Available online at www.sciencedirect.com
SCIENCE DIRECT®
Statistics & Probability Letters 72 (2005) 313–322
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Kernel estimation of a partially linear additive model

Sebastiano Manzan[a], Dawit Zerom[b,*]

[a]CeNDEF, University of Amsterdam, The Netherlands
[b]School of Business, University of Alberta, Edmonton, Canada T6G 2R6

Received 24 September 2004; received in revised form 11 January 2005
Available online 17 March 2005

Abstract

In this paper, we introduce a kernel estimator for the finite-dimensional parameter of a partially linear additive model. Under some regularity conditions, we establish $n^{1/2}$-consistency and asymptotic normality of the estimator. Unlike existing kernel-based estimators: Fan et al. (1998. Ann. Statist. 26, 943–971) and Fan and Li (2003. Statist. Sinica 13, 739–762) our estimator attains the semiparametric efficiency bound of the partially linear additive model under homoscedastic errors. We also show that when the true specification is the partially linear additive model, the proposed estimator is asymptotically more efficient than an estimator that ignores the additive structure.

Keywords: Additivity; Kernel; Partially linear additive model; Semiparametric efficient

1. Introduction

The partially linear additive model (hereafter PLAM) has the following form:

$$Y_i = \beta_0 + X_i'\beta + m_1(Z_{1i}) + \cdots + m_q(Z_{qi}) + u_i \quad (i = 1, \ldots, n), \tag{1}$$

where Y_i is a scalar dependent variable, X_i is a $p \times 1$ vector of explanatory variables, $\beta = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of unknown parameters, β_0 is a scalar parameter, $Z_i = (Z_{1i}, \ldots, Z_{qi})'$ is a $q \times 1$ vector of explanatory variables, $m_1(\cdot), \ldots, m_q(\cdot)$ are unknown real-valued smooth

*Corresponding author. Tel.: +1 780 492 3622; fax: +1 780 492 3325.
E-mail address: dawit@ualberta.ca (D. Zerom).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2005.02.005

functions, and u_i is an unobservable random variable that satisfies $\mathbb{E}[u_i|X_i, Z_i] = 0$. The distribution of the regressors (X, Z) is left completely unspecified.

The PLAM is particularly attractive for the following reasons. On the one hand, with respect to the pure additive model,

$$Y_i = \beta_0 + m_1(Z_{1i}) + \cdots + m_q(Z_{qi}) + u_i, \tag{2}$$

the PLAM provides considerable flexibility by allowing interaction terms among the elements of Z enter as the linear part of the model. This is possible as the PLAM permits X_i to be a deterministic, but non-additive function of $(Z_{1i}, \ldots, Z_{qi})$. Furthermore, the PLAM allows a subset or all of the variables in X to be discrete, while pure additive models admit continuous variables only. On the other hand, compared to the partially linear model with non-structured non-parametric component, i.e.

$$Y_i = X_i'\beta + m(Z_{1i}, \ldots, Z_{qi}) + u_i, \tag{3}$$

the PLAM has more explicit non-parametric components that can be estimated with a one-dimensional non-parametric rate and hence avoid the so-called curse of dimensionality (Stone, 1985, 1986). Furthermore, when the true data generating model is PLAM, simply using model (3) (i.e. ignore the additivity of $m(\cdot)$) to estimate β can lead to an inefficient estimate of β. Note also that model (3) does not allow the intercept β_0; only "slope" coefficients can be estimated.

In the context of (3), estimation of β has been the subject of considerable study (see, Härdle et al., 2000). However, only few results are available for the PLAM. This paper presents a kernel-based estimator for the β of the PLAM that is shown to be $n^{1/2}$-consistent and asymptotically normal. We also show that the proposed estimator which is designed to exploit additivity is asymptotically more efficient than an estimator that ignores the additive structure. To the best of our knowledge, there are two kernel-based estimators for estimating the β of PLAM; one is by Fan et al. (1998) and the other by Fan and Li (2003). When compared with the aforementioned kernel estimators, there are at least two advantages to using the proposed estimator. First, when the error u_i is conditional homoscedastic, our estimator attains the semiparametric efficiency bound of the PLAM; see Chamberlain (1992) on the efficiency bound for PLAM type specifications. Recently, Li (2002) also proposed a series-based estimator (non-kernel type) of β that achieves the semiparametric efficiency bound of the PLAM under homoscedastic errors. Second, our estimator reduces the computational requirement by order of the sample size n. From a practical stand point, this computational advantage can be very significant when n is large and/or when implementing computer-intensive methods such as bootstrap or cross-validation.

2. Description of the estimator

Consider the model $A_i = \theta^{(A)}(Z_i) + v_i$ where A is a vector- or real-valued dependent variable, Z is a $q \times 1$ vector of explanatory variables, $\theta^{(A)}(z) \equiv \mathbb{E}[A_i|Z_i = z]$ is an unknown vector- or real-valued smooth function and v_i is an unobservable noise that satisfies $\mathbb{E}[v_i|Z_i] = 0$. Following Stone (1985), we define additivity as follows: The function $\theta^{(A)}(z)$ is said to belong to an additive class of functions $\mathcal{J}$ ($\theta^{(A)}(\cdot) \in \mathcal{J}$) if $\theta^{(A)}(z) = \theta_1^{(A)}(z_1) + \cdots + \theta_q^{(A)}(z_q)$, and $\mathbb{E}[\theta_j^{(A)}(z_j)] = 0$ for all $j = 1, \ldots, q$. When $\theta^{(A)}(\cdot)$ is vector valued, we say $\theta^{(A)}(\cdot) \in \mathcal{J}$ if each component of $\theta^{(A)}(\cdot)$ belongs to $\mathcal{J}$.

Further, let $\theta^{(A*)}(z) = \theta_1^{(A*)}(z_1) + \cdots + \theta_q^{(A*)}(z_q)$ be a function chosen subject to the constraints $\mathbb{E}[\theta_1^{(A*)}(z_j)] = 0$ (for all $j = 1, \ldots, q$) to minimize

$$\mathbb{E}[(\theta^{(A)}(\cdot) - \theta^{(A*)}(\cdot))(\theta^{(A)}(\cdot) - \theta^{(A*)}(\cdot))']. \tag{4}$$

Then, we say $\theta^{(A*)}(\cdot)$ is the closest (best) additive approximation to $\theta^{(A)}(\cdot)$ in L_2. Stone (1985) proved the existence of $\theta^{(A*)}(\cdot)$ that satisfies (4) when $\theta^{(A)}(\cdot)$ is real valued. For $j = 1, \ldots, q$, let the vector W_j denote the set of all Z variables excluding Z_j, i.e. $W_j = (Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_q)'$. Define

$$\phi(z_j, w_j) = \frac{p_z(z_j)p_w(w_j)}{p(z_j, w_j)},$$

where $p_z(\cdot)$ and $p_w(\cdot)$ are the density functions of Z_j and W_j, respectively, and $p(\cdot)$ is the joint probability function of $Z = (Z_j, W_j)$. Let $h_j^{(A)}(z_j) \equiv \mathbb{E}[\phi(z_j, w_j)A|z_j]$. It is easy to show that

$$h_j^{(A)}(z_j) = \int \theta^{(A)}(z)p_w(w_j)\,\mathrm{d}w_j \quad (j = 1, \ldots, q), \tag{5}$$

where we have used $\mathbb{E}[\phi(z_j, w_j)v|z_j] = 0$. We can observe from (5) that $h_1^{(A)}(z_1), \ldots, h_q^{(A)}(z_q)$ are the $L_2(Q)$ (with Q being a product probability measure) projections of $\theta^{(A)}(z)$ onto the space functions of $z_1, \ldots, z_q$, respectively. Therefore, as per the definition of additivity, we may define the sum

$$h^{(A)}(z) = \sum_{j=1}^{q} h_j^{(A)}(z_j), \tag{6}$$

as the best additive approximation to the function $\theta^{(A)}(z)$, i.e.

$$\mathbb{E}\{[\theta^{(A)}(z) - h^{(A)}(z)][\theta^{(A)}(z) - h^{(A)}(z)]'\} = \inf_{\theta^{(A*)} \in \mathscr{I}} \mathbb{E}\{[\theta^{(A)}(z) - \theta^{(A*)}(z)][\theta^{(A)}(z) - \theta^{(A*)}(z)]'\}. \tag{7}$$

In practice, the conditions $\mathbb{E}[h_j^{(A)}(z_j)] = 0$ $(j = 1, \ldots, q)$ are easy to impose. When $\theta^{(A)}(\cdot)$ is in fact additive $(\theta^{(A)}(\cdot) \in \mathscr{I})$, for all $j = 1, \ldots, q$, $h_j^{(A)}(z_j) = \theta_j^{(A)}(z_j)$. Hence, $h^{(A)}(z) = \theta^{(A)}(z)$. Kim et al. (1999) exploited this property and used $h_j^{(A)}(z_j)$ in identifying the additive components in the context of the pure additive model (2).

In what follows, we extend the foregoing discussion to PLAM (1). The intercept β_0 is set to zero in the sequel without loss of generality (see footnote 1). For identification purposes, let us assume that $\mathbb{E}[m_j(Z_{ji})] = 0$ for all $j = 1, \ldots, q$. Replacing A by Y or X, observe that $\mathbb{E}[\phi(Z_{ji}, W_{ji})Y_i|Z_{ji}]$ and $\mathbb{E}[\phi(Z_{ji}, W_{ji})X_i|Z_{ji}]$ correspond to $h_j^{(Y)}(Z_{ji})$ and $h_j^{(X)}(Z_{ji})$, respectively. Using previous arguments and notations, it follows from (1) that

$$h_j^{(Y)}(Z_{ji}) = m_j(Z_{ji}) + (h_j^{(X)}(Z_{ji}))'\beta \quad (j = 1, \ldots, q). \tag{8}$$

If we add the q-equations in (8) and subtract the result from (1), we obtain

$$Y_i - h^{(Y)}(Z_i) = (X_i - h^{(X)}(Z_i))'\beta + u_i, \tag{9}$$

where we have used the definition in (6). Notice from (9) that we have reduced the PLAM to a linear model where the dependent and independent variables are expressed in deviations form around the best additive approximations, $h^{(Y)}(\cdot)$ and $h^{(X)}(\cdot)$, respectively. Let $\tilde{A}_i$ denote an estimator of $h^{(A)}(Z_i)$ where A can be Y or X. Based on (6), we define $\tilde{A}_i$ as

$$\tilde{A}_i = \sum_{j=1}^{q} \tilde{A}_i^j, \tag{10}$$

where $\tilde{A}_i^j$ denotes an estimate of $h_j^{(A)}(Z_{ji}) \equiv \mathbb{E}[\phi(Z_{ji}, W_{ji})A_i | Z_{ji}]$. We compute $\tilde{A}_i^j$ by

$$\tilde{A}_i^j = \frac{1}{(n-1)b} \sum_{\ell \neq i}^{n} K\left(\frac{Z_{j\ell} - Z_{ji}}{b}\right) \frac{\hat{p}_w(W_{j\ell})}{\hat{p}(Z_{j\ell}, W_{j\ell})} A_\ell \quad (i = 1, \ldots, n;\ j = 1, \ldots, q), \tag{11}$$

where $K(\cdot)$ is a kernel function, b is a bandwidth (or smoothing parameter), and $\hat{p}_w(\cdot)$ and $\hat{p}(\cdot)$ are kernel smoothers of the corresponding densities. Based on (9) and using the kernel smoothers defined above, an estimator of β for the PLAM can be defined as the vector of OLS coefficients $\hat{\beta}$ of the deviation $(Y_i - \tilde{Y}_i)$ on $(X_i - \tilde{X}_i)$.[1] Following the notations of Robinson (1988), this OLS estimator can be formulated as

$$\hat{\beta} = S_{X-\tilde{X}}^{-1} S_{X-\tilde{X}, Y-\tilde{Y}}, \tag{12}$$

where for scalar or column vector sequences C_i and D_i, $S_{C,D} = n^{-1} \sum_{i=1}^{n} C_i D_i'$ and $S_C = S_{C,C}$.

If we were to ignore the additive structure of $m(z_1, \ldots, z_q)$ and assumed, instead the partially linear model (3), we would have estimated β by regressing $Y_i - \hat{\theta}^{(Y)}(Z_i)$ on $X_i - \hat{\theta}^{(X)}(Z_i)$; as proposed by Robinson (1988); see Remark 3 in Section 3. From a practical stand point, the estimator $\hat{\beta}$ is computationally more attractive than other kernel-based β estimators of the PLAM introduced in the literature. A case in point is the estimator proposed by Fan et al. (1998). The reduction in computational cost is due to the way $\tilde{A}_i^j$ is constructed. Close inspection of the definition shows that as compared against the used kernel conditional expectation estimator, $\tilde{A}_i^j$ eliminates the explicit estimation of the density $p_j(Z_{ji})$ in the denominator, see, Jones et al. (1994) for details on such type of estimation. This simplification leads to a reduction in computation of order n.

3. Asymptotic results

We will use $\theta(\cdot)$, $h(\cdot)$ and $h_j(\cdot)$ to denote $\theta^{(X)}(\cdot)$, $h^{(X)}(\cdot)$ and $h_j^{(X)}(\cdot)$, respectively. The latter three functions are as defined in Section 2. Using the new notations, $h(Z_i) = \sum_{j=1}^{q} h_j(Z_{ji})$. Define $v_i = X_i - \theta(Z_i)$, $\eta_i = \theta(Z_i) - h(Z_i)$ and $\varepsilon_i = X_i - h(Z_i)$. Notice that $\varepsilon_i = \eta_i + v_i$. In Section 2, we define the vector W_j as $W_j = (Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_q)'$ for $(j = 1, \ldots, q)$. To simplify notations, we shall suppress the subscript j from W_j and simply denote it by W. Hence, for any $j \in [1, \ldots, q]$, $Z = (Z_j, W)$. We shall denote the density of W by $p_w(\cdot)$, the density of Z_j by $p_z(\cdot)$, and the density of Z by $p(\cdot)$. Let $\bar{\mathscr{G}}_v^\alpha$ denote the class of functions such that if $f \in \bar{\mathscr{G}}_v^\alpha (\alpha > 0$ and $v \geqslant 2)$, then, (i) f is v

[1] In empirical work, one may also be interested in estimating the intercept β_0. It is easy to see that when $\beta_0 \neq 0$, Eq. (9) would become $Y_i - h^{(Y)}(Z_i) = (1-q)\beta_0 + (X_i - h^{(X)}(Z_i))'\beta + u_i$. Hence, we would instead regress $(Y_i - \tilde{Y}_i)$ on $(1, (X_i - \tilde{X}_i)')'$ so as to incorporate the estimation of the intercept.

times differentiable, and f and its derivatives (up to order v) are all bounded by some function that has αth order moments; (ii) For all $s, t \in \mathbb{R}$, $|f(s) - f(t)| \leqslant H_f(t)|s-t|$, where $H_f(t)$ is a continuous function with a finite αth moment. Also let $\mathscr{K}_v$ denote the class of v-order even kernel functions K: $\mathbb{R} \to \mathbb{R}$ satisfying (i) $\int \mathscr{K}_\ell(s)s^k ds = 0,\ k = 1, \ldots, v-1$, and (ii) $\mathscr{K}$ is compactly supported, bounded, and Lipschitz continuous. Below, we impose some regularity conditions that enable us to characterize the asymptotic properties of $\hat{\beta}$.

A1. (i) (Y_i, X_i, Z_i), $i = 1, \ldots, n$ are i.i.d. as (Y, X, Z). The joint densities $p_w(\cdot)$ and $p(\cdot)$ are bounded away from zero and infinity on their compact support. Further, the density function $p_z(\cdot) \in \bar{\mathscr{G}}^{\infty}_{v-1}$;
(ii) $\mathbb{E}[u^2|x,z] = \sigma^2_u(x,z)$ is continuous in z, and both u and v have finite fourth moments. The function $\theta(z)$ is bounded on the support z; for all $j \in [1, \ldots, q]$, $m_j(\cdot) \in \bar{\mathscr{G}}^4_v$, and $h_j(\cdot) \in \bar{\mathscr{G}}^4_v$ where $v \geqslant 2$ is an integer.

A2. $K \in \mathscr{K}_2$ and the bandwidth sequence b is of the form $b = an^{-\kappa}$, where a is some constant and κ is a positive real number satisfying $\frac{1}{4} < \kappa < \frac{1}{3}$.

Assumption A2 gives the conditions on the order of the kernel function and the rate of convergence of the bandwidth. We like to note that only when $q < 4$ where q is the number of additive components that the use of a second-order kernel $\mathscr{K}_2$ is sufficient to attain convergence of $\hat{\beta}$ at the parametric rate. When $q \geqslant 4$, a higher order kernel is needed for $n^{1/2}$-consistency. For example, one may choose $K \in \mathscr{K}_4$ and $b = o(n^{-1/8})$.

Theorem. *Let* $\Phi = \mathbb{E}[\varepsilon_i \varepsilon_i']$. *Under Assumptions* A1 *and* A2, *and provided* Φ *is positive definite*,

$$n^{1/2}(\hat{\beta} - \beta) \to \mathrm{N}(0, \Sigma),$$

where $\Sigma = \Phi^{-1}\Omega\Phi^{-1}$ *and* $\Omega = \mathbb{E}[\sigma^2_u(X_i, Z_i, W_i)\varepsilon_i\varepsilon_i']$. *The variance–covariance matrix* Σ *can be consistently estimated by* $\hat{\Sigma} = \hat{\Phi}^{-1}\hat{\Omega}\hat{\Phi}^{-1}$ *where* $\hat{\Phi} = n^{-1}\sum_i (X_i - \tilde{X}_i)(X_i - \tilde{X}_i)'$, $\hat{\Omega} = n^{-1}\sum_i \hat{u}_i^2(X - \tilde{X}_i)(X_i - \tilde{X}_i)'$, *and* $\hat{u}_i = Y_i - \tilde{Y}_i - (X_i - \tilde{X}_i)'\hat{\beta}$.

The proof of the theorem is given in Section 4. A few remarks are in order.

Remark 1 (*Weaker identification condition*). It is easy to see that $\Phi = \mathbb{E}[\varepsilon_i\varepsilon_i'] = \mathbb{E}[v_i v_i'] + \mathbb{E}[\eta_i\eta_i']$ since $\varepsilon_i = \eta_i + v_i$. Thus, when in fact $m(\cdot)$ is additive, the condition needed to identify β is weaker in the sense that we only require $\mathbb{E}[v_i v_i']$ or $\mathbb{E}[\eta_i\eta_i']$ to be positive definite. When the additivity of $m(\cdot)$ is ignored, we strictly require $\mathbb{E}[v_i v_i']$ to be positive definite (see Robinson, 1988). An interesting consequence of this weaker identification is that X can be allowed to be a deterministic function of Z as long as $\theta(\cdot)$ is not additive. Note that when $\theta(\cdot)$ is not additive, $\eta \neq 0$. Thus β is identified even when $v = 0$ (or $\theta(Z) = X$). Therefore, the PLAM in conjunction with the proposed estimator $\hat{\beta}$ provides additional modeling flexibility by allowing interaction terms among the elements of Z to enter as the linear part of the model, i.e., $X = D(Z)$ where $D(\cdot)$ is some known deterministic non-additive function of Z.

Remark 2 (*Semiparametric efficiency*). Suppose the errors of the PLAM are homoscedastic, i.e., $\sigma^2_u(x,z) = \sigma^2_u$. Under this condition and applying the more general result of Chamberlain (1992), the semiparametric efficiency bound for the inverse of the asymptotic variance of an estimator of β is given by $J_0 = (1/\sigma_u^2)\inf_{\theta^* \in \mathscr{J}} E\{[\theta(Z_i) - \theta^*(Z_i)][\theta(Z_i) - \theta^*(Z_i)]'\}$. But, from Section 2 we know that h is the best additive approximation to θ. Hence, J_0 becomes, $J_0 = (1/\sigma_u^2)E\{[X_i - h(Z_i)][X_i - h(Z_i)]'\} = (1/\sigma_u^2)\Phi = \Sigma^{-1}$. Therefore, our estimator $\hat{\beta}$ is semiparametric efficient in the sense that the inverse of the asymptotic variance Σ^{-1} reaches the semiparametric efficiency bound when the errors are homoscedastic. In contrast, neither the kernel-based β estimator by Fan et al. (1998) nor that of Fan and Li (2003) are semiparametric efficient.

Remark 3 (*Ignoring additivity*). Suppose we ignore the additivity of $m(\cdot)$ (our model set-up) and simply use an approach, say the estimator (denote this by $\hat{\beta}^{\mathrm{R}}$) by Robinson (1988). Under homoscedasticity of the errors u, we know from Robinson (1988) that the asymptotic variance of $n^{1/2}(\hat{\beta}^{\mathrm{R}} - \beta)$ is $\sigma_u^2(E\{[X_i - \theta(Z_i)][X_i - \theta(Z_i)]'\})^{-1}$. Therefore, when $\theta(\cdot)$ is non-additive, $\hat{\beta}$ (see the theorem) has a smaller asymptotic variance and hence is asymptotically more efficient than $\hat{\beta}^{\mathrm{R}}$.

4. Proofs

The theorem is proved in two parts. First, we show that $n^{1/2}(\hat{\beta} - \beta) \to \mathrm{N}(0, \Sigma)$. To simplify the notations in Section 3, we introduce the following short-hand forms: $\theta_i = \theta(Z_i)$, $m_i = m(Z_i)$, $h_i = h(Z_i)$, $m_{ji} = m_j(Z_{ji})$, $h_{ji} = h_j(Z_{ji})$, $v_i = X_i - \theta_i$, $\eta_i = \theta_i - h_i$, and $\varepsilon_i = X_i - h_i$. Using the definition of $\tilde{A}_i$ in (11) and noting that $X_i = \eta_i + v_i + h_i$, it is easy to see that $Y_i - \tilde{Y}_i = (X_i - \tilde{X}_i)'\beta + (m_i - \tilde{m}_i) + (u_i - \tilde{u}_i)$ and $X_i - \tilde{X}_i = \eta_i + v_i + (h_i - \tilde{h}_i) - \tilde{v}_i - \tilde{\eta}_i$. Then, if $S_{X-\tilde{X}}^{-1}$ exists,

$$n^{1/2}(\hat{\beta} - \beta) = S_{X-\tilde{X}}^{-1} n^{1/2} S_{X-\tilde{X}, m-\tilde{m}+u-\tilde{u}}. \tag{13}$$

Therefore, to prove the first result, it is sufficient to verify the following results: (I) $S_{X-\tilde{X}} = \hat{\Phi} = \Phi + o_p(1)$, (II) $n^{1/2}S_{X-\tilde{X},m-\tilde{m}} = o_p(1)$, (III) $n^{1/2}S_{X-\tilde{X},\tilde{u}} = o_p(1)$, and (IV) $n^{1/2}S_{X-\tilde{X},u} \to \mathrm{N}(0, \Omega)$. Combining (13) with results (I)–(IV) gives $n^{1/2}(\hat{\beta} - \beta) = \Phi^{-1}\mathrm{N}(0, \Omega) + o_p(1) \to \mathrm{N}(0, \Phi^{-1}\Omega\Phi^{-1})$ as needed. Below are the proofs of results (I)–(IV). In proving these results, Assumption A2 will be repeatedly used without being explicitly mentioned.

(I) $S_{X-\tilde{X}} = \Phi + o_p(1)$: $S_{X-\tilde{X}} = S_{\eta+v+(h-\tilde{h})-\tilde{\eta}-\tilde{v}} = S_{\eta+v} + S_{(h-\tilde{h})-\tilde{v}-\tilde{\eta}} + 2S_{\eta+v,(h-\tilde{h})-\tilde{v}-\tilde{\eta}}$. Note that $S_{\eta+v} = n^{-1}\sum_i \varepsilon_i\varepsilon_i' = \Phi + o_p(1)$ by the law of large numbers. The second term: $S_{(h-\tilde{h})-\tilde{v}-\tilde{\eta}} \leqslant 3\{S_{h-\tilde{h}} + S_{\tilde{v}} + S_{\tilde{\eta}}\} = o_p(1)$ by Lemmas 2, 4(a) and (c). The final term: $S_{\eta+v,(h-\tilde{h})-\tilde{v}-\tilde{\eta}} \leqslant \{S_{\eta+v}S_{(h-\tilde{h})-\tilde{v}-\tilde{\eta}}\}^{1/2} = \{O_p(1)o_p(1)\}^{1/2} = o_p(1)$. □

(II) $n^{1/2}S_{X-\tilde{X},m-\tilde{m}} = o_p(1)$: $S_{X-\tilde{X},m-\tilde{m}} = S_{\eta+v+(h-\tilde{h})-\tilde{\eta}-\tilde{v},m-\tilde{m}} = S_{\eta+v,m-\tilde{m}} + S_{h-\tilde{h},m-\tilde{m}} - S_{\tilde{v},m-\tilde{m}} - S_{\tilde{\eta},m-\tilde{m}}$. First, note that $S_{\eta+v,m-\tilde{m}} \leqslant \{S_{\eta+v}S_{m-\tilde{m}}\}^{1/2}$, $S_{h-\tilde{h},m-\tilde{m}} \leqslant \{S_{h-\tilde{h}}S_{m-\tilde{m}}\}^{1/2}$, $S_{\tilde{v},m-\tilde{m}} \leqslant \{S_{\tilde{v}}S_{m-\tilde{m}}\}^{1/2}$, and $S_{\tilde{\eta},m-\tilde{m}} \leqslant \{S_{\tilde{\eta}}S_{m-\tilde{m}}\}^{1/2}$. Using Lemmas 2, 4(a) and (c), it follows that $S_{X-\tilde{X},m-\tilde{m}} = o_p(n^{-1/2})$. □

(III) $n^{1/2}S_{X-\tilde{X},\tilde{u}} = o_p(1)$: $S_{X-\tilde{X},\tilde{u}} = S_{\eta+v+(h-\tilde{h})-\tilde{\eta}-\tilde{v},\tilde{u}} = S_{\eta,\tilde{u}} + S_{v,\tilde{u}} + S_{h-\tilde{h},\tilde{u}} - S_{\tilde{v},\tilde{u}} - S_{\tilde{\eta},\tilde{u}}$. By Lemmas 5(c) and (b), respectively, $S_{\eta,\tilde{u}} = o_p(n^{-1/2})$ and $S_{v,\tilde{u}} = o_p(n^{-1/2})$. Noting that $S_{h-\tilde{h},\tilde{u}} \leqslant \{S_{h-\tilde{h}}S_{\tilde{u}}\}^{1/2}$ and using Lemmas 2, and 4(b), $S_{h-\tilde{h},\tilde{u}} = o_p(n^{-1/2})$. Similarly $S_{\tilde{v},\tilde{u}} = o_p(n^{-1/2})$ by Lemmas 4(a) and (b). Finally, $S_{\tilde{\eta},\tilde{u}} = o_p(n^{-1/2})$ by Lemmas 4(b) and (c). Combining all these results, $S_{X-\tilde{X},\tilde{u}} = o_p(n^{-1/2})$. □

(IV) $n^{1/2}S_{X-\tilde{X},u} \to$ N$(0,\Omega)$: $S_{X-\tilde{X},u} = S_{\eta+v+(h-\tilde{h})-\tilde{\eta}-\tilde{v},u} = S_{\eta+v,u} + S_{h-\tilde{h},u} - S_{\tilde{v},u} - S_{\tilde{\eta},u}$. By Lemmas 3, 5(a), and (d), $S_{h-\tilde{h},u} = \mathrm{o_p}(n^{-1})$, $S_{\tilde{v},u} = \mathrm{o_p}(n^{-1/2})$ and $S_{\tilde{\eta},u} = \mathrm{o_p}(n^{-1/2})$. Thus, $n^{1/2}S_{X-\tilde{X},u} = n^{1/2}S_{\eta+v,u} + \mathrm{o_p}(1) = n^{-1/2}\sum_i \varepsilon_i u_i + \mathrm{o_p}(1) \to$ N$(0,\Omega)$ by Levi–Lindberg central limit theorem. □

Now we move to the second part of the theorem, i.e. $\hat{\Sigma} = \Sigma + \mathrm{o_p}(1)$, or $\hat{\Phi}^{-1}\hat{\Omega}\hat{\Phi}^{-1} = \Phi^{-1}\Omega\Phi^{-1} + \mathrm{o_p}(1)$. We have already established $\hat{\Phi} = \Phi + \mathrm{o_p}(1)$; see result (I) above. Thus, it remains to prove $\hat{\Omega} = \Omega + \mathrm{o_p}(1)$. It will be shown that: (V) $X_i - \hat{X}_i = \varepsilon_i + \mathrm{o_p}(1)$ and (VI) $\hat{u}_i = u_i + \mathrm{o_p}(1)$. These two results imply that $\hat{\Omega} = n^{-1}\sum_i \hat{u}_i^2(X_i - \hat{X}_i)(X_i - \hat{X}_i)' = n^{-1}\sum_i u_i^2\varepsilon_i\varepsilon_i' + \mathrm{o_p}(1) = \Omega + \mathrm{o_p}(1)$ by the law of large numbers. This completes the proof. Below are the proofs of results (V) and (VI).

(V) $X_i - \tilde{X}_i = \varepsilon_i + \mathrm{o_p}(1)$: recall that $X_i - \tilde{X}_i = \eta_i + v_i + (h_i - \tilde{h}_i) - \tilde{v}_i - \tilde{\eta}_i$. From Lemmas 2, 4(a) and (c), $(h_i - \tilde{h}_i) = \mathrm{o_p}(1)$, $\tilde{v}_i = \mathrm{o_p}(1)$, and $\tilde{\eta}_i = \mathrm{o_p}(1)$. These imply that $X_i - \tilde{X}_i = \varepsilon_i + \mathrm{o_p}(1)$ where $\varepsilon_i = \eta_i + v_i$. □

(VI) $\hat{u}_i = u_i + \mathrm{o_p}(1)$: replacing $(Y_i - \tilde{Y}_i)$ in $\hat{u}_i = Y_i - \tilde{Y}_i - (X_i - \tilde{X}_i)'\hat{\beta}$ by $(X_i - \tilde{X}_i)'\beta + (m_i - \tilde{m}_i) + (u - \tilde{u}_i)$ gives $u_i - \hat{u}_i = (X_i - \tilde{X}_i)'(\hat{\beta} - \beta) + (\tilde{m}_i - m_i) + \tilde{u}_i$. From the first part of the theorem, we have $\hat{\beta} - \beta = \mathrm{O_p}(n^{-1/2})$. By Lemmas 2 and 4(b), $(m_i - \tilde{m}_i) = \mathrm{o_p}(1)$, and $\tilde{u}_i = \mathrm{o_p}(1)$. These combined with result (V) give $\hat{u}_i = u_i + \mathrm{o_p}(1)$. □

4.1. Intermediate results

Here we state and prove five lemmas that are useful in proving the main theorem of the paper. In all the lemmas, we assume that the conditions of the theorem are satisfied. To simplify our presentation, we denote $C_n \sim B_n$ when $C_n = B_n + \mathrm{o_p}(B_n)$, i.e. C_n equals B_n plus a term that goes to zero in probability faster than B_n. In what follows, g can be h or m. The function g will be a $p \times 1$ vector when it assumes h. Note that v and η are also $p \times 1$ vectors. In proving the lemmas, we only consider the scalar case, $p = 1$. Using the Cauchy inequality, the proofs for the vector cases follow from the corresponding scalar cases. We repeatedly use the following notations: $Z_i^* = Z_{1i}$, $f_i = f(Z_i^*) = g_{1i}$, and $\tilde{f}_i = \tilde{g}_{1i}^1$. Note that $f \in \mathscr{G}_v^4$. We let $K_{ij} = K\{(Z_j^* - Z_i^*)/b\}$, $p_{w,i} = p_w(W_i)$, and $p_{zw,i} = p(Z_i^*, W_i)$. We will use $\mathbb{E}_i(\cdot)$ to denote $\mathbb{E}(\cdot|Z_i^*, W_i)$. Throughout this subsection, M denotes a generic constant.

Lemma 1. $\tilde{A}_i^j \sim (1/nb)\sum_{\ell=1}^n K((Z_{j\ell} - Z_{ji})/b)(p_w(W_\ell)/p(Z_{j\ell}, W_\ell))A_\ell$.

Proof. Due to the uniform convergence of $\hat{p}_w(\cdot)$ and $\hat{p}(\cdot,\cdot)$ to $p_w(\cdot)$ and $p(\cdot,\cdot)$, respectively, the result follows. □

Lemma 2. $S_{g-\hat{g}} = \mathrm{O_p}(n^{-1}b + b^{2v})$.

Proof. Note that $S_{g-\tilde{g}} = S_{g_1-\tilde{g}_1^1+\cdots+g_q-\tilde{g}_q^q}$ since $\tilde{g}_j^k = 0$ for $k \neq j$. Let $C_i = C_{1i} + C_{2i} + \cdots + C_{qi}$. It is easy to see that $S_C \leqslant \sum_{j=1}^q S_{C_j} + 2\sum_{k\neq j}(S_{C_j}S_{C_k})^{1/2}$. Using this inequality and noting the fact that each summand in $S_{g-\tilde{g}}$ plays the same role, it suffices to evaluate the order of one of the summands, say $Sg_1 - \tilde{g}_1^1$,

$$\begin{aligned}\mathbb{E}[S_{g_1-\tilde{g}_1^1}] = \mathbb{E}[(\tilde{f}_1 - f_1)^2] &\sim \frac{1}{n^2b^2}\sum_{i\neq 1}\sum_{j\neq 1}\mathbb{E}\left[(f_i - f_1)K_{1i}\frac{p_{w,i}}{p_{zw,i}}(f_j - f_1)K_{1j}\frac{p_{w,j}}{p_{zw,j}}\right]\\ &= \frac{1}{nb^2}\mathbb{E}\left[(f_2 - f_1)^2K_{12}^2\frac{p_{w,2}^2}{p_{zw,2}^2}\right] + \frac{1}{b^2}\mathbb{E}\left[\mathbb{E}_1\left\{(f_2 - f_1)K_{12}\frac{p_{w,2}}{p_{zw,2}}\right\}\right.\\ &\left.\times\mathbb{E}_1\left\{(f_3 - f_1)K_{13}\frac{p_{w,3}}{p_{zw,3}}\right\}\right]\\ &= D_1 + D_2.\end{aligned}$$

With $Z_2^* = Z_1^* + bs$ and noting that $p_w(\cdot)$ and $p(\cdot,\cdot)$ are uniformly bounded and $f \in \bar{\mathscr{G}}_v^4$,

$$\begin{aligned}D_1 &= \frac{1}{nb}\int\int\int\int(f(Z_1^* + bs) - f(Z_1^*))^2K^2(s)\frac{p_w^2(W_2)}{p(Z_1^* + bs, W_2)}p(Z_1^*, W_1)\,\mathrm{d}Z_1^*\,\mathrm{d}W_1\,\mathrm{d}s\,\mathrm{d}W_2\\ &\leqslant \frac{Mb}{n}\int\int\int p(Z_1^*, W_1)H_f^2(Z_1^*)s^2K^2(s)\,\mathrm{d}Z_1^*, \mathrm{d}W_1\,\mathrm{d}s = \mathrm{O}(n^{-1}b).\end{aligned}$$

Similarly,

$$D_2 \leqslant \frac{M}{b^2}\mathbb{E}[\mathbb{E}_1\{(f_2 - f_1)K_{12}\}\mathbb{E}_1\{(f_3 - f_1)K_{13}\}] \leqslant \frac{M}{b^2}b^{2+2v}\mathbb{E}\{(D_f)^2\} = \mathrm{O}(b^{2v})$$

by Lemma 1 of Li (1996). Combining the above results, $\mathbb{E}[S_{g_1-\tilde{g}_1^1}] = \mathrm{O}(n^{-1}b + b^{2v})$. □

Lemma 3. $S_{g-\hat{g},u} = \mathrm{O_p}(n^{-1}b^{1/2} + n^{-1/2}b^v)$.

Proof. Note that $S_{g-\tilde{g},u} = S_{g_1-\tilde{g}_1^1,u} + \cdots + S_{g_q-\tilde{g}_q^q,u}$. As in Lemma 2, it suffices to evaluate the order of $S_{g_1-\tilde{g}_1^1,u}$. This can be shown by following similar steps as in Lemma 2 with the following in mind: $\mathbb{E}[u_i(f_i - \tilde{f}_i)u_j(f_j - \tilde{f}_j)] = 0$ for $i \neq j$ due to the independence assumption, $\mathbb{E}(u_i|X_i, Z_i^*, W_i) = 0$ and assumption that u and v have finite fourth moments. □

Lemma 4. (a) $S_{\tilde{v}} = \mathrm{O_p}((nb)^{-1})$, (b) $S_{\tilde{u}} = \mathrm{O_p}((nb)^{-1})$ *and* (c) $S_{\tilde{\eta}} = \mathrm{O_p}((nb)^{-1} + b^{2v})$.

Proof. (a) Note that $S_{\tilde{v}} = S_{\tilde{v}^1+\cdots,\tilde{v}^q}$. Reasoning as in Lemma 2, it suffices to evaluate the order of $S_{\tilde{v}^1}$. Using $\mathbb{E}(v_i|Z_i^*, W_i) = 0$ and noting that the densities are uniformly bounded,

$$\begin{aligned}\mathbb{E}[S_{\tilde{v}^1}] = \mathbb{E}[(\tilde{v}^1)^2] &= \mathbb{E}\left\{\frac{1}{nb}\sum_{i\neq 1}K_{1i}\frac{\hat{p}_{w,i}}{\hat{p}_{zw,i}}v_i\frac{1}{nb}\sum_{j\neq 1}K_{1j}\frac{\hat{p}_{w,j}}{\hat{p}_{zw,j}}v_j\right\}\\ &\leqslant \frac{M}{nb^2}\mathbb{E}\{\mathbb{E}_2\{v_2^2K_{12}^2\}\} = \mathrm{O}((nb)^{-1}).\end{aligned}$$

(b) The proof of (b) is the same as (a).

(c) For the same reason as in (a), we only evaluate the order for $S_{\tilde{\eta}^1}$.

$$\begin{aligned}\mathbb{E}[S_{\tilde{\eta}_1}] = \mathbb{E}[(\tilde{\eta}^1)^2] &\sim \mathbb{E}\left\{\frac{1}{nb}\sum_{i\neq 1} K_{1i}\frac{p_{w,i}}{p_{zw,i}}\eta_i \frac{1}{nb}\sum_{j\neq 1} K_{1j}\frac{p_{w,j}}{p_{zw,j}}\eta_j\right\}\\
&= \frac{1}{nb^2}\mathbb{E}\left[\eta_2^2 K_{12}^2 \frac{p_w^2}{p_{zw}^2}\right] + \frac{1}{b}\mathbb{E}_1\left[K_{12}\frac{p_{w,2}}{p_{zw,2}}\{\theta(Z_2^*, W_2) - h_1(Z_1^*) + \mathrm{O}(b^v)\}\right]\\
&\quad \times \frac{1}{b}\mathbb{E}_1\left[K_{13}\frac{p_{w,3}}{p_{zw,3}}\{\theta(Z_3^*, W_3) - h_1(Z_1^*) + \mathrm{O}(b^v)\}\right]\\
&= \frac{1}{nb^2}\mathbb{E}\left[\eta_2^2 K_{12}^2 \frac{p_w^2}{p_{zw}^2}\right] + \left\{\frac{1}{b}\mathbb{E}_1\left[K_{12}\frac{p_{w,2}}{p_{zw,2}}\{\theta(Z_2^*, W_2) - h_1(Z_1^*) + \mathrm{O}(b^v)\}\right]\right\}^2\\
&= D_3 + D_4,\end{aligned}$$

where we have used $\eta_i = \theta(Z_i^*, W_i) - h(Z_i^*, W_i)$ and $(1/nb)\sum_i K_{1i}(p_{w,i}/p_{zw,i})h(Z_i^*, W_i) = h_1(Z_1^*) + \mathrm{O}(b^v)$. Using the same arguments as in Lemmas 2 and 4 and noting that $\theta(\cdot)$ is bounded in its support, it is easy to see that $D_3 = \mathrm{O}((nb)^{-1})$. Moving on to D_4 and using Taylor series expansion, it can be shown that $\mathbb{E}_1[K_{12}p_{w,2}/p_{zw,2}\theta(Z_2^*, W_2)] = bh_1(Z_1^*) + \mathrm{O_p}(b^{v+1})$ and $\mathbb{E}_1[K_{12}p_{w,2}/p_{zw,2}] = b$. Thus, $D_4 = \mathrm{O}(b^{2v})$. Therefore, $\mathbb{E}[S_{\tilde{\eta}_1}] = \mathrm{O}((nb)^{-1} + b^{2v})$. □

Lemma 5. (a) $S_{\tilde{v},u} = \mathrm{O_p}(n^{-1}b^{-1/2})$, (b) $S_{\tilde{u},v} = \mathrm{O_p}(n^{-1}b^{-1/2})$, (c) $S_{\tilde{u},\eta} = \mathrm{O_p}(n^{-1}b^{-1/2})$, *and* (d) $S_{\tilde{\eta},u} = \mathrm{O_p}(n^{-1}b^{-1/2} + n^{-1/2}b^v)$.

Proof. (a) Because $S_{\tilde{v},u} = S_{\tilde{v}^1,u} + \cdots + S_{\tilde{v}^q,u}$, it suffices to evaluate only $S_{\tilde{v}^1,u}$.

$$\begin{aligned}\mathbb{E}(S_{\tilde{v}^1,u}^2) = \frac{1}{n}\mathbb{E}[u_1^2(\tilde{v}_1^1)^2] &\sim \frac{1}{n^3b^2}\sum_{i\neq 1}\sum_{j\neq 1}\mathbb{E}\left[u_1^2 v_i K_{1i}\frac{p_{w,i}}{p_{zw,i}} v_j K_{1j}\frac{p_{w,j}}{p_{zw,j}}\right]\\
&= \frac{1}{n^2b^2}\mathbb{E}\left[u_1^2 v_2^2 K_{12}^2 \frac{p_{w,2}}{p_{zw,2}}\right] + \frac{1}{nb^2}\mathbb{E}\left[u_1^2 v_2 K_{12}\frac{p_{w,2}}{p_{zw,2}} v_3 K_{13}\frac{p_{w,3}}{p_{zw,3}}\right]\\
&\leqslant \frac{M}{n^2b^2}\mathbb{E}[u_1^2 v_2^2 K_{12}^2] = \mathrm{O}((n^2b)^{-1})\end{aligned}$$

by bounded densities and $\mathbb{E}[v_i|Z_i^*, W_i] = 0$. Hence, $S_{\tilde{v}^1,u} = \mathrm{O_p}(n^{-1}b^{-1/2})$.

The proofs for (b) and (c) are the same as (a).

(d) Because $S_{\tilde{\eta},u} = S_{\tilde{\eta}^1,u} + \cdots + S_{\tilde{\eta}^q,u}$, it suffices to evaluate only $S_{\tilde{\eta}^1,u}$. Use the same arguments as in Lemmas 5(a) and 4(c) to obtain the result. □

Acknowledgements

Dawit Zerom acknowledges financial support from the University of Alberta SAS fellowship.

References

Chamberlain, G., 1992. Efficiency bounds for semiparametric regression. Econometrica 60, 567–596.

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

Fan, Y., Li, Q., 2003. A kernel-based method for estimating additive partially linear models. Statistica Sinica 13, 739–762.

Härdle, W., Liang, H., Gao, J., 2000. Partially Linear Models. Springer, Physica-Verlag, Heidelberg.

Jones, M.C., Davies, S.J., Park, B.U., 1994. Versions of kernel-type regression estimators. J. Amer. Statist. Assoc. 89, 825–832.

Kim, W., Linton, O., Hengartner, N., 1999. A computationally efficient oracle estimator for additive nonparametric regression with bootstrap confidence intervals. J. Comput. Graph. Statist. 8, 278–297.

Li, Q., 1996. On the root-N-consistent semiparametric estimation of partially linear models. Econom. Lett. 51, 277–285.

Li, Q., 2002. Efficient estimation of additive partially linear models. Internat. Econom. Rev. 41, 1073–1092.

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

Stone, C.J., 1985. Additive regression and other nonparametric models. Ann. Statist. 13, 685–705.

Stone, C.J., 1986. The dimensionality reduction principle for generalized additive models. Ann. Statist. 14, 592–606.