Journal of Statistical Planning and Inference 139 (2009) 1980 – 1991

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Semiparametric additive isotonic regression

Guang Cheng*

Department of Statistics, Purdue University, West Lafayette, IN 47906, USA

ARTICLE INFO

Article history:
Received 15 November 2007
Received in revised form
4 September 2008
Accepted 4 September 2008
Available online 5 October 2008

Keywords:
Isotonic regression
Semiparametric model
Additive regression
Asymptotic normality
Oracle property

ABSTRACT

We consider the efficient estimation in the semiparametric additive isotonic regression model where each additive nonparametric component is assumed to be a monotone function. We show that the least-square estimator of the finite-dimensional regression coefficient is root-n consistent and asymptotically normal. Moreover, the isotonic estimator of each additive functional component is proved to have the oracle property, which means the additive component can be estimated with the highest asymptotic accuracy as if the other components were known. A fast algorithm is developed by iterating between a cyclic pool adjacent violators procedure and solving a standard ordinary least squares problem. Simulations are used to illustrate the performance of the proposed procedure and verify the oracle property.

1. Introduction

The nonparametric isotonic regression model takes the form

$$Y = H(W) + \varepsilon, \tag{1.1}$$

where random variables $(Y, W) \in \mathbb{R} \times \mathbb{R}^J$ and ε is an random error term. In this paper, we assume the nonlinear monotone effects of W, and thus $H(\cdot)$ is modelled to be monotonic in each coordinate of W. The isotonic estimator $\widehat{H}_n(\cdot)$ at each observation of W is computed through the max–min formula (Hanson et al., 1973). The asymptotic behaviors of the isotonic estimators in the low-dimensional models, i.e. $J = 1, 2$, have been studied in Hanson et al. (1973), Brunk (1970), Barlow et al. (1972) and Robertson et al. (1988).

A natural extension of (1.1) which incorporates the adjustments for covariates is the semiparametric isotonic regression model expressed in the form:

$$Y = X'\beta + H(W) + \varepsilon. \tag{1.2}$$

For simplicity of expositions, we assume that the error term ε is independent of (X, W) and has zero mean in this paper. The use of parametric terms for covariates X, if properly specified, reduces the risk of overfitting the model by decreasing the overall "degrees of freedom" of the fit. For example, the nonparametric regression would be meaningless when X is a dummy variable. Under (1.2) given $J = 1$, Huang (2002) showed that the least-square estimate for β, denoted as $\widehat{\beta}_n$, is asymptotically normal, and the asymptotic distribution of the isotonic estimate $\widehat{H}_n$ is unaffected by the introduction of the parametric terms. However, the asymptotic behaviors of $(\widehat{\beta}_n, \widehat{H}_n)$ are unknown when W has arbitrarily large dimension.

* Tel.: +1 765 496 9549.
E-mail address: chengg@stat.purdue.edu

0378-3758/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.09.009

In Section 2, we first present some preliminary analysis about the convergence rates of $(\widehat{\beta}_n, \widehat{H}_n)$ when $J > 1$, which establishes the estimation consistency for the general semiparametric isotonic model (1.2). The analysis also leads us to conjecture that either $\widehat{\beta}_n$ is not asymptotically normally distributed or there exists curse of dimensionality in the estimation of high-dimensional H. The *curse of dimensionality* was first discovered by Stone (1985) in the nonparametric regression models where he assumes certain smoothness conditions, rather than shape constraints, on H. In order to circumvent the above estimation difficulties we impose an additive structure on the nonparametric function H, i.e. $H(W) = \sum_{j=1}^{J} h_j(W_j)$, under which the asymptotic normality of $\widehat{\beta}_n$ can be established. Additionally, the estimator for h_j, denoted as $\widehat{h}_j$, is shown to possess the oracle property, which means that h_j can be estimated as well as it could be by an isotonic estimator when the other parametric and nonparametric components are known. The oracle property also implies that the nonparametric rate of convergence of $\widehat{h}_j$ is independent of the number of additive components in the model. Similar asymptotic property of $\widehat{h}_j$ is also established in the nonparametric additive isotonic regression model where only the nonparametric relationship between the response and the covariates is assumed (Mammen and Yu, 2007).

The semiparametric isotonic regression model (1.2) thus becomes the form:

$$Y = X'\beta + \sum_{j=1}^{J} h_j(W_j) + \varepsilon, \tag{1.3}$$

where each h_j is assumed to be monotone. The nonparametric version of (1.3), i.e. without the term $X'\beta$, was proposed by Bacchetti (1989) as a generalization of the additive models introduced in Nelder and Wedderburn (1972) and Hastie and Tibshirani (1986), where the isotonicity is replaced by the linearity and the smoothness, respectively. Model (1.3) also covers the possibility of using a (known) link function, which is the case presented in Bacchetti (1989) and in Morton-Jones et al. (2000), although some weights are added to the model in the latter. In the epidemiology area, (1.3) is often used to model the relationship between risk and exposure, which is believed either linear or nondecreasing with increasing exposure. For example, Morton-Jones et al. (2000) employed model (1.3) to analyze some real data about the effect of the father's paternal preconceptional radiation dose on the sex ratio of children. In their studies, the response of interest is the log sex ratio. The dose received in the 90 days prior to conception and the total does received prior to that 90 days period are treated as the isotonic variables. The other explanatory variables are either found linear relationship with the response, e.g. paternal age and birth order (James and Rostron, 1985), or of categorical type, e.g. the social class of fathers. Therefore the linear part in (1.3) is particulary needed for modelling those variables.

In Section 3, an iterative algorithm is proposed to successively compute the backfitting estimates $\hat{h}_j$'s by a cyclic pool adjacent violators (CPAV) procedure (Bacchetti, 1989) while holding β fixed, and compute $\widehat{\beta}_n$ by solving a simple linear regression while holding $(h_1, \ldots, h_J)$ fixed. Section 4 illustrates the performance of the above iterative computation algorithm via simulated examples. Section 5 contains a brief discussion of future research directions, and proofs are given in Appendix A.

2. Main results

2.1. Preliminary analysis

In this subsection, we give some preliminary asymptotic results about the estimator $(\widehat{\beta}_n, \widehat{H}_n)$ defined below, which establish the estimation consistency for the general semiparametric isotonic model (1.2). Furthermore, the asymptotic analysis leads us to conjecture that there exists the curse of dimensionality in the estimation of H under shape constraints. Similar to the definition of an isotonic estimate, $(\widehat{\beta}_n, \widehat{H}_n)$ is defined with respect to the least square deviation criteria, i.e.

$$(\widehat{\beta}_n, \widehat{H}_n) = arg\ \min \left(n^{-1} \sum_{i=1}^{n} (Y_i - X_i'\beta - H(W_i))^2 \right) \tag{2.1}$$

subject to the restrictions that β belongs to some convex subset $\mathscr{B} \in \mathbb{R}^d$ and H is nondecreasing in each coordinate of W. Note that H is an isotonic (order preserving) function with respect to the partial order $\ll$ in $\mathbb{R}^J$ defined as the follows: $V \ll U$ if and only if $V_j \leqslant U_j$, where V_j and U_j are the j-th element of U and V, respectively. The requirement that $H(\cdot)$ has the same monotonic direction in each coordinate can be satisfied easily in practice by the change of sign. The solution of (2.1) is well defined and uniquely determined since $\mathscr{B}$ is a convex subset and the class of isotonic functions forms a closed convex cone. Without loss of generality, we assume that the true function $H_0(\cdot)$ is uniformly bounded, $X \in [-1, 1]^d$ and $W \in [-1, 1]^J$. For convenience, we call an arbitrary function of J real variables as being "J-dimensional" in the following.

Proposition 2.1 first proves the consistency of $(\widehat{\beta}_n, \widehat{H}_n)$ and further illustrates that the rate of convergence of $\widehat{H}_n$ decreases rapidly as its dimension increases. We require the below two primary assumptions in the proof:

$$E(X - E(X|W))^{\otimes 2} \text{ is strictly positive, and,} \tag{2.2}$$

$$E(\exp(\gamma|\varepsilon|)) < C \tag{2.3}$$

for some $\gamma, C > 0$ and the outer product $V^{\otimes 2}$ is defined as VV' for any vector $V \in \mathbb{R}^d$.

Proposition 2.1. *Suppose that conditions* (2.2) *and* (2.3) *hold, we have*

(i) $(\widehat{\beta}_n, \widehat{H}_n)$ *is a consistent estimator*;
(ii) $\widehat{\beta}_n$ *is* n^{δ}*-consistent for* $\delta < \frac{1}{2}$ *when* $J > 1$;
(iii) *Given that* $\|\cdot\|_2$ *is* L_2 *norm*,

$$\|\widehat{H}_n(W) - H_0(W)\|_2 = O_P(n^{-1/3}\log n) \quad \text{for } J = 1, \tag{2.4}$$

$$\|\widehat{H}_n(W) - H_0(W)\|_2 = O_P(n^{-1/4}(\log n)^2) \quad \text{for } J = 2, \tag{2.5}$$

$$\|\widehat{H}_n(W) - H_0(W)\|_2 = O_P(n^{-1/4(J-1)}\log n) \quad \text{for } J \geqslant 3. \tag{2.6}$$

It is well known that the convergence rate of nonparametric estimate depends on the entropy number of the corresponding function classes. Hence, we make use of the bracketing entropy number for the class of multi-dimensional monotone functions (Gao and Wellner, 2007) in the proof of Proposition 2.1. Although we do not give lower bounds for the convergence rates of $\widehat{H}_n$ and $\hat{\beta}_n$ in Proposition 2.1, the tightness of the entropy bounds shown in Lemma A.1 together with the phase change in the entropies between $J = 1$ and 2 leads us to conjecture that (see Gao and Wellner, 2007, Remark 5.2):

$$\|X'(\hat{\beta}_n - \beta_0) + \widehat{H}_n(W) - H_0(W)\|_2 \geqslant O_P(n^{-1/4(J-1)}\log n) \quad \text{for } J \geqslant 3 \tag{2.7}$$

following the line in the proof of Proposition 2.1. Eq. (2.7) further implies that

$$\|\hat{\beta}_n - \beta_0\| \geqslant O_P(n^{-1/4(J-1)}\log n) \quad \text{or} \quad \|\widehat{H}_n - H_0\|_2 \geqslant O_P(n^{-1/4(J-1)}\log n) \tag{2.8}$$

for $J \geqslant 3$ because of the boundedness of X. Combining the upper bound shown in Proposition 2.1 and the conjectured lower bound in (2.8), we believe that either there exists the curse of dimensionality in the estimation of H under shape constraints or $\hat{\beta}_n$ is not asymptotically normally distributed for $J \geqslant 3$.

However, in the context that $H(\cdot)$ is a high-dimensional smooth function, we can construct an asymptotically normally distributed $\widehat{\beta}_n$ through inserting a higher order kernel estimator $\widehat{H}_n$ in the nonlinear orthogonal projection on W (Robinson, 1988) or using series estimation (Newey, 1997). We may explain such loss of asymptotic efficiency due to the fact that the class of J-dimensional monotone functions does not belong to the P-Donsker for $J > 1$ (Gao and Wellner, 2007). In next section, all the concerns raised here can be resolved by imposing the additive structure of $H(\cdot)$.

Remark 2.1. In Proposition 2.1, we assume that the tail of the error term is sub-exponential, i.e. (2.3). However, the estimation consistency in the general semiparametric model (1.2) can still be guaranteed even when we relax the somewhat strong condition. For example, if we assume the moment condition that $E|\varepsilon|^r < \infty$ for some $r > 0$, then the rates of convergence in (2.4) and (2.5) become $o_P(n^{-1/3+1/r})$, $o_P(n^{-1/4+1/r}\log n)$ and $o_P(n^{-1/(4J-4)+1/r})$, respectively. Obviously, we can still achieve the consistency by choosing the proper value of r. Interestingly, we notice that the stricter moment condition on the error term is required for the consistent estimation of the higher-dimensional monotone function. The well-known *curse of dimensionality* phenomenon is again reflected from this point of view.

2.2. Asymptotic distribution

Motivated by the previous analysis, we will decompose $H(\cdot)$ into the sum of one-dimensional monotone h_j's in this section, and then explore the related asymptotic properties. Similarly, the estimators $(\widehat{\beta}_n, \widehat{h}_1, \ldots, \widehat{h}_J)$ are defined as the minimizer of

$$S_n(\beta, h_1, \ldots, h_J) = n^{-1}\sum_{i=1}^{n}\left(Y_i - X_i'\beta - \sum_{j=1}^{J} h_j(W_{ij})\right)^2 \tag{2.9}$$

given that each h_j is a monotone function with bounded derivative. We also assume the norming condition that $\int h_j(w_j)\,dw_j = 0$ for the purpose of parameter identifiability. Under the additive structure of $H(W)$, $\widehat{\beta}_n$ is shown to be asymptotically normal with smaller variance even comparing to that of the semiparametric efficient estimate. Next, we have proved that each additive component h_j can be estimated as well as it could be by an isotonic estimator as if the other components were known.

We now make the following regularity conditions for $1 \leqslant j \leqslant J$:

S1. The function h_j satisfies the condition that

$$\inf_{|w_j - w_j'| \geqslant \delta} |h_j(w_j) - h_j(w_j')| \geqslant C_1\delta^{\gamma} \tag{2.10}$$

for any $\delta > 0$ and some constants $C_1, \gamma > 0$.

S2. The density for W_j, denoted as p_{W_j}, is assumed to be bounded away from zero and infinity, and fulfills the below Lipschitz condition

$$\sup_{-1\leqslant w_j,w_j'\leqslant 1}|p_{W_j}(w_j)-p_{W_j}(w_j')|\leqslant M|w_j-w_j'|^{\rho} \tag{2.11}$$

for some constants $M, \rho > 0$.

S3. The function $\zeta_j(w)\equiv E(X|W_j=w)$ satisfies the condition

$$\|\zeta_j(w_j)-\zeta_j(w_j')\|\leqslant C_2|w_j-w_j'| \tag{2.12}$$

for some constant $C_2 > 0$.

S4. $E(X-\sum_{j=1}^{J}E(X|W_j))^{\otimes 2}$ is strictly positive.

Assumptions S1 and S2 are also used in the estimation of nonparametric additive isotonic regression models (Mammen and Yu, 2007). Specifically, S1 implicitly requires that h_j is strictly monotone, and S2 restricts the variation of the density for W_j up to some order. The condition S4 is weaker than (2.2) assumed in the nonadditive model (1.2) since we can decompose $E(X-\sum_{j=1}^{J}E(X|W_j))^{\otimes 2}$ as the sum of $E(X-E(X|W))^{\otimes 2}$ and $E(E(X|W)-\sum_{j=1}^{J}E(X|W_j))^{\otimes 2}$. One gain of such relaxation is that the parameter β is still identifiable even when X is a deterministic function of W as long as $E(X|W)\neq\sum_{j=1}^{J}E(X|W_j)$. For example, we can allow the interaction term to enter the model parameterically, e.g. $Y=(W_1W_2)\beta+h_1(W_1)+h_2(W_2)+\varepsilon$.

Under the above regularity conditions, Theorem 1 summarizes the asymptotic behaviors of the estimators in the semiparametric additive isotonic regression model.

Theorem 1. *Given that assumptions* (S1)–(S4) *and* (2.3) *hold, and that W is pairwise independent, we have*

$$\sqrt{n}(\widehat{\beta}_n-\beta_0)\xrightarrow{\mathrm{d}}\mathrm{N}(0,\Sigma), \tag{2.13}$$

where $\Sigma=\sigma^2[E(X-\sum_{j=1}^{J}E(X|W_j))^{\otimes 2}]^{-1}$ *and* $\sigma^2=Var(\varepsilon)$. *For* $w_j\in(-1,1)$ *with first derivative* $\dot{h}_j(w_j)>0$, *holds that*

$$n^{1/3}\frac{(2p_{W_j}(w_j))^{1/3}}{\sigma^{2/3}\dot{h}_j(w_j)^{1/3}}[\widehat{h}_j(w_j)-h_j(w_j)] \tag{2.14}$$

converges in distribution to the slope of the greatest convex minorant of $W(t)+t^2$, *where* $W(t)$ *is a two sided Brownian motion.*

Theorem 1 implies that $\widehat{\beta}_n$ is $\sqrt{n}$-consistent and asymptotically normal. We provide an intuitive justification for its asymptotic covariance matrix here. Note that (1.3) can be rewritten as $Y=(X-\sum_{j=1}^{J}E(X|W_j))'\beta+\sum_{j=1}^{J}f_j(W_j)+\varepsilon$, where $f_j(W_j)=h_j(W_j)+E(X|W_j)'\beta$. Since $f_k(W_k)$ is uncorrelated with $(X-\sum_{j=1}^{J}E(X|W_j))$ for any $k=1,\ldots,J$, we can view $\widehat{\beta}_n$ as the ordinary least square (OLS) estimate of simple linear regression $Y\sim(X-\sum_{j=1}^{J}E(X|W_j))$. This intuitively explains the form of Σ. The larger asymptotic variance for each coordinate of $\widehat{\beta}_n$, compared with that of OLS in the regression $Y\sim X$, can be viewed as the effect of nonparametrically modelled covariate W.

Based on the asymptotic normality result (2.13), we can easily construct the $(1-\alpha)$-th level confidence region for β, i.e.

$$[\widehat{\beta}_n-\widehat{\Sigma}^{1/2}(z_{1-\alpha/2}/\sqrt{n}),\widehat{\beta}_n+\widehat{\Sigma}^{1/2}(z_{1-\alpha/2}/\sqrt{n})],$$

where $\widehat{\Sigma}$ is a consistent estimate for Σ and z_β is the standard normal β-th quantile. We can consistently estimate Σ by building upon the consistent estimate $\widehat{E}(X|W_j=w_{ij})$ for the conditional expectation $E(X|W_j=w_{ij})$, i.e.

$$\widehat{\Sigma}=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-x_i'\widehat{\beta}_n-\sum_{j=1}^{J}\widehat{h}_j(w_{ij})\right)^2\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\sum_{j=1}^{J}\widehat{E}(X|W_j=w_{ij})\right)^{\otimes 2}\right)^{-1}. \tag{2.15}$$

Some nonparametric approaches with k nearest neighbor weights or local linear weights can be used to obtain $\widehat{E}(X|W_j=w_{ij})$, see Sections 3 and 4 in Stone (1977). In practice, X is often a categorical type of data. Alternatively, we can estimate $E(X|W_j=w_j)$ using the below kernel method. For simplicity, we assume that X is a dichotomous variable indicating two treatment groups with $P(X=1)=\gamma$ and $P(X=0)=1-\gamma$. Then $\widehat{E}(X|W_j=w_{ij})=(\widehat{\gamma}\widehat{p}_1(w_{ij}))/\widehat{p}_{W_j}(w_{ij})$, where $\widehat{\gamma}$ is the proportion of subjects in treatment group with $X=1$ in all the observations, $\widehat{p}_{W_j}$ and $\widehat{p}_1$ are the corresponding consistent kernel estimates for the density of W_j and the conditional density of W_j given $X=1$. The proper choices of bandwidth and kernel function can be referred to Silverman (1986).

In model (1.3), we further prove that each h_j can be estimated with the highest asymptotic accuracy. Specifically, the limiting distribution of $\hat{h}_j$ at a fixed point is the same as that of a regular isotonic estimator when the other parametric and nonparametric components are known. Such nice property is called the *oracle property* in the literature. The asymptotic distribution of $\hat{h}_j$ is not affected by the fact that the parameter β and other additive functional components are being estimated simultaneously. Intuitively, this is because the convergence rate of $\widehat{\beta}_n$ is faster than that of $\hat{h}_j$ and the localization of the estimation for h_j takes place with respect to other covariates.

Remark 2.2. It is well known that under conditional homoskedasticity the lower bound of the asymptotic variance of an estimator for β equals to $\sigma^2\{E(X - E(X|\mathscr{F}))^{\otimes 2}\}^{-1}$, where $E(X|\mathscr{F})$ is the projection of X onto the sum of closed $L_2(P_{W_j})$ spaces, given that h_j's have no shape restrictions (Chamberlin, 1992). However, considering the inequality that $E(X-\sum_{j=1}^{J} E(X|W_j))^{\otimes 2} \geqslant E(X-E(X|\mathscr{F}))^{\otimes 2}$, we find the asymptotic variance of $\widehat{\beta}_n$ in our paper is even smaller than the above lower bound. This result is not surprising if we interpret it from a geometrical point of view. In this paper, we are actually estimating the projection of Y onto the sum of a linear space and a group of closed convex cones formed by the monotone functions $h_j(\cdot)$'s. Geometrically speaking, a cone can usually be viewed as a half-space. That could explain why $\widehat{\beta}_n$ has less perturbation. Combining with the above, we know that the least square estimate of β in (1.3) will become asymptotically less efficient if the monotonicity of h_j is ignored.

3. Computation algorithm

In this section, we suggest a two-step minimization approach to compute $(\widehat{\beta}_n, \hat{h}_1, \ldots, \hat{h}_J)$. For convenience of expressions, we use $\Lambda(w)$ and $\widehat{\Lambda}_n(w)$ to represent $(h_1(w_1), \ldots, h_J(w_J))$ and its backfitting estimator, respectively. The main idea is to successively minimize $S_n(\beta, \Lambda)$ over Λ for all fixed values of $\beta \in \mathscr{B}$ first to obtain $\widehat{\Lambda}_n(\cdot, \beta)$, and then minimize the profiled function $S_n(\beta, \widehat{\Lambda}_n(\cdot, \beta))$ over $\mathscr{B}$ to find $\widehat{\beta}_n$, and hence $\widehat{\Lambda}_n(w) = \widehat{\Lambda}_n(w, \widehat{\beta}_n)$. In the first minimization stage, we obtain $\widehat{\Lambda}_n(\cdot, \beta)$ for any fixed $\beta \in \mathscr{B}$ through the iterative application of the pool adjacent violator algorithm to the additive components, called CPAV algorithm in Bacchetti (1989). Our computation algorithm in the above can be viewed as an extension of Bacchetti's CPAV procedure to include one more minimization step for adjusting the covariates X.

Algorithm.

Λ-step: Let $k = k + 1$. Minimize $S_n(\beta^{(k-1)}, \Lambda)$ with respect to Λ to obtain $\Lambda^{(k)} \equiv (h_1^{(k)}, \ldots, h_j^{(k)})$ using the CPAV algorithm;
β-step: Minimize $S_n(\beta, \Lambda^{(k)})$ with respect to β to obtain $\beta^{(k)}$. Go back to Λ-step until convergence.

We can interpret the optimization problem in the Λ-step in a backfitting manner. For a given $\beta^{(k-1)}$, we can estimate $h_j^{(k)}$ by solving the isotonic regression: $(y - x'\beta^{(k-1)} - h_1^{(k)} - \cdots - h_{j-1}^{(k)} - h_{j+1}^{(k-1)} - \cdots - h_J^{(k-1)}) \sim w_j$ based on the below max–min formula. Given $\beta^{(0)} = 0$ and $h_j^{(0)} = (0, \ldots, 0)'$ for any $j = 1, \ldots, J$, we have

$$h_j^{(k)}(w_{(i)j}) = \max_{s \leqslant i} \min_{t \geqslant i} \frac{\sum_{l=s}^{t} (y_{[l]} - x'_{[l]}\beta^{(k-1)} - h_{[l]1}^{(k)} - \cdots - h_{[l]j-1}^{(k)} - h_{[l]j+1}^{(k-1)} - \cdots - h_{[l]J}^{(k-1)})}{t - s + 1},$$

where $(y_{[l]}, x_{[l]}, w_{[l]L})$ is the observation corresponding to the l-th ordered w_j, and $h_{[l]L}^{(k)}$ is the value of $h_L^{(k)}$ at $w_{[l]L}$ after the k-th iteration for $j = 1, \ldots, J$ and $L = 1, \ldots, j-1, j+1, \ldots, J$. In the β-step, we essentially solve a simple linear regression $(y - \sum_{j=1}^{J} h_j^{(k)}) \sim x$.

We can view the proposed iterative algorithm as the cyclic projection of the response Y onto the sum of a closed subspace S_1 and a set of closed convex cones $S_{21}, \ldots, S_{2J}$. The convergence in the Λ-step is established by Dykstra (1983), which further implies the convergence of the whole algorithm. The rigorous proof of our algorithm convergence can be adapted from that of Theorem 2 in Mammen and Yu (2007) after some minor revisions. However, the additive component backfitting estimate $\widehat{\Lambda}_n$ is generally not unique when there exists a tuple of vectors $(f_1, \ldots, f_J)$ such that $f_j + \widehat{h}_j$ is monotone and $\sum_{j=1}^{J} f_{ij} = 0$ for any $i = 1, \ldots, n$.

4. Simulation example

We conduct Monte Carlo simulations to evaluate the finite sampling performance of the proposed method. The primary model M1 in this section is designed as:

M1 $Y = (W_1 W_2)\beta + h_1(W_1) + h_2(W_2) + \varepsilon$,

Table 1
Simulation results of β in M1–M1(h_1, h_2) ($\beta_0 = 1$)

n	M1	M1(h_1)	M1(h_2)	M1(h_1, h_2)
100	0.955 (0.358)	0.967 (0.242)	1.029 (0.237)	0.981 (0.169)
300	0.981 (0.215)	1.014 (0.135)	0.987 (0.127)	0.989 (0.089)
600	0.991 (0.189)	1.006 (0.101)	0.994 (0.093)	0.997 (0.067)

Table 2
Simulation results of h_1 in M1–M3 (MISE)

n	M1	M2	M3	M1/M2	M1/M3
100	0.0246	0.0226	0.0222	1.0885	1.1081
300	0.0152	0.0144	0.0146	1.0556	1.0411
600	0.0102	0.0097	0.0098	1.0515	1.0408

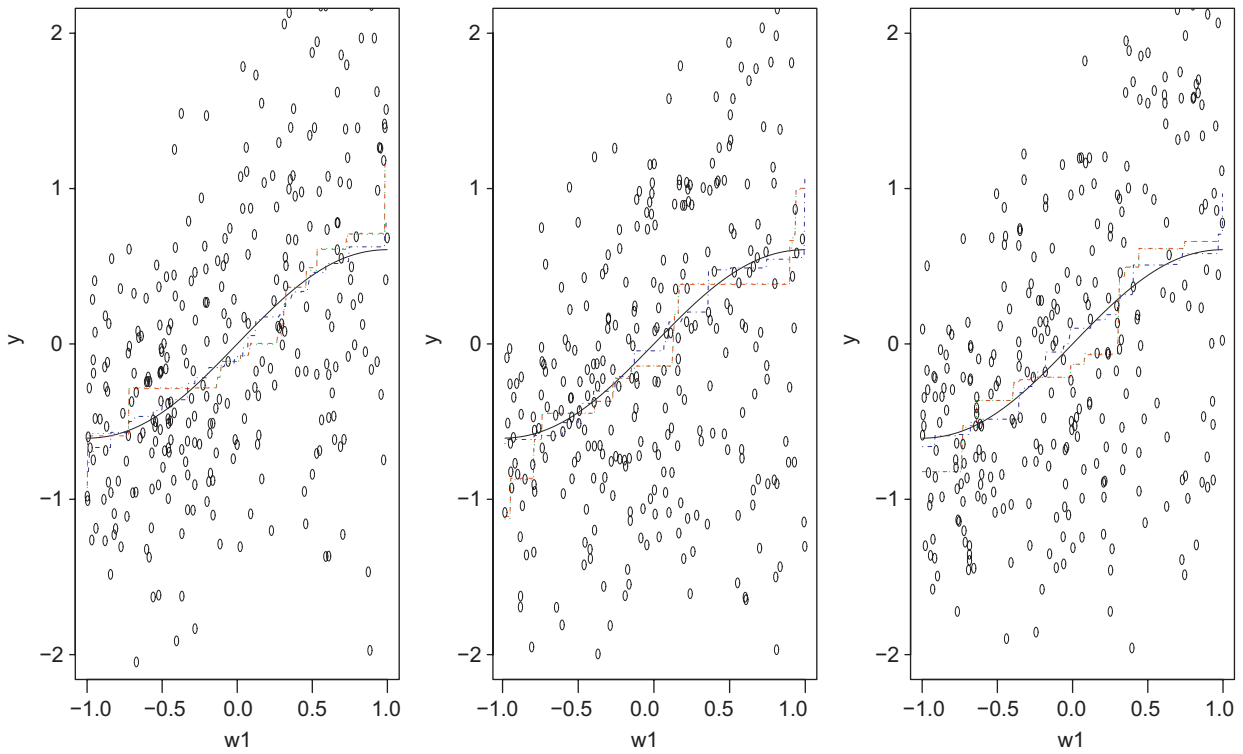

Fig. 1. Isotonic estimate for h_1. The solid lines, dashed lines, dotted lines and dotted-dashed lines represent the true curve, semiparametric additive isotonic estimate, nonparametric additive isotonic estimate and oracle estimate, respectively. The left, center and right graphs show the estimated curves corresponding to 25%, 50% and 75% quantiles for MISE between semiparametric additive isotonic and oracle estimators with 300 observations.

where $W_1 \sim \text{Unif}[-1,1]$, $W_2 \sim$ truncated normal within the interval $[-1,1]$ and $\varepsilon \sim 0.5\text{N}(1, 0.5^2) + 0.5\text{N}(-1, 1^2)$. We set the true values $\beta_0 = 1$, $h_{10}(w_1) = w_1 \exp(-w_1^2/2)$ and $h_{20}(w_2) = \sin(\pi w_2/2)$. The simulations are run for various sample sizes ranging from 100 to 600. For each sample size, 100 datasets were analyzed. In each setting, we iterate 500 times.

In Table 1, we report the estimators for β and its standard deviation (given in the parentheses) in different models specified below. It is easy to notice that β is consistently estimated in all the models and the simulation results in the full model M1 are worse than those in other models where partial or complete information about the additive functional components is provided. However, the standard deviation in each model approaches closer as the sample size n increases, which is consistent with our theoretical results.

$$\begin{aligned}
\text{M1}(\text{h}_1)\quad & Y^{(1)} = (W_1 W_2)\beta + h_2(W_2) + \varepsilon, \text{ where } Y^{(1)} = Y - h_{10}(W_1),\\
\text{M1}(\text{h}_2)\quad & Y^{(2)} = (W_1 W_2)\beta + h_1(W_1) + \varepsilon, \text{ where } Y^{(2)} = Y - h_{20}(W_2),\\
\text{M1}(\text{h}_1, \text{h}_2)\quad & Y^{(12)} = (W_1 W_2)\beta + \varepsilon, \text{ where } Y^{(12)} = Y - h_{10}(W_1) - h_{20}(W_2).
\end{aligned}$$

Table 2 presents the empirical mean integrated squared error (MISE) of the semiparametric additive isotonic estimator in M1, the nonparametric isotonic estimator in M2 and the oracle estimator in M3. The "Oracle Model" M3, in which only one nonparametric additive component is unknown, is included here as a golden rule for comparison. Note that the isotonic estimates in M1–M3 have the same asymptotic distribution according to Corollary 1 in Mammen and Yu (2007) and Theorem 1. The results in Table 2 show that three estimators have very similar finite sample performance. In particular, the ratios (M1/M2) and (M1/M3) are close to one and converge to one as sample size increases.

M2 $Y^{(0)} = h_1(W_1) + h_2(W_2) + \varepsilon$, where $Y^{(0)} = Y - W_1W_2\beta_0$,
M3 $Y^{(02)} = h_1(W_1) + \varepsilon$, where $Y^{(2)} = Y - W_1W_2\beta_0 - h_{20}(W_2)$.

Fig. 1 visually represents the isotonic estimators for h_1 in the models M1–M3 corresponding to 25%, 50% and 75% quantiles of the empirical MISE between its semiparametric additive isotonic and oracle estimators. We observe that all the isotonic estimators produce almost identical curves, which is consistent with our theoretical results.

5. Discussion

We consider the optimal estimation of both parametric and nonparametric components in the semiparametric additive isotonic regression model. The parametric estimator is proved to be $\sqrt{n}$-consistent and asymptotically normal. We also establish the oracle property of the additive isotonic estimators. Although the isotonic estimation posses the above advantages, we realize that carrying out one data isotonization process has some potential drawbacks, e.g., heavy dependence on the behavior of the data at the tails. Hence incorporating the smoothing techniques (Mammen, 1991) in the estimation of our model may be a future research goal. In the future research, we are also interested in generalizing the results in this paper to the semiparametric model with an unknown link function (Horowitz and Mammen, 2005), monotone response model (Pal and Banerjee, 2007) or survival analysis (Huang, 1999). Another worthwhile avenue of research is to relax some assumptions in our paper such that our theoretical results can be applied to more general models. For example, we can relax the pairwise independence assumption on W, though it can be trivially satisfied for fixed design point W. We may assume the shape constraints (Mammen and Thomas-Agnan, 1999), i.e. convexity and monotonicity, on only some of the additive components.

Acknowledgments

I sincerely thank Dr. Kyusang Yu, Dr. Nicoleta Serban, Dr. Fuchang Gao and an associate editor for several insightful comments and suggestions

Appendix A. Proofs

We first state the bracketing entropy number of a class of monotone functions $f : [0,1]^d \mapsto [0,1]$, denoted by $\mathscr{F}_d$, and then present the asymptotic continuity modulus of some empirical processes.

Lemma A.1. *Suppose Q is a probability measure on $[0,1]^d$ with Lebesgue density q satisfying $1/M \leqslant \inf_{x\in[0,1]^d} q(x) \leqslant \sup_{x\in[0,1]^d} q(x) \leqslant M$ for some $M > 0$. Then if $(d-1)p \neq d$, we have*

$$K_2\varepsilon^{-\alpha} \leqslant \log N_{[\,]}(\varepsilon, \mathscr{F}_d, L_p(Q)) \leqslant K_1\varepsilon^{-\alpha}, \tag{A.1}$$

where $\log N_{[\,]}(\varepsilon, \mathscr{F}_d, L_p(Q))$ is the ε-bracketing entropy number and $\alpha = \max\{d, (d-1)p\}$. If $(d-1)p = d$, then

$$K_2\varepsilon^{-d} \leqslant \log N_{[\,]}(\varepsilon, \mathscr{F}_d, L_p(Q)) \leqslant K_1\varepsilon^{-d}(\log(1/\varepsilon))^d \tag{A.2}$$

for constants K_1 and K_2 depending only on d, p and M.

Lemma A.2. *Consider a uniformly bounded class of functions $\mathscr{G}$, with $\sup_{g\in\mathscr{G}}\|g - g_0\|_\infty < \infty$ and $\log N_{[\,]}(\varepsilon, \mathscr{G}, P) \leqslant A\varepsilon^{-\alpha}$ for all $\varepsilon > 0$, where $\alpha \in (0,2)$. Then for $\delta_n = n^{-1/(2+\alpha)}$,*

$$\sup_{g\in\mathscr{G}} \frac{|(\mathbb{P}_n - P)(g - g_0)|}{\|g - g_0\|_2^{1-\alpha/2} \vee \sqrt{n}\delta_n^2} = O_P(n^{-1/2}).$$

Lemma A.3. *Let $\mathscr{F} = \{f_t : t \in T\}$ be a class of functions satisfying $|f_s(x) - f_t(x)| \leqslant d(s,t)F(x)$ for every s and t and some fixed function F. Then, for any norm $\|\cdot\|$,*

$$N_{[\,]}(2\varepsilon\|F\|, \mathscr{F}, \|\cdot\|) \leqslant N(\varepsilon, T, d).$$

Remark A.1. Lemmas A.1 and A.2 are adapted from Corollary 1.3 in Gao and Wellner (2007) and Lemma 5.13 in Van de Geer (2000) with minor revisions, respectively. Lemma A.3 is exactly Theorem 2.7.11 in Van der Vaart and Wellner (1996).

Proof of Proposition 2.1. Let $w_{(i)}$ be the observation of $W_{(i)}$, where $W_{(i)} \equiv \{W \in [-1,1]^J : H(W_{(i)}) = (H(W))_{(i)}\}$. We first show that

$$\max\{|\widehat{H}_n(w_{(1)})|, |\widehat{H}_n(w_{(n)})|\} = O_P(\log n). \tag{A.3}$$

Without loss of generality, we only prove $|\widehat{H}_n(w_{(n)})| = O_P(\log n)$. Let $\mathscr{L}_J$ be the collection of subsets L of $\mathbb{R}^J$ having the property that if $U \in L$ and $U \ll V$ then $V \in L$. Then the min–max formula in Hanson et al. (1973) implies that

$$\widehat{H}_n(w_{(n)}) = \max_{\{L: w_{(n)} \in L \in \mathscr{L}_J\}} \left(\frac{\sum_{\{\alpha: w_\alpha \in L, \alpha \leqslant n\}} (y_\alpha - x'_\alpha \widehat{\beta}_n)}{\mathrm{card}\{\alpha : w_\alpha \in L, \alpha \leqslant n\}} \right), \tag{A.4}$$

where card(A) represents the number of elements in the set A. Following the above characterizations of $\widehat{H}_n(w_{(n)})$, we establish the below set of inequalities:

$$\begin{aligned} |\widehat{H}_n(w_{(n)})| &\leqslant \max_{\{L: w_{(n)} \in L \in \mathscr{L}_J\}} \left(\frac{\sum_{\{\alpha: w_\alpha \in L, \alpha \leqslant n\}} |y_\alpha - x'_\alpha \widehat{\beta}_n|}{\mathrm{card}\{\alpha : w_\alpha \in L, \alpha \leqslant n\}} \right) \\ &\leqslant \max_{\{L: w_{(n)} \in L \in \mathscr{L}_J\}} \left(\frac{\sum_{\{\alpha: w_\alpha \in L, \alpha \leqslant n\}} |x'_\alpha(\beta_0 - \widehat{\beta}_n)| + |H_0(w_\alpha)|}{\mathrm{card}\{\alpha : w_\alpha \in L, \alpha \leqslant n\}} \right) + \max_{\{L: w_{(n)} \in L \in \mathscr{L}_J\}} \left(\frac{\sum_{\{\alpha: w_\alpha \in L, \alpha \leqslant n\}} |\varepsilon_\alpha|}{\mathrm{card}\{\alpha : w_\alpha \in L, \alpha \leqslant n\}} \right) \\ &\leqslant M + |\varepsilon|_{(n)}, \end{aligned}$$

where M is a finite positive constant. The last inequality follows from the assumptions on the parameters (β, H). Considering the sub-exponential tail of ε, we have proved (A.3).

By the definition of $(\widehat{\beta}_n, \widehat{H}_n)$, we have

$$P(X' d_{1n} + d_{2n}(W))^2 \leqslant (\mathbb{P}_n - P)[(Y - X'\beta_0 - H_0)^2 - (Y - X'\widehat{\beta}_n - \widehat{H}_n)^2],$$

where $\mathbb{P}_n$ is the empirical measure of (Y_i, X_i, W_i), $d_{1n} = (\widehat{\beta}_n - \beta_0)$ and $d_{2n}(W) = (\widehat{H}_n - H_0)(W)$. Next, we consider the class of functions

$$\mathscr{F}_n \equiv \{[r_n(y - x'\beta_0) - r_n H_0(w)]^2 - [r_n(y - x'\beta) - G(w)]^2,\ \beta \in \mathscr{B} \text{ and } G \in \mathscr{G}\},$$

where $r_n = (\log n)^{-1}$ and $\mathscr{G}$ is a class of uniformly bounded nondecreasing functions of J-dimension. Note that $r_n \|\widehat{H}_n\|_\infty = O_P(1)$, where $\|\cdot\|_\infty$ is the uniform norm, based on (A.3). We next study the δ-bracketing entropy of $\mathscr{F}_n$ in terms of $L_2(P)$-norm. And we know that δ-bracketing entropy of $\mathscr{G}$ is of the order $1/\delta$ for $J = 1$, $\delta^{-2}(\log(1/\delta))^2$ for $J = 2$, and $\delta^{-(2J-2)}$ for $J > 2$ based on Lemma A.1. Since the functions in $\mathscr{F}_n$ are quadratic in (β, G), δ-bracketing entropy of $\mathscr{F}_n$ is the same as that of $\mathscr{G}$. In view of the discussions in Van der Vaart and Wellner (1996, p. 326), we have the convergence rates that $P(X' d_{1n} + d_{2n}(W))^2 = O_P(\delta_n^2 r_n^{-2})$ based on Theorem 3.4.1 and Lemma 3.4.2 in Van der Vaart and Wellner (1996), where δ_n is determined by

$$\sqrt{n}\delta_n^2 \geqslant \tilde{J}_{[\,]}(\delta_n, \mathscr{F}_n, L_2(P)) \left(1 + \frac{\tilde{J}_{[\,]}(\delta_n, \mathscr{F}_n, L_2(P))}{\sqrt{n}\delta_n^2}\right)$$

and

$$\tilde{J}_{[\,]} = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon, \mathscr{F}_n, L_2(P))}\, \mathrm{d}\varepsilon \quad \text{if } J = 1,$$

$$\tilde{J}_{[\,]} = \int_{c\delta^2}^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon, \mathscr{F}_n, L_2(P))}\, \mathrm{d}\varepsilon \quad \text{if } J \geqslant 2.$$

By considering the parameter identifiability condition (2.2), we have completed the whole proof. □

Proof of Theorem 1. Note that the δ-bracketing entropy number for a class of additive monotone functions

$$\mathscr{Q} \equiv \left\{ H(w) = \sum_{j=1}^J h_j(w_j) : h_j(w_j) \text{ is uniformly bounded and monotone with } \int_{-1}^1 h_j(w_j)\, \mathrm{d}w_j = 0 \right\}$$

is of the order (J^2/δ) based on Lemma A.1. Then, following similar logic in Proposition 2.1, we have

$$\|\widehat{\beta}_n - \beta_0\| = O_P(n^{-1/3}\log n), \tag{A.5}$$

$$\left\|\sum_{j=1}^{J}(\widehat{h}_j - h_j)\right\|_2 = O_P(n^{-1/3}\log n), \tag{A.6}$$

under condition (2.3) and S4. Note that (A.6) implies that

$$\int [(\hat{h}_1(w_1) - h_1(w_1)) + \cdots + (\hat{h}_J(w_J) - h_J(w_J))]^2\, dw = O_P(n^{-2/3}(\log n)^2)$$

by the assumption that the density for W is bounded away from zero. Then based on the normalization condition that $\int h_j(w_j)\,dw_j = 0$, we have

$$\max_{1\leqslant j\leqslant J}\int (\hat{h}_j(w_j) - h_j(w_j))^2\, dw_j = O_P(n^{-2/3}(\log n)^2). \tag{A.7}$$

Therefore, we have

$$\max_{1\leqslant j\leqslant J}\|\widehat{h}_j - h_j\|_2 = O_P(n^{-1/3}\log n) \tag{A.8}$$

by the assumption that the density for W is bounded away from infinity. Since $E(X|\widehat{h}(W_j))$ has the same jump points as $\widehat{h}_j$ by the characterization of the solution to the isotonic regression (see Barlow et al., 1972, Theorem 1.7), the stationary equation implies that

$$\mathbb{P}_n\left(Y - X'\widehat{\beta}_n - \sum_{j=1}^{J}\widehat{h}_j(W_j)\right)\left(X - \sum_{j=1}^{J}E(X|h_j^{-1}(\widehat{h}_j(W_j)))\right) = 0,$$

where h_j^{-1} is the inverse function of h_j. Combining the above equation and the independence assumption of ε, we have

$$\begin{aligned} &E\left[\left(X'(\widehat{\beta}_n - \beta_0) + \sum_{j=1}^{J}(\widehat{h}_j - h_j)(W_j)\right)\left(X - \sum_{j=1}^{J}E(X|\widehat{h}_j(W_j))\right)\right] \\ &\quad = n^{-1/2}\mathbb{G}_n\left[\left(Y - X'\widehat{\beta}_n - \sum_{j=1}^{J}\widehat{h}_j(W_j)\right)\left(X - \sum_{j=1}^{J}E(X|\widehat{h}_j(W_j))\right)\right], \end{aligned} \tag{A.9}$$

where $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$. Note that (A.9) can be rewritten as

$$\Sigma_n(\widehat{\beta}_n - \beta_0) + \Delta_n = n^{-1/2}\mathbb{G}_n\left(Y - X'\beta_0 - \sum_{j=1}^{J}h_j(W_j)\right)\left(X - \sum_{j=1}^{J}E(X|h_j(W_j))\right) + \Pi_n, \tag{A.10}$$

where

$$\Sigma_n = E\left(\left(X - \sum_{j=1}^{J}E(X|\widehat{h}_j(W_j))\right)X'\right),$$

$$\Delta_n = E\left(\sum_{j=1}^{J}(\widehat{h}_j - h_j)(W_j)\left(X - \sum_{k=1}^{J}E(X|\widehat{h}_k(W_k))\right)\right),$$

$$\Pi_n = n^{-1/2}\mathbb{G}_n\left[\left(Y - X'\widehat{\beta}_n - \sum_{j=1}^{J}\widehat{h}_j(W_j)\right)\left(X - \sum_{j=1}^{J}E(X|\widehat{h}_j(W_j))\right) - \left(Y - X'\beta_0 - \sum_{j=1}^{J}h_j(W_j)\right)\left(X - \sum_{j=1}^{J}E(X|h_j(W_j))\right)\right].$$

We first analyze the remaining term Π_n in the right-hand side of (A.10) by the use of Lemma A.2. Following similar logic in the proof of Proposition 2.1, we consider the class of functions

$$\bar{\mathscr{F}}_n \equiv \left\{[r_n(y - x'\beta) - G(w)]\left(x - \sum_{j=1}^{J}E(X|h_j(w_j))\right),\ \beta \in \mathscr{B} \text{ and } G \in \mathscr{Q}\right\},$$

where $r_n = (\log n)^{-1}$ and $\mathscr{Q}$ is defined above. By applying Lemma A.3 to the function $E(X|h_j(W_j))$ and considering the bracketing entropy number for $\mathscr{Q}$, we know the ε-bracketing entropy number for $\bar{\mathscr{F}}_n$ is of the order $1/\varepsilon$. By similar analysis of $|\widehat{H}_n(w_{(n)})|$ in the proof of Proposition 2.1, we can show that

$$\max_{1\leqslant j\leqslant J-1}\sup_{\leqslant w_j\leqslant 1}|\widehat{h}_j(w_j)| = \mathrm{O_P}(\log n)$$

since each h_j is assumed to be strictly increasing. Thus, we have $r_n\sum_{j=1}^J \widehat{h}_j(W_j) \in \mathscr{Q}$. Let $g_n = r_n(Y - X'\widehat{\beta}_n - \sum_{j=1}^J \widehat{h}_j(W_j))(X - \sum_{j=1}^J E(X|\widehat{h}_j(W_j)))$ and $g_0 = r_n(Y - X'\beta_0 - \sum_{j=1}^J h_j(W_j))(X - \sum_{j=1}^J E(X|h_j(W_j)))$. Note that $\|g_n - g_0\|_2 = \mathrm{O_P}(n^{-1/3})$ by (A.5) and (A.6). By applying Lemma A.2 to $\bar{\mathscr{F}}_n$ we can show that

$$\Pi_n = \mathrm{O_P}(n^{-2/3}\log n). \tag{A.11}$$

We next consider the terms Σ_n and Δ_n in the left-hand side of (A.10). We note that,

$$\begin{aligned}\|\Delta_n\| &= \left\|\sum_{j=1}^J E(\widehat{h}_j - h_j)(W_j)(E(X|W_j) - E(X|\widehat{h}_j(W_j)))\right\| \\ &\leqslant \left\|\sum_{j=1}^J |\widehat{h}_j - h_j|\right\|_2^2 \leqslant J\sum_{j=1}^J \|\widehat{h}_j - h_j\|_2^2 \\ &= \mathrm{O_P}(n^{-2/3}(\log n)^2) = \mathrm{o_P}(n^{-1/2})\end{aligned} \tag{A.12}$$

based on the pairwise independence of W, Assumption S3 and (A.8). As for the term Σ_n, we have

$$\begin{aligned}\Sigma_n - E\left(X - \sum_{j=1}^J E(X|W_j)\right)^{\otimes 2} &= E\left(\left(\sum_{j=1}^J E(X|W_j) - \sum_{j=1}^J E(X|\widehat{h}_j(W_j))\right)X'\right) \\ &\leqslant \left\|\sum_{j=1}^J |\widehat{h}_j - h_j|\right\|_2 \leqslant \sqrt{J}\sqrt{\sum_{j=1}^J \|\widehat{h}_j - h_j\|_2^2} \xrightarrow{\mathrm{p}} 0\end{aligned} \tag{A.13}$$

since $\widehat{h}_j$ is a consistent estimator by (A.8). By considering the above analysis about Σ_n, Δ_n and Π_n in (A.10), we can obtain (2.13) by applying the CLT to the first term in the right-hand side of (A.10). This completes the proof of (2.13).

Next, we prove (2.14). The proof (2.14) is adapted from that of Theorem 1 in Mammen and Yu (2007). Let $\tilde{h}_j$ be the standard isotonic estimate of h_j. Its asymptotic distribution is as follows:

For $w_j \in (-1, 1)$ with first derivative $\dot{h}_j(w_j) > 0$,

$$n^{1/3}\frac{(2p_{W_j}(w_j))^{1/3}}{\sigma^{2/3}\dot{h}_j(w_j)^{1/3}}[\tilde{h}_j(w_j) - h_j(w_j)]$$

converges in distribution to the slope of the greatest convex minorant of $W(t) + t^2$, where $W(t)$ is a two sided Brownian motion.

Thus, it suffices to show that

$$\widehat{h}_j(w_j) - \tilde{h}_j(w_j) = \mathrm{o_P}(n^{-1/3}) \tag{A.14}$$

for $w_j \in (-1, 1)$. First, we want to show that

$$\widehat{h}_j(w_j) = \tilde{h}_j(w_j) - \sum_{l\neq j}\int_0^1 [\widehat{h}_l(u_l) - h_l(u_l)]p_{W_l}(u_l)\,\mathrm{d}u_l + \mathrm{o_P}(n^{-1/3}) \tag{A.15}$$

for $n^{-1/3} - 1 \leqslant w_j \leqslant 1 - n^{-1/3}$. In order to prove (A.15), we need to define the following localized versions of $\widehat{h}_j$:

$$\widehat{h}_{j,\mathrm{loc}}(w_j) = \max_{w_j - e_n \leqslant u \leqslant w_j}\min_{w_j \leqslant v \leqslant w_j + e_n}\frac{\sum_{i:u\leqslant W_{ij}\leqslant v}\widehat{Y}_{ij}^s}{\mathrm{card}\{i : u \leqslant W_{ij} \leqslant v\}},$$

$$\widehat{h}^{+}_{j,\mathrm{loc}}(w_j) = \max_{w_j-e_n \leqslant u \leqslant w_j} \min_{w_j+d_n \leqslant v \leqslant w_j+e_n} \frac{\sum_{i:u \leqslant W_{ij} \leqslant v} \widehat{Y}^{s}_{ij}}{\mathrm{card}\{i : u \leqslant W_{ij} \leqslant v\}},$$

$$\widehat{h}^{-}_{j,\mathrm{loc}}(w_j) = \max_{w_j-e_n \leqslant u \leqslant w_j-d_n} \min_{w_j \leqslant v \leqslant w_j+e_n} \frac{\sum_{i:u \leqslant W_{ij} \leqslant v} \widehat{Y}^{s}_{ij}}{\mathrm{card}\{i : u \leqslant W_{ij} \leqslant v\}},$$

where $\widehat{Y}^{s}_{ij} = Y_i - X_i'\widehat{\beta}_n - \sum_{l\neq j}\widehat{h}_l(W_{il})$, $e_n = (\log n)^{1/\gamma} n^{-2/(9\gamma)} c_n$ with $c_n \to \infty$ slowly enough and $d_n = n^{-d}$ with $\frac{1}{3} < d < \frac{4}{9}$. The localized isotonic estimates, i.e. $\tilde{h}_{j,\mathrm{loc}}$, $\tilde{h}^{+}_{j,\mathrm{loc}}$ and $\tilde{h}^{-}_{j,\mathrm{loc}}$, can be defined similarly. These definitions directly imply that

$$\widehat{h}^{-}_{j,\mathrm{loc}}(w_j) \leqslant \widehat{h}_{j,\mathrm{loc}}(w_j) \leqslant \widehat{h}^{+}_{j,\mathrm{loc}}(w_j), \tag{A.16}$$

$$\tilde{h}^{-}_{j,\mathrm{loc}}(w_j) \leqslant \tilde{h}_{j,\mathrm{loc}}(w_j) \leqslant \tilde{h}^{+}_{j,\mathrm{loc}}(w_j) \tag{A.17}$$

for $n^{-1/3} - 1 \leqslant w_j \leqslant 1 - n^{-1/3}$. Recall that $\max_{1\leqslant j\leqslant J}\int_{-1}^{1}(\widehat{h}_j(w_j) - h_j(w_j))^2\,\mathrm{d}w_j = \mathrm{O_P}(n^{-2/3}(\log n)^2)$. Since the derivative of h_j is assumed to be bounded, it is easy to show

$$\max_{1\leqslant j\leqslant J}\left(\sup_{n^{-2/9}-1 \leqslant w_j \leqslant 1-n^{-2/9}} |\widehat{h}_j(w_j) - h_j(w_j)|\right) = \mathrm{O_P}((\log n)n^{-2/9}). \tag{A.18}$$

Combining (A.18) with Assumption S1, we have $\widehat{h}_j(w_j) = \widehat{h}_{j,\mathrm{loc}}(w_j)$ for $0 \leqslant w_j \leqslant 1$ and $j = 1, \ldots, J$ with probability tending to one. Therefore, we can conclude that

$$\widehat{h}^{-}_{j,\mathrm{loc}}(w_j) \leqslant \widehat{h}_j(w_j) \leqslant \widehat{h}^{+}_{j,\mathrm{loc}}(w_j) \tag{A.19}$$

for $n^{-1/3} - 1 \leqslant w_j \leqslant 1 - n^{-1/3}$ with probability tending to one. Similarly, we have

$$\tilde{h}^{-}_{j,\mathrm{loc}}(w_j) \leqslant \tilde{h}_j(w_j) \leqslant \tilde{h}^{+}_{j,\mathrm{loc}}(w_j) \tag{A.20}$$

for $n^{-1/3} - 1 \leqslant w_j \leqslant 1 - n^{-1/3}$ with probability tending to one. Therefore, it suffices to show (A.15) if we can show

$$\widehat{h}^{\pm}_{j,\mathrm{loc}}(w_j) = \tilde{h}^{\pm}_{j,\mathrm{loc}}(w_j) - \sum_{l\neq j}\int_{-1}^{1}[\widehat{h}_l(u_l) - h_l(u_l)]p_{W_l}(u_l)\,\mathrm{d}u_l + \mathrm{o_P}(n^{-1/3}), \tag{A.21}$$

$$\sup_{-1 \leqslant w_j \leqslant 1} \tilde{h}^{+}_{j,\mathrm{loc}}(w_j) - \tilde{h}^{-}_{j,\mathrm{loc}}(w_j) = \mathrm{o_P}(n^{-1/3}) \tag{A.22}$$

for $n^{-1/3} - 1 \leqslant w_j \leqslant 1 - n^{-1/3}$. Claims (A.22) can be established by the well known properties of the isotone least square estimate. In order to show (A.21) we first define the following notations:

$$\widetilde{G}_j(u_j, w_j) = n^{-1}\sum_{i:W_{ij}\leqslant u_j}(h_j(W_{ij}) + \varepsilon_i) - n^{-1}\sum_{i:W_{ij}\leqslant w_j}(h_j(W_{ij}) + \varepsilon_i),$$

$$\widehat{G}_j(u_j, w_j) = n^{-1}\sum_{i:W_{ij}\leqslant u_j}\widehat{Y}^{s}_{ij} - n^{-1}\sum_{i:W_{ij}\leqslant w_j}\widehat{Y}^{s}_{ij}.$$

By Lemma 4 in Mammen and Yu (2007), we have

$$\begin{aligned}\widehat{G}_j(u_j, w_j) - \widetilde{G}_j(u_j, w_j) = &-n^{-1}[\mathrm{card}\{i : W_{ij} \leqslant u_j\} - \mathrm{card}\{i : W_{ij} \leqslant w_j\}]\sum_{l\neq j}\int_{-1}^{1}(\widehat{h}_l - h_l)(u_l)p_{W_l}(u_l)\,\mathrm{d}u_l \\ &- n^{-1}\left(\sum_{i:W_{ij}\leqslant u_j} X_i - \sum_{i:W_{ij}\leqslant w_j} X_i\right)(\widehat{\beta}_n - \beta_0) + \mathrm{O_P}(r_n) + \mathrm{O_P}(s_n),\end{aligned} \tag{A.23}$$

where $r_n = (|u_j - w_j| + n^{-\alpha})^{2/3} n^{-13/27}(\log n)^{\beta}$ and $s_n = (|u_j - w_j| + n^{-1})n^{-2\rho/(9\gamma)}(\log n)^{\beta}$ for some $\alpha, \beta > 0$. For $w_j - e_n \leqslant u_j \leqslant w_j + e_n$, $\widehat{h}_{j,\mathrm{loc}}(w_j)$ and $\tilde{h}_{j,\mathrm{loc}}(w_j)$ are essentially the slopes of the greatest convex minorants of the functions, which map $\mathrm{card}\{i : W_{ij} \leqslant u_j\}$ onto $\widehat{G}_j(u_j, w_j)$ and $\widetilde{G}_j(u_j, w_j)$, at $u_j = w_j$, respectively. Thus, by the proper choice of d_n and e_n above and (A.23), we can obtain

$$\widehat{h}^{\pm}_{j,\mathrm{loc}}(w_j) = \tilde{h}^{\pm}_{j,\mathrm{loc}}(w_j) - \sum_{l\neq j}\int_{-1}^{1}[\widehat{h}_l(u_l) - h_l(u_l)]p_{W_l}(u_l)\,\mathrm{d}u_l - R^{\pm}_j(\widehat{\beta}_n - \beta_0) + \mathrm{o_P}(n^{-1/3}),$$

where

$$R_j^- = \max_{w_j-e_n \leqslant u \leqslant w_j-d_n} \min_{w_j \leqslant v \leqslant w_j+e_n} \frac{(\sum_{i:W_{ij} \leqslant v} - \sum_{i:W_{ij} \leqslant u})X_i}{\mathrm{card}\{i : W_{ij} \leqslant v\} - \mathrm{card}\{i : W_{ij} \leqslant u\}},$$

$$R_j^+ = \max_{w_j-e_n \leqslant u \leqslant w_j} \min_{w_j+d_n \leqslant v \leqslant w_j+e_n} \frac{(\sum_{i:W_{ij} \leqslant v} - \sum_{i:W_{ij} \leqslant u})X_i}{\mathrm{card}\{i : W_{ij} \leqslant v\} - \mathrm{card}\{i : W_{ij} \leqslant u\}}.$$

Since X is in some compact subset, we have $R_j^{\pm} = \mathrm{O_P}(1)$. Therefore, we have proved (A.21) by considering $\widehat{\beta}_n - \beta_0 = \mathrm{O_P}(n^{-1/2})$. This completes the proof of (A.15).

Next, we will prove $\widehat{h}_j(w_j) - \tilde{h}_j(w_j) = \mathrm{o_P}(n^{-1/3})$ for $w_j \in (-1, 1)$, i.e. (A.14), by rewriting (A.15) as follows:

$$\widehat{h} = \widetilde{h} + T(\widehat{h} - h) + \mathrm{o_P}(n^{-1/3}), \tag{A.24}$$

where h, $\widehat{h}$ and $\tilde{h}$ are tuples of functions h_j, $\widehat{h}_j$ and $\tilde{h}_j$, respectively. In (A.24), T is a linear integral operator defined as

$$\begin{aligned} T_{jl}(\widehat{h}_l - h_l) &= -\int_{-1}^{1} (\widehat{h}_l - h_l)(u_l) p_{W_l}(u_l)\,\mathrm{d}u_l \quad \text{for } l \neq j \\ &= 0 \quad \text{for } l = j. \end{aligned}$$

Obviously, (A.24) is the same as (22) in Mammen and Yu (2007). By similar analysis applied to (22) in Mammen and Yu (2007), we can show (2.14). □

References

Bacchetti, P., 1989. Additive isotonic models. J. Amer. Statist. Assoc. 84, 289–294.
Barlow, R.E., Bartholomew, D.J., Bremner, J.M., Brunk, H.D., 1972. Statistical Inference under Order Restrictions. Wiley, New York.
Brunk, H.D., 1970. Estimation of isotonic regression (with discussion). In: Puri, M.L. (Ed.), Nonparametric Techniques in Statistical Inference. Cambridge University Press, Cambridge.
Chamberlin, G., 1992. Efficiency bound for semiparametric regression. Econometrika 60, 567–596.
Dykstra, R., 1983. An algorithm for restricted least squares regression. J. Amer. Statist. Assoc. 78, 837–842.
Gao, F., Wellner, J., 2007. Entropy estimate for high dimensional monotonic functions. J. Multivariate Anal. 98, 1751–1764.
Hanson, D.L., Pledger, G., Wright, F.T., 1973. On consistency in monotonic regression. Ann. Statist. 3, 401–421.
Hastie, T., Tibshirani, R., 1986. Generalized additive models. Statist. Sci. 3, 297–310.
Horowitz, J.L., Mammen, E., 2005. Oracle-efficiency nonparametric estimation of an additive model with an unknown link function. Preprint.
Huang, J., 1999. Efficient estimation of the partly linear additive Cox model. Ann. Statist. 27, 1536–1563.
Huang, J., 2002. A note on estimating a partly linear model under monotonicity constraints. J. Statist. Plann. Inference 107, 345–351.
James, W.H., Rostron, J., 1985. Parental age, parity and sex ratio in births in England and Wales 1968–1977. J. Biosocial Sci. 17, 47–56.
Mammen, E., 1991. Estimating a smooth monotone regression function. Ann. Statist. 19, 724–740.
Mammen, E., Thomas-Agnan, C., 1999. Smoothing splines and shape restrictions. Scand. J. Statist. 26, 239–255.
Mammen, E., Yu, K., 2007. Additive isotone regression. Lect. Notes—Monogr. Ser. 55, 179–195.
Morton-Jones, T., Diggle, P., Parker, L., Dickinson, H.O., Binks, K., 2000. Additive isotonic regression models in epidemiology. Statist. Med. 19, 849–859.
Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized linear models. J. Roy. Statist. Soc. Ser. A (General) 135, 370–384.
Newey, W.K., 1997. Convergence rates and asymptotic normality for series estimators. J. Econometrics 79, 147–168.
Pal, J.K., Banerjee, M., 2007. Estimation of smooth link functions in monotone response models. Technical Report.
Robertson, T., Wright, F.T., Dykstra, R.L., 1988. Order Restricted Statistical Inference. Wiley, New York.
Robinson, P.M., 1988. Root-N-consistent semiparametric regression. Econometrica 56, 931–954.
Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, New York.
Stone, C.J., 1977. Consistent nonparametric regression. Ann. Statist. 5, 595–645.
Stone, C.J., 1985. Additive regression and other nonparametric models. Ann. Statist. 13, 689–705.
Van de Geer, S., 2000. Empirical Processes in M-estimation. Cambridge University Press, Cambridge.
Van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes: With Applications to Statistics. Springer, New York.