Consistent Specification Tests for Semiparametric/Nonparametric Models Based on Series Estimation Methods *

Qi Li
Department of Economics, Texas A&M University
College Station, TX 77843, U.S.A.

Cheng Hsiao
Department of Economics, University of Southern California
Los Angeles, CA 90089, U.S.A.

Joel Zinn
Department of Mathematics, Texas A&M University
College Station, TX 77843, U.S.A.

First version: December 1998, This version: January 2002

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 suggest to use the wild bootstrap method to approximate the critical values of the test. 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.

Key words: Consistent tests; Semiparametric models; Series estimation; Wild bootstrap.

*We would like to thank three referees, an associate editor, and Peter Robinson for very helpful comments that greatly improved the paper. 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.

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 have been limited to cases involving only a small number of variables because of the issue of "curse of dimensionality". Stone (1985,1986), 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,

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

to get around the issue of "curse of dimensionality", where X_{li}'s are scalar variables, $l = 1, ..., L$. Model (1.1) has the advantage that it only involves one-dimensional nonparametric functions $m_l(.)$ 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 $m_l(.)$ functions to have some common overlapping variables, the following more general model has been suggested:

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

where X_{li} is of dimension q_l ($q_l \geq 1$), γ 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 $q \times 1$ non-overlapping variables of $X_{1i}, X_{2i}, ..., X_{Li}$. 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, Härdle and Mammen (1998), Fan and Li (1996a), Li (2000), and Sperlich, Tjostheim and Yang (2002),[1] 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

[1] Although Sperlich, Tjostheim and Yang (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.

not sufficiently large, this first step efficiency loss may not 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 non-additive 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 Fan, Härdle and Mammen (1998), and Fan and Li (1996).[2] On the other hand, the series method is less costly in computation because it only involves least squares. Furthermore, the series method can estimate γ efficiently in the sense that the asymptotic variance of a series estimator of γ attains the semiparametric efficiency bound, while the existing kernel marginal integration based method does not (e.g., Fan, Härdle and Mammen (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 Andrews (1997), Bierens (1982,1990), Bierens and Ploberger (1996), De John (1996), Delgado (1993), Delgado and Manteiga (1998), Delgado and Stengos (1994), Donald (1997), Eubank and Spiegelman (1990), Chen and Fan (1999), Härdle and Mammen (1993), Hong and White (1995), Horowitz and Härdle (1994), 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, except Delgado and Manteiga (1998) and Chen and Fan (1999), who consider 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

[2] Note that Fan, Härdle and Mammen (1998), and Fan and Li (1996) did not consider the case of nonparametric additive interaction terms. Allowing additive interaction terms will make the theoretical analysis more complex.

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 (1998) and Chen and Fan (1999) also share properties (i) and (ii), but as discussed above, kernel methods are 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 discussed in section 4. Conclusions are in section 5. The proofs are given in Appendices A and B.

2 A 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:

$$H_0 : E(Y_i|X_i) = z_0(X_i)'\gamma + \sum_{l=1}^{L} m_l(X_{li}) \qquad \text{a.s. for some } \gamma \in \mathcal{B}, \textstyle\sum_{l=1}^{L} m_l(.) \in \mathcal{G}, \tag{2.1}$$

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

Definition 1.

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

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

The alternative hypothesis H_1 is the negation of H_0, i.e.

$$H_1 : E(Y_i|X_i) \neq z_0(X_i)'\gamma + \sum_{l=1}^{L} m_l(X_{li}) \tag{2.2}$$

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

The null hypothesis H_0 is equivalent to $E(U_i|X_i) = 0$ almost surely (a.s.), where U_i is defined in (1.2). Note that $E(U_i|X_i) = 0$ a.s. if and only if $E[U_i M(X_i)] = 0$ for all $M(\cdot) \in \mathcal{M}$, 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 (1997), and Stute (1997), in this paper we consider the following unconditional moment test[3]

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

where $\mathcal{H}(.,.)$ 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 $\mathcal{H}$.

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

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

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

Stute (1997) uses $\mathcal{H}(X_i, x) = \mathbf{1}(X_i \leq x)$ and the Skorohod topology to study the weak convergence of $J_n^0(.)$. Since J_n^0 is a random element in the Skorohod space $\mathcal{D}(\mathcal{S})$, Stute shows that $J_n^0(.)$ converges to a Gaussian process in $\mathcal{D}(\mathcal{S})$. 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. 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 $\mathcal{L}_2(\mathcal{S}, \nu)$ of all real, Borel measurable functions f on $\mathcal{S}$ such that $\int_{\mathcal{S}} f(x)^2\nu(dx) < \infty$, which is endowed with the L_2-norm $||f||_\nu^2 = \int_{\mathcal{S}} f(x)^2\nu(dx)$. The theory of probability on Banach (or Hilbert) spaces, developed in the 1960's and 1970's, turned the problem of studying the asymptotic distribution of statistics like $||J_n^0||_{L_2(\nu)}$ in an easier task, because central limit theorems for random elements taking values in $\mathcal{L}_2(\mathcal{S}, \nu)$. For example, Araujo and Gine (1980, pp205), or Van der Vaart and Wellner (1996, pp.50) assert that for a sequence $\{Z_n(.)\}_n$ of i.i.d. $\mathcal{L}_2(\mathcal{S}, \nu)$-valued elements that $n^{-1/2}\sum_{i=1}^n Z_i(.)$ converges to $\mathcal{Z}(.)$ in

[3]Similar approaches were used by Chen and Fan (1999), and Delgado and Manteiga (1998) to construct kernel-based consistent model specification tests when the null models contain nonparametric components.

the topology of $(\mathcal{L}_2(\mathcal{S},\nu), ||.||_{L_2(\nu)})$ if and only if $\int_{\mathcal{S}} E[Z_1(x)^2]\nu(dx) < \infty$, where $\mathcal{Z}$ 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 $\nu(\mathcal{S}) < \infty$. Since we will only consider the case that $\mathcal{S}$ is a bounded subset of $\mathcal{R}^d$, we will choose $\nu(.)$ to be the Lebesgue measure on $\mathcal{S}$. Then $J_n^0(\cdot)$ is a Hilbert-valued random element in $\mathcal{L}_2(\mathcal{S},\nu)$. We present a H-valued central limit theorem in a lemma below.

Lemma 2.1 *Let $Z_1(.)$, ..., $Z_n(.)$ be H-valued, independent and identically distributed zero mean random elements on $\mathcal{L}_2(\mathcal{S},\nu)$ such that $E[||Z_i(.)||_\nu^2] < \infty$. Then $n^{-1/2}\sum_{i=1}^n Z_i(\cdot)$ converges weakly*[4] *to a zero mean Gaussian process with the covariance (kernel) function given by $\Omega(x,x') = E[Z_i(x)Z_i(x')]$*

Proof: See Theorem 2.1 of Politis and Romano (1994), or van der Vaart and Wellner (1996, ex. 1.8.5, pp50). Note that $E[||Z_i(.)||_\nu^2] < \infty$ is a sufficient condition that ensures the process $n^{-1/2}\sum_{i=1}^n Z_i(.)$ is tight.

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

$$\begin{aligned} E[||Z_i(\cdot)||_\nu^2] &= E\{\int U_i^2[\mathcal{H}(X_i,x)]^2\nu(dx)\} \\ &= E\{\sigma^2(X_i)\int[\mathcal{H}(X_i,x)]^2\nu(dx)\} \le E[\sigma^2(X_i)]\{C\int_{\mathcal{S}}\nu(dx)\} < \infty, \end{aligned}$$

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

Thus by lemma 2.1, we know that

$$J_n^0(.) \text{ converges weakly to } J_\infty^0(.) \text{ in } \mathcal{L}_2(\mathcal{S},\nu,||.||_\nu),$$

where J_∞^0 is a Gaussian process centered at zero and with covariance function Ω given by

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

where x, $x' \in \mathcal{S}$.

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 (2.12) below) and construct a feasible version of (2.4) as

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

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

We will use series estimation method to construct a consistent test based on (2.6). Obviously the individual functions $m_l(\cdot)$ $(l = 1, ..., L)$ are not identified without some identification conditions. In the kernel estimation literature, a convenient identification condition is $E[m_l(X_{li})] = 0$ when X_{li}'s are all scalars $(l = 2, ..., L)$. 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: $E(Y_i|X_{1i}, ..., X_{qi})$ (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,

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

where $x_j \in R$, we can use $m_j(x_j = 0) \equiv m_j(0) = 0$ as the identification condition. When estimating (2.7) by the series method, one simply chooses approximating base functions that are zero at $x_l = 0$. For example, if one uses power series as the base function, one should use $\{x_j^t\}_{t \in \mathcal{N}_1}$ to approximate $m_j(x_j)$, where $\mathcal{N}_1 = \{1, 2, ...\}$ is the set of positive integers.

One can also use power series functions to define identification conditions. We assume that $m_j(x_j)$ can be approximated arbitrarily well by $\{x_j^t\}_{\{t \in \mathcal{N}_1\}}$ in the sense that, for all $\epsilon > 0$, there exists $K_0 \in \mathcal{N}_1$ and $(\alpha_1, ..., \alpha_K)' \in R^K$ such that $E\{[m_j(X_j) - \sum_{t=1}^K \alpha_t X_j^t]^2\} < \epsilon$ for all $K \geq K_0$, then our identification conditions require that $m_j(x_j = 0) = 0$ for all $j = 1, ..., q$.

Note that the above power series only serves as a formal statement for an identification condition. We do not assume that the unknown additive function $m_j(x_j)$ has a power series expansion, i.e., we *do not* assume that $m_j(x_j) = \sum_{l=0}^{\infty} \alpha_l x_j^l$ for some constants $\{\alpha\}_{j=1}^{\infty}$, but rather we only assume that the additive function can be approximated arbitrarily well, in the mean squares error sense, by the power series functions. Indeed the smoothness conditions we impose on $m_j(.)$'s are much weaker than a power series function of the form $\sum_{l=0}^{\infty} \alpha_l x_j^l$, the later is differentiable up to any finite order, while when x_j's are scalars, we only require that $m_j(.)$'s are twice differentiable when one uses the spline series or the power series in the nonparametric estimations, see assumption (A5) (iii) below on the specific smoothness conditions imposed on

$m_j(.)$'s $(j = 1, ..., q)$ on a general setting.

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

$$g(x_1, ..., x_g) = c_0 + \sum_{j=1}^{q} m_j(x_j) + \sum_{j=1}^{q}\sum_{l>j,}^{q} m_{jl}(x_j, x_l). \tag{2.8}$$

In addition to the condition that $m_j(x_j = 0) = 0$, we assume that $m_{jl}(x_j, x_l)$ can be approximated arbitrarily well by $\{x_j^t x_l^s\}_{t,s\in\mathcal{N}_1}$ in terms of the above mean-square-error sense. This also implies that $m_{jl}(x_j, 0) = 0$ and $m_{jl}(0, x_l) = 0$ for all $1 \le j < l \le q$.

In practice one dose not have to use power series to estimate additive models. If one uses $\{\phi_t(x_j)\}_{t\in\mathcal{N}_1}$, $j = 1, ..., q$, as the base function to approximate the additive function $m_j(x_j)$, the above identification condition implies that one should let $\{\phi_t(x_j)\}_{t\in\mathcal{N}_1}$ to approximate $m_j(x_j)$ with $\phi_t(x_j = 0) = 0$ for all $t \in \mathcal{N}_1$, and use $\{\phi_t(x_j)\phi_s(x_l)\}_{t,s\in\mathcal{N}_1}$ to approximate $m_{jl}(x_j, x_l)$. These conditions can be generalized straightforwardly for the identification of 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 γ 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 γ. 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 γ to be identified, we need to assume that $z_0(.)$ 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.,

$$E\{[\mathcal{A}_i - E_{\mathcal{G}}(\mathcal{A}_i)]^2\} = \inf_{\sum_{l=1}^{L}\xi_l(.)\in\mathcal{G}} E\{[\mathcal{A}_i - \sum_{l=1}^{L}\xi(X_{li})]^2\}. \tag{2.9}$$

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

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

(A1). (i) (Y_1, X_1),...,(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; $F(.)$, *the distribution function of* X_1, *is absolutely continuous with respect to the Lebesgue measure.*[5]

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

(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* $E[P^K(X_i)P^K(X_i)']$ *is bounded away from zero uniformly in* $K \in \mathcal{N}$; *(ii) there is a sequence of constants* $\zeta_0(K)$ *satisfying* $sup_{z\in\mathcal{S}}||P^K(z)|| \leq \zeta_0(K)$ *and* $K = K_n$ *such that* $(\zeta_0(K))^2(K/n) \to 0$ *as* $n \to \infty$, *where* $\mathcal{S}$ *is the support of* X.

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

(A4) (i) The weight function $\mathcal{H}(X_i, x) = w(X_i'x)$ *with* $w(.)$ *being an analytic, non-polynomial function.*[6] *(ii)* $\mathcal{H}(.,.)$ *is bounded on* $\mathcal{S} \times \mathcal{S}$ *and satisfies a Lipschitz condition, for all* $x_1, x_2 \in \mathcal{S}$, $|\mathcal{H}(X_i, x_1) - \mathcal{H}(X_i, x_2)| \leq G(X_i)||x_1 - x_2||$ *with* $E[G^2(X_i)] < \infty$; *(iii)* $\nu(.)$ *is the Lebesgue measure.*

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

(A1) (i) and (ii) are standard conditions in the literature of estimating additive models. (A1) (i) rules out discrete random variables.[7] The bounded conditional variance assumption (A1) (ii) is restrictive, however, it still allows a wide range of conditional heteroskedasticity of the form: $U_i = \mathcal{U}_i h(X_i)$, where $\mathcal{U}_i$ is i.i.d. with mean zero and has a finite second moment (say σ^2), $\mathcal{U}_i$ and X_i are independent for all i and j, $h(x)$ is a continous (or bounded) function on $\mathcal{S}$, then $var(U_i|X_i) = \sigma^2[h(X_i)]^2$ is a bounded function in (the compact set) $\mathcal{S}$. (A1) (iii) is an identification condition for γ, requiring that $z_0(.)$ should not lie in $\mathcal{G}$ because otherwise $z_0(.)$ and

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

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

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

$\sum_l m_l(.)$ cannot be identified separately.

(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. (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 $f \in \mathcal{G}$. Hence (A3) (i) is implied by the following: for all $l = 1, ..., L$, there exists some $\delta_l > 0$, $\beta_{fl} = \beta_{fl,K_l}$ (β_{fl} is the lth component of β_f), such that $sup_{x_l \in \mathcal{S}_l} |f_l(x_l) - p_l^{K_l}(x_l)'\beta_{fl}| = O(K_l^{-\delta_l})$, as $K_l \to \infty$, where $\mathcal{S}_l$ is the support of x_l. It is possible to weaken (A3) (i) to $\int_{\mathcal{S}} [f(x) - P^K(x)'\beta_f]^2 \nu(dx) = O(\sum_{l=1}^{L} K_l^{-2\delta_l})$ since we only work with the L_2-norm (the $||.||_\nu$ 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.

(A4) imply that $P[E(U_i|X_i) = 0) < 1$ if and only if $P[E(U_i\mathcal{H}(X_i, x)) = 0) < 1$ (or $\int \{E[U_i\mathcal{H}(X_i, x)]\}^2 F(dx)\} > 0$). Thus, testing $E(U_i|X_i) = 0$ almost everywhere is equivalent to test $E[U_i\mathcal{H}(X_i, x)] = 0$ for almost all $x \in \mathcal{S}$.

We require the approximation function $p^K(x)$ to have the properties that: (a) $p^K(x) \in \mathcal{G}$ and (b) as K_l grows (for all $l = 1, ..., L$), there is a linear combination of $p^K(x)$ that can approximate any $f \in \mathcal{G}$ 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 $\mathcal{H}(X, x) = exp(X_i'x)$ (Bierens (1990)), or $\mathcal{H}(X_i, x) = 1/[1 + exp(c - X_i'x)]$ with $c \neq 0$, or $\mathcal{H}(X_i, x) = cos(X_i'x) + sin(X_i'x)$, (Stichcombe and White (1997)). (A4) (i) to (iii) are very mild conditions on the weight function $\mathcal{H}(.,.)$. They imply that $\int_{\mathcal{S}} \mathcal{H}(y, x)\nu(dx) \leq C \int_{\mathcal{S}} \nu(dx) = C\nu(\mathcal{S}) < \infty$.

Let

$$\begin{aligned} p_l &= (p_l^{K_l}(X_{l1}), ..., p_l^{K_l}(X_{ln}))' \qquad (l = 1, ..., L), \text{ and} \\ P &= (p_1, ..., p_L). \end{aligned} \tag{2.10}$$

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 ith row given by $z_0(X_i)'$. Then in vector-matrix notation, we can write (1.2) as

$$\begin{aligned} Y &= Z_0\gamma + m_1 + m_2 + ... + m_L + U \equiv Z_0\gamma + m + U \\ &= Z_0\gamma + P\beta + (m - P\beta) + U = (Z_0, P)\begin{pmatrix}\gamma \\ \beta\end{pmatrix} + (m - P\beta) + U \\ &= \mathcal{X}\alpha + (m - P\beta) + U, \end{aligned} \tag{2.11}$$

where $\mathcal{X} = (Z_0, P)$, $\alpha = (\gamma', \beta')'$, Y and U are both $n \times 1$ vectors with the ith component given by Y_i and U_i, respectively, m is $n \times 1$ with the ith 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 $\mathcal{X}$:

$$\hat{\alpha} = \begin{pmatrix} \hat{\gamma} \\ \hat{\beta} \end{pmatrix} = (\mathcal{X}'\mathcal{X})^{-}\mathcal{X}'Y, \tag{2.12}$$

where $(\mathcal{X}'\mathcal{X})^{-}$ is a generalized inverse of $(\mathcal{X}'\mathcal{X})$. Li (2000) showed that $\hat{\gamma} - \gamma = O_p(n^{-1/2})$. Also $\hat{m}(x) - m(x) = O_p((K/n)^{1/2} + \sum_{l=1}^{L} K_l^{-\delta_l})$ 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

$$\hat{U}_i = Y_i - z_0(X_i)'\hat{\gamma} - p^K(X_i)'\hat{\beta}. \tag{2.13}$$

Our test statistic for H_0 is based on

$$\hat{J}_n(x) = \frac{1}{\sqrt{n}} \sum_i \mathcal{H}(X_i, x)\hat{U}_i, \tag{2.14}$$

where $\hat{U}_i$ is given in (2.12). With $\hat{J}_n(x)$ we can construct a Cramer-von Mises (CM) type statistic for testing H_0.

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

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

The next theorem establishes the weak convergence of $\hat{J}_n(.)$ and CM_n under H_0.

THEOREM 2.2 *Suppose that assumptions (A1) to (A4) hold, then under H_0,*

(i) $\hat{J}_n(.)$ converges weakly to $J_\infty(.)$ in $\mathcal{L}_2(\mathcal{S}, \nu, ||.||_\nu)$,

where J_∞ is a Gaussian process with zero mean and covariance function given by

$\Sigma_1(x, x') = E[\sigma^2(X_i)\eta_i(x)\eta_i(x')]$, where $\eta_i(x) = \mathcal{H}(X_i, x) - \phi_i(x) - \psi_i(x)$, with $\phi_i(x) = E_{\mathcal{G}}[\mathcal{H}(X_i, x)]$, $\psi_i(x) = E[\mathcal{H}(X_i, x)\epsilon_i]\{E[\epsilon_i\epsilon_i']\}^{-1}\epsilon_i$ and $\epsilon_i = z_0(X_i) - E_{\mathcal{G}}(z_0(X_i))$.

(ii) CM_n converges to $\int [J_\infty(x)]^2 F(dx)$ in distribution,

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

The proof of Theorem 2.2 (i) is given in the Appendix A. Theorem 2.2 (ii) follows from Theorem 2.3 (i), the continuous mapping theorem, and the fact that $F_n(.)$ is close to $F(.)$ ($F_n(.)$ is the empirical distribution of $\{X_i\}_{i=1}^n$). The idea underlying the proof of Theorem 2.3 (i) is very simple. First we write $\hat{J}_n(.) = J_n(.) + [\hat{J}_n(.) - J_n(.)]$, where $J_n(x) = n^{-1/2}\sum_i U_i[\mathcal{H}(X_i, x) - \phi_i(x) - \psi_i(x)]$. $J_n(.)$ converges weakly to the Gaussian process $J_\infty(.)$ by lemma 2.1 (because $E[||J_n(.)||_\nu^2] < \infty$). Next, we show that $||\hat{J}_n(.) - J_n(.)||_\nu = o_p(1)$. This implies that $\hat{J}_n(.)$ and $J_n(.)$ have the same limiting distribution. Therefore, $\hat{J}_n(.)$ converges weakly to $J_\infty(.)$ in $\mathcal{L}_2(\mathcal{S}, \nu, ||.||_\nu)$.

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

$$H_L : E(Y_i|X_i) = z_0(X_i)'\gamma + m(X_i) + \frac{g(X_i)}{\sqrt{n}} \quad a.s., \tag{2.15}$$

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

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

$$E\{[\mathcal{A}_i - E_{\mathcal{G}^+}(\mathcal{A}_i)]^2\} = \inf_{\gamma\in\mathcal{B} \text{ and } \sum_{l=1}^L \xi_l(.)\in\mathcal{G}} E\{[\mathcal{A}_i - z_0(X_i)'\gamma - \sum_{l=1}^L \xi(X_{li})]^2\}. \tag{2.16}$$

Let $Y_i = \theta(X_i) + U_i$ under H_1, then $\theta(.)$ does not belong to $\mathcal{G}^+$. Using similar arguments as in the proof of Theorem 2.3 (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 $E_{\mathcal{G}^+}(Y_i) = E_{\mathcal{G}^+}(\theta(X_i))$ because $E_{\mathcal{G}^+}(U_i) = 0$. Hence, $\hat{U}_i = \theta(X_i) + U_i - z_0(X_i)'\hat{\gamma} - p^K(X_i)'\hat{\beta} = \theta(X_i) - E_{\mathcal{G}^+}(\theta(X_i)) + U_i + o_p(1)$ under H_1. The following theorem gives the asymptotic distribution of $\hat{J}_n(.)$ under the local alternative H_L and the fixed alternative H_1.

THEOREM 2.3 *Suppose that (A1) to (A4) hold,*

(i) Under the local alternative H_L defined in (2.15),

$\hat{J}_n(.)$ converges weakly to $J_\infty(.) + \mu_0(.) \equiv J_{L,\infty}(.)$ in $\mathcal{L}_2(\mathcal{S}, \nu, ||.||_\nu)$, where $\mu_0(x) = E[g(X_i)\mathcal{H}(X_i, x)]$.

(ii) Under the fixed alternative H_1 defined in (2.2),

$n^{-1/2}\hat{J}_n(x)$ converges to $\mu_1(.)$ in probability in $\mathcal{L}_2(\mathcal{S}, \nu, ||.||_\nu)$, where $\mu_1(x) = E\{[\theta(X_i) - E_{\mathcal{G}^+}(\theta(X_i))]\mathcal{H}(X_i, x)\}$.

The proof of Theorem 2.3 is given in the 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 $\int [J_\infty(x) + \mu_0(x)]^2 F(dx)$ under H_L, and $CM_n = n \int [\mu_1(x)]^2 F(dx) + o_p(n)$ 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 $\int [J_\infty(x)]^2 F(dx)$ can be written as an infinite sum of weighted (independent) χ_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 the 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

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

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

$$Y_i^* = z_0(X_i)'\hat{\gamma} + p^K(X_i)'\hat{\beta} + U_i^*, \qquad i = 1, ..., n \tag{2.17}$$

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

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

where $\mathcal{X} = (Z_0, P)$ and $\mathcal{Y}^*$ is an $n \times 1$ vector with jth 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

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

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

Many authors show that some bootstrap method works using the concept of $(\xi_n|Z_1, Z_2, ...)$ converges to $(\xi|Z_1, Z_2, ...)$ 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. See Gine and Zinn (1990) on more detailed discussions on 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 H_0,*

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

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

Theorem 2.4 is proved in Appendix A.

Assumptions (A2) and (A3) used in Theorems 2.2 to 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 Schumaker (1981).

(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 $l = 1, ..., L$, $f_l(x_l)$ is continuously differentiable of order c_l on the the support of X_{il}, where $f_l(\cdot) = m_l(\cdot)$ or $f_l(.) = h_l(.)$ $(l = 1, ..., L)$, where $h_l(.)$ is defined from $E_{\mathcal{G}}[z_0(X_i)] = \sum_{l=1}^L h_l(x_l)$; (iii) $\sqrt{n}\sum_{l=1}^L K^{-c_l/r_l} = o(1)$, where r_l is the dimension of x_l (i.e., $x_l \in \mathcal{R}^{r_l}$).

(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.

(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 2) and that $\zeta_0(K) = O(K)$. Also it follows from assumption (A5) (ii) and Lorentz (1966) that assumption (A3) (i) holds with $\delta_l = c_l/r_l$, $l = 1, ..., L$. 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 assumption (A1) and assumption (A5) are satisfied and $K^3/n \to 0$, then Theorem 2.2, Theorem 2.3 and Theorem 2.4 hold.*

Corollary 2.6 *For B-splines, if assumptions (A1), (A5) and (A6) are satisfied and $K^2/n \to 0$, then Theorem 2.2, Theorem 2.3 and Theorem 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):

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

where $m_1(x_1) = x_1 + x_1^2 - .5x^3$, $m_2(x_2) = x_2 + sin(x_2\pi)$, X_{li} is uniformly distributed on the interval $[0, 2]$ ($l = 1, 2$), U_i is i.i.d. $N(0, \sigma^2)$ with $\sigma = 0.5$.

The alternative DGP is

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

We use piece-wise linear splines as base functions to estimate the additive functions $m_1(.)$ and $m_2(.)$. 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 $m_1(.) + m_2(.)$.

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 formula $K = C[n^\alpha]$, where [.] denotes the integer part of ., C and α are some positive constants. If we choose $C = 3$ and

$\alpha = 1/4$, we will get $3[n^{1/4}] = 9$, 11, and 14 for $n = 100$, 200 and 500, respectively. Obviously other choices of C and α 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.

Table 1: Estimated Sizes

	$K = 8$			$K = 10$			$K = 12$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
$n = 100$	.011	.071	.167	.012	.068	.160	.015	.087	.165
	$K = 10$			$K = 12$			$K = 14$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
$n = 200$	.006	.056	.121	.012	.063	.131	.009	.057	.139
	$K = 12$			$K = 14$			$K = 16$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
$n = 500$	.010	.058	.102	.006	.054	.116	.012	.056	.108

Table 2: Estimated Powers

	$K = 8$			$K = 10$			$K = 12$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
$n = 100$	.270	.620	.746	.298	.602	.742	.318	.662	.774
	$K = 10$			$K = 12$			$K = 14$		
	1%	5%	10%	1%	5%	10%	1%	5%	10%
$n = 200$	.812	.934	.964	.828	.951	.975	.821	.950	.982

From Table 1 we can see that for $n = 100$, the test is over sized. However, we observe that the estimated sizes improve as sample size increases although the estimated sizes are still somewhat oversized for n=200. 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 $n = 200$. 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 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 A Consistent Test For An Additive Model

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

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

The null hypothesis for testing an additive model is

$H_0^b : E(Y_i|X_i) = m_1(X_{1i}) + ... + m_L(X_{Li})$ *a.s.*

and the alternative hypothesis H_1^b is the negation of H_0^b. From Theorem 2.2 we immediately have the following corollary.

Corollary 4.1 *Under the same conditions as in Theorem 2.3 but with* $\gamma = 0$*, define* $\hat{J}_n^b(x) = \frac{1}{\sqrt{n}} \sum_i \hat{U}_{b,i} \mathcal{H}(X_i, x)$ *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) $\hat{J}_n^b(.)$ *converges weakly to* $J_\infty^b(.)$*, and*

(ii) CM_n^b *converges to* $\int [\hat{J}_\infty^b(x)]^2 F(dx)$ *in distribution,*

where J_∞^b *is a Gaussian process with zero mean and the covariance function given by*

$\Sigma_{x,x'} = E[\sigma^2(X_i)\eta_{b,i}(x)\eta_{b,i}(x')]$ *with* $\eta_{b,i}(x) = \mathcal{H}(X_i, x) - \phi_i(x)$ *(*$\phi_i(x)$ *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 $\lambda = 0$ and remove the part related to $\hat{\gamma}$, this amounts to remove the $\psi(.)$ term in Theorem 2.2 (i). (ii) follows from (i), the continuous mapping theorem, and similar arguments as in the proof of Theorem 2.2 (ii).

Similar to the bootstrap statistic $\hat{J}_n^*(.)$, one can define a bootstrap statistic $\hat{J}_n^{*b}(.)$: $\hat{J}_n^{*b}(x) = \frac{1}{\sqrt{n}} \sum_i \hat{U}_{b,i}^* \mathcal{H}(X_i, x)$, 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 the proof of Theorem 2.4, one can show that $\hat{J}_n^{*b}(.)$ converges weakly to $J_\infty^{b*}(.)$ ($J_\infty^{b*}(.)$ has the same distribution as $J_\infty^b(.)$). With $\hat{J}_n^{b*}(.)$ 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 $O_p(n^{-1/2})$, 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 χ_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 A 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) and Stock (1989)).

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

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

$$H_0^c : E(Y_i|X_i) = Z_i'\gamma + g(W_i) \quad a.s.,$$

and the alternative H_1^c is the negation of 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 $P_c = (p_c^K(W_1), ..., p_c^K(W_n))'$. Also let Z be the $n \times r$

matrix with its ith row given by Z_i'. Finally define an $n \times (r+K)$ matrix $\mathcal{X}_c$ by $\mathcal{X}_c = (Z, P_c)$. 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 $\begin{pmatrix} \hat{\gamma}_c \\ \hat{\beta}_c \end{pmatrix} = (\mathcal{X}_c'\mathcal{X}_c)^{-}\mathcal{X}_c'Y$, 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 $g(.)$ satisfies the same conditions as $m_1(.)$. Let $X_i = (Z_i', W_i')'$ and $x = (z', w')'$ and $\mathcal{H}(X_i, x) = \mathcal{H}(Z_i, z)\mathbf{1}(W_i \le w)$. Define $\hat{J}_n^c(x) = \frac{1}{\sqrt{n}}\sum_i \hat{U}_{c,i}\mathcal{H}(X_i, x)$.*

Then $\hat{J}_n^c(.)$ converges weakly to $J_\infty^c(.)$ under H_0^c,

where $\hat{J}_\infty^c$ is a Gaussian process with zero mean and the covariance function given by

$$\Sigma_{x,x'}^c = Cov(J_\infty^c(x), J_\infty^c(x')) = E[\sigma^2(X_i)\eta_{c,x}(X_i)\eta_{c,x'}(X_i)],$$

with $\eta_{c,x}(X_i) = \mathcal{H}(X_i, x) - \phi_{c,x}(X_i) - \psi_{c,x}(X_i)$, where $\phi_{c,x}(X_i) = E[\mathcal{H}(X_i, x)|W_i]$, $\psi_{c,x}(X_i) = E[\mathcal{H}(X_i, x)\epsilon_{c,i}]\{E[\epsilon_{c,i}\epsilon_{c,i}']\}^{-1}\epsilon_{c,i}$ and $\epsilon_{c,i} = Z_i - E(Z_i|W_i)$.

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

A test statistic for H_0^c can be based on $CM_n^c = n^{-1}\sum_i[\hat{J}_n^c(X_i)]^2$ and the bootstrap critical values can be obtained from $CM_n^{*c} = n^{-1}\sum_i[\hat{J}_n^{*c}(X_i)]^2$, where $J_n^{*c}(x) = n^{-1/2}\sum_i \hat{U}_{c,i}^*\mathcal{H}(X_i, x)$, $\hat{U}_{c,i}^* = Y_{c,i}^* - p_c^K(X_i)'\hat{\beta}_c^*$, $\hat{\beta}_c^* = (P'P)^{-}P'Y_c^*$, $Y_{c,i}^* = p_c^K(X_i)'\hat{\beta}_c + U_{c,i}^*$, $\hat{\beta}_c = (P_c'P_c)^{-}P_c'Y$ and $U_{c,i}^*$ is the two point wild bootstrap error obtained from $\{\hat{U}_{c,i}\}_{i=1}^n$ ($U_{c,i}$ 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.

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, $\mathcal{H}(X,x)$, U, $\hat{U}$, m and $\phi(x)$ are $n \times 1$ vectors with the ith element given by $\mathcal{H}(X_i,x)$, U_i, $\hat{U}_i$, $m(X_i)$ and $\phi_i(x)$, respectively. Also for an $n \times 1$ (or $d \times 1$) vector $\mathcal{A}$, we use $||\mathcal{A}||$ 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

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

Using equation (A.1), we get

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

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

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

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

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

$$J_n(.) \text{ converges weakly to } J_\infty(.) \text{ in } \mathcal{L}_2(\mathcal{S},\nu,||.||_\nu), \tag{A.4}$$

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

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

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

Proof of Theorem 2.2 (ii).

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

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

$$\begin{aligned} A_n &= CM_n - h(\hat{J}_n^2) = \int [\hat{J}_n(x)]^2 F_n(dx) - \int [\hat{J}_n(x)]^2 F(dx) \\ &= n^{-2} \sum_i \sum_j \sum_k \hat{U}_i \hat{U}_j \{\mathcal{H}(X_i, X_k)\mathcal{H}(X_j, X_k) - E[\mathcal{H}(X_i, X_k)\mathcal{H}(X_j, X_k)|X_i, X_j]\} \\ &\equiv n^{-2} \sum_i \sum_j \sum_k \hat{U}_i \hat{U}_j V_{ijk}, \end{aligned} \quad \text{(A.5)}$$

where $V_{ijk} = \mathcal{H}(X_i, X_k)\mathcal{H}(X_j, X_k) - E[\mathcal{H}(X_i, X_k)\mathcal{H}(X_j, X_k)|X_i, X_j]$. Let $g_i \equiv g(X_i) \stackrel{def}{=} z_0(X_i)'\gamma + m(X_i)$ ($m(X_i)$ 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

$$\begin{aligned} A_n &= n^{-2} \sum_i \sum_j \sum_k [U_i + (g_i - \hat{g}_i)][U_j + (g_j - \hat{g}_j)] V_{ijk} \\ &= n^{-2} \sum_i \sum_j \sum_k U_i U_j V_{ijk} + 2n^{-2} \sum_i \sum_j \sum_k U_i (g_j - \hat{g}_j) V_{ijk} \\ &\quad + n^{-2} \sum_i \sum_j \sum_k (g_i - \hat{g}_i)(g_j - \hat{g}_j) V_{ijk} \\ &\equiv A_{1n} + 2A_{2n} + A_{3n}, \end{aligned} \quad \text{(A.6)}$$

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

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

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

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

$$E[A_{1n,1}^2] = \frac{1}{n^4} \sum_i \sum_{j \neq i} \sum_{k \neq i, k \neq j} \sum_{i'} \sum_{j' \neq i'} \sum_{k' \neq i', k' \neq j'} E[U_i U_j V_{ijk} U_{i'} U_{j'} V_{i'j'k'}] = \frac{1}{n^4} O(n^3) = O(n^{-1}). \text{(A.8)}$$

(A.7) and (A.8) imply that $A_{1n} = O_p(n^{-1/2}) = o_p(1)$.

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

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

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

$$A_{2n} = n^{-2}\sum_i\sum_j\sum_k U_i\{z_0(X_j)'(\gamma-\hat{\gamma}) + (m(X_j)-p(X_j)'\beta) + p(X_j)'(\beta-\hat{\beta})\}V_{ijk}$$
$$\equiv A_{2n,1}(\gamma-\hat{\gamma}) + A_{2n,2} + A_{2n,3}(\beta-\hat{\beta}), \qquad (A.10)$$

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

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

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

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

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

Summarizing the above we have shown that $A_{2n} = o_p(1)$.

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

$$A_{3n} = n^{-2}\sum_i\sum_j\sum_k[(\gamma-\hat{\gamma})'z_0(X_i) + (\beta-\hat{\beta})'p(X_i) + (m(X_i)-p(X_i)'\beta)]V_{ijk}$$
$$\times[z_0(X_j)'(\gamma-\hat{\gamma}) + p(X_j)'(\beta-\hat{\beta}) + (m(X_j)-p(X_j)'\beta)]$$
$$= (\gamma-\hat{\gamma})A_{3n,1}(\gamma-\hat{\gamma}) + A_{3n,2} + A_{3n,3} + \text{ other terms}, \qquad (A.11)$$

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

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

$|A_{3n,2}| = |n^{-3/2}\sum_i\sum_j(\hat{\beta}-\beta)'p(X_i)p(X_j)'(\hat{\beta}-\beta)'[n^{-1/2}\sum_k V_{ijk}]||$
$\le n^{-1/2}\sum_i[(\hat{\beta}-\beta)'p(X_i)p(X_i)'(\hat{\beta}-\beta)][\,sup_{x,x'\in\mathcal{S}}\,|\,n^{-1/2}\sum_k V_{x,x',X_k}|\,]$
$\le \{n^{1/2}(\hat{\beta}-\beta)'(P'P/n)(\hat{\beta}-\beta)\}O_p(1)$
$= n^{1/2}(\hat{\beta}-\beta)'[I_K + o_p(1)](\hat{\beta}-\beta)O_p(1)$
$= 2n^{1/2}O_p(K/n + \sum_{l=1}^L K^{-2\delta_l}) = o_p(1)$.
where in the above $V_{x,x',X_k} = \mathcal{H}(x,X_k)\mathcal{H}(x',X_k) - E[\mathcal{H}(x,X_k)\mathcal{H}(x',X_k)]$, and we used $sup_{x,x'\in\mathcal{S}}|n^{-1/2}\sum_k V_{x,x',X_k}| = O_p(1)$ by lemma A.4.

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

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

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

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

$\hat{J}_n(.)$ converges weakly to $J_\infty(.) + \mu_0(.)$ in $\mathcal{L}_2(\mathcal{S}, \nu, ||.||_\nu)$.

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

$||n^{-1/2}\hat{J}_n(.) - n^{-1}\sum_i[\theta(X_i) - E_{\mathcal{G}^+}(\theta(X_i))]\mathcal{H}(X_i,.)||_\nu = o_p(1)$, and that

$||n^{-1}\sum_i[\theta(X_i) - E_{\mathcal{G}^+}(\theta(X_i))]\mathcal{H}(X_i,x) - \mu_1(.)||_\nu = o_p(1)$. These imply that

$||n^{-1/2}\hat{J}_n(.) - \mu_1(.)||_\nu = o_p(1)$.

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 $J_n^*(.)$ 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, $J_n^*(.) = J_{n1}^*(.) + \Delta_n^*(.)$, and we show that $(J_{n1}^*(.)|Z_1, Z_2, ...) \to J_\infty^*(.)$ in distribution in probability, and that $E^*[||\Delta_n(.)^*||_\nu^2] = o_p(1)$. 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 equation (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,

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

Using equation (A.13) we have

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

Similar to the proofs of lemma A.1 and lemma A.2, one can show that (i), $||J_{n2}^*(.) - n^{-1/2}\phi(.)U^*||_\nu^2 = o_p(1)$, and (ii) $||J_{n3}^*(.) - n^{-1/2}\psi(.)'U^*||_\nu^2 = o_p(1)$. Here, for any two random elements $A_n^*(.)$ and $B_n^*(.)$, $||A_n^*(.)-B_n^*(.)||_\nu^2 = o_p(1)$ means that $plim_{n\to\infty}\{E^*[||A_n^*(.)-B_n^*(.)||_\nu^2]\} = 0$. 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 ith 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., $V_i(x) = \mathcal{H}(X_i,x) - \phi_i(x)$. Define $\tilde{\phi}(x) = M_n\phi(x)$ and $\tilde{V}(x) = M_nV(x)$. Then we have

$E^*\{||J_{n2}^*(x) - n^{-1/2}\phi(x)'U^*||_\nu^2\} = E^*\{\int[J_{n2}^*(x) - n^{-1/2}\phi(x)'U^*]^2\nu(dx)\}$
$= E^*\{\int[n^{-1/2}[(\mathcal{H}(X,x))'M_nU^* - \phi(x)'U^*]]^2\nu(dx)\}$
$= n^{-1}\int\{[M_n\mathcal{H}(X,x)-\phi(x)]'E^*[U^*U^{*'}][M_n\mathcal{H}(X,x)-\phi(x)]\}\nu(dx)$
$= n^{-1}\int\{[M_n(\phi(x)+V(x))-\phi(x)]'D(\hat{U}^2)[M_n(\phi(x)+V(x))-\phi(x)]\}\nu(dx)$
$= n^{-1}\int\{[\tilde{\phi}(x)+\tilde{V}(x)-\phi(x)]'D(\hat{U}^2)[\tilde{\phi}(x)+\tilde{V}(x)-\phi(x)]\}\nu(dx)$
$= n^{-1}\sum_i \hat{U}_i^2\int[\tilde{\phi}_i(x)+\tilde{V}_i(x)-\phi_i(x)]^2\nu(dx)$.

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

$n^{-1}\sum_i \hat{U}_i^2\int[\tilde{\phi}_i(x)+\tilde{V}_i(x)-\phi_i(x)]^2\nu(dx)$
$= n^{-1}\sum_i U_i^2\int[(\tilde{\phi}_i(x)+\tilde{V}_i(x)-\phi_i(x)]^2\nu(dx) + o_p(1)$
$= n^{-1}\sum_i \sigma^2(X_i)\int[(\tilde{\phi}_i(x)+\tilde{V}_i(x)-\phi_i(x)]^2\nu(dx) + o_p(1)$
$\leq Cn^{-1}\sum_i\int[(\tilde{\phi}_i(x)+\tilde{V}_i(x)-\phi_i(x)]^2\nu(dx) + o_p(1)$
$= C\int[||\tilde{\phi}(.)-\phi(.)||_\nu^2 + ||\tilde{V}(.)||_\nu^2] + o_p(1)$
$= O_p(\sum_{l=1}^L K_l^{-2\delta_l} + K/n) + o_p(1) = o_p(1)$,

by lemma B.1 and B.2.

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

$E^*[||J_n^*(.)||_\nu^2] = n^{-1}\sum_i \hat{U}_i^2\int[\mathcal{H}(X_i,.)-\phi(.)-\psi(.)]^2\nu(dx) \leq C[n^{-1}\sum_i\hat{U}_i^2]$.

When n is large we can replace $\hat{U}_i$ by U_i, and applying a weak law of large numbers gives that

$C[n^{-1}\sum_i\hat{U}_i^2] = C[n^{-1}\sum_i U_i^2] + o_p(1) \xrightarrow{p} CE[\sigma^2(X)] < \infty$.

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

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

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

$cov^*(Z_1^*(x), Z_1^*(x')) = n^{-1}\sum_i[\mathcal{H}(X_i,x)-\phi(x)-\psi(x)][\mathcal{H}(X_i,x')-\phi(x')-\psi(x')]U_i^2$

$= E\{\sigma^2(X_i)[\mathcal{H}(X_i,x)-\phi(x)-\psi(x)][\mathcal{H}(X_i,x')-\phi(x')-\psi(x')]\} + o_p(1) = \Sigma(x,x') + o_p(1)$.

Next, we consider the finite dimensional distribution of $J_n^*(x)$. Let $f(.) \in L^2(\mathcal{S}, ||.||_\nu)$, and $<.,.>$ denote the inner product. Define $B(X_i,x) = [\mathcal{H}(X_i,x)-\phi(x)-\psi(x)]$. Then $< J_n^*(.), f(.) > = \int J_n^*(x)f(x)\nu(dx) = n^{-1/2}\sum_i U_i^* \int B(X_i,x)f(x)\nu(dx) \equiv n^{-1/2}\sum_i W_i U_i^*$. 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, pp148-149) shows that the Linderberg's condition indeed holds. Thus, $< J_n^*(.), f(.) >$ converges to a normal variable with zero mean and variance $E[\sigma^2(X_i)\int B(X_i,x)^2 f^2(x)\nu(dx)]$. By Cramer-Wold device we obtain the finite dimensional convergence result.

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

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

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 skechy 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

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

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

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

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

$\le n^{1/2}(\hat{\beta}^* - \hat{\beta})(P'P/n)(\hat{\beta}^* - \hat{\beta})O_p(1)$

$= n^{-1/2}[U^{*'}P(P'P)^{-1}(P'P)(P'P)^{-1}P'U^*]O_p(1)$

$= n^{1/2}[n^{-1}U^{*'}P](P'P/n)^{-1}[n^{-1}P'U^*]O_p(1)$

$= n^{1/2}O_p((K/n) + \sum_l K^{-2\delta_l})O_p(1) = o_p(1)$,

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

$E^*[\,||n^{-1}P'U^*||^2] = n^{-2}E^*[U^{*'}PP'U^*] = n^{-2}[\hat{U}'PP'\hat{U}]$

$= n^{-2}\sum_i\sum_j \hat{U}_i\hat{U}_j P(X_i)'P(X_j) = n^{-2}\sum_i\sum_j \hat{U}_i\hat{U}_j P(X_i)'P(X_j)$

$= n^{-2}\sum_i\sum_j[U_iU_j + 2U_i(g_j - \hat{g}_j) + (g_i - \hat{g}_i)(g_j - \hat{g}_j)]$
$= O_p(n^{-1}) + O_p(n^{-1/2}[(K/n)^{1/2} + \sum_l K^{-\delta_l}]) + O_p((K/n) + \sum_l K^{-2\delta_l})$
$= O_p((K/n) + \sum_l K^{-2\delta_l})$.

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

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

Thus, we have

$$\begin{aligned} CM_n^* &= \int [J_n^*(x)]^2 dF_n(x) = \int [J_n^*(x)]^2 dF(x) + A_n^* = \int [J_n^*(x)]^2 dF(x) + o_p(1) \\ &\to \int [J_\infty^*(x)]^2 dF(x) \end{aligned}$$

in distribution by the continuous mapping theorem.

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

Lemma A.1 $||J_{n2}(.) - n^{-1/2}\phi(.)'U||_\nu^2 = o_p(1)$.
where $\phi(x)$ is a $n \times 1$ vector with the ith element given by $\phi_i(x) = E_{\mathcal{G}}[\mathcal{H}(X_i, x)]$.

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

$E[||J_{n2}(.) - n^{-1/2}\phi(.)'U||_\nu^2|X]$
$= n^{-1}\int[(\mathcal{H}(X, x))'M_n - \phi(x)']E(UU'|X)[M_n\mathcal{H}(X, x) - \phi(x)]\nu(dx)$
$\leq Cn^{-1}\int[M_n\mathcal{H}(X, x) - \phi(x)]'[M_n\mathcal{H}(X, x) - \phi(x)]\nu(dx)$
$= Cn^{-1}||\, M_n\mathcal{H}(X, x) - \phi(x)\,||_\nu^2 = Cn^{-1}||\, M_n(\phi(x) + V(x) - \phi(x)\,||_\nu^2$
$\leq 2Cn^{-1}\{\,||\, M_n\phi(x) - \phi(x)\,||_\nu^2] + ||\, M_nV(x)\,||_\nu^2\,\}$
$= O_p(\sum_{l=1}^L K_l^{-2\delta_l} + K/n) = o_p(1)$ by lemma B.1 and lemma B.2.

Lemma A.2 $||J_{n3}(.) - n^{-1/2}\psi(.)U||_\nu^2 = o_p(1)$.

Proof: Note that $z_0(X_i) - \tilde{z}_0(X_i)$ estimates $\epsilon_i = z_0(X_i) - E_{\mathcal{G}}[z_0(X_i)]$, or in matrix notation $Z_0 - M_nZ_0$ estimates ϵ. From lemma B.3 we know that $(\hat{\gamma} - \gamma) = \{E[\epsilon_i\epsilon_i']\}^{-1}n^{-1}\sum_i \epsilon_iU_i] + o_p(n^{-1/2})$. Using lemmas B.1 and B.2 we have $||n^{-1}\mathcal{H}(X, .)'(I_n - M_n)Z_0 - E[\mathcal{H}(X_i, .)'\epsilon_i]||_\nu^2 = o_p(1)$. Hence,

$J_{n3}(.) = n^{-1/2}\mathcal{H}(X, .)'(I_n - M_n)Z_0(\hat{\gamma} - \gamma) = E[\mathcal{H}(X_i, .)\epsilon_i][n^{1/2}(\hat{\gamma} - \gamma)] + o_p(1)$
$= E[\mathcal{H}(X_i, .)\epsilon_i]\{E[\epsilon_i\epsilon_i']\}^{-1}[n^{-1/2}\sum_i \epsilon_iU_i] + o_p(1) = n^{-1/2}\psi(.)U + o_p(1)$.

Lemma A.3 $||J_{n4}(.)||_\nu^2 = o_p(1)$.

Proof: $[||J_{n4}(.)||_\nu^2 \le n^{-1}\sum_i\sum_j\int \mathcal{H}(X_i,x)\mathcal{H}(X_j,x)(m_i-\tilde{m}_i)(m_j-\tilde{m}_j)\nu(dx) \le C\sum_i\int[(m_i-\tilde{m}_i)^2]\nu(dx) = C[||m-\tilde{m}||_\nu^2] = nO_p(\sum_{l=1}^L K_l^{-2\delta_l}) = o_p(1)$ by lemma B.1.

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

$$sup_{z\in\mathcal{S}\times\mathcal{S}}|n^{-1/2}\sum_i V_{z,X_i}| = O_p(1).$$

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 $z\in\mathcal{S}\times\mathcal{S}$ under the sup-norm if $|V_{z_1,X_i}-V_{z_2,X_i}|\le A(X_i)||z_1-z_2||$, with $E(A^2(X_i))<\infty$. By the assumption that $\mathcal{H}(.,.)$ 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 $E[V_{z,X_i}V_{z',X_i}]$. Consequently, $sup_{z\in\mathcal{S}\times\mathcal{S}}[n^{-1/2}\sum_i V_{z,X_i}]$ converges weakly to $sup_{z\in\mathcal{S}\times\mathcal{S}}V_\infty(z)=O_p(1)$, which in turn implies that $sup_{z\in\mathcal{S}\times\mathcal{S}}[n^{-1/2}\sum_i V_{z,X_i}]=O_p(1)$.

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 non-singular transformations of $p^K(\cdot)$. Also we will assume $Q \overset{def}{=} E[p^K(X_i)p^K(X_i)'] = I$. This is because, for a symmetric square root $Q^{-1/2}$ of Q^{-1}, $Q^{-1/2}p^K(\cdot)$ is a non-singular transformation of $p^K(\cdot)$, and using (A2) (i), it is easy to show that $\tilde{\zeta}_0(K) \overset{def}{=} sup_{x\in\mathcal{S}}||Q^{-1/2}p^K(\cdot)|| \le C\zeta_0(K)$. 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* $f_i(x)\equiv f_0(x,X_i)\in\mathcal{G}$ *(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* $\tilde{f}_X(x)=M_nf_X(x)$. *Then*

$$n^{-1}||f_X(x)-M_nf_X(x)||_\nu^2 = O_p(\sum_l K_l^{-2\delta_l}) = o_p(1).$$

Proof: $n^{-1}E[||f_X(x)-M_nf_X(x)||_\nu^2] \equiv n^{-1}E[||f_X(x)-\tilde{f}_X(x)||_\nu^2] = n^{-1}\int E[||f_X(x)-\tilde{f}_X(x)||^2]\nu(dx) = O(\sum_l K_l^{-2\delta_l})$ 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* $E_{\mathcal{G}}(v_i(x)) = 0$ *and* $E_{\mathcal{G}}([v_i(x)]^2)$ *uniformly bounded in* $x \in \mathcal{S}$. *Define* $V(x) = (v_1(x), ..., v_n(x))'$ *and* $\tilde{V}(x) = M_n V(x)$. *Then we have*

$$n^{-1}||M_n V(.)||_\nu^2 = n^{-1}||\tilde{V}(.)||_\nu^2 = O_p(K/n) = o_p(1).$$

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

$$E[||P'V(.)/n||_\nu^2] = n^{-2}\{\sum_i \int E[v_i(x)^2 p^K(X_i)'p^K(X_i)]\nu(dx)$$
$$+ \sum_i \sum_{j \neq i} \int E[v_i(x)p^K(X_i)']E[v_j(x)p^K(X_j)]\nu(dx)\} = n^{-1}\int E[v_1(x)^2 p^K(X_1)'p^K(X_1)]\nu(dx)$$
$$\leq Cn^{-1}E[p^K(X_1)'p^K(X_1)] = O(K/n).$$ This implies that

$$||P'V(x)/n||_\nu^2 = O_p(K/n) = o_p(1). \tag{B.1}$$

$$n^{-1}||M_n V(.)||_\nu^2 = n^{-1}\int V(x)'M_n V(x)\nu(dx) = \int (V(x)'P/n)(P'P/n)^-(P'V(x)/n)\nu(dx)$$
$$= \int (V(x)'P/n)[I + (P'P/n)^- - I](P'V(x)/n)\nu(dx) = \int ||P'V(x)/n||^2[1 + o_p(1)]\nu(dx)$$
$$= \int O_p(K/n)[1 + o_p(1)]\nu(dx) = O_p(K/n) = o_p(1)$$

by equation (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, pp161-162)).

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

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

Anderson, N.T., E. Gine, M. Ossiander and J. Zinn (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. (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.

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. and W. Ploberger, (1997) "Asymptotic Theory of Integrated Conditional Moment Tests," *Econometrica* 65, 1129-1151.

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

Chen, X. and Shen, X., (1998) "Sieve extremum estimates for weakly dependent data," *Econometrica* 66, 289-314.

Dechevsky, L. and S. Penez (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., and W.G. Manteiga, (1998) "Significance testing in nonparametric regression based on the bootstrap," manuscript.

Delgado, M.A. and T. Stengos, (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. and W.K. Newey (1994). "Series estimation of semilinear regression," *Journal of Multivariate Analysis* 50, 30-40.

Eubank, R. and S. Spiegelman, (1990) "Testing the Goodness of Fit of a Linear Model via Nonparametric Regression Techniques," *Journal of the American Statistical Association* 85, 387-392.

Fan, J., Härdle, W., and Mammen, E., (1998), "Direct estimation of low dimensional components in additive models," *Annals of Statistics* 26, 943-971.

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

Fan, Y. and Q. Li, (1996b) "Consistent Model Specification Tests: Omitted Variables, Parametric and Semiparametric Functional Forms," *Econometrica* 64, 865-890.

Gine, E. and J. Zinn (1990). "Bootstraping general empirical measures," *The Annals of Probability* 18, 851-869.

Gozalo, P. and O. Linton, (2001), "Testing additivity in generalized nonparametric regression models with estimated parameters," *Journal of Econometrics* 104, 1-48.

Härdle, W. and E. Mammen, (1993) "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, 21, 1926-1947.

Hong, Y. and H. White, (1995) "Consistent Specification Testing via Nonparametric Series Regression," *Econometrica* 63, 1133-1159.

Horowitz, J.L. and W. Härdle, (1994) "Testing a Parametric Model Against a Semiparametric Alternative," *Econometric Theory* 10, 821-848.

Lavergne, P. and Q. Vuong, (1996) "Nonparametric Selection of Regressors: The Nonnested Case," *Econometrica* 64, 207-219.

Ledoux, M. and M. Talagrand (1991) Probability in Banach Space. New York: Spring-Verlag.

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 models," *International Economic Reviews* 41, 1073-1092.

Li, Q. and S. Wang (1998), "A simple consistent bootstrap test for a parametric regression functional form," *Journal of Econometrics* 87, 145-165.

Linton, O.B., and 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.

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.

Ossiander, M. (1987) "A central limit theorem under metric entropy with L_2 bracketing," *Annals of Probability* 15, 897-919.

Politis, D.N. and J.P. Romano (1994) "The stationary bootstrap," *Journal of the American Statistical Association* 89, 1303-1313.

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.

Sperlich, Tjostheim and Yang (2002), "Nonparametric estimation and testing of interaction in additive models," *Econometric Theory* 18, 197-251.

Stinchcombe, M.B. and H. White, (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," *The Annals of Statistics* 13, 685-705.

Stone, C.J., (1986), "The dimensionality reduction principle for generalized additive models," *The Annals of Statistics* 14, 592-606.

Stute, W., (1997), "Nonparametric model checks for regression," *Annals of Statistics* 25, 613-641.

Stute, W., W.G. Gonzalez and M. Presedo, (1998), "Bootstrap approximation in model checks for regression," *Journal of American Statistical Association* 93, 141-149.

Tjostheim, D. and Auestad, B.H. (1994), "Nonparametric identification of nonlinear time series: projections," *Journal of American Statistical Association* 89, 1398-1409.

Van der Vaart, A.W. and J.A. Wellner, 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.