ELSEVIER

Journal of Econometrics 112 (2003) 295–325

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

# Consistent specification tests for semiparametric/nonparametric models based on series estimation methods

Qi Li[a,*], Cheng Hsiao[b], Joel Zinn[c]

[a] *Department of Economics, Texas A&M University, College Station, TX 77843, USA*
[b] *Department of Economics, University of Southern California, Los Angeles, CA 90089, USA*
[c] *Department of Mathematics, Texas A&M University, College Station, TX 77843, USA*

**Abstract**

This paper considers the problem of consistent model specification tests using series estimation methods. The null models we consider in this paper all contain some nonparametric components. A leading case we consider is to test for an additive partially linear model. The null distribution of the test statistic is derived using a central limit theorem for Hilbert-valued random arrays. The test statistic is shown to be able to detect local alternatives that approach the null models at the order of $O_p(n^{-1/2})$. We show that the wild bootstrap method can be used to approximate the *null* distribution of the test statistic. A small Monte Carlo simulation is reported to examine the finite sample performance of the proposed test. We also show that the proposed test can be easily modified to obtain series-based consistent tests for other semiparametric/nonparametric models.


*JEL classification:* C12; C14

*Keywords:* Consistent tests; Semiparametric models; Series estimation; Wild bootstrap

## 1. Introduction

Semiparametric/nonparametric methods have become increasingly popular because they avoid imposing many strong a priori assumptions associated with a parametric approach. Nevertheless, most applications of semiparametric/nonparametric methods

* Corresponding author. Tel.: + 1-979-845-9954; fax: +1-979-847-8757.
*E-mail address:* qi@econ.tamu.edu (Q. Li).

0304-4076/02/$ - see front matter 
PII: S0304-4076(02)00198-7

have been limited to cases involving only a small number of variables because of the issue of "curse of dimensionality". Stone (1985, 1986), Andrews (1991), Andrews and Whang (1990), Tjostheim and Auestad (1994), Newey (1994, 1995), Linton and Nielsen (1995), etc., have proposed estimating an additive model of the form, [1]

$$Y_i = m_1(X_{1i}) + m_2(X_{2i}) + \cdots + m_L(X_{Li}) + U_i, \tag{1.1}$$

to get around the issue of "curse of dimensionality", where $X_{li}$'s are scalar variables, $l = 1, \ldots, L$. Model (1.1) has the advantage that it only involves one-dimensional nonparametric functions $m_l(\cdot)$ and hence the "curse of dimensionality" is greatly reduced. However, one restrictive assumption of model (1.1) is that it does not allow any interaction terms among the $X_{li}$'s.

To maintain the simplicity of the additive model while allowing for the presence of interaction terms and different $m_l(\cdot)$ functions to have some common overlapping variables, the following more general model has been suggested:

$$Y_i = z_0(X_i)'\gamma + m_1(X_{1i}) + m_2(X_{2i}) + \cdots + m_L(X_{Li}) + U_i, \tag{1.2}$$

where $X_{li}$ is now of dimension $q_l$ ($q_l \geqslant 1$), $\gamma$ is an $s \times 1$ unknown parameter, $z_0(X_i)$ is an $s \times 1$ known function of $X_i$, $X_i$ is a $q \times 1$ nonoverlapping variables of $X_{1i}, X_{2i}, \ldots, X_{Li}$. Model (1.2) is an additive partially linear model and it allows interaction terms among $X_{li}$ to enter as the linear part of the model. For instance, consider the simple case of $L = 2$, where $X_{1i}$, $X_{2i}$ and $z_0(X_i)$ are all scalars, we can let $z_0(X_i) = X_{1i}X_{2i}$.

Both kernel and series methods have been proposed to estimate models (1.2), e.g., Fan et al. (1998), Fan and Li (1996a), Li (2000), and Sperlich et al. (2002), [2] among others. The kernel marginal integration method is to first estimate a nonparametric model with high dimension (ignoring the additive structure) and then to use the method of marginal averaging to obtain an estimator of a function with low dimension (utilizing the additive structure). This approach may lead to some *finite sample* efficiency loss due to the use of inefficient estimation procedure in the first stage. This is because we estimate a high-dimensional model less accurately compared with estimating a low-dimensional model. If the sample size is not sufficiently large, this first step efficiency loss may not be totally recovered by the second step marginal integration computation in *finite* sample applications. The kernel marginal integration method is also computationally costly. The computation time of estimating an additive partially linear model is about $n$ ($n$ is the sample size) times that of estimating a nonadditive partially linear model. Also, the asymptotic analysis (of estimating an additive partially linear model) using kernel methods is quite complex as can be seen from the works of

[1] For recent development on using backfitting method to estimate an additive model, see Mammen et al. (1999), Linton (2000), Nielsen and Linton (1998), and Opsomer and Ruppert (1997).

[2] Although Sperlich et al. (2002) did not consider the parametric component $z_0(X_i)$, they allow the additive functions to contain second-order pairwise interactions, i.e., allowing $X_{li}$ contain overlapping variables. Sperlich et al. (2002) proposed several tests for testing zero interaction terms.

Fan et al. (1998), and Fan and Li (1996a, b).[3] On the other hand, the series method is less costly in computation because it only involves least squares. Furthermore, the series method can estimate $\gamma$ efficiently in the sense that the asymptotic variance of a series estimator of $\gamma$ attains the semiparametric efficiency bound, while the *existing* kernel marginal integration-based method does not (e.g., Fan et al., 1998; Fan and Li, 1996a). It is also fairly straightforward to impose restrictions such as additive separability or shape preserving by using series method (e.g., Chen and Shen, 1998; Dechevsky and Penez, 1997).

Because the relative ease of implementing series method in nonparametric estimation, in this paper we consider constructing test statistics based on series estimation method, in particular, the problem of testing the adequacy of Model (1.2) using the series method as an alternative to the kernel method. We will also show that our testing procedure can be easily generalized to testing other semiparametric econometric models.

There is an abundance of literature on constructing consistent model specification tests using various estimation techniques, see Ait-Sahalia et al. (2001), Andrews (1997), Bierens (1982, 1990), Bierens and Ploberger (1997), Chen and Fan (1999), De John (1996), Delgado (1993), Delgado and Manteiga (2001), Delgado and Stengos (1994), Donald (1997), Ellison and Ellison (2000), Eubank and Spiegelman (1990), Fan and Li (1999), Härdle and Mammen (1993), Hong and White (1995), Horowitz and Härdle (1994), Lavergne (2001), Lavergne and Vuong (1996), Lewbel (1995), Li and Wang (1998), Robinson (1989, 1991), Stute (1997), Wooldridge (1992), Yatchew (1992), and Zheng (1996), among others. Most authors consider the problem of testing a parametric null model. Ait-Sahalia et al. (2001), Delgado and Manteiga (2001) and Chen and Fan (1999), have considered the case of testing nonparametric/semiparametric null models. They use nonparametric kernel methods to estimate the null models. We use nonparametric series methods to estimate the null model (1.2). Both their test statistics and ours are nuisance parameter dependent. Some bootstrap methods are therefore needed to compute the critical values of the test statistics. However, estimating the additive model (1.2) by kernel methods in conjunction with the use of bootstrap methods to evaluate the critical values of the test statistics could be computationally more burdensome than by the series method.

The test statistic considered in this paper has the properties that: (i) it avoids estimating the alternative model nonparametrically so as to partially circumvent the "curse of dimensionality" problem, (ii) it can detect local alternatives of the order of $O_p(n^{-1/2})$, and (iii) it is computational simple. We note that the test statistics of Delgado and Manteiga (2001) and Chen and Fan (1999) also share properties (i) and (ii), but as discussed above, kernel methods could be computationally costly when estimating *additive* models.

In Section 2 we develop a consistent model specification test for additive partially linear models based on series estimation method. Section 3 reports simulation results to examine the finite sample performance of the proposed test. Generalizations are

[3] Note that Fan et al. (1998), and Fan and Li (1996a, b) did not consider the case of nonparametric additive interaction terms. Allowing additive interaction terms will make the theoretical analysis more complex.

discussed in Section 4. Conclusions are in Section 5. The proofs are given in Appendices A and B.

## 2. Consistent test for an additive partially linear model

In this section, we propose a consistent test for the additive partially linear model (1.2). The null hypothesis is

$$\mathrm{H}_0:\ \mathrm{E}(Y_i|X_i) = z_0(X_i)'\gamma + \sum_{l=1}^{L} m_l(X_{li}) \quad \text{a.s. for some } \gamma \in \mathscr{B}, \sum_{l=1}^{L} m_l(\cdot) \in \mathscr{G}, \quad (2.1)$$

where $\mathscr{B}$ is a compact subset of $R^r$ and $\mathscr{G}$ is the class of additive functions defined below.

**Definition 1.** We say that a function $\xi(z)$ belongs to an additive class of function $\mathscr{G}$ ($\xi \in \mathscr{G}$) if (i) $\xi(z) = \sum_{l=1}^{L} \xi_l(z_l)$, $\xi_l(z_l)$ is continuous in its support $\mathscr{S}_l$, where $\mathscr{S}_l$ is a compact subset of $R^{q_l}$ ($l = 1,\ldots,L$; $L \geqslant 2$ is a finite positive integer); (ii) $\sum_{l=1}^{L} \mathrm{E}[\xi_l(Z_l)]^2 < \infty$ and (iii) $\xi_l(0) = 0$ for $l = 2,\ldots,L$.

When $\xi(z)$ is a vector-valued function, we say $\xi \in \mathscr{G}$ if each component of $\xi$ belongs to $\mathscr{G}$.

The alternative hypothesis $\mathrm{H}_1$ is the negation of $\mathrm{H}_0$, i.e.

$$\mathrm{H}_1:\ \mathrm{E}(Y_i|X_i) \neq z_0(X_i)'\gamma + \sum_{l=1}^{L} m_l(X_{li}) \quad (2.2)$$

on a set with positive measure for any $\gamma \in \mathscr{B}$, and any $\sum_{l=1}^{L} m_l(\cdot) \in \mathscr{G}$.

The null hypothesis $\mathrm{H}_0$ is equivalent to $\mathrm{E}(U_i|X_i) = 0$ almost surely (a.s.), where $U_i$ is defined in (1.2). Note that $\mathrm{E}(U_i|X_i) = 0$ a.s. if and only if $\mathrm{E}[U_i M(X_i)] = 0$ for all $M(\cdot) \in \mathscr{M}$, the class of bounded $\sigma(X_i)$-measurable functions. Instead of considering the conditional moment test of (2.1), following Bierens and Ploberger (1997), Stinchcombe and White (1998), and Stute (1997), in this paper we consider the following unconditional moment test: [4]

$$\mathrm{E}[U_i \mathscr{H}(X_i, x)] = 0 \quad \text{for almost all } x \in \mathscr{S} \subset \mathscr{R}^q, \quad (2.3)$$

where $\mathscr{H}(\cdot,\cdot)$ is a proper choice of a weight function so as to make (2.3) equivalent to (2.1), see Assumption (A4)(i) and (ii) below on the specific conditions on $\mathscr{H}$.

We assume that the weight function $\mathscr{H}(\cdot,\cdot)$ is bounded on $\mathscr{S} \times \mathscr{S}$. Stinchcombe and White (1998) have shown that there exists a wide class of weight functions $\mathscr{H}(\cdot,\cdot)$ that makes (2.3) equivalent to $\mathrm{E}(U_i|X_i)=0$ a.s. Choices of weight functions include the exponential function $\mathscr{H}(X_i,x)=\exp(X_i'x)$, the logistic function $\mathscr{H}(X_i,x)=1/[1+\exp(c-X_i'x)]$ with $c \neq 0$, and $\mathscr{H}(X_i,x) = \cos(X_i'x) + \sin(X_i'x)$, see Stinchcombe and White (1998), and Bierens and Ploberger (1997) for more discussion on this. By switching

[4] Similar approaches were used by Chen and Fan (1999), and Delgado and Manteiga (2001) to construct kernel-based consistent model specification tests when the null models contain nonparametric components.

a conditional moment test (2.1) to an unconditional moment test of (2.3), we avoid having to estimate the alternative model nonparametrically, as in Chen and Fan (1999), and Delgado and Manteiga (2001).

Multiplying by $\sqrt{n}$ the sample analogue of $\mathrm{E}[U_i\mathscr{H}(X_i,x)]$, we have

$$J_n^0(x)=\sqrt{n}\left[\frac{1}{n}\sum_i U_i\mathscr{H}(X_i,x)\right]=\frac{1}{\sqrt{n}}\sum_i U_i\mathscr{H}(X_i,x). \tag{2.4}$$

Stute (1997) uses $\mathscr{H}(X_i,x)=\mathbf{1}(X_i\leqslant x)$ and the Skorohod topology to study the weak convergence of $J_n^0(\cdot)$. Since $J_n^0$ is a random element in the Skorohod space $\mathscr{D}(\mathscr{S})$, Stute shows that $J_n^0(\cdot)$ converges to a Gaussian process in $\mathscr{D}(\mathscr{S})$. Central limit theorems for goodness-of-fit tests, like Stute (1997), are usually based on weak convergence of empirical process interpreted as random elements taking values in the space of continuous functions endowed with uniform topology or cadlag functions endowed with the Skorohod topology. When a test statistic involves nonparametric estimations, establishing its weak convergence with uniform topology can be quite challenging. However, there is a natural way to study the asymptotic properties of statistics of Cramer–von Mises type. $J_n^0$ can be viewed as a random element taking values in the separable space $\mathscr{L}_2(\mathscr{S},v)$ of all real, Borel measurable functions $f$ on $\mathscr{S}$ such that $\int_{\mathscr{S}} f(x)^2 v(\mathrm{d}x)<\infty$, which is endowed with the $L_2$-norm $\|f\|_v^2=\int_{\mathscr{S}} f(x)^2 v(\mathrm{d}x)$. The theory of probability on Banach (or Hilbert) spaces, developed in the 1960s and 1970s, turned the problem of studying the asymptotic distribution of statistics like $\|J_n^0\|_{L_2(v)}$ to an easier task, because sufficient conditions of central limit theorems for random elements taking values in $\mathscr{L}_2(\mathscr{S},v)$ are well established and are easy to check. For example, Araujo and Gine (1980, p. 205), Ledoux and Talagrand (1991), Van der Vaart and Wellner (1996, p. 50), Chen and White (1997) assert that for a sequence $\{Z_n(\cdot)\}_n$ of i.i.d. $\mathscr{L}_2(\mathscr{S},v)$-valued elements that $n^{-1/2}\sum_{i=1}^n Z_i(\cdot)$ converges to $\mathscr{Z}(\cdot)$ in the topology of $(\mathscr{L}_2(\mathscr{S},v), \|\cdot\|_{L_2(v)})$ if and only if $\int_{\mathscr{S}}\mathrm{E}[Z_1(x)^2]v(\mathrm{d}x)<\infty$, where $\mathscr{Z}$ is a Gaussian element with the same covariance function as $Z_1$, we will formally summarize this result in a lemma below for ease of reference.

We assume $v(\mathscr{S})<\infty$. Since we will only consider the case that $\mathscr{S}$ is a bounded subset of $\mathscr{R}^d$, we will choose $v(\cdot)$ to be the Lebesgue measure on $\mathscr{S}$. Then $J_n^0(\cdot)$ is a Hilbert-valued random element in $\mathscr{L}_2(\mathscr{S},v)$. We present a $H$-valued central limit theorem in a lemma below.

**Lemma 2.1.** *Let $Z_1(\cdot),\dots,Z_n(\cdot)$ be $H$-valued, independent and identically distributed zero mean random elements on $\mathscr{L}_2(\mathscr{S},v)$ such that $\mathrm{E}[\|Z_i(\cdot)\|_v^2]<\infty$. Then $n^{-1/2}\sum_{i=1}^n Z_i(\cdot)$ converges weakly*[5] *to a zero mean Gaussian process with the covariance (kernel) function given by $\Omega(x,x')=\mathrm{E}[Z_i(x)Z_i(x')]$.*

**Proof.** See Theorem 2.1 of Politis and Romano (1994), or Van der Vaart and Wellner (1996, ex. 1.8.5, p. 50). Note that $\mathrm{E}[\|Z_i(\cdot)\|_v^2]<\infty$ is a sufficient condition that ensures the process $n^{-1/2}\sum_{i=1}^n Z_i(\cdot)$ is tight.

[5] A sequence of $H$-valued random element $\mathscr{Z}_n$ *converges weakly* to $\mathscr{Z}$ if $\mathrm{E}[h(\mathscr{Z}_n)]\to\mathrm{E}[h(\mathscr{Z})]$ for all real-valued bounded continuous function $h$.

It is straightforward to check that $J_n^0(\cdot)$ is tight using Lemma 2.1. Letting $Z_i(\cdot)=U_i\mathscr{H}(X_i,\cdot)$, we have

$$\mathrm{E}[\|Z_i(\cdot)\|_v^2]=\mathrm{E}\left\{\int U_i^2[\mathscr{H}(X_i,x)]^2 v(\mathrm{d}x)\right\}$$

$$=\mathrm{E}\left\{\sigma^2(X_i)\int[\mathscr{H}(X_i,x)]^2 v(\mathrm{d}x)\right\}\leqslant \mathrm{E}[\sigma^2(X_i)]\left\{C\int_{\mathscr{S}} v(\mathrm{d}x)\right\}<\infty,$$

where $\sigma^2(X_i)=\mathrm{E}(U_i^2|X_i)$.

Thus by Lemma 2.1, we know that

$J_n^0(\cdot)$ converges weakly to $J_\infty^0(\cdot)$ in $\mathscr{L}_2(\mathscr{S},v,\|\cdot\|_v)$,

where $J_\infty^0$ is a Gaussian process centered at zero and with covariance function $\Omega$ given by

$$\Omega(x,x')=\mathrm{E}[Z_i(x)Z_i(x')]=\mathrm{E}[\sigma^2(X_i)\mathscr{H}(X_i,x)\mathscr{H}(X_i,x')], \tag{2.5}$$

where $x,x'\in\mathscr{S}$. □

Since $U_i$ is unobservable, we need to replace $U_i$ by some estimate of it, say $\hat{U}_i$, (the definition of $\hat{U}_i$ is given in (2.16) below) and construct a feasible version of (2.4) as

$$\hat{J}_n(x)=\frac{1}{\sqrt{n}}\sum_i \hat{U}_i\mathscr{H}(X_i,x). \tag{2.6}$$

We will use series estimation method to construct a consistent test based on (2.6). Obviously the individual functions $m_l(\cdot)$ ($l=1,\dots,L$) are not identified without some identification conditions. In the kernel estimation literature, a convenient identification condition is $\mathrm{E}[m_l(X_{li})]=0$ when $X_{li}$'s are all scalars ($l=2,\dots,L$). When the arguments in the additive functions contain pairwise interaction terms of $X_{li}$, some additional identification conditions are required, see Sperlich et al. (2002) for the detailed marginal-integration-based identification conditions in this case. In practice the kernel marginal integration method is to first estimate a high-dimensional nonparametric function: $\mathrm{E}(Y_i|X_{1i},\dots,X_{qi})$ (without imposing additive structure). In the second stage, the marginal integration method is used to obtain estimated additive functions.

When using series estimation methods, the additive structure can be imposed on the series approximating base functions. Therefore, one does not need to estimate a high-dimensional nonparametric model. Only the one-step least squares estimation method is needed to obtain all the estimated additive functions. The identification conditions for series estimation methods can be obtained by choosing some normalization rules that are easy to impose on series approximating base functions. For instance, in the case of an additive model without interaction terms,

$$g(x_1,\dots,x_q)=c+m_1(x_1)+\cdots+m_q(x_q), \tag{2.7}$$

where $x_j\in R$, we can use $m_j(x_j=0)\equiv m_j(0)=0$ as the identification condition. This is because for an arbitrary additive function $g(x_1,\dots,x_q)=\tilde{c}+\tilde{m}_1(x_1)+\cdots+\tilde{m}_q(x_q)$, we can always re-write it as

$$g(x_1,\ldots,x_q) = \left[\tilde{c} + \sum_{j=1}^{q} \tilde{m}_j(0)\right] + [\tilde{m}_1(x_1) - \tilde{m}_1(0)] + \cdots + [m_q(x_q) - m_q(0)]$$
$$\equiv c + m_1(x_1) + \cdots + m_q(x_q), \tag{2.8}$$

where $c = \tilde{c} + \sum_{j=1}^{q} \tilde{m}_j(0)$ and $m_j(x_j) = \tilde{m}_j(x_j) - \tilde{m}_j(0)$. Therefore, $m_j(0) = 0$ is satisfied for all $j = 1,\ldots,q$.

Similarly, for the case of an additive model with second-order interactions, we have

$$g(x_1,\ldots,x_q) = c + \sum_{j=1}^{q} m_j(x_j) + \sum_{j=1}^{q-1}\sum_{l>j}^{q} m_{jl}(x_j,x_l). \tag{2.9}$$

The representation in (2.9) is not unique, but can be made so by imposing the following identification conditions:

$$m_j(x_j = 0) \equiv m_j(0) = 0, \quad j = 1,\ldots,q$$

and

$$m_{jl}(0,x_l) = m_{jl}(x_j,0) = 0 \quad \text{for all values of } x_j,\ x_l\ (1 \leqslant j < l \leqslant q). \tag{2.10}$$

To see these conditions indeed uniquely identify the $m_j(\cdot)$ and $m_{jl}(\cdot,\cdot)$ functions, consider an arbitrary additive function with second-order interactions: $g(x_1,\ldots,x_q) = \tilde{c} + \sum_{j=1}^{q} \tilde{m}_j(x_j) + \sum_{j=1}^{q-1} \sum_{l>j}^{q} \tilde{m}_{jl}(x_j,x_l)$. Then $m_{jl}(x_j,x_l)$, $m_j(x_j)$ and $c$ can be obtained by

$$m_{jl}(x_j,x_l) = \tilde{m}_{jl}(x_j,x_l) - \tilde{m}_{jl}(0,x_l) - \tilde{m}_{jl}(x_j,0) + \tilde{m}_{jl}(0,0),$$

$$m_j(x_j) = \bar{m}_j(x_j) - \bar{m}_j(0), \quad \text{where}$$

$$\bar{m}_j(x_j) = \tilde{m}_j(x_j) + \sum_{l=j+1}^{q} \tilde{m}_{jl}(x_j,0) + \sum_{l=1}^{j-1} \tilde{m}_{lj}(0,x_j),$$

$$c = \tilde{c} + \sum_{j=1}^{q} \tilde{m}_j(0) + \sum_{l>j}^{q}\sum_{j=1}^{q-1} \tilde{m}_{jl}(0,0). \tag{2.11}$$

Obviously, $m_{jl}(x_j,x_l)$ and $m_j(x_j)$ defined in (2.11) satisfy the identification condition (2.10), and it is straightforward to check that $c + \sum_{j=1}^{q} m_j(x_j) + \sum\sum_{q \geqslant l > j \geqslant 1} m_{jl}(x_j,x_l) = \tilde{c} + \sum_{j=1}^{q} \tilde{m}_j(x_j) + \sum\sum_{q \geqslant l > j \geqslant 1} \tilde{m}_{jl}(x_j,x_l)$. Thus, (2.10) indeed provides unique conditions for identifying the additive functions $m_j(\cdot)$ and $m_{jl}(\cdot,\cdot)$.

In principle one can always impose the identification conditions on the approximating base functions. Let $\mathcal{N}_1 = \{1,2,\ldots\}$ denote the set of positive integers. If one uses a finite linear combination of $\{\phi_t(x_j)\}_{t\in\mathcal{N}_1}$, $j = 1,\ldots,q$, as the base function to approximate the additive function $m_j(x_j)$, the above identification condition implies that one should use a finite linear combination of $\{\phi_t(x_j)\}_{t\in\mathcal{N}_1}$ to approximate $m_j(x_j)$ with $\phi_t(x_j = 0) = \phi_t(0) = 0$ for all $t \in \mathcal{N}_1$. Then one can use a finite linear combination of $\{\phi_t(x_j)\phi_s(x_l)\}_{s>t\in\mathcal{N}_1}$

to approximate $m_{jl}(x_j,x_l)$. For example, consider the case of polynomial (power) series with $\phi_t(x_j)=x_j^t$. The approximation function for $m_j(x_j)$ is a finite linear combination of $\{x_j^t\}_{t\in\mathcal{N}_1}=\{x_j,x_j^2,x_j^3,\ldots\}$ (without the constant term) so that $\phi_t(0)=0^t=0$ is satisfied. And the base function for approximating $m_{jl}(x_j,x_l)$ is a finite linear combination of $\{x_j^t x_l^s\}_{s>t\in\mathcal{N}_1}=\{x_jx_l,x_j^2x_l,x_jx_l^2,x_j^2x_l^2,\ldots\}$. The approximating functions have the property that $\phi_t(x_j=0)\phi_s(x_l)=\phi_t(x_j)\phi_s(x_l=0)=0$ as imposed by (2.10). It is straightforward to generalize the above identification conditions to additive models with higher-order interaction terms.

Obviously, the above identification conditions rule out an intercept term in the additive functions. Therefore, an intercept term should be included in the parametric part of the model. We can assume the first element of $z_0(X_i)$ to be one, then the first element of $\gamma$ is the intercept term. Note that using series estimation method, the additive structure is automatically imposed in the one-step least squares estimation procedure. While the kernel marginal integration method ignores the additive structure in the initial estimation stage. Therefore, a second stage of marginal integration method is needed to obtain additive function estimations. We also need an identification condition for the other components of $\gamma$. For example, we cannot allow $z_0(X_i)$ to have an additive structure like $z_0(X_i)=\sum_{l=1}^{L} z_{0,l}(X_{li})$, because then $m_l(x_l)$ and $z_{0,l}(x_l)$ cannot be identified separately since the functional form of $m_l(x_l)$ is not specified. In order for the parameter $\gamma$ to be identified, we need to assume that $z_0(\cdot)$ does not belong to the class of additive functions $\mathcal{G}$ as defined in Assumption (A1)(iii) below using the definition of projection matrix.

For any random variable (vector) $\mathcal{A}_i$, let $E_{\mathcal{G}}(\mathcal{A}_i)$ denote the projection of $\mathcal{A}_i$ onto the linear additive functional space $\mathcal{G}$. Then $E_{\mathcal{G}}(\mathcal{A}_i)$ is the closest function to (in the mean square error sense) $\mathcal{A}_i$ among all functions in the class of additive functions $\mathcal{G}$, i.e.,

$$\mathrm{E}\{[\mathcal{A}_i - E_{\mathcal{G}}(\mathcal{A}_i)]^2\} = \inf_{\sum_{l=1}^{L}\ \xi_l(\cdot)\in\mathcal{G}} \mathrm{E}\left\{\left[\mathcal{A}_i - \sum_{l=1}^{L}\xi(X_{li})\right]^2\right\}. \tag{2.12}$$

**Remark 2.1.** Let $\mathcal{V}_i=\mathcal{A}_i - E_{\mathcal{G}}(\mathcal{A}_i)$, then $E_{\mathcal{G}}(\mathcal{V}_i)=0$. That is, for any random variable (vector) $\mathcal{A}_i$, we have the orthogonal decomposition of $\mathcal{A}_i=E_{\mathcal{G}}(\mathcal{A}_i)+\mathcal{V}_i$ with $E_{\mathcal{G}}(\mathcal{A}_i)\in\mathcal{G}$ and $\mathcal{V}_i\perp\mathcal{G}$. When $\mathcal{V}_i\perp\mathcal{G}$, we also write $\mathcal{V}_i\in\mathcal{G}^{\perp}$.

We use a linear combination of $K_l$ functions: $p_l^{K_l}(x_l)=(p_{l1}^{K_l}(x_l),\ldots,p_{lK_l}^{K_l}(x_l))'$ to approximate $m_l(x_l)$ $(l=1,\ldots,L)$. That is, we use a linear combination of $K=\sum_{l=1}^{L}K_l$ functions $(p_1^{K_1}(x_1)',\ldots,p_L^{K_L}(x_L)')\equiv p^K(x)'$ to approximate an additive function $\sum_{l=1}^{L} m_l(x_l)$.

We will use $\|\cdot\|$ to denote the usual Euclidean norm ($\|\cdot\|_v$ denotes the $L_2$ norm). We assume that:

(A1). (i) $(Y_1,X_1),\ldots,(Y_n,X_n)$ are independent and identically distributed as $(Y_1,X_1)$, $\mathcal{S}$, the support of $X$, is a compact subset of $R^d$; $F(\cdot)$, the distribution function of $X_1$, is absolutely continuous with respect to the Lebesgue measure.[6]

[6] $F$ is said to be absolutely continuous with respect to a measure $v$ if for any set $A$, $\int \mathbf{1}(x\in A)v(\mathrm{d}x)=0$ implies that $\int \mathbf{1}(x\in A)F(\mathrm{d}x)=0$, $\mathbf{1}(x\in A)=1$ if $x\in A$, 0 otherwise.

(ii) $var(Y_i|X_i)$ is a bounded function on the support of $X_i$; (iii) $z_0(X_i) \notin \mathscr{G}$ in the sense that $\mathrm{E}(\varepsilon_i\varepsilon_i')$ is positive definite, where $\varepsilon_i = z_0(X_i) - E_{\mathscr{G}}[z_0(X_i)]$.

(A2). (i) For every $K$ there is a nonsingular matrix $B$ such that for $P^K(X_i)=Bp^K(X_i)$: the smallest eigenvalue of $\mathrm{E}[P^K(X_i)P^K(X_i)']$ is bounded away from zero uniformly in $K \in \mathscr{N}$;

(ii) there is a sequence of constants $\zeta_0(K)$ satisfying $\sup_{z\in\mathscr{S}}\|P^K(z)\| \leqslant \zeta_0(K)$ and $K = K_n$ such that $(\zeta_0(K))^2(K/n) \to 0$ as $n \to \infty$, where $\mathscr{S}$ is the support of $X$.

(A3). (i) For any $f \in \mathscr{G}$, there exist some positive $\delta_l\,(>1)$ $(l=1,\ldots,L)$, $\beta_f=\beta_{fK}=(\beta'_{fK_1},\ldots,\beta'_{fK_L})'$, $\sup_{x\in\mathscr{S}}|f(x)-P^K(x)'\beta_f|=\mathrm{O}(\sum_{l=1}^L K_l^{-\delta_l})$ as $\min\{K_1,\ldots,K_L\} \to \infty$; (ii) $\sqrt{n}(K/n + \sum_{l=1}^L K_l^{-\delta_l}) \to 0$ as $n \to \infty$.

(A4). (i) The weight function $\mathscr{H}(X_i,x) = w(X_i'x)$ with $w(\cdot)$ being an analytic, non-polynomial function.[7] (ii) $\mathscr{H}(\cdot,\cdot)$ is bounded on $\mathscr{S}\times\mathscr{S}$ and satisfies a Lipschitz condition, for all $x_1,x_2 \in \mathscr{S}$, $|\mathscr{H}(X_i,x_1) - \mathscr{H}(X_i,x_2)| \leqslant G(X_i)\|x_1 - x_2\|$ with $\mathrm{E}[G^2(X_i)] < \infty$; (iii) $v(\cdot)$ is the Lebesgue measure.

**Remark 2.2.** We give some remarks on the above regularity conditions.

Conditions (A1)(i) and (ii) are standard in the literature of estimating additive models. Condition (A1)(i) rules out discrete random variables.[8] The bounded conditional variance assumption (A1)(ii) is restrictive, however, it still allows a wide range of conditional heteroskedasticity of the form: $U_i = \mathscr{U}_i h(X_i)$, where $\mathscr{U}_i$ is i.i.d. with mean zero and has a finite second moment (say $\sigma^2$), $\mathscr{U}_i$ and $X_i$ are independent for all $i$ and $j$, $h(x)$ is a continuous (or bounded) function on $\mathscr{S}$, then $var(U_i|X_i) = \sigma^2[h(X_i)]^2$ is a bounded function in (the compact set) $\mathscr{S}$. Condition (A1)(iii) is an identification condition for $\gamma$, requiring that $z_0(\cdot)$ should not lie in $\mathscr{G}$ because otherwise $z_0(\cdot)$ and $\sum_l m_l(\cdot)$ cannot be identified separately.

Condition (A2)(i) ensures that $(P'P)$ is asymptotically nonsingular. Note that in (A2)(i) we do not assume that $p^K(x)$ is an *orthogonal* base function since the density function of $X_i$ is unknown, therefore it is not feasible to orthogonalize the base function in practice. Condition (A2)(ii) is a standard condition for the consistency of the series estimator.

We can write $f(x) = \sum_{l=1}^L f_l(x_l)$ for any $f \in \mathscr{G}$. Hence, (A3)(i) is implied by the following: for all $l = 1,\ldots,L$, there exists some $\delta_l > 0$, $\beta_{fl} = \beta_{fl,K_l}$ ($\beta_{fl}$ is the $l$th component of $\beta_f$), such that $\sup_{x_l\in\mathscr{S}_l}|f_l(x_l) - p_l^{K_l}(x_l)'\beta_{fl}| = \mathrm{O}(K_l^{-\delta_l})$, as $K_l \to \infty$, where $\mathscr{S}_l$ is the support of $x_l$. It is possible to weaken (A3)(i) to $\int_{\mathscr{S}}[f(x) - P^K(x)'\beta_f]^2 v(\mathrm{d}x) = \mathrm{O}(\sum_{l=1}^L K_l^{-2\delta_l})$ since we only work with the $L_2$-norm (the $\|\cdot\|_v$ norm), but this will require many more new notations and a much longer proof with little practical implications. Therefore, we will not pursue this generality.

Condition (A4) implies that $P[\mathrm{E}(U_i|X_i) = 0] < 1$ if and only if $P[\mathrm{E}(U_i\mathscr{H}(X_i,x)) = 0] < 1$ (or $\int\{\mathrm{E}[U_i\mathscr{H}(X_i,x)]\}^2F(\mathrm{d}x) > 0$). Thus, testing $\mathrm{E}(U_i|X_i) = 0$ almost

[7] An analytic function is one locally equal to its Taylor expansion at each point of its point of its domain, such as $\exp(\cdot)$, the logistic, the hyperbolic tangent, the sine and cosine, etc.

[8] If a discrete variable $X$ takes only finitely many different values, it becomes a parametric model since only finitely many series-based functions are needed to estimate an unknown function $\theta(\cdot)$.

everywhere is equivalent to test $\mathrm{E}[U_i\mathscr{H}(X_i,x)]=0$ for almost all $x\in\mathscr{S}$, see Stinchcombe and White (1998) for a detailed discussion on this equivalence.

We require the approximation function $p^K(x)$ to have the properties that: (a) $p^K(x)\in\mathscr{G}$ and (b) as $K_l$ grows (for all $l=1,\dots,L$), there is a linear combination of $p^K(x)$ that can approximate any $f\in\mathscr{G}$ arbitrarily well in the mean square error sense. While (A2) and (A3) are not primitive conditions, it is known that many series functions satisfy these conditions. Newey (1997) gives primitive conditions for power series and splines such that (A2) and (A3) hold (see assumptions (A5) and (A6) below).

(A4)(i) is similar to the assumption of Chen and Fan (1999), it allows $\mathscr{H}(X,x)=\exp(X_i'x)$ (Bierens, 1990), or $\mathscr{H}(X_i,x)=1/[1+\exp(c-X_i'x)]$ with $c\neq 0$, or $\mathscr{H}(X_i,x)=\cos(X_i'x)+\sin(X_i'x)$, (Stichcombe and White, 1997). (A4)(i)–(iii) are very mild conditions on the weight function $\mathscr{H}(\cdot,\cdot)$. They imply that $\int_{\mathscr{S}}\mathscr{H}(y,x)v(\mathrm{d}x)\leqslant C\int_{\mathscr{S}}v(\mathrm{d}x)=Cv(\mathscr{S})<\infty$.

Let

$$p_l=(p_l^{K_l}(X_{l1}),\dots,p_l^{K_l}(X_{ln}))' \quad (l=1,\dots,L), \quad \text{and}$$
$$P=(p_1,\dots,p_L). \tag{2.13}$$

Note that $p_l$ is of dimension $n\times K_l$ and $P$ is of dimension $n\times K$. Let $Z_0$ be the $n\times r$ matrix with the $i$th row given by $z_0(X_i)'$. Then in vector–matrix notation, we can write (1.2) as

$$\begin{aligned}Y&=Z_0\gamma+m_1+m_2+\cdots+m_L+U\equiv Z_0\gamma+m+U\\&=Z_0\gamma+P\beta+(m-P\beta)+U=(Z_0,P)\begin{pmatrix}\gamma\\ \beta\end{pmatrix}+(m-P\beta)+U\\&=\mathscr{X}\alpha+(m-P\beta)+U,\end{aligned}\tag{2.14}$$

where $\mathscr{X}=(Z_0,P)$, $\alpha=(\gamma',\beta')'$, $Y$ and $U$ are both $n\times 1$ vectors with the $i$th component given by $Y_i$ and $U_i$, respectively, $m$ is $n\times 1$ with the $i$th component given by $m_i=\sum_{l=1}^L m_l(X_{li})$. $P$ is of dimension $n\times K$ and $\beta=\beta_m$ is a $K\times 1$ vector that satisfies assumption (A3) (with $f=m$).

We estimate $\alpha=(\gamma',\beta')'$ by the least squares method of regressing $Y$ on $\mathscr{X}$:

$$\hat\alpha=\begin{pmatrix}\hat\gamma\\ \hat\beta\end{pmatrix}=(\mathscr{X}'\mathscr{X})^-\mathscr{X}'Y, \tag{2.15}$$

where $(\mathscr{X}'\mathscr{X})^-$ is a generalized inverse of $(\mathscr{X}'\mathscr{X})$. Li (2000) showed that $\hat\gamma-\gamma=\mathrm{O}_p(n^{-1/2})$. Also $\hat m(x)-m(x)=\mathrm{O}_p((K/n)^{1/2}+\sum_{l=1}^L K_l^{-\delta_l})$ by the results of Andrews and Whang (1990) and Newey (1995,1997), where $\hat m(x)=p^K(x)'\hat\beta$. Hence, we estimate $U_i$ by

$$\hat U_i=Y_i-z_0(X_i)'\hat\gamma-p^K(X_i)'\hat\beta. \tag{2.16}$$

Our test statistic for $\mathrm{H}_0$ is based on

$$\hat J_n(x)=\frac{1}{\sqrt n}\sum_i\mathscr{H}(X_i,x)\hat U_i, \tag{2.17}$$

where $\hat{U}_i$ is given in (2.16). With $\hat{J}_n(x)$ we can construct a Cramer–von Mises $(CM)$-type statistic for testing $\mathrm{H}_0$.

$$CM_n = \int [\hat{J}_n(x)]^2 F_n(\mathrm{d}x) = \frac{1}{n}\sum_i [\hat{J}_n(X_i)]^2,$$

where $F_n(\cdot)$ is the empirical distribution of $X_1,\ldots,X_n$.

The next theorem establishes the weak convergence of $\hat{J}_n(\cdot)$ and $CM_n$ under $\mathrm{H}_0$.

**Theorem 2.2.** *Suppose that assumptions* (A1)–(A4) *hold, then under* $\mathrm{H}_0$,

(i) $\hat{J}_n(\cdot)$ converges weakly to $J_\infty(\cdot)$ in $\mathscr{L}_2(\mathscr{S}, v, \|\cdot\|_v)$,

*where $J_\infty$ is a Gaussian process with zero mean and covariance function given by*

$$\Sigma_1(x,x') = \mathrm{E}[\sigma^2(X_i)\eta_i(x)\eta_i(x')],$$

*where $\eta_i(x) = \mathscr{H}(X_i,x) - \phi_i(x) - \psi_i(x)$, with $\phi_i(x) = E_{\mathscr{G}}[\mathscr{H}(X_i,x)]$, $\psi_i(x) = \mathrm{E}[\mathscr{H}(X_i,x)\varepsilon_i]\{\mathrm{E}[\varepsilon_i\varepsilon_i']\}^{-1}\varepsilon_i$ and $\varepsilon_i = z_0(X_i) - E_{\mathscr{G}}(z_0(X_i))$.*

(ii) *$CM_n$ converges to* $\int [J_\infty(x)]^2 F(\mathrm{d}x)$ *in distribution,*

*where $F(\cdot)$ is the distribution function of $X_i$.*

Proof of Theorem 2.2(i) is given in the Appendix A. Theorem 2.2(ii) follows from Theorem 2.2(i), the continuous mapping theorem, and the fact that $F_n(\cdot)$ is close to $F(\cdot)$ ($F_n(\cdot)$ is the empirical distribution of $\{X_i\}_{i=1}^n$). The idea underlying the proof of Theorem 2.2(i) is very simple. First we write $\hat{J}_n(\cdot) = J_n(\cdot) + [\hat{J}_n(\cdot) - J_n(\cdot)]$, where $J_n(x) = n^{-1/2}\sum_i U_i[\mathscr{H}(X_i,x) - \phi_i(x) - \psi_i(x)]$. $J_n(\cdot)$ converges weakly to the Gaussian process $J_\infty(\cdot)$ by Lemma 2.1 (because $\mathrm{E}[\|J_n(\cdot)\|_v^2] < \infty$). Next, we show that $\|\hat{J}_n(\cdot) - J_n(\cdot)\|_v = o_p(1)$. This implies that $\hat{J}_n(\cdot)$ and $J_n(\cdot)$ have the same limiting distribution. Therefore, $\hat{J}_n(\cdot)$ converges weakly to $J_\infty(\cdot)$ in $\mathscr{L}_2(\mathscr{S}, v, \|\cdot\|_v)$.

Next we study the asymptotic distribution of $\hat{J}_n$ and $CM_n$ under the Pitman local alternative and the fixed alternative. The Pitman local alternative is given by

$$\mathrm{H}_L\colon\ \mathrm{E}(Y_i|X_i) = z_0(X_i)'\gamma + m(X_i) + \frac{g(X_i)}{\sqrt{n}} \quad \text{a.s.}, \tag{2.18}$$

where $g(\cdot) \in \mathscr{G}^\perp$ and $0 < \mathrm{E}\{[g(X_i)]^2\} < \infty$. Note that since $m(x) = \sum_l m_l(x_l) \in \mathscr{G}$ and the functional forms of $m_l(\cdot)$'s are not specified, only $g(\cdot) \in \mathscr{G}^\perp$ should be considered in the local alternative $\mathrm{H}_L$.

For any (vector) random variable $\mathscr{A}_i$, we use $E_{\mathscr{G}^+}(\mathscr{A}_i)$ to denote the projection of $\mathscr{A}_i$ onto the space $\mathscr{G}^+ = \{f(X_i) = z_0'(X_i)\gamma + g(X_i) : \gamma \in \mathscr{B}, g \in \mathscr{G}\}$. More specifically $E_{\mathscr{G}^+}(\mathscr{A}_i)$ is the optimal predictor of $\mathscr{A}_i$ (in the mean square sense) in the class of functions $\mathscr{G}^+$, i.e.,

$$\mathrm{E}\{[\mathscr{A}_i - E_{\mathscr{G}^+}(\mathscr{A}_i)]^2\}$$
$$= \inf_{\gamma\in\mathscr{B} \text{ and } \sum_{l=1}^L \xi_l(\cdot)\in\mathscr{G}} \mathrm{E}\left\{\left[\mathscr{A}_i - z_0(X_i)'\gamma - \sum_{l=1}^L \xi(X_{li})\right]^2\right\}. \tag{2.19}$$

Let $Y_i = \theta(X_i) + U_i$ under $H_1$, then $\theta(\cdot)$ does not belong to $\mathscr{G}^+$. Using similar arguments as in the proof of Theorem 2.2(i), one can show that, under $H_1$, $z_0(X_i)'\hat{\gamma} + p^K(X_i)'\hat{\beta}$ is a consistent estimator of $E_{\mathscr{G}^+}(Y_i) = E_{\mathscr{G}^+}(\theta(X_i))$ because $E_{\mathscr{G}^+}(U_i) = 0$. Hence, $\hat{U}_i = \theta(X_i) + U_i - z_0(X_i)'\hat{\gamma} - p^K(X_i)'\hat{\beta} = \theta(X_i) - E_{\mathscr{G}^+}(\theta(X_i)) + U_i + o_p(1)$ under $H_1$. The following theorem gives the asymptotic distribution of $\hat{J}_n(\cdot)$ under the local alternative $H_L$ and the fixed alternative $H_1$.

**Theorem 2.3.** *Suppose that* (A1)–(A4) *hold,*

(i) *Under the local alternative* $H_L$ *defined in* (2.18), $\hat{J}_n(\cdot)$ *converges weakly to* $J_\infty(\cdot) + \mu_0(\cdot) \equiv J_{L,\infty}(\cdot)$ *in* $\mathscr{L}_2(\mathscr{S}, v, \|\cdot\|_v)$,
*where* $\mu_0(x) = \mathrm{E}[g(X_i)\mathscr{H}(X_i, x)]$.

(ii) *Under the fixed alternative* $H_1$ *defined in* (2.2),
$n^{-1/2}\hat{J}_n(x)$ *converges to* $\mu_1(\cdot)$ *in probability in* $\mathscr{L}_2(\mathscr{S}, v, \|\cdot\|_v)$,
*where* $\mu_1(x) = \mathrm{E}\{[\theta(X_i) - E_{\mathscr{G}^+}(\theta(X_i))]\mathscr{H}(X_i, x)\}$.

Proof of Theorem 2.3 is given in Appendix A.

A consequence of Theorem 2.3 is that our statistic $CM_n$ can detect local alternatives that reach the null model at rate $n^{1/2}$ and that $CM_n$ is a consistent test. This is because by the continuous mapping theorem and the arguments similar to the proof of Theorem 2.2, we know that $CM_n$ converges to $\int [J_\infty(x) + \mu_0(x)]^2 F(\mathrm{d}x)$ under $H_L$, and $CM_n = n\int [\mu_1(x)]^2 F(\mathrm{d}x) + o_p(n)$ under $H_1$, which diverges to $+\infty$ at the rate of $n$ under $H_1$.

Similar to Bierens and Ploberger (1997), and Chen and Fan (1999), one can show that $\int [J_\infty(x)]^2 F(\mathrm{d}x)$ can be written as an infinite sum of weighted (independent) $\chi_1^2$ random variables with weights depending on the unknown distribution of $(X_i, Y_i)$. Therefore, it is difficult to obtain critical values. We suggest using a residual-based wild bootstrap method to approximate the critical values for the null limiting distribution of $CM_n$. The wild bootstrap error $U_i^*$ is generated via a two point distribution: $U_i^* = [(1-\sqrt{5})/2]\hat{U}_i$ with probability $(1+\sqrt{5})/[2\sqrt{5}]$ and $U_i^* = [(\sqrt{5}+1)/2]\hat{U}_i$ with probability $(\sqrt{5}-1)/[2\sqrt{5}]$. Note that $U_i^*$ satisfies

$$\mathrm{E}^*(U_i^*) = 0, \quad \mathrm{E}^*(U_i^{*2}) = \hat{U}_i^2, \quad \text{and} \quad \mathrm{E}^*(U_i^{*3}) = \hat{U}_i^3,$$

where $\mathrm{E}^*(\cdot) = \mathrm{E}(\cdot|\mathscr{W}_n)$ and $\mathscr{W}_n = \{Y_i, X_i\}_{i=1}^n$. From $\{U_i^*\}_{i=1}^n$, we generate $Y_i^*$ according to the null model

$$Y_i^* = z_0(X_i)'\hat{\gamma} + p^K(X_i)'\hat{\beta} + U_i^*, \quad i = 1, \ldots, n. \tag{2.20}$$

Then using the bootstrap sample $\{(Y_i^*, X_i)\}_{i=1}^n$, we obtain

$$\begin{pmatrix} \hat{\gamma}^* \\ \hat{\beta}^* \end{pmatrix} = (\mathscr{X}'\mathscr{X})^- \mathscr{X}'\mathscr{Y}^*,$$

where $\mathscr{X} = (Z_0, P)$ and $\mathscr{Y}^*$ is an $n \times 1$ vector with $j$th element given by $Y_i^*$. The bootstrap residual is given by $\hat{U}_i^* = Y_i^* - z_0(X_i)'\hat{\gamma}^* - p^K(X_i)'\hat{\beta}^*$ and the bootstrap statistic $\hat{J}_n^*(x)$ is obtained by replacing $\hat{U}_i$ in $\hat{J}_n(x)$ by $\hat{U}_i^*$, that is

$$\hat{J}_n^*(x) = \frac{1}{\sqrt{n}} \sum_i \hat{U}_i^* \mathscr{H}(X_i, x).$$

Using $\hat{J}_n^*(\cdot)$ we can compute a bootstrap version of the $CM_n$ statistic, i.e.,

$$CM_n^* = \frac{1}{n}\sum_i [\hat{J}_n^*(X_i)]^2.$$

To show that the bootstrap statistic $CM_n^*$ can be used to approximate the null distribution of $CM_n$, we first give a definition of convergence in distribution in probability.

**Definition 2.** Let $\xi_n$ denote a statistic that depends on the random sample $\{Z_i\}_{i=1}^n$, we say that $(\xi_n|Z_1,Z_2,\dots)$ converges to $(\xi\,|\,Z_1,Z_2,\dots)$ *in distribution in probability* if for any subsequence $\xi_{n'}$, there exists a further subsequence $\xi_{n''}$ such that $(\xi_{n''}|Z_1,Z_2,\dots)$ converges to $(\xi|Z_1,Z_2,\dots)$ for almost every sequence $(Z_1,Z_2,\dots)$.

Many authors show that some bootstrap method works using the concept of $(\xi_n|Z_1,Z_2,\dots)$ converges to $(\xi|Z_1,Z_2,\dots)$ in distribution with probability one (e.g., Stute et al., 1998). The 'with probability one' result is difficult to establish with the series nonparametric estimation method we adopted here. We choose to work with the concept of convergence in distribution in probability in this paper. Equivalently, one can also describe the weak convergence in probability of the bootstrap statistic using the dual bounded Lipschitz metric on probability measures as in Gine and Zinn (1990, p. 861). But the concept of convergence in distribution in probability as defined above may be easier to understand, see Gine and Zinn (1990) on more detailed discussions of these concepts.

The next theorem shows that the wild bootstrap works.

**Theorem 2.4.** *Under the same conditions as in Theorem* 2.3, *we have under* $\mathrm{H}_0$,

$$CM_n^* \text{ converges to } \int [J_\infty^*(x)]^2 F(\mathrm{d}x) \text{ in distribution in probability,}$$

*where* $J_\infty^*(\cdot)$ *has the same distribution as* $J_\infty(\cdot)$.

Theorem 2.4 is proved in Appendix A.

Assumptions (A2) and (A3) used in Theorems 2.2–2.4 are not primitive conditions. Newey (1997) gives primitive conditions for power series and regression spline (B-splines) such that the above Assumptions (A2) and (A3) hold. For readers' convenience we re-state these primitive conditions below. For the construction of B-spline functions, see Schumakers (1980).

(A5). (i) The support of $X_i$ is a Cartesian product of compact connected intervals on which $X_i$ has an absolutely continuous probability density function that is bounded above by a positive constant and bounded away from zero; (ii) for $l=1,\dots,L$, $f_l(x_l)$ is continuously differentiable of order $c_l$ on the support of $X_{il}$, where $f_l(\cdot)=m_l(\cdot)$ or $f_l(\cdot)=h_l(\cdot)(l=1,\dots,L)$, where $h_l(\cdot)$ is defined from $E_{\mathscr{G}}[z_0(X_i)]=\sum_{l=1}^L h_l(x_l)$; (iii) $\sqrt{n}\sum_{l=1}^L K^{-c_l/r_l}=\mathrm{o}(1)$, where $r_l$ is the dimension of $x_l$ (i.e., $x_l\in\mathscr{R}^{r_l}$).

(A6). The support of $X_i$ is $[-1,1]^q$.

When the support of $X_i$ is known and assumption (A5)(i) is satisfied, $X_i$ can always be rescaled so that assumption (A6) holds.

Condition (A5) is restrictive because it rules out random regressors with unbounded support (e.g., Gaussian $X_i$) or discrete regressors. It may be possible to relax the bounded support assumption in (A5)(i) by introducing some bounded transformation of the regressors (e.g., Bierens, 1982; Newey, 1994) provided some additional regularity conditions hold. Newey (1997, p. 167) has shown that for power series, Assumption (A5)(i) implies that the smallest eigenvalue of $E[P^K(Z_i)P^K(Z_i)']$ is bounded for all $K$ ($P^K(z)=Bp^K(z)$), see assumption (A2) and that $\zeta_0(K)=O(K)$. Also it follows from assumption (A5)(ii) and Lorentz (1966) that Assumption (A3)(i) holds with $\delta_l=c_l/r_l$, $l=1,\dots,L$. Thus, Assumption (A5) gives primitive conditions for Assumptions (A2) and (A3) for power series. Newey (1997) has also shown that Assumptions (A5) and (A6) imply that Assumptions (A2) and (A3) hold for B-splines with $\zeta_0(K)=O(\sqrt{K})$. We summarize the above results in the two corollaries below.

**Corollary 2.5.** *For power series, if Assumptions* (A1) *and* (A5) *are satisfied and* $K^3/n\to 0$, *then Theorems* 2.2–2.4 *hold.*

**Corollary 2.6.** *For B-splines, if Assumptions* (A1), (A5) *and* (A6) *are satisfied and* $K^2/n\to 0$, *then Theorems* 2.2–2.4 *hold.*

## 3. Monte Carlo experiments

In this section, we report a small Monte Carlo experiment to examine the finite sample performance of the proposed test. We consider the following null data generating process (DGP):

$$\text{DGP0:}\quad Y_i=X_{1i}X_{2i}\gamma+m_1(X_{1i})+m_2(X_{2i})+U_i, \tag{3.1}$$

where $(\gamma_1,\gamma_2)=(1,1)$, $m_1(x_1)=x_1+x_1^2-0.5x^3$, $m_2(x_2)=x_2+\sin(x_2\pi)$, $X_{li}=V_{1i}+0.5V_{2i}$ $X_{li}=V_{1i}+0.5V_{3i}$, $V_{1i}$, $V_{2i}$ and $V_{3i}$ are independent and uniformly distributed on the interval $[0,2]$, $U_i$ is i.i.d. $N(0,\sigma^2)$ with $\sigma=0.5$.

The alternative DGP is

$$\text{DGP1:}\quad Y_i=\gamma_1+X_{1i}^2X_{2i}\gamma_2+m_1(X_{1i})+m_2(X_{2i})+U_i. \tag{3.2}$$

$\gamma_1,\gamma_2$, $m_1(x_1)$ and $m_2(x_2)$ are defined the same as in DGP0. We use piece-wise cubic splines as base functions to estimate the additive functions $m_1(\cdot)$ and $m_2(\cdot)$. Both exponential and logistic weight functions have been used. The results are similar, so we only report the results based on an exponential weight function: $\mathscr{H}(X_i,x)=\exp(X_i'x)$. The number of replications is 2000, and within each replication 1000 bootstrap test statistics ($CM_n^*$) are computed to yield the bootstrap critical values for the $CM_n$ test. The sample sizes are $n=100$, 200, 500 for size estimation, and $n=100$ and 200 for power estimation. We choose $K_1=K_2$ and use $K=K_1+K_2$ spline functions to approximate the additive function $m_1(\cdot)+m_2(\cdot)$.

We choose some add hoc values of $K$ in the simulations and allow three different values of $K$ for each sample size. We can also justify these choices of $K$ by the

Table 1
Estimated sizes

| | $K=8$ | | | $K=10$ | | | $K=12$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n=100$ | 0.011 | 0.067 | 0.147 | 0.011 | 0.066 | 0.143 | 0.014 | 0.072 | 0.150 |
| | $K=10$ | | | $K=12$ | | | $K=14$ | | |
| | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n=200$ | 0.008 | 0.056 | 0.121 | 0.009 | 0.059 | 0.122 | 0.008 | 0.061 | 0.130 |
| | $K=12$ | | | $K=14$ | | | $K=16$ | | |
| | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n=500$ | 0.010 | 0.054 | 0.103 | 0.008 | 0.052 | 0.108 | 0.012 | 0.057 | 0.107 |

Table 2
Estimated powers

| | $K=8$ | | | $K=10$ | | | $K=12$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n=100$ | 0.326 | 0.726 | 0.846 | 0.362 | 0.742 | 0.858 | 0.376 | 0.7502 | 0.875 |
| | $K=10$ | | | $K=12$ | | | $K=14$ | | |
| | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| $n=200$ | 0.752 | 0.948 | 0.976 | 0.784 | 0.965 | 0.990 | 0.779 | 0.956 | 0.985 |

formula $K=C[n^\alpha]$, where $[\cdot]$ denotes the integer part of $\cdot$, $C$ and $\alpha$ are some positive constants. If we choose $C=3$ and $\alpha=\frac{1}{4}$, we will get $3[n^{1/4}]=9$, 11, and 14 for $n=100$, 200 and 500, respectively. Obviously other choices of $C$ and $\alpha$ can also lead to the $K$ values we used in Table 1.

The estimated sizes and powers are reported in Table 1 and Table 2, respectively.

From Table 1 we can see that for $n=100$, the test is somewhat over sized. However, we observe that the estimated sizes improve as sample size increases, the estimated sizes seem to be satisfactory for $n=200$, and for $n=500$, the estimated sizes are quite close to their nominal sizes.

Table 2 shows that our test is quite powerful in detecting the derivation from the null additive partially linear model as given in DGP1. This is expected since our test is consistent against all deviations from the null model. Also the power increases drastically as the sample size increases from $n=100$ to 200. For $n=500$, the power of our test equal to one for all cases.

Further efforts are needed to investigate the sensitivity of our test to different choices of base functions such as piece-wise higher-order polynomial splines, the choice of different weight functions, and the possibilities of using different bootstrap methods to approximate the null distribution of the test statistic, and using data-driven method to choose series expansion term $K$. We leave all these to future research.

## 4. Some generalizations

In this section, we show that the result of Section 2 can be easily generalized to provide series-based consistent model specification tests for other types of semiparametric models.

### 4.1. Consistent test for an additive model

When $\gamma = 0$ model (1.2) reduces to an additive model without the linear component:

$$Y_i = m_1(X_{1i}) + \cdots + m_L(X_{Li}) + U_i. \tag{4.1}$$

The null hypothesis for testing an additive model is

$$\mathrm{H}_0^b: \ \mathrm{E}(Y_i|X_i) = m_1(X_{1i}) + \cdots + m_L(X_{Li}) \quad \text{a.s.}$$

and the alternative hypothesis $\mathrm{H}_1^b$ is the negation of $\mathrm{H}_0^b$. From Theorem 2.2 we immediately have the following corollary.

**Corollary 4.1.** *Under the same conditions as in Theorem* 2.2 *but with* $\gamma = 0$, *define* $\hat{J}_n^b(x) = \frac{1}{\sqrt{n}} \sum_i \hat{U}_{b,i} \mathscr{H}(X_i, x)$ *and* $CM_n^b = n^{-1} \sum_i [\hat{J}_n^b(X_i)]^2$, *where* $\hat{U}_{b,i} = Y_i - p^K(X_i)'\hat{\beta}_b$ *and* $\hat{\beta}_b = (P'P)^- P'Y$. *Then*

(i) $\hat{J}_n^b(\cdot)$ *converges weakly to* $J_\infty^b(\cdot)$, *and*
(ii) $CM_n^b$ *converges to* $\int [\hat{J}_\infty^b(x)]^2 F(\mathrm{d}x)$ *in distribution*,

*where* $J_\infty^b$ *is a Gaussian process with zero mean and the covariance function given by* $\Sigma_{x,x'} = \mathrm{E}[\sigma^2(X_i)\eta_{b,i}(x)\eta_{b,i}(x')]$ *with* $\eta_{b,i}(x) = \mathscr{H}(X_i, x) - \phi_i(x)$ ($\phi_i(x)$ *is the same as defined in Theorem* 2.2).

**Proof.** (i) is the same as the proof of Theorem 2.2(i) except that one needs to use $\gamma = 0$ and remove the part related to $\hat{\gamma}$, this amounts to remove the $\psi(\cdot)$ term in Theorem 2.2(i). (ii) follows from (i), the continuous mapping theorem, and similar arguments as in Proof of Theorem 2.2(ii).

Similar to the bootstrap statistic $\hat{J}_n^*(\cdot)$, one can define a bootstrap statistic $\hat{J}_n^{*b}(\cdot)$: $\hat{J}_n^{*b}(x) = (1/\sqrt{n}) \sum_i \hat{U}_{b,i}^* \mathscr{H}(X_i, x)$, where $\hat{U}_{b,i}^* = Y_{b,i}^* - p^K(X_i)'\hat{\beta}_b^*$, $\hat{\beta}_b^* = (P'P)^- P'Y_b^*$, $Y_{b,i}^* = p^K(X_i)'\hat{\beta}_b + U_{b,i}^*$, $\hat{\beta}_b = (P'P)^- P'Y$ and $U_{b,i}^*$ is the two point wild bootstrap error obtained from $\hat{U}_{b,i}$. Similar to Proof of Theorem 2.4, one can show that $\hat{J}_n^{*b}(\cdot)$ converges weakly to $J_\infty^{b*}(\cdot)$ ($J_\infty^{b*}(\cdot)$ has the same distribution as $J_\infty^b(\cdot)$). With $\hat{J}_n^{b*}(\cdot)$ one gets the bootstrap version of the CM-type statistic: $CM_n^{*b} = n^{-1} \sum_i [\hat{J}_n^{*b}(X_i)]^2]$ which can be used to approximate the finite sample *null* distribution of $CM_n^b = n^{-1} \sum_i [\hat{J}_n^b(X_i)]^2$.

Gozalo and Linton (2001) propose a consistent test for an additive model in which they estimate both the null and the alternative models nonparametrically by the kernel method. In contrast our $\hat{J}_n^b$ statistic only estimates the *null model* nonparametrically, hence, it partially circumvents the "curse of dimensionality" problem. Also our test can detect Pitman local alternatives that approach the null at a rate of $\mathrm{O}_p(n^{-1/2})$, while the test in Gozalo and Linton (2001) can only detect such local alternatives that are $O_p((n^{1/2}h^{q/4})^{-1})$ apart from the null model, where $h$ is the smoothing parameter, $h \to 0$ as $n \to \infty$. Thus, our test is asymptotically more powerful than that of Gozalo and Linton (2001) against Pitman local alternatives. One advantage of Gozalo and Linton's (2001) test is that it has a simple asymptotic limiting distribution (standard normal), while in our case, the test statistic has a complicated asymptotic distribution (an infinite sum of weighted $\chi_1^2$ random variables), therefore some bootstrap methods are required to approximate the finite sample critical values of the null distribution. On the other hand, bootstrap tests usually give better estimated sizes than the asymptotic tests. The bootstrap method can also be used to approximate the finite sample null distribution of Gozalo and Linton's (2001) test. However, using bootstrap method combined with the kernel marginal integration method to estimate and test for an additive model is computationally costly. In this respect, series-based testing is computationally much less costly.

### 4.2. Test for partially linear models

The result of Section 2 can be used to obtain a consistent test for a partially linear model (without imposing additive structure). Consider the following partially linear model (e.g., Robinson, 1988; Stock, 1989).

$$Y_i = Z_i'\gamma + g(W_i) + U_i, \tag{4.2}$$

where $W_i \in \mathscr{R}^p$, $Z_i \in \mathscr{R}^{q-p}$ $(1 \leqslant p \leqslant q-1)$, and the functional form of $g(\cdot)$ is not specified. Note that since $g(\cdot)$ may not have an additive structure, we cannot allow $Z_i$ to be a deterministic function of $X_i$. Define $V_i = X_i - \mathrm{E}(X_i|Z_i)$, we need to assume that $\mathrm{E}(V_iV_i')$ is positive definite. The null hypothesis is

$$\mathrm{H}_0^c\colon\ \mathrm{E}(Y_i|X_i) = Z_i'\gamma + g(W_i) \quad \text{a.s},$$

and the alternative $\mathrm{H}_1^c$ is the negation of $\mathrm{H}_0^c$.

Estimating partially linear model by series methods are discussed in Donald and Newey (1994), and Newey (1997) among others. Let $p_c^K(w)$ denote a $K \times 1$ series approximating functions. Note that $p_c^K(w)$ does not have an additive structure since $g(w)$ may not be an additive separable function. Define an $n \times K$ matrix $P_c = (p_c^K(W_1), \ldots, p_c^K(W_n))'$. Also let $Z$ be the $n \times r$ matrix with its $i$th row given by $Z_i'$. Finally define an $n \times (r+K)$ matrix $\mathscr{X}_c$ by $\mathscr{X}_c = (Z, P_c)$. Then we estimate $U_i$ by

$$\hat{U}_{c,i} = Y_i - Z_i'\hat{\gamma}_c - p_c^K(X_i)'\hat{\beta}_c, \tag{4.3}$$

where $\hat{\gamma}_c$ and $\hat{\beta}_c$ are given by $\begin{pmatrix}\hat{\gamma}_c \\ \hat{\beta}_c\end{pmatrix} = (\mathscr{X}_c'\mathscr{X}_c)^{-}\mathscr{X}_c'Y$, $Y$ is an $n \times 1$ vector with a typical element given by $Y_i$.

From Theorem 2.2 we immediately have the following Theorem.

**Theorem 4.2.** *Assume that $g(\cdot)$ satisfies the same conditions as $m_1(\cdot)$. Let $X_i = (Z_i', W_i')'$ and $x = (z', w')'$ and $\mathscr{H}(X_i, x) = \mathscr{H}(Z_i, z)\mathscr{H}(W_i, w)$. Define $\hat{J}_n^c(x) = (1/\sqrt{n}) \sum_i \hat{U}_{c,i}\mathscr{H}(X_i, x)$.*

*Then $\hat{J}_n^c(\cdot)$ converges weakly to $J_\infty^c(\cdot)$ under $\mathrm{H}_0^c$,* *where* $\hat{J}^c_\infty$ *is a Gaussian process with zero mean and the covariance function given by*

$$\Sigma^c_{x,x'} = cov(J^c_\infty(x), J^c_\infty(x')) = \mathrm{E}[\sigma^2(X_i)\eta_{c,x}(X_i)\eta_{c,x'}(X_i)],$$

*with* $\eta_{c,x}(X_i) = \mathscr{H}(X_i,x) - \phi_{c,x}(X_i) - \psi_{c,x}(X_i)$, *where* $\phi_{c,x}(X_i) = \mathrm{E}[\mathscr{H}(X_i,x)|W_i]$, $\psi_{c,x}(X_i) = \mathrm{E}[\mathscr{H}(X_i,x)\varepsilon_{c,i}]\{\mathrm{E}[\varepsilon_{c,i}\varepsilon'_{c,i}]\}^{-1}\varepsilon_{c,i}$ *and* $\varepsilon_{c,i} = Z_i - \mathrm{E}(Z_i|W_i)$.

**Proof.** The proof is the same as that of Theorem 2.2(i) except that one needs to replace $z_0(X_i)$ and $E_{\mathscr{G}}(X_i)$ by $Z_i$ and $\mathrm{E}(X_i|W_i)$, respectively, whenever they occur. □

A test statistic for $\mathrm{H}^c_0$ can be based on $CM^c_n = n^{-1}\sum_i [\hat{J}^c_n(X_i)]^2$ and the bootstrap critical values can be obtained from $CM^{*c}_n = n^{-1}\sum_i [\hat{J}^{*c}_n(X_i)]^2$, where $J^{*c}_n(x) = n^{-1/2}\sum_i \hat{U}^*_{c,i}\mathscr{H}(X_i,x)$, $\hat{U}^*_{c,i} = Y^*_{c,i} - Z'_i\hat{\gamma}^*_c - p^K_c(X_i)'\hat{\beta}^*_c$, $\hat{\gamma}^*_c$ and $\hat{\beta}^*_c$ are given by $\begin{pmatrix}\hat{\gamma}^*_c \\ \hat{\beta}^*_c\end{pmatrix} = (X'_cX_c)^{-1}X'_cY^*_c$, $Y^*_{c,i} = Z'_i\hat{\gamma} + p^K_c(X_i)'\hat{\beta} + U^*_{c,i}$, and $U^*_{c,i}$ is the two point wild bootstrap error obtained from $\{\hat{U}_{c,i}\}^n_{i=1}$ ($U_{c,i}$ is given in (4.3)).

## 5. Conclusion

In this paper, we propose to use a series method to construct consistent model specification tests when null models have nonparametric components. The series method is convenient in imposing restrictions such as additive separability. The series method is also convenient to test such restrictions. A leading case we consider is to test for an additive partially linear model. The asymptotic distribution of the test statistic is obtained using a central limit theorem for Hilbert-valued random arrays. We suggest using wild bootstrap methods to approximate the finite sample null distribution of the test statistic. A small Monte Carlo simulation is reported to examine the finite sample performances of the proposed test. We also show that the proposed test can be easily modified to obtain series-based consistent tests for other semiparametric/nonparametric models.

## Acknowledgements

We would like to thank three referees, an associate editor, and Peter Robinson for very helpful comments that greatly improved the paper. We would also like to thank Sittisak Leelahanon for his assistant in carrying out the simulations. C. Hsiao's research is supported by National Science Foundation grant SBR 96-19330. Q. Li's research is supported by Bush Program in the Economics of Public Policy, and the Private Enterprise Research Center, Texas A& M University. J. Zinn's research is partially supported by a NSA grant MDA 904-01-1-0027.

## Appendix A. proofs of the main results

In Appendices A and B, we use (usually capital) letters without subscript $i$ to denote vectors or matrices. For example, $\mathscr{H}(X,x)$, $U$, $\hat{U}$, $m$ and $\phi(x)$ are $n \times 1$ vectors with the $i$th element given by $\mathscr{H}(X_i,x)$, $U_i$, $\hat{U}_i$, $m(X_i)$ and $\phi_i(x)$, respectively. Also for an $n\times 1$ (or $d\times 1$) vector $\mathscr{A}$, we use $\|\mathscr{A}\|$ to denote its Euclidean norm.

**Proof of Theorem 2.2.** (i) Note that $\hat{U}_i = Y_i - z_0(X_i)'\hat{\gamma} - p^K(X_i)\hat{\beta} = U_i - z_0(X_i)'(\hat{\gamma}-\gamma) + m(X_i) - \hat{m}(X_i)$ and $\hat{m}(X_i) \equiv p_K(X_i)'\hat{\beta} = p^K(X_i)'(P'P)^- P'(Y - Z_0\hat{\gamma})$. Hence, we have, in vector–matrix notation, $\hat{m} = P(P'P)^- P'(Y - Z_0\hat{\gamma}) = M_n(Y - Z_0\hat{\gamma}) = M_n[U - Z_0(\hat{\gamma}-\gamma) + m]$ and

$$\hat{U} = U - M_n U - (I_n - M_n)Z_0(\hat{\gamma}-\gamma) + (I_n - M_n)m. \tag{A.1}$$

Using equation (A.1), we get

$$\begin{aligned}\hat{J}_n(x) &= n^{-1/2}\sum_i \mathscr{H}(X_i,x)\hat{U}_i = n^{-1/2}(\mathscr{H}(X,x))'\hat{U}\\ &= n^{-1/2}(\mathscr{H}(X,x))'U - n^{-1/2}(\mathscr{H}(X,x))'M_nU\\ &\quad - n^{-1/2}(\mathscr{H}(X,x))'(I_n - M_n)Z_0(\hat{\gamma}-\gamma) + n^{-1/2}(\mathscr{H}(X,x))'(I_n - M_n)m\\ &\equiv J_{n1}(x) - J_{n2}(x) - J_{n3}(x) + J_{n4}(x).\end{aligned} \tag{A.2}$$

Lemma A.1 of Appendix A shows that, $\|J_{n2}(\cdot) - n^{-1/2}\phi(\cdot)'U\|_v = o_p(1)$, where $\phi(x)$ is an $n\times 1$ vector with the $i$th component given by $\phi_i(x) = E_{\mathscr{G}}[\mathscr{H}(X_i,x)]$. Lemma A.3 establishes that $\|J_{n4}(\cdot) - n^{-1/2}\psi(\cdot)'U\|_v = o_p(1)$, where $\psi(x)$ is an $n\times 1$ vector with the $i$th component given by $\psi_i(x) = \mathrm{E}[\mathscr{H}(X_i,x)\varepsilon_i']\{\mathrm{E}[\varepsilon_i\varepsilon_i']\}^{-1}\varepsilon_i$ and $\varepsilon_i = z_0(X_i) - E_{\mathscr{G}}[z_0(X_i)]$. Lemma A.2 proves that $\|J_{n3}(\cdot)\|_v = o_p(1)$

Define $J_n(x) = n^{-1/2}\sum_i [\mathscr{H}(X_i,x) - \phi_i(x) - \psi_i(x)]U_i \equiv n^{-1/2}\sum_i Z_i(x)$. Then by Lemmas A.1–A.3, we have

$$\|\hat{J}_n(\cdot) - J_n(\cdot)\|_v = o_p(1). \tag{A.3}$$

It is easy to see that $\mathrm{E}[\|J_n(\cdot)\|_v^2] < \infty$, i.e., $J_n(\cdot)$ is tight. Hence, by the central limit theorem for Hilbert-valued random arrays (Lemma 2.1) we have that

$$J_n(\cdot) \text{ converges weakly to } J_\infty(\cdot) \text{ in } \mathscr{L}_2(\mathscr{S}, v, \|\cdot\|_v), \tag{A.4}$$

where $J_\infty(\cdot)$ is a Gaussian process with mean zero and covariance function given by

$$\begin{aligned}\Sigma(x,x') &= cov(J_n(x), J_n(x')) = \mathrm{E}[Z_i(x)Z_i(x')]\\ &= \mathrm{E}\{\sigma^2(X_i)[\mathscr{H}(X_i,x) - \phi_i(x) - \psi_i(x)][\mathscr{H}(X_i,x') - \phi_i(x') - \psi_i(x')]\}.\end{aligned}$$

(A.3) implies that $\hat{J}_n(\cdot)$ and $J_n(\cdot)$ have the same limiting distribution, this and (A.4) imply that $\hat{J}_n(\cdot)$ converges weakly to $J_\infty(\cdot)$. This completes the Proof of Theorem 2.2(i). $\square$

(ii) Obviously $h(J) \stackrel{\text{def}}{=} \int [J(x)]^2 F(\mathrm{d}x)$ is a continuous function in $\mathscr{L}_2(S,F)$. Given that $F$ is absolutely continuous with respect to the Lebesgue measure $v$, $h(J)$ is also continuous in $\mathscr{L}_2(S,v)$. Therefore, by Theorem 2.2(i) and the continuous mapping theorem, we have $\int [\hat{J}_n(x)]^2 F(\mathrm{d}x)$ converges to $\int [J_\infty(x)]^2 F(\mathrm{d}x)$ in distribution.

Now, define $A_n = CM_n - h(\hat{J}_n^2)$. Then

$$A_n = CM_n - h(\hat{J}_n^2) = \int [\hat{J}_n(x)]^2 F_n(\mathrm{d}x) - \int [\hat{J}_n(x)]^2 F(\mathrm{d}x)$$

$$= n^{-2} \sum_i \sum_j \sum_k \hat{U}_i \hat{U}_j \{\mathscr{H}(X_i,X_k)\mathscr{H}(X_j,X_k) - \mathrm{E}[\mathscr{H}(X_i,X_k)\mathscr{H}(X_j,X_k)|X_i,X_j]\}$$

$$\equiv n^{-2} \sum_i \sum_j \sum_k \hat{U}_i \hat{U}_j V_{ijk}, \tag{A.5}$$

where $V_{ijk} = \mathscr{H}(X_i,X_k)\mathscr{H}(X_j,X_k) - \mathrm{E}[\mathscr{H}(X_i,X_k)\mathscr{H}(X_j,X_k)|X_i,X_j]$. Let $g_i \equiv g(X_i) \stackrel{\mathrm{def}}{=} z_0(X_i)'\gamma + m(X_i)$ ($m(X_i)$ is an additive function), and $\hat{g}_i = z_0(X_i)'\hat{\gamma} + p^K(X_i)'\hat{\beta}$, then $\hat{U}_i = Y_i - \hat{g}_i = U_i + (g_i - \hat{g}_i)$. Substituting this into (A.5), yields

$$A_n = n^{-2} \sum_i \sum_j \sum_k [U_i + (g_i - \hat{g}_i)][U_j + (g_j - \hat{g}_j)] V_{ijk}$$

$$= n^{-2} \sum_i \sum_j \sum_k U_i U_j V_{ijk} + 2n^{-2} \sum_i \sum_j \sum_k U_i (g_j - \hat{g}_j) V_{ijk}$$

$$+ n^{-2} \sum_i \sum_j \sum_k (g_i - \hat{g}_i)(g_j - \hat{g}_j) V_{ijk}$$

$$\equiv A_{1n} + 2A_{2n} + A_{3n}, \tag{A.6}$$

where the definitions of $A_{jn}$ $(j = 1,2,3)$ should be apparent.

Using $\mathrm{E}(U_i|X_i) = 0$ and $\mathrm{E}(V_{ijk}|X_i,X_j) = 0$, it is easy to see that

$$A_{1n} = n^{-2} \sum_i \sum_{j \neq i} \sum_{l \neq i, k \neq j} U_i U_j V_{ijk} + \mathrm{O}_p(n^{-1/2}) \equiv A_{1n,1} + \mathrm{O}_p(n^{-1/2}), \tag{A.7}$$

where $A_{1n,1} = n^{-2} \sum_i \sum_{j \neq i} \sum_{k \neq i, k \neq j} U_i U_j V_{ijk}$. It is easy to see that

$$\mathrm{E}[A_{1n,1}^2] = \frac{1}{n^4} \sum_i \sum_{j \neq i} \sum_{k \neq i, k \neq j} \sum_{i'} \sum_{j' \neq i'} \sum_{k' \neq i', k' \neq j'} \mathrm{E}[U_i U_j V_{ijk} U_{i'} U_{j'} V_{i'j'k'}]$$

$$= \frac{1}{n^4} \mathrm{O}(n^3) = \mathrm{O}(n^{-1}). \tag{A.8}$$

(A.7) and (A.8) imply that $A_{1n} = \mathrm{O}_p(n^{-1/2}) = \mathrm{o}_p(1)$.

Next, we show that $A_{2n} = \mathrm{o}_p(1)$. Let $\beta = \beta_f$ satisfy assumption (A3)(i) with $f = m$ (the additive function). Then we have

$$g_i - \hat{g}_i = z_0(X_i)(\gamma - \hat{\gamma}) + m(X_i) - p(x_i)'\beta + p(X_i)'(\beta - \hat{\beta}). \tag{A.9}$$

Substituting (A.9) into $A_{2n}$ we get

$$A_{2n} = n^{-2}\sum_i\sum_j\sum_k U_i\{z_0(X_j)'(\gamma-\hat\gamma)+(m(X_j)-p(X_j)'\beta)+p(X_j)'(\beta-\hat\beta)\}V_{ijk}$$

$$\equiv A_{2n,1}(\gamma-\hat\gamma)+A_{2n,2}+A_{2n,3}(\beta-\hat\beta), \tag{A.10}$$

where $A_{2n,1} = n^{-2}\sum_i\sum_j\sum_k U_i z_0(X_j)'V_{ijk}$, $A_{2n,2} = n^{-2}\sum_i\sum_j\sum_k U_i(m(X_j)-p(X_j)'\beta)V_{ijk}$, and $A_{2n,3} = n^{-2}\sum_i\sum_j\sum_k U_i p(X_j)'V_{ijk}$.

$\mathrm{E}[\|A_{2n,1}\|^2] = n^{-4}\sum_i\sum_j\sum_k\sum_{j'}\mathrm{E}[U_i^2 z_0(X_j)'z_0(X_{j'})V_{ijk}V_{ij'k}] = \mathrm{O}(1)$. Hence, $A_{2n,1}(\gamma-\hat\gamma) = \mathrm{O}_p(1)\mathrm{O}_p(n^{-1/2}) = \mathrm{o}_p(1)$.

$\mathrm{E}[\|A_{2n,2}\|^2] \leqslant n^{-4}\sum_i\sum_j\sum_k\sum_{j'}\mathrm{E}[\sigma^2(X_i)(m(X_j)-p(X_j)'\beta)^2 V_{ijk}V_{ij'k}] \leqslant C\sup_{x\in\mathscr{S}}|m(x)-p(x)'\beta|^2 = \mathrm{O}(\sum_l K^{-2\delta_l}) = \mathrm{o}(1)$. Thus, $A_{2n,2} = \mathrm{o}_p(1)$.

$$\mathrm{E}[\|A_{2n,3}\|^2] = n^{-4}\sum_i\sum_j\sum_k\sum_{j'}\mathrm{E}[\sigma^2(X_i)p(X_j)'p(X_{j'})V_{ijk}V_{ij'k}]$$

$$\leqslant C\mathrm{E}[p(X_i)'p(X_i)]$$

$$= C\,tr\,\mathrm{E}[p(X_i)'p(X_i)] = C\mathrm{E}\{tr[p(X_i)p(X_i)']\} = CK.$$

This implies that $A_{2n,3} = \mathrm{O}_p(K^{1/2})$. Hence, $A_{2n,3}(\beta-\hat\beta) = \mathrm{O}_p(K^{1/2})\mathrm{O}_p(K^{1/2}/n^{1/2}+\sum_{l=1}^L K^{-\delta_l}) = \mathrm{o}_p(1)$.

Summarizing the above we have shown that $A_{2n} = \mathrm{o}_p(1)$.

Using (A.9) $A_{3n}$ can be written as

$$A_{3n} = n^{-2}\sum_i\sum_j\sum_k[(\gamma-\hat\gamma)'z_0(X_i)+(\beta-\hat\beta)'p(X_i)+(m(X_i)-p(X_i)'\beta)]V_{ijk}$$

$$\times[z_0(X_j)'(\gamma-\hat\gamma)+p(X_j)'(\beta-\hat\beta)+(m(X_j)-p(X_j)'\beta)]$$

$$= (\gamma-\hat\gamma)A_{3n,1}(\gamma-\hat\gamma)+A_{3n,2}+A_{3n,3}+\text{other terms}, \tag{A.11}$$

where $A_{3n,1} = n^{-2}\sum_i\sum_j\sum_k z_0(X_i)'z_0(X_j)V_{ijk}$, $A_{3n,2} = n^{-2}\sum_i\sum_j\sum_k(\hat\beta-\beta)'p(X_i)p(X_j)'(\hat\beta-\beta)V_{ijk}$, $A_{3n,3} = n^{-2}\sum_i\sum_j\sum_k(m(X_i)-p(X_i)'\beta)(m(X_j)-p(X_j)'\beta)V_{ijk}$.

$\mathrm{E}[\|A_{3n,1}\|] = n^{-4}\sum_i\sum_j\sum_k\sum_{i'}\sum_{j'}\mathrm{E}[z_0(X_i)z_0(X_j)z_0(X_{i'})Z_0(X_{j'})V_{ijk}V_{i'j'k}] = \mathrm{O}(n)$. Hence, $A_{3n,1} = \mathrm{O}_p(n^{1/2})$ and $(\gamma-\hat\gamma)A_{3n,1}(\gamma-\hat\gamma) = \mathrm{O}_p(n^{-1/2})$ because $\gamma-\hat\gamma = \mathrm{O}_p(n^{-1/2})$.

$$|A_{3n,2}| = \left|n^{-3/2}\sum_i\sum_j(\hat\beta-\beta)'p(X_i)p(X_j)'(\hat\beta-\beta)'\left[n^{-1/2}\sum_k V_{ijk}\right]\right|$$

$$\leqslant n^{-1/2}\sum_i[(\hat\beta-\beta)'p(X_i)p(X_i)'(\hat\beta-\beta)]\left[\sup_{x,x'\in\mathscr{S}}\left|n^{-1/2}\sum_k V_{x,x',X_k}\right|\right]$$

$$\leqslant \{n^{1/2}(\hat\beta-\beta)'(P'P/n)(\hat\beta-\beta)\}\mathrm{O}_p(1)$$

$$= n^{1/2}(\hat{\beta}-\beta)'[I_K + o_p(1)](\hat{\beta}-\beta)O_p(1)$$

$$= 2n^{1/2}O_p\left(K/n + \sum_{l=1}^{L} K^{-2\delta_l}\right) = o_p(1),$$

where in the above $V_{x,x',X_k} = \mathscr{H}(x,X_k)\mathscr{H}(x',X_k) - E[\mathscr{H}(x,X_k)\mathscr{H}(x',X_k)]$, and we used $\sup_{x,x'\in\mathscr{S}} |n^{-1/2}\sum_k V_{x,x',X_k}| = O_p(1)$ by Lemma A.4.

$E[\|A_{3n,3}\|^2] = n^{-4}\sum_i\sum_j\sum_k\sum_{i'}\sum_{j'} E[(m_i - p_i'\beta)(m_j - p_j'\beta)(m_{i'} - p_{i'}'\beta)(m_{j'} - p_{j'}'\beta)V_{ijk}V_{i'j'k}] \leqslant Cn\sup_{x\in\mathscr{S}}|m(x) - p(x)'\beta|^4 = O(n)O(\sum_l K^{-4\delta_l}) = o(1)$, where $m_i = m(X_i)$ and $p_i = p(X_i)$. Hence, $A_{3n,3} = o_p(1)$.

Similarly, one can show that all the other terms in $A_{3n}$ are $o_p(1)$. Thus, we have shown that $A_n = A_{1n} + 2A_{2n} + A_{3n} = o_p(1)$. Therefore, we have

$$CM_n = \int [\hat{J}_n(x)]^2 F(dx) + A_n = \int [\hat{J}_n(x)]^2 F(dx) + o_p(1) \to \int [J_\infty(x)]^2 F(dx), \tag{A.12}$$

in distribution by the result of Theorem 2.2(i) and the continuous mapping theorem. This completes the Proof of Theorem 2.2(ii). □

**Proof of Theorem 2.3.** (i) Following the same proof as that of Theorem 2.2(i), one can show that, under $H_L$, $\|\hat{J}_n(\cdot) - [J_n(\cdot) + n^{-1}\sum_i g(X_i)\mathscr{H}(X_i,\cdot)]\|_v = o_p(1)$. Also, $\|n^{-1}\sum_i g(X_i)\mathscr{H}(X_i,\cdot) - E[g(X_i)\mathscr{H}(X_i,\cdot)]\|_v = o_p(1)$. These imply that

$$\|\hat{J}_n(\cdot) - [J_n(\cdot) + \mu_0(\cdot)]\|_v = o_p(1).$$

Hence, by the same arguments as in the Proof of Theorem 2.2(i) we have that (the tightness of $J_n(\cdot) + \mu_0(\cdot)$ follows from Lemma 2.1)

$\hat{J}_n(\cdot)$ converges weakly to $J_\infty(\cdot) + \mu_0(\cdot)$ in $\mathscr{L}_2(\mathscr{S}, v, \|\cdot\|_v)$.

(ii) Using the similar arguments as in the proof of Theorem 2.2(i), one can show that, under $H_1$,

$\|n^{-1/2}\hat{J}_n(\cdot) - n^{-1}\sum_i [\theta(X_i) - E_{\mathscr{G}^+}(\theta(X_i))]\mathscr{H}(X_i,\cdot)\|_v = o_p(1)$, and that $\|n^{-1}\sum_i [\theta(X_i) - E_{\mathscr{G}^+}(\theta(X_i))]\mathscr{H}(X_i,x) - \mu_1(\cdot)\|_v = o_p(1)$. These imply that $\|n^{-1/2}\hat{J}_n(\cdot) - \mu_1(\cdot)\|_v = o_p(1)$. □

**Proof of Theorem 2.4.** The idea underlying the proof of Theorem 2.4 is very simple. In order to show that a statistic converges to a limiting distribution in *distribution in probability*, we verify that certain conditions hold in probability. Hence, for any subsequence, there is a further subsequence that those conditions hold almost surely. For our $J_n^*(\cdot)$ statistic, we write it in two parts, a leading term converges to a zero mean Gaussian process, and a remainder term that converges to zero in probability. Say, $J_n^*(\cdot) = J_{n1}^*(\cdot) + \Delta_n^*(\cdot)$, and we show that $(J_{n1}^*(\cdot)|Z_1, Z_2, \ldots) \to J_\infty^*(\cdot)$ in distribution in probability, and that $E^*[\|\Delta_n(\cdot)^*\|_v^2] = o_p(1)$. Then $\int [J_n^*(x)]^2\, dF(x) \to \int [J_\infty^*(x)]^2\, dF(x)$ in distribution by the continuous mapping theorem. Finally, we show that $CM_n^* - \int [J_n^*(x)]^2\, dF(x) = o_p(1)$. Thus, $CM_n^* \to \int [J_\infty^*(x)]^2\, dF(x)$ in distribution in probability.

Now we turn to the proof of Theorem 2.4.

$\hat{U}_i^* = Y_i^* - z_0(X_i)'\hat{\gamma}^* - p^K(X_i)'\hat{\beta}^* = U_i^* - z_0(X_i)'(\hat{\gamma}^* - \hat{\gamma}) - p^K(X_i)'(\hat{\beta}^* - \hat{\beta})$. Similar to the derivation of Eq. (A.1), we have $p^K(X_i)'\hat{\beta}^* = p^K(X_i)'(P'P)^{-1}P'(Y^* - Z_0\hat{\gamma}^*)$, and in matrix notation, $P\hat{\beta}^* = P(P'P)^{-1}P'(U^* - Z_0(\hat{\gamma}^* - \hat{\gamma}) + P\hat{\beta}) = M_n(U^* - Z_0(\hat{\gamma}^* - \hat{\gamma}) + P\hat{\beta})$. Hence,

$$\hat{U}^* = U^* - M_n U^* - (I_n - M_n)Z_0(\hat{\gamma}^* - \hat{\gamma}). \tag{A.13}$$

Using Eq. (A.13) we have

$$\begin{aligned}
\hat{J}_n^*(x) &= n^{-1/2}\sum_i \mathscr{H}(X_i,x)\hat{U}_i^* = n^{-1/2}(\mathscr{H}(X,x))'\hat{U}^* \\
&= n^{-1/2}(\mathscr{H}(X,x))'U^* - n^{-1/2}(\mathscr{H}(X,x))'M_n U^* \\
&\quad - n^{-1/2}(\mathscr{H}(X,x))'(I_n - M_n)Z_0(\hat{\gamma}^* - \gamma^*) \\
&\equiv J_{n1}^*(x) - J_{n2}^*(x) - J_{n3}^*(x).
\end{aligned}$$

Similar to Proofs of Lemmas A.1 and A.2, one can show that (i), $\|J_{n2}^*(\cdot) - n^{-1/2}\phi(\cdot) U^*\|_v^2 = \mathrm{o}_p(1)$, and (ii) $\|J_{n3}^*(\cdot) - n^{-1/2}\psi(\cdot)'U^*\|_v^2 = \mathrm{o}_p(1)$. Here, for any two random elements $A_n^*(\cdot)$ and $B_n^*(\cdot)$, $\|A_n^*(\cdot) - B_n^*(\cdot)\|_v^2 = \mathrm{o}_p(1)$ means that $\operatorname{plim}_{n\to\infty}\{\mathrm{E}^*[\|A_n^*(\cdot) - B_n^*(\cdot)\|_v^2]\} = 0$. We provide a proof for (i) below. (ii) can be proved similarly.

Define an $n \times n$ diagonal matrix $D(\hat{U}^2)$ with the $i$th diagonal element given by $\hat{U}_i^2$. Also let $V_i(x)$ be defined as in the proof of Lemma A.1, i.e., $V_i(x) = \mathscr{H}(X_i,x) - \phi_i(x)$. Define $\tilde{\phi}(x) = M_n\phi(x)$ and $\tilde{V}(x) = M_n V(x)$. Then we have

$$\begin{aligned}
\mathrm{E}^*\{\|J_{n2}^*(x) - n^{-1/2}\phi(x)'U^*\|_v^2\} &= \mathrm{E}^*\left\{\int [J_{n2}^*(x) - n^{-1/2}\phi(x)'U^*]^2 v(\mathrm{d}x)\right\} \\
&= \mathrm{E}^*\left\{\int [n^{-1/2}[(\mathscr{H}(X,x))'M_n U^* - \phi(x)'U^*]]^2 v(\mathrm{d}x)\right\} \\
&= n^{-1}\int \{[M_n\mathscr{H}(X,x) - \phi(x)]'\mathrm{E}^*[U^*U^{*'}][M_n\mathscr{H}(X,x) - \phi(x)]\} v(\mathrm{d}x) \\
&= n^{-1}\int \{[M_n(\phi(x) + V(x)) - \phi(x)]'D(\hat{U}^2)[M_n(\phi(x) + V(x)) - \phi(x)]\} v(\mathrm{d}x) \\
&= n^{-1}\int \{[\tilde{\phi}(x) + \tilde{V}(x) - \phi(x)]'D(\hat{U}^2)[\tilde{\phi}(x) + \tilde{V}(x) - \phi(x)]\} v(\mathrm{d}x) \\
&= n^{-1}\sum_i \hat{U}_i^2 \int [\tilde{\phi}_i(x) + \tilde{V}_i(x) - \phi_i(x)]^2 v(\mathrm{d}x).
\end{aligned}$$

Now we consider the case that $n$ is large, it is easy to see that

$$\begin{aligned}
& n^{-1}\sum_i \hat{U}_i^2 \int [\tilde{\phi}_i(x) + \tilde{V}_i(x) - \phi_i(x)]^2 v(\mathrm{d}x) \\
&= n^{-1}\sum_i U_i^2 \int [(\tilde{\phi}_i(x) + \tilde{V}_i(x) - \phi_i(x)]^2 v(\mathrm{d}x) + \mathrm{o}_p(1) \\
&= n^{-1}\sum_i \sigma^2(X_i) \int [(\tilde{\phi}_i(x) + \tilde{V}_i(x) - \phi_i(x)]^2 v(\mathrm{d}x) + \mathrm{o}_p(1)
\end{aligned}$$

$$\leqslant Cn^{-1}\sum_i \int [(\tilde{\phi}_i(x)+\tilde{V}_i(x)-\phi_i(x)]^2 v(\mathrm{d}x)+\mathrm{o}_p(1)$$

$$= C\int [\|\tilde{\phi}(\cdot)-\phi(\cdot)\|_v^2+\|\tilde{V}(\cdot)\|_v^2]+\mathrm{o}_p(1)$$

$$= \mathrm{O}_p\left(\sum_{l=1}^{L} K_l^{-2\delta_l}+K/n\right)+\mathrm{o}_p(1)=\mathrm{o}_p(1),$$

by Lemmas B.1 and B.2.

Therefore, we have shown that $\hat{J}_n^*(\cdot)=n^{-1/2}\sum_i [\mathscr{H}(X_i,\cdot)-\phi(\cdot)-\psi(\cdot)]U_i^*+\mathrm{o}_p(1)\equiv J_n^*(\cdot)+\mathrm{o}_p(1)$, where $J_n^*(\cdot)=n^{-1/2}\sum_i [\mathscr{H}(X_i,\cdot)-\phi(\cdot)-\psi(\cdot)]U_i^*\equiv n^{-1/2}\sum_i Z_i^*(\cdot)$. Lemma 2.1 gives the tightness of $J_n^*(\cdot)$, i.e.,

$$\mathrm{E}^*[\|J_n^*(\cdot)\|_v^2]=n^{-1}\sum_i \hat{U}_i^2\int [\mathscr{H}(X_i,\cdot)-\phi(\cdot)-\psi(\cdot)]^2 v(\mathrm{d}x)\leqslant C\left[n^{-1}\sum_i \hat{U}_i^2\right].$$

When $n$ is large we can replace $\hat{U}_i$ by $U_i$. Applying a weak law of large numbers yields

$$C\left[n^{-1}\sum_i \hat{U}_i^2\right]=C\left[n^{-1}\sum_i U_i^2\right]+\mathrm{o}_p(1)\xrightarrow{p} C\mathrm{E}[\sigma^2(X)]<\infty.$$

The conditional covariance function of $J_n^*(\cdot)$ is

$$cov^*(Z_1^*(x),Z_1^*(x'))=n^{-1}\sum_i [\mathscr{H}(X_i,x)-\phi(x)-\psi(x)][\mathscr{H}(X_i,x')$$
$$-\phi(x_s)-\psi(x')]\hat{U}_i^2.$$

For $n$ large, we can replace $\hat{U}_i$ by $U_i$ and by a weak law of large numbers, we get

$$\begin{aligned}cov^*(Z_1^*(x),Z_1^*(x'))&=n^{-1}\sum_i [\mathscr{H}(X_i,x)-\phi(x)-\psi(x)]\\&\quad[\mathscr{H}(X_i,x')-\phi(x')-\psi(x')]U_i^2\\&=\mathrm{E}\{\sigma^2(X_i)[\mathscr{H}(X_i,x)-\phi(x)-\psi(x)]\\&\quad\mathscr{H}(X_i,x')-\phi(x')-\psi(x')]\}+\mathrm{o}_p(1)=\Sigma(x,x')+\mathrm{o}_p(1).\end{aligned}$$

Next, we consider the finite-dimensional distribution of $J_n^*(x)$. Let $f(\cdot)\in L^2(\mathscr{S},\|\cdot\|_v)$, and $\langle\cdot,\cdot\rangle$ denote the inner product. Define $B(X_i,x)=[\mathscr{H}(X_i,x)-\phi(x)-\psi(x)]$. Then $\langle J_n^*(\cdot),f(\cdot)\rangle=\int J_n^*(x)f(x)v(\mathrm{d}x)=n^{-1/2}\sum_i U_i^*\int B(X_i,x)f(x)v(\mathrm{d}x)\equiv n^{-1/2}\sum_i W_iU_i^*$. $W_i$ only depends on the original data. The $U_i^*$ are conditional independent and have zero means, we only need to verify the Lindeberg's condition. Exactly the same arguments as in Stute et al. (1998, pp. 148–149) shows that the Lindeberg's condition indeed holds. Thus, $\langle J_n^*(\cdot),f(\cdot)\rangle$ converges to a normal variable with zero mean and variance $\mathrm{E}[\sigma^2(X_i)\int B(X_i,x)^2f^2(x)v(\mathrm{d}x)]$. By Cramer–Wold device we obtain the finite-dimensional convergence result.

Summarizing the above we have shown that $J_n^*(\cdot)$ converges weakly to $J_\infty^*(\cdot)$, where $J_\infty^*(\cdot)$ is a zero mean Gaussian process, with covariance function identical to that of $J_\infty(\cdot)$.

Define $A_n^* = CM_n^* - \int [\hat{J}_n^*(x)]^2 \, \mathrm{d}F(x)$.

Then by similar arguments as we did in the proof of Theorem 2.2, one can show that $A_n^* = o_p(1)$. We provide a sketchy proof below. Note that $A_n^*$ can be obtained from $A_n$ given in (A.5) with $\hat{U}_i \hat{U}_j$ replaced by $\hat{U}_i^* \hat{U}_j^*$. Also note that $\hat{U}_i^* = Y_i^* - z_0(X_i)'\hat{\gamma}^* - p(X_i)'\hat{\beta}^* = U_i^* + z_0(X_i)'(\hat{\gamma} - \hat{\gamma}^*) + p(X_i)'(\hat{\beta} - \hat{\beta}^*)$, we have

$$
\begin{aligned}
A_n^* &= n^{-2} \sum_i \sum_j \sum_k \hat{U}_i^* \hat{U}_j^* V_{ijk} \\
&= n^{-2} \sum_i \sum_j \sum_k U_i^* U_j^* V_{ijk} + (\hat{\gamma} - \hat{\gamma}^*)' \left[ n^{-2} \sum_i \sum_j \sum_k z_0(X_i) z_0(X_i)' V_{ijk} \right] \\
&\quad \times (\hat{\gamma} - \hat{\gamma}^*) + n^{-2} \sum_i \sum_j \sum_k (\hat{\beta} - \hat{\beta}^*)' p(X_i) p(X_i)' V_{ijk} (\hat{\beta} - \hat{\beta}^*) + \text{other terms} \\
&\equiv B_{1n} + B_{2n} + B_{3n} + \text{other terms}.
\end{aligned} \tag{A.14}
$$

We will consider $B_{3n}$ first. Similar to the analysis of the $A_{3n}$ term, we have

$$
\begin{aligned}
|B_{3n}| &\leqslant n^{-3/2} \sum_i \sum_j |(\hat{\beta}^* - \beta^*)' p(X_i) p(X_j)' (\hat{\beta}^* - \beta^*)| \left[ \left| n^{-1/2} \sum_k V_{ijk} \right| \right] \\
&\leqslant n^{-1/2} \sum_i \{[(\hat{\beta}^* - \beta^*)' p(X_i) p(X_i)' (\hat{\beta}^* - \beta^*)]^2\} \left[ \sup_{x, x' \in \mathscr{S}} \left| n^{-1/2} \sum_k V_{x, x', X_k} \right| \right] \\
&\leqslant n^{1/2} (\hat{\beta}^* - \hat{\beta})(P'P/n)(\hat{\beta}^* - \hat{\beta}) O_p(1) \\
&= n^{-1/2} [U^{*'} P (P'P)^{-1} (P'P) (P'P)^{-1} P' U^*] O_p(1) \\
&= n^{1/2} [n^{-1} U^{*'} P] (P'P/n)^{-1} [n^{-1} P' U^*] O_p(1) \\
&= n^{1/2} O_p\left( (K/n) + \sum_l K^{-2\delta_l} \right) O_p(1) = o_p(1),
\end{aligned}
$$

where we have used $n^{-1} P' U^* = O_p((K/n)^{1/2} + \sum_l K^{-\delta_l})$ and $P'P/n = O_p(1)$. This is because

$$
\begin{aligned}
&E^*[\|n^{-1} P' U^*\|^2] = n^{-2} E^*[U^{*'} P P' U^*] = n^{-2} [\hat{U}' P P' \hat{U}] \\
&= n^{-2} \sum_i \sum_j \hat{U}_i \hat{U}_j P(X_i)' P(X_j) = n^{-2} \sum_i \sum_j \hat{U}_i \hat{U}_j P(X_i)' P(X_j) \\
&= n^{-2} \sum_i \sum_j [U_i U_j + 2U_i (g_j - \hat{g}_j) + (g_i - \hat{g}_i)(g_j - \hat{g}_j)]
\end{aligned}
$$

$$=\mathrm{O}_p(n^{-1})+\mathrm{O}_p\left(n^{-1/2}\left[(K/n)^{1/2}+\sum_l K^{-\delta_l}\right]\right)+\mathrm{O}_p\left((K/n)+\sum_l K^{-2\delta_l}\right)$$

$$=\mathrm{O}_p\left((K/n)+\sum_l K^{-2\delta_l}\right).$$

Hence, $n^{-1}P'U^*=\mathrm{O}_p((K/n)^{1/2}+\sum_l K^{-\delta_l})$.

Using the fact that $\mathrm{E}^*(U_i^*)=0$ and $(\hat{\gamma}^*-\hat{\gamma})=\mathrm{O}_p(n^{-1/2})$, it is easy to show that $B_{1n}$ and $B_{2n}$ are $\mathrm{o}_p(1)$. Similarly one can show all the other terms are $\mathrm{o}_p(1)$. Therefore, $A_n^*=\mathrm{o}_p(1)$.

Thus, we have

$$CM_n^*=\int [J_n^*(x)]^2\,\mathrm{d}F_n(x)=\int [J_n^*(x)]^2\,\mathrm{d}F(x)+A_n^*$$

$$=\int [J_n^*(x)]^2\,\mathrm{d}F(x)+\mathrm{o}_p(1)\to\int [J_\infty^*(x)]^2\,\mathrm{d}F(x)$$

in distribution by the continuous mapping theorem.

Below we give some lemmas that are used in proofs of Theorems 2.2 and 2.3. For an $n\times d$ matrix $A$, we denote $\tilde{A}=M_nA$ and $\tilde{A}_i$ is the $i$th component of $\tilde{A}$. For example $\tilde{m}=M_nm$ and $\tilde{Z}_0=M_nZ_0$.

**Lemma A.1.** $\|J_{n2}(\cdot)-n^{-1/2}\phi(\cdot)'U\|_v^2=\mathrm{o}_p(1)$, *where* $\phi(x)$ *is a* $n\times 1$ *vector with the ith element given by* $\phi_i(x)=E_{\mathscr{G}}[\mathscr{H}(X_i,x)]$.

**Proof.** Define $V_i(x)=\mathscr{H}(X_i,x)-\phi_i(x)$. Then $E_{\mathscr{G}}(V_i(x))=0$ and $E_{\mathscr{G}}(V_i^2(x))$ is bounded for any $x\in\mathscr{S}$. We have,

$$\mathrm{E}[\|J_{n2}(\cdot)-n^{-1/2}\phi(\cdot)'U\|_v^2|X]$$

$$=n^{-1}\int [(\mathscr{H}(X,x))'M_n-\phi(x)']\mathrm{E}(UU'|X)[M_n\mathscr{H}(X,x)-\phi(x)]v(\mathrm{d}x)$$

$$\leqslant Cn^{-1}\int [M_n\mathscr{H}(X,x)-\phi(x)]'[M_n\mathscr{H}(X,x)-\phi(x)]v(\mathrm{d}x)$$

$$=Cn^{-1}\|M_n\mathscr{H}(X,x)-\phi(x)\|_v^2=Cn^{-1}\|M_n(\phi(x)+V(x)-\phi(x)\|_v^2$$

$$\leqslant 2Cn^{-1}\{\|M_n\phi(x)-\phi(x)\|_v^2]+\|M_nV(x)\|_v^2\}$$

$$=\mathrm{O}_p\left(\sum_{l=1}^{L}K_l^{-2\delta_l}+K/n\right)=\mathrm{o}_p(1)\quad\text{by Lemmas B.1 and B.2.}$$

**Lemma A.2.** $\|J_{n3}(\cdot)-n^{-1/2}\psi(\cdot)U\|_v^2=\mathrm{o}_p(1)$.

**Proof.** Note that $z_0(X_i)-\tilde{z}_0(X_i)$ estimates $\varepsilon_i=z_0(X_i)-E_{\mathscr{G}}[z_0(X_i)]$, or in matrix notation $Z_0-M_nZ_0$ estimates $\varepsilon$. From Lemma B.3 we know that $(\hat{\gamma}-\gamma)=\{\mathrm{E}[\varepsilon_i\varepsilon_i']\}^{-1}n^{-1}\sum_i \varepsilon_iU_i]+$

$\mathrm{o}_p(n^{-1/2})$. Using Lemmas B.1 and B.2 we have $\|n^{-1}\mathscr{H}(X,\cdot)'(I_n - M_n)Z_0 - \mathrm{E}[\mathscr{H}(X_i,\cdot)'\varepsilon_i]\|_v^2 = \mathrm{o}_p(1)$. Hence,

$$\begin{aligned} J_{n3}(\cdot) &= n^{-1/2}\mathscr{H}(X,\cdot)'(I_n - M_n)Z_0(\hat{\gamma} - \gamma) = \mathrm{E}[\mathscr{H}(X_i,\cdot)\varepsilon_i][n^{1/2}(\hat{\gamma} - \gamma)] + \mathrm{o}_p(1) \\ &= \mathrm{E}[\mathscr{H}(X_i,\cdot)\varepsilon_i]\{\mathrm{E}[\varepsilon_i\varepsilon_i']\}^{-1}\left[n^{-1/2}\sum_i \varepsilon_i U_i\right] + \mathrm{o}_p(1) = n^{-1/2}\psi(\cdot)U + \mathrm{o}_p(1). \end{aligned}$$

**Lemma A.3.** $\|J_{n4}(\cdot)\|_v^2 = \mathrm{o}_p(1)$.

**Proof.** $[\|J_{n4}(\cdot)\|_v^2 \leqslant n^{-1}\sum_i\sum_j \int \mathscr{H}(X_i,x)\mathscr{H}(X_j,x)(m_i - \tilde{m}_i)(m_j - \tilde{m}_j)v(\mathrm{d}x) \leqslant C \sum_i \int [(m_i - \tilde{m}_i)^2]v(\mathrm{d}x) = C[\|m - \tilde{m}\|_v^2] = n\mathrm{O}_p(\sum_{l=1}^{L} K_l^{-2\delta_l}) = \mathrm{o}_p(1)$ by Lemma B.1.

**Lemma A.4.** *Denotes* $z = (x,x') \in \mathscr{S} \times \mathscr{S}$ *and define* $V_{z,X_i} = \mathscr{H}(x,X_i)\mathscr{H}(x',X_i) - \mathrm{E}[\mathscr{H}(x,X_i)\mathscr{H}(x',X_i)]$. *Then*

$$\sup_{z\in\mathscr{S}\times\mathscr{S}}\left|n^{-1/2}\sum_i V_{z,X_i}\right| = \mathrm{O}_p(1).$$

**Proof.** By Theorem 3.1 of Ossiander (1987), or a more general result from Anderson et al. (1988), we know that $n^{-1/2}\sum_i V_{z,X_i}$ is tight in $z \in \mathscr{S}\times\mathscr{S}$ under the sup-norm if $|V_{z_1,X_i} - V_{z_2,X_i}| \leqslant A(X_i)\|z_1 - z_2\|$, with $\mathrm{E}(A^2(X_i)) < \infty$. By the assumption that $\mathscr{H}(\cdot,\cdot)$ is bounded and satisfies a Lipschitz condition (see Assumption A.4), it is easy to check that the above conditions hold. Hence, we know that $n^{-1/2}\sum_i V_{z,X_i}$ is tight under the *sup-norm*. The finite-dimensional convergence of $n^{-1/2}\sum_i V_{z,X_i}$ is trivial to check. Thus, $n^{-1/2}\sum_i V_{\cdot,X_i}$ converges weakly to a zero mean Gaussian process (say, $V_\infty(\cdot)$) with covariance structure given by $\mathrm{E}[V_{z,X_i}V_{z',X_i}]$. Consequently, $\sup_{z\in\mathscr{S}\times\mathscr{S}}[n^{-1/2}\sum_i V_{z,X_i}]$ converges weakly to $\sup_{z\in\mathscr{S}\times\mathscr{S}}V_\infty(z) = \mathrm{O}_p(1)$, which in turn implies that $\sup_{z\in\mathscr{S}\times\mathscr{S}}[n^{-1/2}\sum_i V_{z,X_i}] = \mathrm{O}_p(1)$. □

## Appendix B. some useful lemmas

Following the same arguments as in Newey (1997), we will assume $B = I$ in Assumption (A2). Hence $p^K(\cdot) = P^K(\cdot)$, This is because all nonparametric series estimators are invariant to nonsingular transformations of $p^K(\cdot)$. Also we will assume $Q \stackrel{\text{def}}{=} \mathrm{E}[p^K(X_i)p^K(X_i)'] = I$. This is because, for a symmetric square root $Q^{-1/2}$ of $Q^{-1}$, $Q^{-1/2}p^K(\cdot)$ is a nonsingular transformation of $p^K(\cdot)$, and using (A2)(i), it is easy to show that $\tilde{\zeta}_0(K) \stackrel{\text{def}}{=} \sup_{x\in\mathscr{S}}\|Q^{-1/2}p^K(\cdot)\| \leqslant C\zeta_0(K)$. Also if we change $p^K(\cdot)$ to $\bar{p}^K(\cdot) \equiv Q^{-1/2}p^K(\cdot)$ and define $\bar{\beta} = Q^{1/2}\beta$, assumption (A3)(i) is satisfied since $|g(\cdot) - p^K(\cdot)'\beta| = |f(\cdot) - \bar{p}^K(\cdot)'\bar{\beta}|$. Thus, all the assumptions still hold when $p^K(\cdot)$ is changed to $Q^{-1/2}p^K(\cdot)$.

**Lemma B.1.** *Let $f_i(x) \equiv f_0(x, X_i) \in \mathscr{G}$ (the class of additive functions), $f_0(x, X_i)$ is of dimension $d \times 1$ ($d$ is a finite positive integer). Let $f_X(x)$ denote the $n \times d$ matrix with the ith row given by $f_i(x)'$. Define $\tilde{f}_X(x) = M_n f_X(x)$. Then*

$$n^{-1}\|f_X(x) - M_n f_X(x)\|_v^2 = \mathrm{O}_p\left(\sum_l K_l^{-2\delta_l}\right) = \mathrm{o}_p(1).$$

**Proof.** $n^{-1}\mathrm{E}[\|f_X(x) - M_n f_X(x)\|_v^2] \equiv n^{-1}\mathrm{E}[\|f_X(x) - \tilde{f}_X(x)\|_v^2] = n^{-1}\int \mathrm{E}[\|f_X(x) - \tilde{f}_X(x)\|^2]v(\mathrm{d}x) = \mathrm{O}(\sum_l K_l^{-2\delta_l})$ by the result of Andrews and Whang (1990) and Newey (1995, 1997), or see Lemma A.4 of Li (2000) for a proof of this result.

**Lemma B.2.** *Denotes $v_i(x) \equiv V(x, X_i)$ with $E_{\mathscr{G}}(v_i(x)) = 0$ and $E_{\mathscr{G}}([v_i(x)]^2)$ uniformly bounded in $x \in \mathscr{S}$. Define $V(x) = (v_1(x), \ldots, v_n(x))'$ and $\tilde{V}(x) = M_n V(x)$. Then we have*

$$n^{-1}\|M_n V(\cdot)\|_v^2 = n^{-1}\|\tilde{V}(\cdot)\|_v^2 = \mathrm{O}_p(K/n) = \mathrm{o}_p(1).$$

**Proof.** Without loss of generality we can assume $\mathrm{E}[p^K(X_i)p^K(X_i)'] = I_K$ (see the arguments in the beginning of Appendix B). First we show that $\mathrm{E}[\|P'V(\cdot)/n\|_v^2] = \mathrm{O}((K/n)^{1/2})$. Note that $p^K(X_i) \in \mathscr{G}$ and $v_i(\cdot) \perp \mathscr{G}$ imply that $\mathrm{E}[p^K(X_i)v_i(\cdot)] = 0$. We have

$$\begin{aligned}
\mathrm{E}[\|P'V(\cdot)/n\|_v^2] &= n^{-2}\left\{\sum_i \int \mathrm{E}[v_i(x)^2 p^K(X_i)' p^K(X_i)]v(\mathrm{d}x)\right.\\
&\quad \left. + \sum_i \sum_{j \neq i} \int \mathrm{E}[v_i(x) p^K(X_i)']\mathrm{E}[v_j(x) p^K(X_j)]v(\mathrm{d}x)\right\}\\
&= n^{-1}\int \mathrm{E}[v_1(x)^2 p^K(X_1)' p^K(X_1)]v(\mathrm{d}x)\\
&\leqslant Cn^{-1}\mathrm{E}[p^K(X_1)' p^K(X_1)] = \mathrm{O}(K/n).
\end{aligned}$$

This implies that

$$\|P'V(x)/n\|_v^2 = \mathrm{O}_p(K/n) = \mathrm{o}_p(1). \tag{B.1}$$

$$\begin{aligned}
n^{-1}\|M_n V(\cdot)\|_v^2 &= n^{-1}\int V(x)' M_n V(x) v(\mathrm{d}x)\\
&= \int (V(x)'P/n)(P'P/n)^-(P'V(x)/n)v(\mathrm{d}x)\\
&= \int (V(x)'P/n)[I + (P'P/n)^- - I](P'V(x)/n)v(\mathrm{d}x)
\end{aligned}$$

$$= \int \|P'V(x)/n\|^2[1 + o_p(1)]\nu(dx)$$

$$= \int O_p(K/n)[1 + o_p(1)]\nu(dx) = O_p(K/n) = o_p(1)$$

by Eq. (B.1), and the fact that $\|(P'P/n)^- - I\| = O_p(\zeta_0(K)\sqrt{K}/\sqrt{n}) = o_p(1)$ (see the proof of Theorem 1 of Newey (1997, pp. 161–162)).

**Lemma B.3.** $(\hat{\gamma} - \gamma) = \{E[\varepsilon_i\varepsilon_i']\}^{-1}\{n^{-1}\sum_i \varepsilon_i U_i\} + o_p(n^{-1/2})$, *where* $\varepsilon_i = z_0(X_i) - E_{\mathscr{G}}(z_0(X_i))$.

This was proved in Theorem 2.1 of Li (2000). Note that Lemma B.3 implies that $\hat{\gamma} - \gamma = O_p(n^{-1/2})$.

## References

Ait-Sahalia, Y., Bickel, P., Stoker, T.M., 2001. Goodness-of-fit tests for kernel regression with an application to option implied volatilities. Journal of Econometrics 105, 363–412.

Anderson, N.T., Gine, E., Ossiander, M., Zinn, J., 1988. The Central limit theorems and the law of iterative logarithm for empirical processes under local conditions. Probability Theory and Related Fields 77, 271–303.

Andrews, D.W.K., 1991. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. Econometrica 59, 307–345.

Andrews, D.W.K., 1997. A conditional Kolmogorov test. Econometrica 65, 1097–1128.

Andrews, D.W.K., Whang, Y.J., 1990. Additive interactive regression models: circumvention of the curse of dimensionality. Econometric Theory 6, 466–479.

Araujo, A., Gine, E., 1980. The Central Limit Theorem and Banach Valued Random Variables. Wiley, New York.

Bierens, H.J., 1982. Consistent model specification tests. Journal of Econometrics 20, 105–134.

Bierens, H.J., 1990. A consistent conditional moment test of functional form. Econometrica 58, 1443–1458.

Bierens, H.J., Ploberger, W., 1997. Asymptotic theory of integrated conditional moment tests. Econometrica 65, 1129–1151.

Chen, X., Fan, Y., 1999. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. Journal of Econometrics 91, 373–401.

Chen, X., Shen, X., 1998. Sieve extremum estimates for weakly dependent data. Econometrica 66, 289–314.

Chen, X., White, H., 1997. Central limit theorems and functional central limit theorems for Hilbert-valued dependent heterogeneous arrays with applications. Econometric Theory 14, 289–314.

Dechevsky, L., Penez, S., 1997. On shape-preserving probabilistic wavelet approximators. Stochastic Analysis and Applications 15 (2), 187–215.

De John, R.M., 1996. The Bierens test under data dependence. Journal of Econometrics 72, 1–32.

Delgado, M.A., 1993. Testing the equality of nonparametric curves. Probability and Statistics Letters 17, 199–204.

Delgado, M.A., Manteiga, W.G., 2001. Significance testing in nonparametric regression based on the bootstrap. Annals of Statistics 29, 1469–1507.

Delgado, M.A., Stengos, T., 1994. Semiparametric testing of non-nested econometric models. Review of Economic Studies 75, 345–367.

Donald, S.G., 1997. Inference concerning the number of factors in a multivariate nonparametric relationship. Econometrica 65, 103–131.

Donald, S.G., Newey, W.K., 1994. Series estimation of semilinear regression. Journal of Multivariate Analysis 50, 30–40.

Ellison, G., Ellison, S.F., 2000. A simple framework for nonparametric specification testing. Journal of Econometrics 96, 1–23.

Eubank, R., Spiegelman, S., 1990. Testing the goodness of fit of a linear model via nonparametric regression techniques. Journal of the American Statistical Association 85, 387–392.

Fan, Y., Li, Q., 1996a. On estimating additive partially linear models, unpublished manuscript.

Fan, Y., Li, Q., 1996b. Consistent model specification tests: omitted variables, parametric and semiparametric functional forms. Econometrica 64, 865–890.

Fan, Y., Li, Q., 1999. Central limit theorem for degenerate U-statistics of absolutely regular processes with applications to model specification tests. Journal of Nonparametric Statistics 10, 245–271.

Fan, J., Härdle, W., Mammen, E., 1998. Direct estimation of low dimensional components in additive models. Annals of Statistics 26, 943–971.

Gine, E., Zinn, J., 1990. Bootstraping general empirical measures. Annals of Probability 18, 851–869.

Gozalo, P., Linton, O., 2001. Testing additivity in generalized nonparametric regression models with estimated parameters. Journal of Econometrics 104, 1–48.

Härdle, W., Mammen, E., 1993. Comparing nonparametric versus parametric regression fits. Annals of Statistics 21, 1926–1947.

Hong, Y., White, H., 1995. Consistent specification testing via nonparametric series regression. Econometrica 63, 1133–1159.

Horowitz, J.L., Härdle, W., 1994. Testing a parametric model against a semiparametric alternative. Econometric Theory 10, 821–848.

Lavergne, P., 2001. An equality test across nonparametric regressions. Journal of Econometrics 103, 307–344.

Lavergne, P., Vuong, Q., 1996. Nonparametric selection of regressors: the nonnested case. Econometrica 64, 207–219.

Ledoux, M., Talagrand, M., 1991. Probability in Banach Space. Springer, New York.

Lewbel, A., 1995. Consistent nonparametric testing with an application to testing Slusky symmetry. Journal of Econometrics 67, 379–401.

Li, Q., 2000. Efficient estimation of additive partially linear 5 models. International Economic Reviews 41, 1073–1092.

Li, Q., Wang, S., 1998. A simple consistent bootstrap test for a parametric regression functional form. Journal of Econometrics 87, 145–165.

Linton, O.B., 2000. Efficient estimation of generalized additive nonparametric regression models. Econometric Theory 16, 502–523.

Linton, O.B., Nielsen, J.P., 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika 82, 91–100.

Lorentz, G.G., 1966. Approximation of Functions. Chelsea, New York.

Mammen, E., Linton, O., Nielsen, J.P., 1999. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. Annals of Statistics 27, 1443–1490.

Newey, W.K., 1994. Kernel estimation of partial means in a general variance estimator. Econometric Theory 10, 233–253.

Newey, W.K., 1995. Convergence rates for series estimators. In: Maddala, G.S., Phillips, P.C.B., Srinavsan, T.N. (Eds.), Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao. Blackwell, Cambridge, USA, pp. 254–275.

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

Nielsen, J.P., Linton, O.B., 1998. An optimal interpretation of integration and backfitting estimators for separable nonparametric models. Journal of the Royal Statistical Society, Series B 60, 217–222.

Opsomer, J.D., Ruppert, D., 1997. Fitting a bivariate additive model by local polynomial regression. Annals of Statistics 25, 186–211.

Ossiander, M., 1987. A central limit theorem under metric entropy with $L_2$ bracketing. Annals of Probability 15, 897–919.

Politis, D.N., Romano, J.P., 1994. Limit theorems for weakly dependent Hilbert space valued random variables with applications to stationary bootstrap. Statistica Sinica 4, 461–476.

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

Robinson, P.M., 1989. Hypothesis testing in semiparametric and nonparametric models for econometric time series. Review of Economic Studies 56, 511–534.

Robinson, P.M., 1991. Consistent nonparametric entropy-based testing. Review of Economic Studies 58, 437–453.

Schumakers, L.L., 1980. Spline Functions: Basic Theories. Wiley, New York.

Sperlich, S., Tjostheim, D., Yang, L., 2002. Nonparametric estimation and testing of interaction in additive models. Econometric Theory 18, 197–251.

Stinchcombe, M.B., White, H., 1998. Consistent specification testing with nuisance parameters present only under the alternative. Econometric Theory 14, 295–324.

Stock, J.H., 1989. Nonparametric policy analysis. Journal of the American Statistical Association 84, 567–575.

Stone, C.J., 1985. Additive regression and other nonparametric models. Annals of Statistics 13, 685–705.

Stone, C.J., 1986. The dimensionality reduction principle for generalized additive models. Annals of Statistics 14, 592–606.

Stute, W., 1997. Nonparametric model checks for regression. Annals of Statistics 25, 613–641.

Stute, W., Gonzalez, W.G., Presedo, M., 1998. Bootstrap approximation in model checks for regression. Journal of American Statistical Association 93, 141–149.

Tjostheim, D., Auestad, B.H., 1994. Nonparametric identification of nonlinear time series: projections. Journal of American Statistical Association 89, 1398–1409.

Van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes: With Applications to Statistics. Springer, New York.

Wooldridge, J., 1992. A test for functional form against nonparametric alternatives. Econometric Theory 8, 452–475.

Yatchew, A.J., 1992. Nonparametric regression tests based on least squares. Econometric Theory 8, 435–451.

Zheng, J.X., 1996. A consistent test of functional form via nonparametric estimation technique. Journal of Econometrics 75, 263–289.