well as exogenous random variables as predictors. We investigate the finite-sample performance of the test through a simulation study. We also illustrate usefulness of the method through a data example.

The paper is organized as follows. In the next section we propose the test statistic. In Section 3, we give the asymptotic null distribution. Section 4 is devoted to a discussion of the asymptotic power and rate-optimality of the proposed test. In Section 5, we propose a new method of obtaining critical values for our test, illustrate the finite-sample properties of the testing procedure, and then apply the method to a body fat dataset. All technical details are contained in Section 6.

2. Proposed Test Statistic

For a given $s \in \{1, \ldots, d\}$, we want to test the null hypothesis $H_0 : f_s \in \mathcal{F}_0$, where

$$\mathcal{F}_0 \equiv \{g(\cdot, \theta) : \theta \in \Theta,\ g \text{ is a known function}\} \tag{2.1}$$

is a parametric family and Θ is a subset of $\mathbb{R}^k$. The alternative hypothesis we consider is given by $H_1 : f_s \in \mathcal{F}_1$, where $\mathcal{F}_1$ is a nonparametric family that is apart from $\mathcal{F}_0$ at a certain distance. An example of $\mathcal{F}_0$ is the polynomial model where $g(x, \theta) = \theta_0 + \theta_1 x + \cdots + \theta_{k-1} x^{k-1}$.

The proposed test statistic is based on a nonparametric estimator of f_s. We employ the marginal integration method with local polynomial estimation proposed by Yang et al. (2006). To describe estimation of f_s at x_s, write

$$\begin{aligned}\mathbf{X}^i_{-s} &= (X^i_1, \ldots, X^i_{s-1}, X^i_{s+1}, \ldots, X^i_d)^\top,\\ \mathbf{T}^i_{-s} &= (T^i_1, \ldots, T^i_{s-1}, T^i_{s+1}, \ldots, T^i_d)^\top.\end{aligned}$$

First, we fit the model $m(\mathbf{x}, \mathbf{t}) = \sum_{j=1}^d f_j(x_j) t_j$ locally at $\mathbf{x} = (x_s, \mathbf{x}_{-s})$ for a $(d-1)$-dimensional point $\mathbf{x}_{-s}$. We approximate f_s by a pth order polynomial in a neighborhood of x_s, and the other coefficient functions f_j $(j \neq s)$ by constants in a neighborhood $\mathbf{x}_{-s}$. We use a univariate kernel K with a bandwidth h for the x_s-direction and a $(d-1)$-dimensional kernel L with a bandwidth vector $\mathbf{b}_{-s} = (b_1, \ldots, b_{s-1}, b_{s+1}, \ldots, b_d)$ for the others. This yields *full-dimensional* estimators, $\hat{\alpha}_s(\mathbf{x}) \equiv (\hat{\alpha}_{s,0}(\mathbf{x}), \ldots, \hat{\alpha}_{s,p}(\mathbf{x}))$ and $\hat{\gamma}_j(\mathbf{x})$, $j \neq s$, that minimize

$$\sum_{i=1}^n \left\{ Y^i - \sum_{l=0}^p \alpha_{s,l} (X^i_s - x_s)^l T^i_s - \sum_{j \neq s} \gamma_j T^i_j \right\}^2 K\left(\frac{X^i_s - x_s}{h}\right) L\left(\frac{\mathbf{X}^i_{-s} - \mathbf{x}_{-s}}{\mathbf{b}_{-s}}\right),$$

where $\mathbf{u}/\mathbf{v} = (u_1/v_1, \ldots, u_k/v_k)^\top$ for k-dimensional vectors $\mathbf{u}$ and $\mathbf{v}$.

Note that $\hat{\alpha}_{s,0}(\mathbf{x})$ is given by

$$\hat{\alpha}_{s,0}(\mathbf{x}) = \mathbf{e}_0^\top \left[\mathbf{Z}_s(x_s)^\top \mathbf{W}_s(\mathbf{x})\mathbf{Z}_s(x_s)\right]^{-1} \mathbf{Z}_s(x_s)^\top \mathbf{W}_s(\mathbf{x})\mathbf{Y}, \tag{2.2}$$

where $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$, $\mathbf{e}_0 = (1, 0, \ldots, 0)^\top$ is the $(p + d)$-dimensional unit vector,

$$\mathbf{W}_s(\mathbf{x}) = \text{diag}\left[K\left(\frac{X_s^i - x_s}{h}\right) L\left(\frac{\mathbf{X}_{-s}^i - \mathbf{x}_{-s}}{\mathbf{b}_{-s}}\right)\right]_{1\le i\le n},$$

$$\mathbf{Z}_s(x_s) = \left[\mathbf{p}\left(\frac{X_s^i - x_s}{h}\right)^\top T_s^i,\ \mathbf{T}_{-s}^{i\top}\right]_{1\le i\le n},$$

and $\mathbf{p}(u) = (1, u, \ldots, u^p)^\top$. This depends not only on x_s, but also on $\mathbf{x}_{-s}$, thus is not a relevant estimator of $f_s(x_s)$. We estimate $f_s(x_s)$ by integrating $\hat{\alpha}_{s,0}(x_s, \mathbf{x}_{-s})$ with respect to $\mathbf{x}_{-s}$ at the observed data $\mathbf{X}_{-s}^i$. Thus, the *marginal integration estimator* of $f_s(x_s)$ is

$$\hat{f}_s(x_s) = \left\{\sum_{i=1}^n w_{-s}(\mathbf{X}_{-s}^i)\right\}^{-1} \sum_{i=1}^n w_{-s}(\mathbf{X}_{-s}^i)\hat{\alpha}_{s,0}(x_s, \mathbf{X}_{-s}^i), \tag{2.3}$$

where the weight function $w_{-s}(\cdot)$ has a compact support with nonempty interior, introduced to avoid technical difficulties that may arise when the density of $\mathbf{X}_{-s}^i$ has an unbounded support. The marginal integration estimator at (2.3) needs the values of $\hat{\alpha}_{s,0}(x_s, \cdot)$ only at the observed data points $\mathbf{X}_{-s}^i$. This is certainly computationally less intensive than

$$\bar{f}_s(x_s) = \left[\int w_{-s}(\mathbf{x}_{-s})\, d\mathbf{x}_{-s}\right]^{-1} \int w_{-s}(\mathbf{x}_{-s})\hat{\alpha}_{s,0}(\mathbf{x})\, d\mathbf{x}_{-s}.$$

The latter requires evaluation of $\hat{\alpha}_{s,0}(x_s, \cdot)$ on a fine grid of $\mathbf{x}_{-s}$ in $\mathbb{R}^{d-1}$ and thus the amount of computation increases rapidly as the dimension d gets high. Yang et al. (2006) showed that the estimators at (2.3) have the univariate rate of convergence, and that they are asymptotically normally distributed.

Our test statistic for testing $H_0 : f_s \in \mathcal{F}_0$, where $\mathcal{F}_0$ is as given at (2.1), is based on a minimum distance principle. Let w_s is a weight function supported on a compact set in $\mathbb{R}$. We propose to use

$$V_n = n^{-1} \min_{\theta\in\Theta} \sum_{i=1}^n \left\{\hat{f}_s(X_s^i) - g(X_s^i, \theta)\right\}^2 w_s(X_s^i).$$

The statistic V_n is an estimator of $\min_{\theta\in\Theta} E\{f_s(X_s) - g(X_s,\theta)\}^2 w_s(X_s)$. The value of the latter is zero if $f_s \in \mathcal{F}_0$. If V_n is large, then one would reject the null hypothesis H_0. A critical value may be obtained from the null distribution of V_n or its estimate. In the next two sections we see that V_n, under the null and alternative hypotheses, converges to normal.

3. Asymptotic Null Distribution

We treat the case where the vector process $\{(\mathbf{X}^i, \mathbf{T}^i)\}_{i=1}^n$ is strictly stationary and β-mixing, with mixing coefficients satisfying $\beta(j) \le C\rho^j$ for some constants $C > 0$ and $0 < \rho < 1$. Here

$$\beta(j) = \sup_k E\left[\sup\left\{\left|P(A|\mathcal{G}_{-\infty}^k) - P(A)\right| : A \in \mathcal{G}_{j+k}^{\infty}\right\}\right],$$

where $\mathcal{G}_l^u$ is the σ-algebra generated by $(\mathbf{X}_l, \mathbf{T}_l), (\mathbf{X}_{l+1}, \mathbf{T}_{l+1}), \ldots, (\mathbf{X}_u, \mathbf{T}_u)$ for $l < u$.

Define

$$S_n(\theta) = n^{-1}\sum_{i=1}^n \left\{\hat{f}_s(X_s^i) - g(X_s^i,\theta)\right\}^2 w_s(X_s^i), \tag{3.1}$$

so that $V_n = \min_{\theta\in\Theta} S_n(\theta)$. Let $\hat\theta$ denote the minimizer of $S_n(\theta)$ in Θ. Theorem 1 below gives a higher-order stochastic expansion of the test statistic V_n under the null hypothesis H_0. To state the theorem, we need to introduce more notation.

Let $K^*(u,\mathbf{t},\mathbf{x}) = \mathbf{e}_0^\top D_s^{-1}(\mathbf{x})\mathbf{q}_s(u,\mathbf{t})^\top K(u)$, where $\mathbf{q}_s(u,\mathbf{t})^\top = \big(t_s\mathbf{p}(u)^\top, \mathbf{t}_{-s}\big)$ and

$$D_s(\mathbf{x}) = \int E\Big[\mathbf{q}_s(u,\mathbf{T})\mathbf{q}_s(u,\mathbf{T})^\top\big|\mathbf{X} = \mathbf{x}\Big]K(u)\,du. \tag{3.2}$$

It can be seen that $\hat\alpha_{s,0}(\mathbf{x})$ defined at (2.2) satisfies:

$$\hat\alpha_{s,0}(\mathbf{x}) \simeq n^{-1}h^{-1}b_{s,\mathrm{prod}}^{-1}\sum_{i=1}^n Y^i K^*\left(\frac{X_s^i - x_s}{h}\right) L\left(\frac{\mathbf{X}_{-s}^i - \mathbf{x}_{-s}}{\mathbf{b}_{-s}}\right),$$

where $b_{s,\mathrm{prod}} = b_1 \times\cdots\times b_{s-1}\times b_{s+1}\times\cdots\times b_d$. Let

$$\tilde{K}^*(u,\mathbf{t},\mathbf{t}',\mathbf{x},\mathbf{x}'_{-s}) = \int K^*(w,\mathbf{t},\mathbf{x})K^*(w+u,\mathbf{t}',x_s,\mathbf{x}'_{-s})\,dw.$$

Let $w_{-s}^*(\mathbf{x}_{-s}) = w_{-s}(\mathbf{x}_{-s})/E\{w_{-s}(\mathbf{X}_{-s})\}$ so that $E\{w_{-s}^*(\mathbf{X}_{-s})\} = 1$, and let

$$\kappa(x,\theta) = E\left[w_{-s}^*(\mathbf{X}_{-s})T_s\int u^{p+1}K^*(u;\mathbf{T},\mathbf{X})\,du\Big|X_s = x\right] \times \frac{1}{(p+1)!}\,g^{(p+1)}(x,\theta),$$

where $g^{(j)}(x,\theta)$ denotes the jth derivative of $g(x,\theta)$ with respect to x. Let $\dot{g}(x,\theta)$ denote the vector of the first partial derivative of $g(x,\theta)$ with respect to θ. Denote the density function of $(\mathbf{X},\mathbf{T})$ by ψ, and the density function of $\mathbf{X}$ by φ. Let φ_{-s} and φ_s be the marginal density functions of $\mathbf{X}_{-s}$ and X_s, respectively. Define $\eta(\mathbf{x},\mathbf{t}) = w_{-s}^{*2}(\mathbf{x}_{-s})\varphi_{-s}^2(\mathbf{x}_{-s})\varphi_s(x_s)\varphi^{-2}(\mathbf{x})\sigma^2(\mathbf{x},\mathbf{t})$, and

$$
\begin{aligned}
c_1(\theta) &= \left\|\Pi\big(\kappa(\cdot,\theta)\,\big|\,[\dot{g}(\cdot,\theta)]^{\perp}\big)\right\|_s^2,\\
c_2 &= E\left[w_s(X_s)\eta(\mathbf{X},\mathbf{T})\int K^{*2}(u;\mathbf{T},\mathbf{X})\,du\right],
\end{aligned} \tag{3.3}
$$

where $\Pi(\cdot|\mathcal{S})$ denotes the projection operator onto $\mathcal{S}$, $[\dot{g}(\cdot,\theta)]$, the linear span of $\dot{g}(\cdot,\theta)$ in a Hilbert space with inner product $\langle q_1,q_2\rangle_s = E\,[w_s(X_s)q_1(X_s)q_2(X_s)]$ and $\|\cdot\|_s$ the corresponding norm. Specifically,

$$
\begin{aligned}
\Pi\big(\kappa(\cdot,\theta)\,\big|\,[\dot{g}(\cdot,\theta)]\big)(x) &= \left(E\,[w_s(X_s)\kappa(X_s,\theta)\dot{g}(X_s,\theta)]\right)^{\top}\\
&\quad\times\left(E\left[w_s(X_s)\dot{g}(X_s,\theta)\dot{g}(X_s,\theta)^{\top}\right]\right)^{-1}\dot{g}(x,\theta).
\end{aligned}
$$

Also, we define

$$
\begin{aligned}
\gamma_{11} &= \int\left[\eta(\mathbf{x},\mathbf{t})\eta(x_s,\mathbf{x}'_{-s},\mathbf{t}')\int\tilde{K}^{*2}(u;\mathbf{t},\mathbf{t}',\mathbf{x},\mathbf{x}'_{-s})\,du\right]\\
&\qquad\times\psi(\mathbf{x},\mathbf{t})\psi(x_s,\mathbf{x}'_{-s},\mathbf{t}')\,d\mathbf{x}\,d\mathbf{x}'_{-s}\,d\mathbf{t}\,d\mathbf{t}',\\
\gamma_{22}(\theta) &= E\Bigg[w_s^2(X_s)\eta(\mathbf{X},\mathbf{T})\varphi_s(X_s)\left\{\kappa(X_s,\theta)-2\Pi\big(\kappa(\cdot,\theta)\,\big|\,[\dot{g}(\cdot,\theta)]\big)(X_s)\right\}^2\\
&\qquad\times\left\{\int K^*(u,\mathbf{T},\mathbf{X})\,du\right\}^2\Bigg].
\end{aligned}
$$

We note that $c_1(\theta)$, c_2, γ_{11}, $\gamma_{22}(\theta)\geq 0$.

Theorem 1. *Suppose that $f_s = g(\cdot,\theta_0)$ for some $\theta_0\in\Theta$, and that the assumptions* (A1)$-$(A9) *in the Appendix hold. Then, one has*

$$
\begin{aligned}
V_n &= c_1(\theta_0)h^{2p+2} + c_2n^{-1}h^{-1} + \gamma_{11}^{1/2}n^{-1}h^{-1/2}Z_{1,n} + \gamma_{22}(\theta_0)^{1/2}n^{-1/2}h^{p+1}Z_{2,n}\\
&\quad + o_p\left(h^{2p+2}+n^{-1}h^{-1/2}+n^{-1/2}h^{p+1}\right),
\end{aligned}
$$

where $Z_{1,n}$ and $Z_{2,n}$ are uncorrelated and asymptotically $N(0,1)$ as $n\to\infty$.

The deterministic term $c_1(\theta)h^{2p+2}$ in the expansion of V_n comes from the bias of the estimator $\hat{f}_s$. Note that the constant factor $c_1(\theta_0)=0$ if p is even and the kernel K is symmetric. Thus, in the discussion below we assume p is odd.

Note also that $c_1(\theta_0)$ does not depend on other component functions, since the smoothing bias of the estimator $\hat{f}_s$ due to local averaging for other components f_j, $j \neq s$, is made negligible by choosing the bandwidth b_j small enough and using a higher-order kernel L; see the conditions on the bandwidth b_j and the kernel L in Section 6.1.

From Theorem 1, one can deduce that, if f_s is a constant,

$$nh^{1/2}V_n - h^{-1/2}c_2 \xrightarrow{d} N(0, \gamma_{11}).$$

This coincides with the results of Theorem 5 in Yang et al. (2006). If $f_s = g(\cdot, \theta_0)$ and the $(p+1)$th derivative of $g(\cdot, \theta_0)$ is not identically zero on the support of w_s, then the smoothing bias incurred by the pth order local polynomial fitting produces the stochastic term of order $O_p(n^{-1/2}h^{p+1})$ and the non-stochastic term of order $O(h^{2p+2})$.

One can derive the asymptotic null distribution of the test statistic, depending on the size of the bandwidth h. For simplicity we state only the result for the case $nh^{2p+3} \to 0$, in which the stochastic term $O_p(n^{-1/2}h^{p+1})$ is negligible. We note that $h \sim n^{-2/(4p+5)}$ is the order of the bandwidth h that gives a rate-optimal test, see Section 4.

Corollary 1. *Suppose that $nh^{2p+3} \to 0$ as $n \to \infty$. Under the conditions of Theorem 1,*

$$nh^{1/2}V_n - c_1(\theta_0)nh^{2p+(5/2)} - c_2h^{-1/2} \xrightarrow{d} N(0, \gamma_{11}).$$

In the corollary, the term $c_1(\theta_0)nh^{2p+(5/2)}$, which is negligible in comparison with the term $c_2h^{-1/2}$ under the condition that $nh^{2p+3} \to 0$ as $n \to \infty$, should not be removed, since it may not converge to zero as $n \to \infty$. We also note that the limit distribution of V_n depends on σ^2, ψ, and only on f_s among the coefficient functions f_j.

For each level $0 < \alpha < 1$, one can construct a test that has the level α asymptotically. Define $c_1^* = \sup_{\theta \in \Theta} c_1(\theta)$. Let $\hat{c}_1^*$, $\hat{c}_2$, and $\hat{\gamma}_{11}$ be estimates of c_1^*, c_2, and γ_{11}, respectively, such that

$$\hat{c}_1^* - c_1^* = o_p(n^{-1}h^{-2p-(5/2)}), \quad \hat{c}_2 - c_2 = o_p(h^{1/2}), \quad \hat{\gamma}_{11} - \gamma_{11} = o_p(1). \tag{3.4}$$

Define

$$v_{n,\alpha} = \hat{c}_1^* h^{2p+2} + \hat{c}_2 n^{-1}h^{-1} + \hat{\gamma}_{11}^{1/2} n^{-1}h^{-1/2} z_{1-\alpha}, \tag{3.5}$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of $N(0,1)$. Let F denote the triple (f_s, ψ, σ^2), and P_F be the probability measure associated with F. Then, we see from Corollary 1 that $\lim_{n\to\infty} \sup_{F \in H_0} P_F(V_n > v_{n,\alpha}) = \alpha$. Thus, the test

$$T_n(\alpha): \text{ Reject } H_0 \text{ if } V_n > v_{n,\alpha} \tag{3.6}$$

has the asymptotic size and level α.

If one uses the optimal bandwidth $h \sim n^{-2/(4p+5)}$, then the conditions at (3.4) are simply

$$\hat{c}_1^* - c_1^* = o_p(1), \quad \hat{c}_2 - c_2 = o_p(n^{-1/(4p+5)}), \quad \hat{\gamma}_{11} - \gamma_{11} = o_p(1).$$

Estimation of the constants c_1^*, c_2, and γ_{11} involves that of D_s at (3.2), the conditional expectation in the definition of κ, the density function ψ, and the variance function σ^2. These are typical nonparametric function estimation problems. Thus, one may construct their estimators that satisfy the conditions at (3.4), sufficient smoothness of ψ and σ^2 being permitted.

4. Properties Under Local Alternatives

4.1. Asymptotic alternative distribution

In this section, we derive the asymptotic distribution of the test statistic when f_s lies in a nonparametric alternative $\mathcal{F}_1$ that is apart from the null class $\mathcal{F}_0$. It can be verified that the test statistic converges to a strictly positive constant in probability when f_s ($\notin \mathcal{F}_0$) is fixed. Also, from the results of the previous section, we know that the critical value at a certain level that is derived from the null distribution converges to zero as n tends to infinity. This means that the power of the test at a fixed f_s not in $\mathcal{F}_0$ converges to one as the sample size increases to infinity. Thus, it would be of interest to calculate the asymptotic power of the test in a neighborhood of $\mathcal{F}_0$ that shrinks to $\mathcal{F}_0$ at a certain rate.

Specifically, we take as an alternative

$$\mathcal{F}_1 = \left\{ g(\cdot, \theta) + \rho_n \Delta(\cdot) : \theta \in \Theta, \Delta \in \mathcal{G}, \rho_n \geq C r_n \right\}, \tag{4.1}$$

where C is a positive constant, r_n is a sequence of real numbers converging to zero as $n \to \infty$, and $\mathcal{G}$ is the class of functions Δ that have p derivatives, a Lipschitz condition of order 1 on the pth derivative, and satisfy $\|\Delta\|_s = 1$, $\|\Delta^2\|_s < \infty$ and $\Delta \perp \mathcal{F}_0$. Note that without loss of generality we may assume $\|\Delta\|_s = 1$. To derive the asymptotic alternative distribution and compute the asymptotic power of the test, we take an arbitrary local alternative

$$f_{s,n}(x) = g(x, \theta_1) + \rho_n \Delta(x), \tag{4.2}$$

where $\theta_1 \in \Theta$, $\Delta \in \mathcal{G}$, and ρ_n is a sequence of real numbers converging to zero as the sample size n tends to infinity.

Define

$$c_{1n}(\theta, \Delta) = \left\| \rho_n \Delta + h^{p+1} \Pi \left(\kappa(\cdot, \theta) \mid [\dot{g}(\cdot, \theta)]^{\perp} \right) \right\|_s^2,$$

$$\gamma_{23}(\theta,\Delta) = 2\,E\Bigg[w_s^2(X_s)\eta(\mathbf{X},\mathbf{T})\varphi_s(X_s)\Delta(X_s)\left\{\int K^*(u,\mathbf{T},\mathbf{X})\,du\right\}^2 \times \left\{\kappa(X_s,\theta) - 2\Pi\big(\kappa(\cdot,\theta)\,\big|\,[\dot{g}(\cdot,\theta)]\big)(X_s)\right\}\Bigg],$$

$$\gamma_{33}(\Delta) = 4\,E\left[w_s^2(X_s)\eta(\mathbf{X},\mathbf{T})\varphi_s(X_s)\Delta^2(X_s)\left\{\int K^*(u,\mathbf{T},\mathbf{X})\,du\right\}^2\right].$$

We are now ready to state a theorem that details the asymptotic behavior of V_n when the sth coefficient function takes the form at (4.2).

Theorem 2. *Suppose that* (4.2) *and the assumptions* (A1)−(A9) *in the Appendix hold. Assume that* $\|\Delta^2\|_s < \infty$, $\Delta \perp \mathcal{F}_0$, $\|\Delta\|_s = 1$, *and* Δ *has* p *derivatives and a Lipschitz condition of order 1 on the* p*th derivative. Then*

$$\begin{aligned} V_n &= c_{1n}(\theta_1,\Delta) + c_2\, n^{-1}h^{-1} \\ &\quad + \gamma_{11}^{1/2} n^{-1}h^{-1/2} Z_{1,n} + \gamma_{22}(\theta_1)^{1/2} n^{-1/2} h^{p+1} Z_{2,n} + \gamma_{33}(\Delta)^{1/2} n^{-1/2}\rho_n Z_{3,n} \\ &\quad + o_p\left(h^{2p+2} + h^{p+1}\rho_n + n^{-1}h^{-1/2} + n^{-1/2}h^{p+1} + n^{-1/2}\rho_n\right), \end{aligned}$$

where the marginal distributions of $Z_{1,n}$, $Z_{2,n}$ *and* $Z_{3,n}$ *are asymptotically* $N(0,1)$, $Z_{1,n}$, *and* $(Z_{2,n}, Z_{3,n})$ *are uncorrelated, and*

$$\mathrm{Cov}(Z_{2,n}, Z_{3,n}) = \gamma_{22}(\theta_1)^{-1/2}\gamma_{33}(\Delta)^{-1/2}\gamma_{23}(\theta_1,\Delta) + o(1).$$

4.2. Asymptotic power analysis

Theorem 2 enables us to analyze the power of the test given at (3.6) in Section 3. For this, we fix θ_1 and Δ. Let $\beta_n(\theta_1,\Delta,\rho_n)$ denote the power of the test when the sth coefficient function is given by $f_{n,s}(x) = g(x,\theta_1) + \rho_n\Delta(x)$. Note that the asymptotic distribution of the test statistic V_n depends only on the sth coefficient function f_s among $\{f_j\}_{j=1}^d$. In the discussion below, we assume that the $\hat{c}_1^*$, $\hat{c}_2$, and $\hat{\gamma}_{11}$ in the definition of $v_{n,\alpha}$ at (3.5) satisfy (3.4).

To avoid too much complication, we confine ourselves, as in Corollary 1, to the case where $nh^{2p+3} \to 0$. This covers a bandwidth range that gives an optimal test as we discuss in the next subsection. We also assume that $n^{1/2}\rho_n \to \infty$ as $n \to \infty$. Note that $\rho_n \sim n^{-1/2}$ is the minimal distance between a null hypothesis and a local alternative that are 'distinguishable' from each other in parametric models. Certainly, it is more difficult to discriminate one from the other in nonparametric models. Therefore, the ρ_n that makes the corresponding local alternative distinguishable from H_0 is of larger order than $n^{-1/2}$.

(i) *The case that* $nh\rho_n^2 \to 0$ *as* $n \to \infty$: Here $\gamma_{11}^{1/2} n^{-1} h^{-1/2} Z_{1,n}$ in the expansion of V_n in Theorem 2 is the leading stochastic term. Thus,

$$\beta_n(\theta_1, \Delta, \rho_n) = P\left[Z_{1,n} > z_{1-\alpha} + \gamma_{11}^{-1/2}\left(c_1^* nh^{2p+(5/2)} - nh^{1/2} c_{1n}(\theta_1, \Delta)\right)\right] + o(1). \tag{4.3}$$

Recall that $c_{1n}(\theta_1, \Delta)$ is of order $O(h^{2p+2} + \rho_n^2)$.

Suppose that $\rho_n << h^{p+1}$ ($a_n << b_n$ means $a_n/b_n \to 0$ as $n \to \infty$). Then the asymptotic power at (4.3) reduces to

$$\beta_n(\theta_1, \Delta, \rho_n) = P\left[Z_{1,n} > z_{1-\alpha} + \gamma_{11}^{-1/2} nh^{2p+(5/2)}(c_1^* - c_1(\theta_1))\right] + o(1).$$

By definition, $c_1^* \geq c_1(\theta_1)$ for all $\theta_1 \in \Theta$. Thus, $\beta_n(\theta_1, \Delta, \rho_n) \to \beta_*$ as $n \to \infty$ for some β_* such that $0 \leq \beta_* \leq \alpha$. The constant β_* depends on the value of θ_1 and the speed at which the bandwidth $h \to 0$.

Suppose that $\rho_n >> h^{p+1}$. Then ρ_n in $c_{1n}(\theta_1, \Delta)$ is the dominating term, so that

$$\begin{aligned}\beta_n(\theta_1, \Delta, \rho_n) &= P\left[Z_{1,n} > z_{1-\alpha} - \gamma_{11}^{-1/2} nh^{1/2}\rho_n^2\right] + o(1)\\ &\to \begin{cases} \alpha & \text{if } nh^{1/2}\rho_n^2 \to 0,\\ 1 & \text{if } nh^{1/2}\rho_n^2 \to \infty,\\ \beta_*' & \text{if } nh^{1/2}\rho_n^2 \to a \ (0 < a < \infty),\end{cases}\end{aligned}$$

for some β_*' such that $\alpha < \beta_*' < 1$. The constant β_*' in this case depends on a, the limit of $nh^{1/2}\rho_n^2$.

Finally, suppose that $\rho_n \sim h^{p+1}$. If we write $\rho_n = Ch^{p+1}$ for some constant C, then $c_{1n}(\theta_1, \Delta) = c_1'(C, \theta_1, \Delta) h^{2p+2}$, where

$$c_1'(C, \theta_1, \Delta) = \|C\Delta + \Pi\left(\kappa(\cdot, \theta_1) \,\middle|\, [\dot{g}(\cdot, \theta_1)]^\perp\right)\|_s^2. \tag{4.4}$$

It holds that

$$\beta_n(\theta_1, \Delta, \rho_n) = P\left[Z_{1,n} > z_{1-\alpha} + \gamma_{11}^{-1/2} nh^{2p+(5/2)}\left(c_1^* - c_1'(C, \theta_1, \Delta)\right)\right] + o(1). \tag{4.5}$$

From this we deduce that

$$\beta_n(\theta_1, \Delta, \rho_n) \to \begin{cases} \alpha & \text{if } nh^{(4p+5)/2} \to 0 \text{ or } c_1^* = c_1'(C, \theta_1, \Delta),\\ 1 & \text{if } nh^{(4p+5)/2} \to \infty \text{ and } c_1^* < c_1'(C, \theta_1, \Delta),\\ 0 & \text{if } nh^{(4p+5)/2} \to \infty \text{ and } c_1^* > c_1'(C, \theta_1, \Delta),\\ \beta_+ & \text{if } nh^{(4p+5)/2} \to a' \ (0 < a' < \infty) \text{ and } c_1^* < c_1'(C, \theta_1, \Delta),\\ \beta_- & \text{if } nh^{(4p+5)/2} \to a' \ (0 < a' < \infty) \text{ and } c_1^* > c_1'(C, \theta_1, \Delta),\end{cases}$$

for some β_+ and β_- such that $0 < \beta_- < \alpha < \beta_+ < 1$. The two constants β_+ and β_- depend on the constant a', the limit of $nh^{(4p+5)/2}$, and the value of $c_1^* - c_1'(C, \theta_1, \Delta)$.

(ii) *The case that* $nh\rho_n^2 \to \infty$ *or* $nh\rho_n^2 \to a''$ *for some* $0 < a'' < \infty$ *as* $n \to \infty$: Here ρ_n dominates h^{p+1}, since we assume $nh^{2p+3} \to 0$ as $n \to \infty$. When $nh\rho_n^2 \to \infty$, we have

$$\beta_n(\theta_1, \Delta, \rho_n) = P[Z_{3,n} > -\gamma_{33}(\Delta)^{-1/2} n^{1/2} \rho_n] + o(1) \to 1$$

since $n^{1/2}\rho_n \to \infty$ as $n \to \infty$. On the other hand, if $nh\rho_n^2 \to a''$ for some $0 < a'' < \infty$, then

$$\beta_n(\theta_1, \Delta, \rho_n) = P\left[Z_{4,n} > \frac{z_{1-\alpha} - \gamma_{11}^{-1/2} nh^{1/2} \rho_n^2}{\sqrt{1 + a''(\gamma_{33}(\Delta)/\gamma_{11})}}\right] + o(1) \to \beta_*''$$

for some β_*'' such that $\alpha < \beta_*'' \le 1$, where $Z_{4,n}$ is asymptotically $N(0,1)$. Here $\beta_*'' = 1$ is achieved only when $nh^{1/2}\rho_n^2 \to \infty$ as $n \to \infty$, and β_*'' in the range $(\alpha, 1)$ depends on a'' and the limit of $nh^{1/2}\rho_n^2$.

In some of the cases above, the asymptotic power of the test is smaller than the level of the test. One is the case where ρ_n is too small, and the other is the case where $\rho_n \sim h^{p+1}$ but $c_1^* > c_1'(C, \theta_1, \Delta)$. This is intrinsic to all nonparametric methods and can be explained as follows. Note that the acceptance region for the testing procedure is

$$\mathcal{A}_n = \{f : \min_{\theta \in \Theta} \|f - g(\cdot, \theta)\|_{s,n} \le v_{n,\alpha}^{1/2}\} \supset \mathcal{F}_0,$$

where $\|\cdot\|_{s,n}$ is an empirical version of the Hilbert norm $\|\cdot\|_s$, defined by $\|f\|_{s,n}^2 = n^{-1}\sum_{i=1}^n f(X_s^i)^2 w_s(X_s^i)$. The distance $v_{n,\alpha}^{1/2}$ from $\mathcal{F}_0$ is the limit within which $\hat{f}_s$ lies asymptotically with probability $(1-\alpha)$ when the true function actually belongs to $\mathcal{F}_0$, and it accommodates the smoothing bias of $\hat{f}_s$, the term of order h^{p+1}. When the true function is away from $\mathcal{F}_0$ at a distance ρ_n, the smoothing bias interacts with the model bias ρ_n. It may push $\hat{f}_s$ further away from $\mathcal{F}_0$, or take $\hat{f}_s$ to the opposite direction toward $\mathcal{F}_0$. In the latter case, $\hat{f}_s$ has better chance to belong to the acceptance region $\mathcal{A}_n$ than under the null.

4.3. Rate-optimality

We consider the class of functions given at (4.1) for the alternative hypothesis. Let $T_n(\alpha)$ denote the test given at (3.6). Quality of a test is often measured by the minimum distance between the null and alternative hypotheses that is necessary to achieve a specified level of power, see Ingster

(1993), Lepski and Tsybakov (2000), Pouet (2001), and Emarkov (2003), for example. Under this approach, we seek the fastest possible rate of r_n for the test $T_n(\alpha)$ such that

$$\text{for any number } 0 < \beta_0 < 1 \text{ there exists } C > 0 \text{ for which} \\ \lim_{n\to\infty} \inf_{F\in H_1} P_F[T_n(\alpha) = 1] \geq \beta_0. \quad (4.6)$$

From the results of the previous subsection the only cases where a specified power may be achieved are: (i) $nh\rho_n^2 << 1$ and $\rho_n >> h^{p+1}$; (ii) $nh\rho_n^2 << 1$ and $\rho_n \sim h^{p+1}$; (iii) $nh\rho_n^2 >> 1$ or $nh\rho_n^2 \sim 1$. For each given bandwidth h, the second case involves the smallest ρ_n. Also, when $\rho_n \sim h^{p+1}$, the case where $nh^{(4p+5)/2} \to \infty$ gives a slower rate for the bandwidth, thus larger ρ_n, than the case where $nh^{(4p+5)/2}$ converges to a positive constant. For simplicity, we consider $h = n^{-2/(4p+5)}$ without a constant factor. One can show that

$$\lim_{\varepsilon\to 0} \inf_{\theta\in\Theta} \|\Delta + \varepsilon\Pi\left(\kappa(\cdot,\theta)\,\big|\,[\dot{g}(\cdot,\theta)]^\perp\right)\|_s \to \|\Delta\|_s = 1.$$

This means $\pi(C) \equiv \inf_{\theta\in\Theta,\Delta\in\mathcal{G}} c_1'(C,\theta_1,\Delta)$ converges to infinity as C tends to infinity. This, together with (4.5), shows that with $r_n = n^{-(2p+2)/(4p+5)}$ and $h = n^{-2/(4p+5)}$,

$$\lim_{n\to\infty} \inf_{F\in H_1} P_F[T_n(\alpha) = 1] = 1 - \Phi\left(z_{1-\alpha} - \gamma_{11}^{-1/2}(\pi(C) - c_1^*)\right) \to 1$$

as C tends to infinity, where Φ denotes the distribution function of the standard normal distribution. This establishes the fact that $r_n = n^{-(2p+2)/(4p+5)}$ is the fastest possible rate for the test $T_n(\alpha)$.

Theorem 3. *Suppose that the assumptions* (A1)−(A9) *hold. Assume that* (3.4) *holds. Then, with* $r_n = n^{-(2p+2)/(4p+5)}$*, the test given at* (3.6) *with a bandwidth choice* $h \sim n^{-2/(4p+5)}$ *satisfies* (4.6) *for testing* $H_0 : f_s \in \mathcal{F}_0$ *versus* $H_1 : f_s \in \mathcal{F}_1$*, where* $\mathcal{F}_0$ *and* $\mathcal{F}_1$ *are given at* (2.1) *and* (4.1)*, respectively.*

One may verify that the rate $r_n = n^{-(2p+2)/(4p+5)}$ is a lower bound, which means that there is no test that achieves a specified level of power if $r_n << n^{-(2p+2)/(4p+5)}$. This can be proved similarly as in Ingster (1993) or Pouet (2001). In this sense, the proposed test with a bandwidth choice $h \sim n^{-2/(4p+5)}$ is minimax-optimal.

If the mean of $\hat{f}_s - g(\cdot,\hat{\theta})$ is available, then one might use a bias-corrected version of the test statistic V_n. One can show that the bias-corrected test statistic does not have the terms h^{2p+2} and $n^{-1/2}h^{p+1}$ in its expansions analogous to Theorems 1 and 2. Also, one can argue along the lines of Sections 4.2 and 4.3 that it can achieve a rate r_n arbitrarily close to $n^{-1/2}$ by letting the bandwidth h tend to zero slowly enough. However, the mean of $\hat{f}_s - g(\cdot, \hat{\theta})$ is not available, but needs to be estimated. The magnitude of the estimation error depends the smoothness of f_s. If we assume that f_s is $(k+1)$-times continuously differentiable, then the estimation error contributes $h^{2k+2} + n^{-1/2}h^{k+1}$ to the expansion of the corresponding bias-corrected test statistic. Thus, one might get better asymptotic power with the bias-corrected test than with V_n for a function class with higher-order smoothness ($k > p$). However, the same rate can be also achieved without bias-correction if one applies kth order, rather than pth order, polynomial fitting in the construction of V_n.

5. The Methods in Practice

5.1. Obtaining critical values

The critical value $v_{n,\alpha}$ given in Section 3 needs to be obtained in practice. One might think of a bootstrap procedure such as the one proposed by Yang et al. (2006). As was pointed out by Yang et al. (2006) and observed by Wu (1986), Liu (1988), Härdle and Mammen (1993), and Sperlich, Tjøstheim, and Yang (2002), among others, methods based on ordinary bootstrapping fail when the errors are heteroscedastic. As an alternative, the wild-bootstrap, introduced by Wu (1986), has been used in many problems, see, for example, Härdle and Mammen (1993), Franke, Kreiss, and Mammen (2002), Franke, Kreiss, Mammen, and Neumann (2002), Li and Wang (1998), Härdle et al. (2004), and Yang et al. (2006).

If one follows the procedure of Yang et al. (2006), one would fit $m(\mathbf{x}, \mathbf{t}) = g(x_s, \theta)t_s + \sum_{j \neq s}^d f_j(x_j)t_j$ to the data under the null model and find $\tilde{m}(\mathbf{x}, \mathbf{t}) = g(x_s, \tilde{\theta})t_s + \sum_{j \neq s}^d \tilde{f}_j(x_j)t_j$; generate bootstrap residuals e_*^i and the bootstrap responses $Y_*^i = \tilde{m}(\mathbf{X}^i, \mathbf{T}^i) + e_*^i$; compute a bootstrap version V_n^* of V_n using the bootstrap sample $\{(Y_*^i, \mathbf{X}^i, \mathbf{T}^i)\}_{i=1}^n$; and then obtain the $(1-\alpha)$ quantile of the distribution of V_n^*. Recall that our aim of bootstrapping in the current problem is to find an estimate of $\sup_{\theta \in \Theta} v_{n,\alpha}(\theta)$, where

$$v_{n,\alpha}(\theta) = c_1(\theta)h^{2p+2} + c_2 n^{-1}h^{-1} + \gamma_{11}^{-1/2} n^{-1}h^{-1/2} z_{1-\alpha}.$$

The latter is the asymptotic $(1-\alpha)$ quantile of the distribution of V_n when $f_s = g(\cdot, \theta)$. The problem is different from the one treated in Yang et al. (2006), since $v_{n,\alpha}(\theta)$ depends on θ under the null model. The point is that one should mimic different null models for different θ to generate bootstrap samples, while the approach taken by Yang et al. (2006) uses the same bootstrap null model

$\tilde{m}(\mathbf{x}, \mathbf{t}) = g(x_s, \tilde{\theta})t_s + \sum_{j \neq s}^d \tilde{f}_j(x_j)t_j$ regardless of the values of $\theta \in \Theta$. Certainly, the latter approach would choose a cut-off value that gives an actual level that is higher than a nominal level.

Our proposal is to first estimate $v_{n,\alpha}(\theta)$ for each $\theta \in \Theta$ by using bootstrap samples generated from an estimated null model that incorporates $f_s = g(\cdot, \theta)$, and then take the maximum of the estimates over $\theta \in \Theta$. To do this, we suggest the following procedure.

1. Fit $\hat{m}(\mathbf{x}, \mathbf{t}) = \sum_{j=1}^d \hat{f}_j(x_j)t_j$ to the data and obtain the residuals $e^i = Y^i - \hat{m}(\mathbf{X}^i, \mathbf{T}^i)$, where $\hat{f}_j$ are the marginal integration estimates defined at (2.3).
2. Generate i.i.d. random variables $Z_W^1, \ldots, Z_W^n$ such that $EZ_W^i = 0, E(Z_W^i)^2 = 1$ and $E(Z_W^i)^3 = 1$, and then compute the (wild-)bootstrap residuals $e_*^i = e^i Z_W^i$.
3. For each $\theta \in \Theta$, put $Y_*^i(\theta) = g(X_s^i, \theta)T_s^i + \sum_{j=1, j \neq s}^d \hat{f}_j(X_j^i)T_j^i + e_*^i$
4. Calculate the bootstrap version of V_n, say $V_n^*(\theta)$, based on the wild-bootstrap sample $\{(Y_*^i(\theta), \mathbf{X}^i, \mathbf{T}^i)\}_{i=1}^n$.
5. Repeat Steps 2, 3 and 4 many times, and find the $(1 - \alpha)$ quantile, denoted by $\hat{v}_{n,\alpha}(\theta)$, of the distribution of $V_n^*(\theta)$ for each $\theta \in \Theta$.
6. Obtain $v_{n,\alpha} = \sup_{\theta \in \Theta} \hat{v}_{n,\alpha}(\theta)$.

For the distribution of Z_W^i, one can choose any distribution that satisfies the above moment conditions in the step (2), see Mammen (1992) for some examples. In the simulation study of the current paper, we used a two points distribution such that $Z_W = (1 - \sqrt{5})/2$ with probability $(5 + \sqrt{5})/10$ and $Z_W = (1 + \sqrt{5})/2$ with probability $(5 - \sqrt{5})/10$.

5.2. Simulation study

We conducted a simulation study to see the performance of the proposed testing procedure. In this simulation we used the wild-bootstrap method described in the previous subsection with 200 bootstrap replications. We chose $p = 1$ (local linear fitting) and the quartic kernel $K(x) = 0.9375(1 - x^2)^2 I_{(-1,1)}(x)$. For simplicity we used the bandwidth $h = n^{-2/(4p+5)}$ and took $b_j = (\log n)^{-1} h^{(p+1)/q} = (\log n)^{-1} n^{-(2p+2)/\{q(4p+5)\}}$. Although the procedure is sensitive to the choice of the bandwidths, we found that this simple choice worked very well. If one wants to use a more deliberate choice, one can develop a data-driven procedure, suitable for the testing problem, along the lines of Yang et al. (2006). For the weight functions, we chose $w_s(x) = I_{[0,1]}(x)$ and $w_{-s}(\mathbf{x}_{-s}) = I_{[0,1]^{d-1}}(\mathbf{x}_{-s})$. In each simulation setting, we generated a total of 200 independent datasets of sizes $n = 100$ and 400.

We simulated two cases. One is the case where $(\mathbf{X}^i, \mathbf{T}^i)$ are independent and identically distributed, and the other is the case where they are lagged observations of the response Y. For the iid case, we generated the data as

$$Y^i = f_1(X_1^i)T_1^i + f_2(X_2^i)T_2^i + \sigma(\mathbf{X}^i, \mathbf{T}^i)\epsilon^i,$$

where ϵ^i were iid standard normal random variables that are independent of $(\mathbf{X}^i, \mathbf{T}^i)$, $\mathbf{X}^i$ were from the uniform distribution on the unit cube $[0,1]^2$, $T_1^i \equiv 1$, T_2^i were from the standard normal distribution independently of $\mathbf{X}^i$, and

$$\sigma(\mathbf{x}, \mathbf{t}) = \frac{1}{2} + \frac{t_2^2}{t_1^2 + t_2^2} \exp\left(-2 + \frac{x_1}{2}\right).$$

For the kernel function L in this case, we used $L = K$. For the time series case, we generated the data from a VCAR (Varying Coefficient Autoregressive) model as

$$Y^t = f_1(Y^{t-3})Y^{t-1} + f_2(Y^{t-4})Y^{t-2} + 0.2\epsilon^t,$$

where ϵ^t were iid standard normal random variables. We took $L = K$ in this case, too.

For the parametric families to test, we took

$$\mathcal{F}_0 = \{f_1 : f_1(x) = \theta_0 + \theta_1 x,\ \theta_j \in \mathbb{R} \text{ for } j = 0, 1\}$$

for both the iid and the time series cases. The rejection probabilities for the tests at the nominal levels $\alpha = 0.05$ and 0.10 were obtained based on the 200 pseudo-samples. In these experiments, we took the models

$$f_1(x) = \theta_0 + \theta_1 x + \rho\cos(2\pi x)$$

for various choices of θ_0, θ_1, and ρ in the ranges $-1 \le \theta_0 \le 1$, $0 \le \theta_1 \le 2$, and $0 \le \rho \le 1$. Note that $\rho = 0$ corresponds to the case where H_0 is true. For the second coefficient function f_2, we took $f_2(x) = \cos(2\pi x^2)$.

Tables 1 and 2 contain the rejection probabilities in the case of $(\theta_0, \theta_1) = (1, 1)$ and $(-1, 2)$. The results for other choices of (θ_0, θ_1) are not reported here, but give similar lessons. The results suggest that the test procedure works quite well. The actual levels are fairly close to the corresponding nominal levels for moderate sample sizes. As the sample size increases, the actual levels approach the corresponding nominal levels. Also, the power of the test increases rapidly as the sample size increases, or as the true function f_1 gets more distant from the null, which is consistent with the theory.

Table 1. Rejection probabilities of $H_0 : f_1 \in \mathcal{F}_0$ in the i.i.d. case.

(θ_0, θ_1)	ρ	level α	$n = 100$	$n = 400$
$(1, 1)$	0	0.05	0.065	0.055
		0.10	0.110	0.095
	0.5	0.05	0.255	0.565
		0.10	0.380	0.690
	1	0.05	0.420	0.780
		0.10	0.575	0.885
$(-1, 2)$	0	0.05	0.070	0.040
		0.10	0.110	0.105
	0.5	0.05	0.295	0.510
		0.10	0.370	0.615
	1	0.05	0.375	0.635
		0.10	0.455	0.810

Table 2. Rejection probabilities of $H_0 : f_1 \in \mathcal{F}_0$ in the time series case.

(θ_0, θ_1)	ρ	level α	$n = 100$	$n = 400$
$(1, 1)$	0	0.05	0.035	0.060
		0.10	0.115	0.095
	0.5	0.05	0.210	0.555
		0.10	0.325	0.660
	1	0.05	0.390	0.710
		0.10	0.560	0.845
$(-1, 2)$	0	0.05	0.055	0.060
		0.10	0.120	0.110
	0.5	0.05	0.285	0.515
		0.10	0.350	0.060
	1	0.05	0.370	0.660
		0.10	0.515	0.775

5.3. A data example

There is interest in healthy lifestyles, and body fat contents can be used to assess health, at least in part. Body fat can be predicted from age, weight, height, measurements on abdominal and thigh circumferences, skin-fold measurements, etc. We applied the proposed testing procedure to build a simple, yet accurate, model that predicts numerical body fat percentage values.

The body fat data we used can be found in the “fat” data set in the “UsingR” package in R software. The data are composed of the percentage of body fat determined by underwater weighing and various body circumference measurements for 252 men. Although abdomen circumference is often considered as the main indicator of body fat, we did not include this variable to see how well other variables can be used to predict body fat percentage. We chose the five predictors which were age(years), weight(lbs), height(inches), thigh and wrist circumferences(cm).

The varying coefficient model we considered was

$$Y = f_1(X_1)T_1 + f_2(X_2)T_2 + f_3(X_3)T_3 + \epsilon,$$

where Y is the body fat percentage, X_1 age, X_2 height, X_3 wrist circumference, $T_1 \equiv 1$, T_2 weight, and T_3 thigh circumference. We set the influence of age as a purely nonparametric additive component; put weight and height together, and the two circumference measurements as a pair, as the other two additive components. Since people with higher weight typically have higher body fat percentage but height is understood not to have such a monotone property, we model the influence of (weight, height) as an additive component $f_2(X_2)T_2$. With this model, an increase of weight by one unit with height being fixed, at x_2, as well as other variables, increases the body fat percentage by $f_2(x_2)$ on the average. When weight is fixed, the influence of height is determined by the function f_2. The same interpretation is valid for the pair (thigh circumference, wrist circumference).

In the analysis of the data, we removed the 39th and 42th observations which were considered as outliers. For the parametric families to test, we considered

$$\begin{aligned}
\mathcal{F}_0^{(0)} &= \{f : f(x) = \theta_0,\ \theta_0 \in \mathbb{R}\},\\
\mathcal{F}_0^{(1)} &= \{f : f(x) = \theta_0 + \theta_1 x,\ \theta_j \in \mathbb{R}\},\\
\mathcal{F}_0^{(2)} &= \{f : f(x) = \theta_0 + \theta_1 x + \theta_2 x^2,\ \theta_j \in \mathbb{R}\}.
\end{aligned}$$

We chose $p = 1$ (local linear fitting), the quartic kernel for K, and the product kernel $L(x_1, x_2) = K(x_1)K(x_2)$. We used the same bandwidths h and b_j as in the simulation study; the weight functions w_s and w_{-s} were also the same. Testing $f_1 \in \mathcal{F}_0^{(j)}$, we found that the null hypothesis was rejected for $j = 0$, but accepted for $j = 1$ at the level $\alpha = 0.05$. The hypothesis $f_2 \in \mathcal{F}_0^{(j)}$ was rejected for $j = 0$, but accepted for $j = 1$. Finally, the hypothesis $f_3 \in \mathcal{F}_0^{(j)}$ was rejected for $j = 0$ and $j = 1$, but accepted for $j = 2$. These results suggested that we could take f_1 and f_2 as a linear and f_3 as a quadratic function.

The nonparametric estimates of the f_j are depicted in the three panels of Figure 1, where the parametric estimates of the functions obtained by least squares fitting of the chosen parametric model

$$Y = (\theta_{10} + \theta_{11}X_1)T_1 + (\theta_{20} + \theta_{21}X_2)T_2 + (\theta_{30} + \theta_{31}X_3 + \theta_{32}X_3^2)T_3 + \epsilon \quad (5.1)$$

are also displayed. One can see that the nonparametric estimates are fairly well approximated by the corresponding parametric estimates. The results also suggest that people tend to have more body fat as they get older, and the body

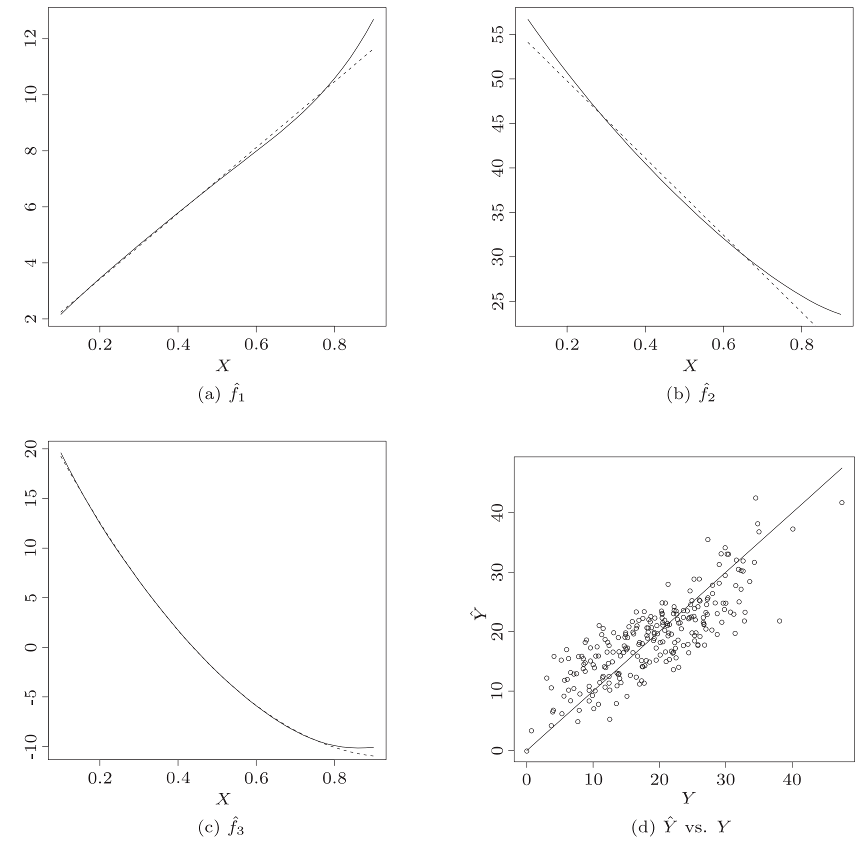

Figure 1. In the panels (a)–(c), the solid curves are the nonparametric estimates of the component functions f_j in the analysis of the body fat data, and the dotted are the parametric estimates of the functions $f_1(x_1) = \theta_{10} + \theta_{11}x_1$, $f_2(x_2) = \theta_{20} + \theta_{21}x_2$ and $f_3(x_3) = \theta_{30} + \theta_{31}x_3 + \theta_{32}x_3^2$, under the chosen parametric model (5.1). The panel (d) is the scatter plot of $(Y_i, \hat{Y}_i)$, where $\hat{Y}^i$ are the predicted values of Y_i under the model (5.1).

fat of people who are taller and have larger wrist circumference is less affected by weight and thigh circumference.

To see how well the chosen parametric model (5.1) fits the data, we plotted $(Y^i, \hat{Y}^i)$ in the bottom-right panel of Figure 1, where $\hat{Y}^i$ are the predicted values of Y_i under the model (5.1). One can see that the model fits the data fairly well. We also computed the R^2 value for the model (5.1) and compared it with the R^2 value for the full linear model that contains all twelve predictors in the regression equation. We found that the model (5.1) with only five predictors yielded $R^2 = 0.6582$, which was higher than 0.6469 of the full linear model with twelve predictors.

6. Technical Details

6.1. Assumptions

Below we collect the assumptions for Theorems 1 and 2. For $j < k < l < m$, let $\psi_{j,k,l,m}$ be the joint probability density function of ($\mathbf{X}^j, \mathbf{T}^j$; $\mathbf{X}^k, \mathbf{T}^k$; $\mathbf{X}^l, \mathbf{T}^l$; $\mathbf{X}^m, \mathbf{T}^m$). Let $\mathcal{X} = \{\mathbf{x} : x_s \in \text{supp}(w_s), \mathbf{x}_{-s} \in \text{supp}(w_{-s})\}$, and for $\varepsilon > 0$ define $\mathcal{X}_\varepsilon = \{\mathbf{x}$: there exists $\mathbf{z} \in \mathcal{X}$ such that $\|\mathbf{z} - \mathbf{x}\| \le \varepsilon\}$.

(A1) The parameter space Θ is compact, and

$$E\{g(X_s, \theta) - g(X_s, \theta')\}^2 w_s(X_s) > 0 \quad \text{for all } \ \theta \neq \theta'.$$

(A2) The function $g(\cdot, \theta)$ has $(p + 1)$ continuous derivatives on supp(w_s); the functions f_j for $j \neq s$ have q bounded derivatives.

(A3) The function $g(x, \theta)$ is twice partially continuously differentiable with respect to θ, and $\dot{g}(x, \theta)$ and $\ddot{g}(x, \theta)$ are continuous in $x \in$ supp(w_s) and $\theta \in \Theta$. For g and its partial derivatives up to order two, all represented by the generic function G, there exists a function $M(\cdot)$ such that

$$\sup_{\theta \in \Theta} G(x, \theta)^2 \le M(x) \quad \text{for all } \ x \ \text{ and } \ Ew_s(X_s)M(X_s) < \infty,$$

and $E\left[w_s(X_s)\dot{g}(X_s, \theta)\dot{g}(X_s, \theta)^\top\right]$ is invertible for all $\theta \in \Theta$.

(A4) The univariate kernel K and $(d - 1)$-variate kernel L are symmetric, Lipschitz continuous, have compact supports with nonempty interiors, and satisfy $\int K(u)\, du = 1 = \int L(\mathbf{u})\, d\mathbf{u}$. While K is nonnegative, the kernel L is of order q.

(A5) The vector process $\{(\mathbf{X}^i, \mathbf{T}^i)\}_{i=1}^n$ is strictly stationary and β-mixing, with mixing coefficients $\beta(k) \le C_1 \rho^k,\ 0 < C_1 < \infty,\ 0 < \rho < 1$.

(A6) On the supports of weight functions w_{-s} and w_s, the functions φ_{-s} and φ_s are uniformly bounded away from zero and infinity. The marginal density φ and $E(T_s T_{s'}|\mathbf{X} = \cdot)$, $1 \le s, s' \le d$, are Lipschitz continuous. Also, $\sigma^2(\cdot, \mathbf{t})$ and $\psi(\cdot, \mathbf{t})$ are equicontinuous. All the above functions have continuous first derivatives on the interior of supports of w_s and w_{-s}.

(A7) The weight functions w_{-s} and w_s are nonnegative, have compact supports with nonempty interiors, and are continuous on their supports, and differentiable on the interior of their supports.

(A8) The error term ϵ^i satisfies $E|\epsilon^i|^{4+c} < \infty$ for some $c > 0$. There exist $\varepsilon > 0$, $\tilde{\sigma}(\mathbf{t})$, and $\tilde{\varphi}_{j,k,l,m}(\mathbf{t}^j, \mathbf{t}^k, \mathbf{t}^l, \mathbf{t}^m)$ such that $\sigma(\mathbf{x}, \mathbf{t}) \le \tilde{\sigma}(\mathbf{t})$, $\psi(\mathbf{x}, \mathbf{t}) \le \tilde{\varphi}(\mathbf{t})$ for all $\mathbf{x} \in \mathcal{X}_\varepsilon$,

$$\psi_{j,k,l,m}(\mathbf{x}^j, \mathbf{t}^j; \mathbf{x}^k, \mathbf{t}^k; \mathbf{x}^l, \mathbf{t}^l; \mathbf{x}^m, \mathbf{t}^m) \le \tilde{\varphi}_{j,k,l,m}(\mathbf{t}^j, \mathbf{t}^k, \mathbf{t}^l, \mathbf{t}^m)$$

for all $\mathbf{x}^j, \mathbf{x}^k, \mathbf{x}^l, \mathbf{x}^m$ in $\mathcal{X}_\epsilon$, and

$$\begin{aligned} &\int \|\mathbf{t}^j\|^{2+c} \tilde{\sigma}(\mathbf{t}_j)^{2+c} \tilde{\varphi}(\mathbf{t}^j)\, d\mathbf{t}^j \ \le\ C < \infty, \\ &\int (\|\mathbf{t}^j\| \|\mathbf{t}^k\| \|\mathbf{t}^l\| \|\mathbf{t}^m\|)^{2+c} \{\tilde{\sigma}(\mathbf{t}^j)\tilde{\sigma}(\mathbf{t}^k)\tilde{\sigma}(\mathbf{t}^l)\tilde{\sigma}(\mathbf{t}^m)\}^{2+c} \\ &\qquad \times \tilde{\varphi}_{j,k,l,m}(\mathbf{t}^j, \mathbf{t}^k, \mathbf{t}^l, \mathbf{t}^m)\, d\mathbf{t}^j d\mathbf{t}^k d\mathbf{t}^l d\mathbf{t}^m \ \le\ C < \infty \end{aligned}$$

for some $c > 0$ and $C > 0$.

(A9) $nh^3 \to \infty$, $h^{2p+1} b_{s,\text{prod}}^{-1} \to 0$, $(nhb_{s,\text{prod}})^{-1/2} \log n = O(n^{-a})$ for some $a > 0$, $(nh \log n)^{1/2} b_{s,\max}^q \to 0$, and $h^{-1} b_{s,\text{prod}}^{-1} b_{s,\max}^{2q} \to 0$ as $n \to \infty$, where $b_{s,\max} = \max\{b_j : j \ne s\}$.

Remark 1. To see how the conditions in (A9) can be reduced to a simpler one on q, the order of the kernel L, suppose $h \sim n^{-2/(4p+5)}$ and $b_j \equiv b \sim (\log n)^{-1} h^{(p+1)/q}$. As discussed in Section 4, the bandwidth order $n^{-2/(4p+5)}$ gives an optimal test. With the magnitude for b_js, the biases that are produced by the local constant fitting in the directions x_j $(j \ne s)$ with a qth order kernel L have the same order of magnitude as the bias due to the pth order local polynomial fitting in the direction of x_s. With these choices the condition $nh^3 \to \infty$ is satisfied for all $p > 0$. Also, $(nh \log n)^{1/2} b_{s,\max}^q \to 0$ for all $q \ge 1$. The condition that $(nhb_{s,\text{prod}})^{-1/2} \log n = O(n^{-a})$ for some $a > 0$ holds if $q > \{2(p+1)(d-1) - d\}/(4p+3)$. The last two conditions and $h^{2p+1} b_{s,\text{prod}}^{-1} \to 0$ and $h^{-1} b_{s,\text{prod}}^{-1} b_{s,\max}^{2q} \to 0$ hold if $q > (p+1)(d-1)/(2p+1)$. Thus, all the conditions in (A9) are satisfied if $q > (p+1)(d-1)/(2p+1)$.

Remark 2. Suppose that we use bandwidths h and b_j that are of the same order, and fit pth order local polynomials for all varying coefficients with both K and L being of second order. Set $b_j = h$, $1 \le j \ne s \le d$, for simplicity. Assume, instead of (A2), that all the coefficient functions are $(p+1)$-times continuously differentiable. Then, one can verify that the results in the theorems are still valid if, instead of (A9), the dimension d is smaller than $2(p+1)$ and the universal bandwidth h satisfies $nh^d \to \infty$ and $nh^{2p+3} \to 0$ as $n \to \infty$.

6.2. Proof of Theorems 1

To prove the theorem, we need to establish that $\hat{\theta}$ converges to θ_0 in probability, as $n \to \infty$. The proof of this is based on the following two lemmas. Let

$$K_h^*(u, \mathbf{t}, \mathbf{x}) = h^{-1} K^*(\frac{u}{h}, \mathbf{t}, \mathbf{x}) \quad \text{and} \quad L_{\mathbf{b}_{-s}}(\mathbf{u}_{-s}) = b_{s,\text{prod}}^{-1} L(\frac{\mathbf{u}_{-s}}{\mathbf{b}_{-s}}).$$

Writing $\xi^i = (\mathbf{X}^i, \mathbf{T}^i, \epsilon^i)$, define

$$\begin{aligned} q_{n,1}(\xi^i) = \epsilon^i \int & \frac{w_s(x_s) w_{-s}(\mathbf{z}_{-s}) w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s, \mathbf{z}_{-s})\varphi(\mathbf{x})} \; \sigma(\mathbf{X}^i, \mathbf{T}^i) \\ & \times K_h^*(X_s^i - x_s, \mathbf{T}^i, x_s, \mathbf{z}_{-s}) K_h^*(u_s - x_s, \mathbf{t}, \mathbf{x}) \\ & \times L_{\mathbf{b}_{-s}}(\mathbf{X}_{-s}^i - \mathbf{z}_{-s}) L_{\mathbf{b}_{-s}}(\mathbf{u}_{-s} - \mathbf{x}_{-s}) \\ & \times \left[g(u_s, \theta_0) - \sum_{k=0}^{p} g^{(k)}(x_s, \theta_0) \frac{(u_s - x_s)^k}{k!} \right] t_s \varphi_{-s}(\mathbf{z}_{-s}) \\ & \times \varphi_s(x_s) \varphi_{-s}(\mathbf{x}_{-s}) \psi(\mathbf{u}, \mathbf{t}) \, d\mathbf{u} \, d\mathbf{t} \, d\mathbf{x} \, d\mathbf{z}_{-s}. \end{aligned} \tag{6.1}$$

Lemma 1. *Under the conditions of Theorem* 1*,*

$$\begin{aligned} S_n(\theta_0) = & h^{2p+2} \|\kappa(\cdot, \theta_0)\|_s^2 + n^{-1} h^{-1} c_2 + n^{-1} h^{-1/2} \gamma_{11}^{1/2} Z_{1,n} \\ & + n^{-1/2} h^{p+1} \nu^{-2} W_{1,n} + o_p(h^{2p+2} + n^{-1} h^{-1/2} + n^{-1/2} h^{p+1}), \end{aligned}$$

where $\nu = E[w_{-s}(\mathbf{X}_{-s})]$*, and* $Z_{1,n}$ *is asymptotically* $N(0,1)$ *and is uncorrelated with* $W_{1,n} \equiv n^{-1/2} h^{-(p+1)} \sum_{i=1}^n q_{n,1}(\xi^i)$*.*

Proof. Define

$$\delta = \left(h + b_{s,\max}^q + \frac{1}{\sqrt{nhb_{s,\text{prod}}}} \right) \log n,$$

and let c be an integer such that $\delta^{c+1} = o(h^{p+2})$. Put $\mathbf{Z}_s = \mathbf{Z}_s(x_s)$, $\mathbf{W}_s^i = (hb_{s,\text{prod}})^{-1} \mathbf{W}_s(x_s^i, \mathbf{X}_{-s}^i)$. Then, we can write

$$\begin{aligned} \hat{f}_s(x_s) - g(x_s, \theta_0) = & \sum_{j=1}^{3} \left[p_{j,n}(x_s) + \sum_{l=1}^{c} r_{j,l,n}(x_s) + \pi_{s,j,n}(x_s) \right] \\ & \times \left[n^{-1} \sum_{i=1}^{n} w_{-s}(\mathbf{X}_{-s}^i) \right]^{-1}, \end{aligned}$$

where, for $1 \le j \le 3$ and $1 \le l \le c$,

$$p_{j,n}(x_s) = n^{-2} \sum_{i=1}^{n} \sum_{i'=1}^{n} \frac{w_{-s}(\mathbf{X}_{-s}^i)}{\varphi(x_s, \mathbf{X}_{-s}^i)} K_h^*(X_s^{i'} - x_s, \mathbf{T}^{i'}, x_s, \mathbf{X}_{-s}^i)$$

$$
\begin{aligned}
&\times L_{\mathbf{b}_{-s}}(\mathbf{X}_{-s}^{i'}-\mathbf{X}_{-s}^{i})H_j^{i,i'}(x_s),\\
r_{j,l,n}(x_s) &= n^{-1}\sum_{i=1}^{n}\frac{w_{-s}(\mathbf{X}_{-s}^{i})}{\varphi(x_s,\mathbf{X}_{-s}^{i})}\,\mathbf{e}_0^\top D_s^{-1}(x_s,\mathbf{X}_{-s}^{i})\\
&\times\left[I_{p+d}-\frac{\mathbf{Z}_s^\top\mathbf{W}_s^i\mathbf{Z}_sD_s^{-1}(x_s,\mathbf{X}_{-s}^{i})}{\varphi(x_s,\mathbf{X}_{-s}^{i})}\right]^l\\
&\times\mathbf{Z}_s^\top\mathbf{W}_s^i\big(H_j^{i,k}(x_s)\big)_{k=1}^n,\\
\pi_{s,j,n}(x_s) &= n^{-1}\sum_{i=1}^{n}w_{-s}(\mathbf{X}_{-s}^{i})\mathbf{e}_0^\top\lambda_{s,n}(x_s,\mathbf{X}_{-s}^{i})\mathbf{Z}_s^\top\mathbf{W}_s^i\big(H_j^{i,k}(x_s)\big)_{k=1}^n,\\
H_j^{i,i'}(x_s) &= \begin{cases}\sigma(\mathbf{X}^{i'},\mathbf{T}^{i'})\epsilon^{i'} & j=1,\\ T_s^{i'}\left\{g(X_s^{i'},\theta_0)-\sum_{k=0}^{p}g^{(k)}(x_s,\theta_0)\frac{(X_s^{i'}-x_s)^k}{k!}\right\} & j=2,\\ \sum_{s'\neq s}^{d}\{f_{s'}(X_{s'}^{i'})-f_{s'}(X_{s'}^{i})\}T_{s'}^{i'} & j=3.\end{cases}
\end{aligned}
$$

In the above, $\lambda_{s,n}(\mathbf{x})$ depends on $(\mathbf{X}^i,\mathbf{T}^i)$ only, which is $o(h^{p+2})$ a.s. uniformly for $\mathbf{x}\in\text{supp}(w_s)\times\text{supp}(w_{-s})$, and $H_j^{i,i'}(x_s)$ does not depend on i for $j=1,2$.

Using the theory of von Mises' differential statistics, see Yoshihara (1993) for example, we may write

$$
\begin{aligned}
n^{-1}\sum_{k=1}^{n}p_{1,n}^2(X_s^k)w_s(X_s^k) &= n^{-1}h^{-1}\nu^2c_2+n^{-1}h^{-1/2}\nu^2\gamma_{11}^{1/2}Z_{1,n},\\
n^{-1}\sum_{k=1}^{n}p_{3,n}^2(X_s^k)w_s(X_s^k) &= o_p(n^{-1}h),\\
n^{-1}\sum_{k=1}^{n}r_{1,l,n}^2(X_s^k)w_s(X_s^k) &= o_p(n^{-1}),\\
n^{-1}\sum_{k=1}^{n}r_{3,l,n}^2(X_s^k)w_s(X_s^k) &= o_p(\delta^2b_{s,\max}^{2q}),\\
n^{-1}\sum_{k=1}^{n}\pi_{s,1,n}^{2m}(X_s^k)\pi_{s,3,n}^{2(1-m)}(X_s^k)w_s(X_s^k) &= o_p(h^{2p+4}),\\
n^{-1}\sum_{k=1}^{n}p_{1,n}(X_s^k)p_{3,n}(X_s^k)w_s(X_s^k) &= o_p(n^{-1}),\\
n^{-1}\sum_{k=1}^{n}p_{1,n}(X_s^k)r_{1,l,n}(X_s^k)w_s(X_s^k) &= o_p(n^{-1}h^{-1/2}\delta),\\
n^{-1}\sum_{k=1}^{n}p_{1,n}(X_s^k)r_{3,l,n}(X_s^k)w_s(X_s^k) &= o_p(n^{-1/2}\delta\, b_{s,\max}^{q}),\\
n^{-1}\sum_{k=1}^{n}p_{1,n}(X_s^k)\pi_{s,1,n}^{m}(X_s^k)\pi_{s,3,n}^{1-m}(X_s^k)w_s(X_s^k) &= o_p(n^{-1/2}h^{p+(3/2)})
\end{aligned}
$$

for $m = 0, 1$ and $1 \le l \le c$, where $Z_{1,n} \xrightarrow{d} N(0,1)$. In fact, derivation of the first approximation starts with the representation

$$n^{-1}\sum_{k=1}^{n} p_{1,n}^2(X_s^k)w_s(X_s^k) = n^{-5}\sum_{i_1,\ldots,i_5} q_n(\xi^{i_1},\ldots,\xi^{i_5})$$

for some function q_n that is symmetric in its arguments, where $\xi^i = (\mathbf{X}^i, \mathbf{T}^i, \epsilon^i)$. From the theory of von Mises' differential statistics, the leading term of the V-statistic is its second-order projection, given by $V_{n,2} \equiv n^{-2}\sum_{i,j}^{n} q_{n,2}(\xi^i, \xi^j)$, where

$$q_{n,2}(\xi^i,\xi^j) = \int q_n(\xi^i,\xi^j,\xi^{k_1},\xi^{k_2},\xi^{k_3})\, dF(\xi^{k_1})dF(\xi^{k_2})dF(\xi^{k_3}), \tag{6.2}$$

and F is the distribution function of ξ^i. The first term in the approximation of the V-statistic comes from $n^{-2}\sum_{i=1}^{n} q_{n,2}(\xi^i,\xi^i)$, the diagonal sum of $V_{n,2}$, and the second term from its off-diagonal sum $n^{-2}\sum_{i\ne j}^{n} q_{n,2}(\xi^i,\xi^j)$.

Using the same theory again, we can prove

$$\begin{aligned}
n^{-1}\sum_{k=1}^{n} p_{2,n}^2(X_s^k)w_s(X_s^k) &= h^{2p+2}\nu^2 E[w_s(X_s)\kappa^2(X_s,\theta_0)] + o_p(h^{2p+2}),\\
n^{-1}\sum_{k=1}^{n} p_{1,n}(X_s^k)p_{2,n}(X_s^k)w_s(X_s^k) &= n^{-1/2}h^{p+1}W_{1,n} + o_p(n^{-1/2}h^{p+1}),\\
n^{-1}\sum_{k=1}^{n} r_{2,l,n}^2(X_s^k)w_s(X_s^k) &= o_p(\delta^2 h^{2p+2}),\\
n^{-1}\sum_{k=1}^{n} \pi_{s,2,n}^2(X_s^k)w_s(X_s^k) &= o_p(h^{2p+4}),
\end{aligned}$$

for $1 \le l \le c$. The fact that $W_{1,n}$ and $Z_{1,n}$ are uncorrelated follows from the fact that the covariances of $q_{n,1}(\xi^i)$ at (6.1) and $q_{n,2}(\xi^j,\xi^k)$ at (6.2), for all i and all $j < k$, are zero. Now Lemma 1 follows from the assumptions on the bandwidths h and $\mathbf{b}_{-s}$ in (A9), since

$$\begin{aligned}
S_n(\theta_0) = n^{-1}\sum_{i=1}^{n}&\left[\sum_{j=1}^{3}\left\{p_{j,n}(X_s^i) + \sum_{l=1}^{c} r_{j,l,n}(X_s^i) + \pi_{s,j,n}(X_s^i)\right\}\right]^2 w_s(X_s^i)\\
&\times\left\{n^{-1}\sum_{i=1}^{n} w_{-s}(\mathbf{X}_{-s}^i)\right\}^{-2}.
\end{aligned}$$

Lemma 2. *Under the conditions of Theorem* 1*,*

$$n^{-1}\sum_{i=1}^{n}\left[g(X_s^i,\theta_0)-g(X_s^i,\theta)\right]^2 w_s(X_s^i)=S(\theta)+o_p(1)$$

uniformly for $\theta\in\Theta$*, where* $S(\theta)=E\{g(X_s,\theta_0)-g(X_s,\theta)\}^2 w_s(X_s)$*.*

Proof. By (A3), there exists an envelope function $G\in L_1(\varphi_s)$ such that

$$\sup_{\theta\in\Theta}\left[g(x,\theta)-g(x,\theta_0)\right]^2 w_s(x)\ \le\ G(x_s)$$

for all $x\in\mathrm{supp}(w_s)$. Since the lemma holds for independent and identically distributed X_s^i, it is valid, too, for the β-mixing process $\{X_s^i\}_{i=1}^n$, by Theorem 1 of Nobel and Dembo (1993).

Lemma 3. *Under the conditions of Theorem* 1*,* $\hat\theta\to\theta_0$ *in probability as* $n\to\infty$*.*

Proof. By Lemma 1,

$$n^{-1}\sum_{i=1}^{n}\{\hat f_s(X_s^i)-g(X_s^i,\theta_0)\}^2 w_s(X_s^i)=o_p(1). \tag{6.3}$$

Also, from Lemma 2

$$\sup_{\theta\in\Theta}\left|n^{-1}\sum_{i=1}^{n}\{g(X_s^i,\theta_0)-g(X_s^i,\theta)\}^2 w_s(X_s^i)-S(\theta)\right|=o_p(1). \tag{6.4}$$

Hölder's inequality, together with (6.3) and (6.4), gives uniform convergence of $S_n(\theta)$ to $S(\theta)$:

$$\sup_{\theta\in\Theta}|S_n(\theta)-S(\theta)|=o_p(1).$$

Since $S(\theta)$ is minimized uniquely at θ_0 and is continuous in θ, the lemma follows.

Define

$$\begin{aligned}q_{n,3}(\xi^i)=\epsilon^i\int \dot g_s(x_s,\theta_0)w_s(x_s)\frac{w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s,\mathbf{x}_{-s})}\,K_h^*(X_s^i-x_s,\mathbf{T}^i,\mathbf{x})\\ \times L_{\mathbf{b}_{-s}}(\mathbf{X}_{-s}^i-\mathbf{x}_{-s})\sigma(\mathbf{X}^i,\mathbf{T}^i)\varphi_{-s}(\mathbf{x}_{-s})\varphi(x_s)\,d\mathbf{x}.\end{aligned} \tag{6.5}$$

Let $\dot S_n(\theta)$ and $\ddot S_n(\theta)$ denote the gradient and the Hessian matrix of $S_n(\theta)$, respectively.

Lemma 4. *Suppose that* (A1)−(A9) *hold. Then, under the null hypothesis,*

$$\dot S_n(\theta_0)=-2\,h^{p+1}\tau(\theta_0)-2\,n^{-1/2}\nu^{-1}W_{2,n}+o_p(h^{p+1}+n^{-1/2}),$$

where $W_{2,n}\equiv n^{-1/2}\sum_{i=1}^n q_{n,3}(\xi^i)$ *is uncorrelated with* $Z_{1,n}$ *in Lemma 1, and* $\tau(\theta)=E[w_s(X_s)\kappa(X_s,\theta)\dot g(X_s,\theta)]$*.*

Proof. First, we observe that

$$\dot{S}_n(\theta_0) = -2\,n^{-1}\sum_{i=1}^{n}\left[\sum_{j=1}^{3}\left\{p_{j,n}(X_s^i) + \sum_{l=1}^{c} r_{j,l,n}(X_s^i) + \pi_{s,j,n}(X_s^i)\right\}\right] \times \dot{g}(X_s^i,\theta_0)w_s(X_s^i)\left\{n^{-1}\sum_{i=1}^{n} w_{-s}(\mathbf{X}_{-s}^i)\right\}^{-1}.$$

As in the proof of Lemma 1, we can verify that

$$n^{-1}\sum_{i=1}^{n} r_{1,l,n}(X_s^i)\dot{g}(X_s^i,\theta_0)w_s(X_s^i) = o_p(n^{-1/2}\delta),$$
$$n^{-1}\sum_{i=1}^{n} r_{2,l,n}(X_s^i)\dot{g}(X_s^i,\theta_0)w_s(X_s^i) = o_p(\delta\, h^{p+1}),$$
$$n^{-1}\sum_{i=1}^{n} r_{3,l,n}(X_s^i)\dot{g}(X_s^i,\theta_0)w_s(X_s^i) = o_p(\delta\, b_{s,\max}^q),$$
$$n^{-1}\sum_{i=1}^{n} \pi_{s,j,n}(X_s^i)\dot{g}(X_s^i,\theta_0)w_s(X_s^i) = o_p(h^{p+2}),$$
$$n^{-1}\sum_{i=1}^{n} p_{2,n}(X_s^i)\dot{g}(X_s^i,\theta_0)w_s(X_s^i) = h^{p+1}\nu\tau(\theta_0) + o_p(h^{p+1}),$$
$$n^{-1}\sum_{i=1}^{n} p_{3,n}(X_s^i)\dot{g}(X_s^i,\theta_0)w_s(X_s^i) = o_p(n^{-1/2}h^{1/2}).$$

Now we derive an expansion of the term in $\dot{S}_n(\theta_0)$ that involves $p_{1,n}(X_s^i)$. One can write

$$n^{-1}\sum_{i=1}^{n} p_{1,n}(X_s^i)\dot{g}(X_s^i,\theta_0)w_s(X_s^i) = n^{-3}\sum_{i,j,k} q_n^*(\xi_i,\xi_j,\xi_k), \tag{6.6}$$

where $q_n^*(\xi^i,\xi^j,\xi^k) = \sum_{i',j',k'}\tilde{q}_n(\xi^{i'},\xi^{j'},\xi^{k'})/3!$, with the summation being for all permutations (i',j',k') of (i,j,k), and

$$\tilde{q}_n(\xi^i,\xi^j,\xi^k) = \dot{g}(X_s^k,\theta_0)w_s(X_s^k)\frac{w_{-s}(\mathbf{X}_{-s}^i)}{\varphi(X_s^k,\mathbf{X}_{-s}^i)}\,K_h^*(X_s^j - X_s^k,\mathbf{T}^j,X_s^k,\mathbf{X}_{-s}^i) \times L_{\mathbf{b}_{-s}}(\mathbf{X}_{-s}^j - \mathbf{X}_{-s}^i)\sigma(\mathbf{X}^j,\mathbf{T}^j)\epsilon^j.$$

Note that $q_{n,3}(\xi^i)$ defined at (6.5) has mean zero. Furthermore, from the theory of von Mises' differential statistics, the leading term of the V-statistic at (6.6) is its first-order projection, so that

$$n^{-1}\sum_{i=1}^{n} p_{1,n}(X_s^i)\dot{g}(X_s^i,\theta_0)w_s(X_s^i) = n^{-1}\sum_{j=1}^{n} q_{n,1}(\xi^j) + o_p(n^{-1}).$$

As in the proof of Lemma 1, one can verify that $q_{n,3}(\xi^i)$ at (6.5) and $q_{n,2}(\xi^j,\xi^k)$ at (6.2) are uncorrelated for all i and all $j<k$, so that $W_{2,n}$ is also uncorrelated with $Z_{1,n}$. This completes the proof of the lemma.

Lemma 5. *Define* $\Gamma(\theta) = E[w_s(X_s)\dot{g}(X_s,\theta)\dot{g}(X_s,\theta)^\top]$. *Under the conditions of Theorem* 1, $\ddot{S}_n(\check{\theta}) = 2\,\Gamma(\theta_0) + o_p(1)$ *for any* $\check{\theta}$ *that lies on the line segment between* $\hat{\theta}$ *and* θ_0.

Proof. Since $\ddot{S}_n(\check{\theta}) - \ddot{S}_n(\theta_0) = o_p(1)$ by Lemma 3, and $\ddot{S}_n(\theta_0) = 2\,\Gamma(\theta_0) + o_p(1)$, we establish the lemma.

The following lemma gives the asymptotic joint distribution of $W_{1,n}$ in Lemma 1 and $W_{2,n}$ in Lemma 4. To state the lemma, we need more notation. Define

$$V_{11}(\theta) = E\left[w_s^2(X_s)\eta(\mathbf{X},\mathbf{T})\kappa(X_s,\theta)^2\varphi_s(X_s)\left\{\int K^*(u,\mathbf{T},\mathbf{X})\,du\right\}^2\right],$$

$$V_{12}(\theta) = E\left[w_s^2(X_s)\dot{g}(X_s,\theta)\eta(\mathbf{X},\mathbf{T})\kappa(X_s,\theta)\varphi_s(X_s)\left\{\int K^*(u,\mathbf{T},\mathbf{X})\,du\right\}^2\right],$$

$$V_{22}(\theta) = E\left[w_s^2(X_s)\dot{g}(X_s,\theta)\dot{g}(X_s,\theta)^\top\eta(\mathbf{X},\mathbf{T})\varphi_s(X_s)\left\{\int K^*(u,\mathbf{T},\mathbf{X})\,du\right\}^2\right].$$

Lemma 6. *Under the conditions of Theorem* 1, $W_{1,n}$ *and* $W_{2,n}$ *are jointly asymptotically* $N(\mathbf{0},\Sigma)$, *where*

$$\Sigma = \begin{pmatrix} \nu^4 V_{11}(\theta_0) & \nu^3 V_{12}(\theta_0)^\top \\ \nu^3 V_{12}(\theta_0) & \nu^2 V_{22}(\theta_0) \end{pmatrix}.$$

Proof. The asymptotic normality can be proved by using Lemma 3.2 of Hjellvik, Yao, and Tjøstheim (1998), for example. The formula for the variance matrix of $(W_{1,n}, W_{2,n}^\top)^\top$ can be obtained by employing the change of variable technique.

Proof of Theorem 1. Since $\dot{S}_n(\hat{\theta}) = 0$, it follows that

$$\begin{aligned} S_n(\theta_0) &= S_n(\hat{\theta}) + \frac{1}{2}(\hat{\theta}-\theta_0)^\top\ddot{S}_n(\tilde{\theta})(\hat{\theta}-\theta_0) \\ &= S_n(\hat{\theta}) + \frac{1}{2}\dot{S}_n(\theta_0)^\top\ddot{S}_n(\bar{\theta})^{-1}\ddot{S}_n(\tilde{\theta})\ddot{S}_n(\bar{\theta})^{-1}\dot{S}_n(\theta_0), \qquad (6.7) \end{aligned}$$

where $\tilde{\theta}$ and $\bar{\theta}$ lie on the line segment joining $\hat{\theta}$ and θ_0. By Lemma 1, we have

$$\begin{aligned} S_n(\theta_0) = \|\kappa(\cdot,\theta_0)\|_s^2 h^{2p+2} + c_2 n^{-1}h^{-1} \\ +\gamma_{11}^{1/2} n^{-1}h^{-1/2} Z_{1,n} + \nu^{-2} n^{-1/2} h^{p+1} W_{1,n} \\ +o_p(h^{2p+2} + n^{-1}h^{-1/2} + n^{-1/2}h^{p+1}), \end{aligned} \tag{6.8}$$

where ν, $Z_{1,n}$ and $W_{1,n}$ are defined in the lemma. Lemma 1 also asserts that $Z_{1,n}$ and $W_{1,n}$ are uncorrelated. Furthermore, by Lemmas 4 and 5, the following expansions hold:

$$\dot{S}_n(\theta_0) = -2\,h^{p+1}\tau(\theta_0) - 2\,\nu^{-1}n^{-1/2}W_{2,n} + o_p(h^{p+1} + n^{-1/2}), \tag{6.9}$$

$$\ddot{S}_n(\check{\theta}) = 2\,\Gamma(\theta_0) + o_p(1), \tag{6.10}$$

where $\tau(\theta)$ and $W_{2,n}$ are defined in Lemma 4, $\Gamma(\theta)$ in Lemma 5, and $\check{\theta}$ is an arbitrary stochastic term that lies on the line segment joining $\hat{\theta}$ and θ_0. Lemma 4 also asserts that $W_{2,n}$ and $Z_{1,n}$ are uncorrelated. The last two expansions yield

$$\begin{aligned} &\dot{S}_n(\theta_0)^\top \ddot{S}_n(\bar{\theta})^{-1}\ddot{S}_n(\tilde{\theta})\ddot{S}_n(\bar{\theta})^{-1}\dot{S}_n(\theta_0) \\ &\quad = 2\,\tau(\theta_0)^\top\Gamma(\theta_0)^{-1}\tau(\theta_0)h^{2p+2} + 4\,\nu^{-1}\tau(\theta_0)^\top\Gamma(\theta_0)^{-1}n^{-1/2}h^{p+1}W_{2,n} \\ &\quad\quad + O_p(n^{-1}) + o_p(h^{2p+2} + n^{-1/2}h^{p+1}). \end{aligned} \tag{6.11}$$

The expansions (6.8), (6.11), and Lemma 6 conclude the theorem.

6.3. Proof of Theorem 2

Define $\mu(\theta,\Delta) = E[w_s(X_s)\kappa(X_s,\theta)\Delta(X_s)]$. Let

$$\begin{aligned} q_{n,4}(\xi^i) = \epsilon^i \int \Delta(x_s)w_s(x_s)\frac{w_{-s}(\mathbf{x}_{-s})}{\varphi(x_s,\mathbf{x}_{-s})}\, K_h^*(X_s^i - x_s, \mathbf{T}^i, \mathbf{x}) \\ \times L_{\mathbf{b}_{-s}}(\mathbf{X}_{-s}^i - \mathbf{x}_{-s})\sigma(\mathbf{X}^i,\mathbf{T}^i)\varphi_{-s}(\mathbf{x}_{-s})\varphi(x_s)\,d\mathbf{x}. \end{aligned}$$

Lemma 7. *Under the conditions of Theorem 2,*

$$\begin{aligned} &2\,n^{-1}\rho_n \sum_{i=1}^n \Big\{\hat{f}_s(X_s^i) - g(X_s^i,\theta_1) - \rho_n\Delta(X_s^i)\Big\}\,\Delta(X_s^i) \\ &\quad = \ 2\,\mu(\theta_1,\Delta)h^{p+1}\rho_n + 2\,n^{-1/2}\rho_n\nu^{-1}W_{3,n} + o_p(h^{p+1}\rho_n + n^{-1/2}\rho_n), \end{aligned}$$

where $W_{3,n} \equiv n^{-1/2}\sum_{i=1}^n q_{n,4}(\xi^i)$ *is uncorrelated with* $Z_{1,n}$.

Proof. The proof is the same as that of Lemma 4, except that $g(\cdot,\theta_0)$ and $\dot{g}(\cdot,\theta_0)$ there are replaced by $g(\cdot,\theta_1) + \rho_n\Delta$ and Δ, respectively.

In the following lemma, let $W'_{1,n}$ and $W'_{2,n}$ denote those in Lemmas 1 and 4, respectively, with θ_0 being replaced by θ_1. Define

$$V_{13}(\Delta) = E\left[w_s^2(X_s)\Delta(X_s)\eta(\mathbf{X},\mathbf{T})\kappa(X_s)\varphi_s(X_s)\left\{\int K^*(u,\mathbf{T},\mathbf{X})\,du\right\}^2\right],$$

$$V_{23}(\theta,\Delta) = E\left[w_s^2(X_s)\dot{g}(X_s,\theta)\Delta(X_s)\eta(\mathbf{X},\mathbf{T})\varphi_s(X_s)\left\{\int K^*(u,\mathbf{T},\mathbf{X})\,du\right\}^2\right],$$

$$V_{33}(\Delta) = E\left[w_s^2(X_s)\Delta(X_s)^2\eta(\mathbf{X},\mathbf{T})\varphi_s(X_s)\left\{\int K^*(u,\mathbf{T},\mathbf{X})\,du\right\}^2\right],$$

Lemma 8. *Under the conditions of Theorem* 2, $W'_{1,n}$, $W'_{2,n}$, *and* $W_{3,n}$ *are jointly asymptotically* $N(\mathbf{0},\Sigma')$, *where*

$$\Sigma' = \begin{pmatrix} \nu^4 V_{11}(\theta_1) & \nu^3 V_{12}(\theta_1)^\top & \nu^3 V_{13}(\Delta) \\ \nu^3 V_{12}(\theta_1) & \nu^2 V_{22}(\theta_1) & \nu^2 V_{23}(\theta_1,\Delta) \\ \nu^3 V_{13}(\Delta) & \nu^2 V_{23}(\theta_1,\Delta)^\top & \nu^2 V_{33}(\Delta) \end{pmatrix}.$$

Proof. The proof is similar to that of Lemma 6.

Proof of Theorem 2. Under the assumptions of the theorem, one can show, similarly as in the proof of Lemma 3, that $\hat{\theta}_n \to \theta_1$ in probability. Also, the expression at (6.7) is still valid with θ_0 being replaced by θ_1. For an expansion of $S_n(\theta_1)$, we note that

$$S_n(\theta_1) = n^{-1}\sum_{i=1}^n \left\{\hat{f}_s(X_s^i) - f_{s,n}(X_s^i)\right\}^2 w_s(X_s^i) + n^{-1}\rho_n^2\sum_{i=1}^n \Delta(X_s)^2 w_s(X_s^i)$$
$$+\,2\,n^{-1}\rho_n\sum_{i=1}^n\left\{\hat{f}_s(X_s^i) - f_{s,n}(X_s^i)\right\}\Delta(X_s).$$

Call these three sums $S_{n,1}$, $S_{n,2}$, and $S_{n,3}$, respectively. Expansion of the first term $S_{n,1}$ is the same as that in Lemma 1, with $\kappa(\theta_0)$ being replaced by $\kappa(\theta_1)$. In the proof of the lemma, we only need to replace $g(\cdot,\theta_0)$ by $g(\cdot,\theta_1)+\rho_n\Delta$. The additional term $\rho_n\Delta$ contributes $O_p(h^{2p+2}\rho_n)$ only in the expansion, which is absorbed into $o_p(h^{2p+2})$. For the second term $S_{n,2}$, we have $S_{n,2} = \rho_n^2 + O_p(n^{-1/2}\rho_n^2)$. By Lemma 7, the third term has the expansion

$$S_{n,3} = 2\,\mu(\theta_1,\Delta)h^{p+1}\rho_n + 2\,\nu^{-1}n^{-1/2}\rho_n W_{3,n} + o_p(h^{p+1}\rho_n + n^{-1/2}\rho_n),$$

where $W_{3,n}$ defined in the lemma is uncorrelated with $Z_{1,n}$.

For an expansion of $\dot{S}_n(\theta_1)$, we observe that

$$\dot{S}_n(\theta_1) = -2\,n^{-1}\sum_{i=1}^{n}\left\{\hat{f}_s(X_s^i) - f_{s,n}(X_s^i)\right\}\dot{g}(X_s^i,\theta_1)w_s(X_s^i)$$
$$-2n^{-1}\rho_n\sum_{i=1}^{n}\Delta(X_s)\dot{g}(X_s^i,\theta_1)w_s(X_s^i).$$

Call these two terms $\dot{S}_{n,1}$ and $\dot{S}_{n,2}$. The first term has the same expansion as the one at (6.9) with $\tau(\theta_0)$ being replaced by $\tau(\theta_1)$. Since $\Delta \perp \dot{g}_s(\cdot,\theta_1)$, we obtain $\dot{S}_{n,2} = O_p(n^{-1/2}\rho_n)$. Since (6.10) continues to hold with $\Gamma(\theta_1)$ replacing $\Gamma(\theta_0)$, we get

$$\begin{aligned}V_n &= c_1(\theta_1)h^{2p+2} + 2\,\mu(\theta_1,\Delta)h^{p+1}\rho_n + \rho_n^2 + c_2n^{-1}h^{-1} + \gamma_{11}^{1/2}n^{-1}h^{-1/2}Z_{1,n}\\ &\quad + n^{-1/2}h^{p+1}\left[\nu^{-2}W_{1,n}' - 2\,\nu^{-1}\tau(\theta_1)^{\top}\Gamma(\theta_1)^{-1}W_{2,n}'\right] + 2\,\nu^{-1}n^{-1/2}\rho_nW_{3,n}\\ &\quad + o_p\left(h^{2p+2} + h^{p+1}\rho_n + n^{-1}h^{-1/2} + n^{-1/2}h^{p+1} + n^{-1/2}\rho_n\right),\end{aligned}$$

where $Z_{1,n}$ is the same as the one in the proof of Theorem 1. Recall that $W_{j,n}'$ for $j=1,2$ and $W_{3,n}$ are uncorrelated with $Z_{1,n}$. The asymptotic normality of $Z_{j,n}$, $j=1,2,3$, and the formula for the covariance of $Z_{2,n}$ and $Z_{3,n}$ follow from Lemma 8.

Acknowledgement

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2008-314-C00046).

References

Cai, Z., Fan, J. and Li, R. (2000). Efficient estimation and inferences for varying-coefficient models. *J. Amer. Statist. Assoc.* **95**, 888-902.

Cai, Z., Fan, J. and Yao, Q. (2000). Functional-coefficient regression models for nonlinear times series. *J. Amer. Statist. Assoc.* **95**, 941-956.

Chen, R. and Tsay, R. S. (1993). Functional-coefficient autoregressive models. *J. Amer. Statist. Assoc.* **88**, 298-308.

Emarkov, M. (2003). On asymptotic minimaxity of kernel-based tests. *ESAIM: Probab. Statist.* **7**, 279-312.

Franke, J., Kreiss, J. -P. and Mammen, E. (2002). Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli* **8**, 1-39.

Franke, J., Kreiss, J. -P., Mammen, E. and Neumann, M. H. (2002). Properties of the nonparametric autoregressive bootstrap. *J. Time Ser. Anal.* **23**, 555-585.

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. *J. Amer. Statist. Assoc.* **76**, 817-823.

Härdle, W., Huet, S., Mammen, E. and Sperlich, S. (2004). Bootstrap inference in semiparametric generalized additive models. *Econom. Theory* **20**, 265-300.

Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21**, 1926-1947.

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

Hastie, T. J. and Tibshirani, R. J. (1993). Varying-coefficient models. *J. Roy. Statist. Soc. Ser. B* **55**, 757-796.

Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998). Linearity testing using local polynomial approximation. *J. Statist. Plann. Inference* **68**, 295-321.

Ingster, Y. I. (1993). Asymptotically minimax hypothesis testing for non-parametric alternatives I, II, III. *Math. Methods Statist.* **2**, 85-114, 171-189, 264-268.

Lepski, O. V. and Tsybakov, A. B. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields* **117**, 17-48.

Li, Q. and Wang, S. (1998). A simple consistent bootstrap test for a parametric regression function. *J. Econometrics* **87**, 145-165.

Liu, R. (1988). Bootstrap Procedures under some non i.i.d. models. *Ann. Statist.* **16**, 1696-1708.

Mammen, E. (1992). *When Does Bootstrap Work: Asymptotic Results and Simulations.* Lecture Notes in Statistics **77**. Springer-Verlag, Berlin.

Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Ann. Statist.* **27**, 1443-1490.

Nobel, A. and Dembo, A. (1993). A Note on uniform laws of averages for dependent processes. *Statist. Probab. Lett.* **17**, 169-172.

Pouet, C. (2001). An asymptotically optimal test for a parametric set of regression functions against a non-parametric alternative. *J. Statist. Plann. Inference* **98**, 177-189.

Sperlich, S. Tjøstheim, D. and Yang, L. (2002). Nonparametric estimation and testing of interaction in additive models. *Econom. Theory* **18**, 197-251.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.* **14**, 1261-1350.

Yang, L., Park, B. U., Xue, L. and Härdle, W. (2006). Estimation and testing for varying coefficients in additive models with marginal integration. *J. Amer. Statist. Assoc.* **101**, 1212-1227.

Yoshihara, K. (1993), *Weakly Dependent Stochastic Sequences and Their Applications.* Sanseido, Tokyo.

Department of Statistics, Seoul National University, Seoul 151-747, Korea.

E-mail: bupark@stats.snu.ac.kr

Hyundai Capital, 15-21 Youido-dong Yeongoeungpo-Gu, Seoul, 150-706, Korea.

E-mail: musigma@lycos.co.kr

Department of Statistics, Seoul National University, Seoul 151-747, Korea.

E-mail: pmsfly@gmail.com

(Received March 2009; accepted October 2009)