Identifiability of single-index models and additive-index models

By WEI LIN

Department of Mathematics, Ohio University

Athens, Ohio 45701, U.S.A.

linwei@math.ohiou.edu

K. B. KULASEKERA

Department of Mathematical Sciences, Clemson University

Clemson, South Carolina 29631, U.S.A.

kk@clemson.edu

Summary

We provide a proof for the identifiability for both single-index models and partially linear single-index models assuming only the continuity of the regression function, a condition much weaker than the differentiability conditions assumed in the existing literature. Our discussion is then extended to the identifiability of the additive-index models.

Some key words: Additive multiple-index model; Identifiability; Partially linear single-index model; Single-index model.

1. Introduction

Suppose data (Y_i, X_i) follow the regression model

$$Y_i = m(X_i) + \epsilon_i, \quad i = 1, \cdots, n, \tag{1}$$

for some function $m(\cdot)$ where the covariate X is a p-dimensional vector and $m(\cdot)$ is unknown. We say that the model follows a single-index model if

$$m(x) = g(\alpha' x), \tag{2}$$

where $g(\cdot)$ is an unknown univariate smooth function and $\alpha \in \mathbb{R}^p$ is an unknown vector. For identifiability purposes, we typically assume that $\|\alpha\| = 1$ with its first nonzero element being positive. The model is said to be a partially linear single-index model if the mean function includes an extra linear component $\theta' x$, i.e.

$$m(x) = \theta' x + g(\alpha' x). \tag{3}$$

To ensure identifiability, $\theta \in \mathbb{R}^p$ is assumed to be perpendicular to α because otherwise we can take $\tilde{\theta} = \theta - (\theta'\alpha)\alpha$ and thus $m(x) = \tilde{\theta}' x + \tilde{g}(\alpha' x)$, where $\tilde{\theta} \perp \alpha$ and $\tilde{g}(\alpha' x) = (\theta'\alpha)\alpha' x + g(\alpha' x)$. These two classes of models have been extensively investigated because of their capacity for dimension reduction (Powell et al., 1989; Duan & Li, 1991; Ichimura, 1993; Härdle et al., 1993; Xia et al., 1999; Hristache & Spokoiny, 2001; Yu & Ruppert, 2002; Stute & Zhu, 2005).

A generalisation of the above two classes of models is provided by the additive-index models (Chiou & Müller, 2004), where

$$m(x) = \sum_{k=1}^{K} g_k(\theta_k' x). \tag{4}$$

One special case of model (4) has mean function

$$m(x) = g_1(\alpha' x_1) + g_2(\beta' x_2). \tag{5}$$

Here $x' = (x'_1, x'_2)'$. This is a special case of model (4) with $K = 2$, $\theta'_1 = (\alpha', 0')'$ and $\theta'_2 = (0', \beta')'$. This model is called the additive single-index model by Naik & Tsai (2001).

Most results in the literature concentrate on the estimation of the index vector(s) and the regression function. It is important to make sure that the representations in (2),(3), (4) and (5) are unique. Ichimura (1993) gives a proof for the identifiability of model (2) assuming that $g(\cdot)$ is differentiable and Xia et al. (1999) prove the identifiability of model (3) assuming that $g(\cdot)$ is twice differentiable. Chiou & Müller (2004) give the identifiability of model (4) under conditions that are not only very restrictive, (for example the component functions have to be monotone), but also very difficult to verify. Their method is as follows. Two functions $f(x)$ and $g(x)$ cannot be identical if we can find two points x_1 and x_2 such that $f(x_1) = f(x_2)$ while $g(x_1) \neq g(x_2)$. Chiou & Müller (2004) make an assumption, M6 in their paper, that essentially states the existence of such two points. There is no direct proof of identifiability for model (5) except that it is a special case of model (4).

In this paper we prove the identifiability for both single-index models (2) and partially linear single-index models (3) under a much weaker condition, just the continuity of $g(\cdot)$, followed by a proof for the identifiability of the additive single-index models (5). In addition, we provide a necessary condition for identifiability in an additive-index model (4).

It is noteworthy that the identifiability discussed in this paper is different from the uniqueness of the regression function $m(\cdot)$ itself. Xia et al. (1999) impose an extra requirement, which we call Assumption 0, for the identifiability of model (3).

Assumption 0. The covariate vector X has a positive density function on an open convex set $\mathcal{A} \subset \mathbb{R}^p$.;

This is actually a redundant condition provided the support S of $m(\cdot)$ is nondegenerate; see Assumption 1 in § 2.1. They give a counterexample when Assumption 0 does not hold. Suppose $y = 2x_1x_2$ and $x = (x_1, x_2, x_3, x_4)$, where the covariates are related by $x_3 = x_1^2$ and $x_4 = x_2^2$ so that the support S is degenerate. Then the regression function has two representations,

$$m(x) = (-x_3 - x_4) + (x_1 + x_2)^2, \quad m(x) = (x_3 + x_4) - (x_1 - x_2)^2.$$

In this case we have multiple expressions for the $m(\cdot)$ function but this does not disprove the identifiability of the partially linear single-index models. It is a uniqueness issue of $m(\cdot)$ itself and thus is beyond the scope of our discussion. This is somewhat similar to the case in classical linear regression analysis where a special method should be applied in the presence of severe multicollinearity.

2. Identifiability

2·1. *Single-index models*

Suppose we have data (X_i, Y_i) coming from model (1), where the mean function $m(\cdot)$ is a p-variate smooth function, the errors are independent with zero mean and finite common variance σ^2, and the following assumption holds.

Assumption 1. The support S of $m(\cdot)$ is a bounded convex set with at least one interior point.

Definition 1. *A p-variate function $m(\cdot) \in L_2(S)$ is said to be a partially linear single-index function with index vectors (θ, α) if $m(x) = \theta' x + g(\alpha' x)$ almost everywhere for some nonlinear function g, $\theta \in \mathbb{R}^p$, $\theta \perp \alpha$ and $\alpha \in D$ where*

$$D = \left\{ \alpha \in \mathbb{R}^p \mid \|\alpha\| = 1 \text{ with first nonzero element positive} \right\}.$$

The function is said to be a single-index function if $\theta = 0$.

With this definition we can define a model to be a single-index model if the regression function is a linear single-index function and to be a partially linear single-index model if the regression function is a partially linear single-index function.

THEOREM 1. *Suppose that Assumption 1 holds and* $m(\cdot)$ *is a nonconstant continuous function on* S. *If*

$$m(x) = g(\alpha' x) = h(\beta' x), \quad \textit{for all } x \in S, \tag{6}$$

for some continuous functions g *and* h, *and some* $\alpha, \beta \in D$, *then* $\alpha = \beta$ *and* $g \equiv h$ *on* $\{\alpha' x | x \in S\}$.

Proof. It is clear that $g \equiv h$ if $\alpha = \beta$. Hence it suffices to show $\alpha = \beta$.

Suppose $\alpha \neq \beta$. Since $m(\cdot)$ is continuous and nonconstant on S, there exists a sphere $B = B(x_0, r) \subset S$ for some x_0 such that $m(\cdot)$ is non-constant on B. By (6) and the fact that $\alpha'\alpha = 1$ we have, for all $t \in (-r, r)$, that $x_0 + t\alpha \in S$,

$$g(\alpha' x_0 + t) = g\big\{\alpha'(x_0 + t\alpha)\big\} = h\big\{\beta'(x_0 + t\alpha)\big\} = h\big\{\beta' x_0 + t(\beta'\alpha)\big\}$$

and

$$h(\beta' x_0 + t) = h\big\{\beta'(x_0 + t\beta)\big\} = g\big\{\alpha'(x_0 + t\beta)\big\} = g\big\{\alpha' x_0 + t(\beta'\alpha)\big\}.$$

By the requirement that the first nonzero components of α and β are positive, we have $\alpha \neq -\beta$. Hence $|\alpha'\beta| < 1$ and, by the continuity of g,

$$\begin{aligned} g(\alpha' x_0 + t) &= h\big\{\beta' x_0 + t(\beta'\alpha)\big\} = g\big\{\alpha' x_0 + t(\beta'\alpha)^2\big\} \\ &= \cdots = g\big\{\alpha' x_0 + t(\beta'\alpha)^{2n}\big\} = \cdots = g(\alpha' x_0), \quad \textit{for all } t \in (-r, r). \end{aligned} \tag{7}$$

Pick any $x \in B(x_0, r)$. Then $x = x_0 + t\gamma$ for some unit vector γ and $t \in (-r, r)$. Hence, by (7), $m(x) = g(\alpha' x) = g(\alpha' x_0 + t\alpha'\gamma) = g(\alpha' x_0)$. This indicates that $m(\cdot)$ is constant on B, which is a contradiction. Thus $\alpha = \beta$ and the proof is completed. □

2·2. *Partially linear single-index models*

THEOREM 2. *Let Assumption 1 hold. Suppose that $m(\cdot)$ is a nonconstant continuous function on S and that*

$$m(x) = \theta_1' x + g(\alpha_1' x) = \theta_2' x + h(\alpha_2' x), \quad \textit{for all } x \in S, \tag{8}$$

for some continuous nonlinear functions g and h, $\theta_k \in \mathbb{R}^p$ and $\alpha_k \in D$ with $\theta_k \perp \alpha_k$, $k = 1, 2$. Then $\theta_1 = \theta_2$, $\alpha_1 = \alpha_2$ and $g = h$.

Proof. Without loss of generality we may assume that the origin is an interior point of S; otherwise use the transformation $z = x - x_0$, where x_0 is any interior point of S. Let $\theta = \theta_1 - \theta_2$ and a be some generic constant.

First we consider $p = 2$. If $\alpha_1 \neq \alpha_2$, then (α_1, α_2) form a basis for $\mathbb{R}^2$; writing θ_1 and θ_2 in terms of the α_k's, we obtain from Theorem 1 that $\alpha_1 = \alpha_2$, which is a contradiction. Hence $\alpha_1 = \alpha_2 = \alpha$ and $\theta \perp \alpha$. Then (8) gives that $f(\alpha' x) = \theta' x$, where $f = h - g$. Since the origin is an interior point of S, $x = t\theta \in S$ and therefore $f(0) = t\|\theta\|^2$ for sufficiently small t. Thus, $\theta = 0$, i.e. $\theta_1 = \theta_2$, and the problem is reduced to the case of Theorem 1.

For $p > 2$, we consider the following four cases.

Case 1. Suppose that $\alpha_1 = a\theta$ for some constant a. Then, by (8),

$$\tilde{m}(x) = m(x) - \theta_2' x = \theta' x + g(a\theta' x) = h(\alpha_2' x), \quad \textit{for all } x \in S.$$

Since $g(\cdot)$ is nonlinear, $\tilde{m}(\cdot)$ cannot be constant on S. By Theorem 1 we have that $\alpha_2 = b\theta$ for some constant b. Since $\alpha_1 \in D$ and $\alpha_2 \in D$, we have that

$\alpha_1 = \alpha_2 = a\theta$. Note that $\theta_k \perp \alpha_k$, $k = 1, 2$. Hence $\theta \perp \theta$, which implies that $\theta_1 = \theta_2$ and thus $g = h$.

Case 2. Suppose that $\alpha_2 = a\theta$ for some constant a. This is identical to Case 1 above.

Case 3. Suppose that $(\alpha_1 - \alpha_2) = (\theta_1 - \theta_2)/a$ for some constant a. Then

$$\tilde{m}(x) = m(x) - \theta_2' x + a\alpha_2' x = a\alpha_1' x + g(\alpha_1' x) = h(\alpha_2' x) + a\alpha_2' x.$$

The result now follows by the same argument as for Case 1.

Case 4. Suppose none of the above cases holds, and let $\theta = \theta_1 - \theta_2$. Equation (8) becomes

$$\theta' x = h(\alpha_2' x) - g(\alpha_1' x). \tag{9}$$

Since $p > 2$ and in this case α_k, $k = 1, 2$, cannot be parallel to θ or $\alpha_1 - \alpha_2$, there exists $x_0 \in S$ such that

$$x_0 \perp \operatorname{span}(\theta, \alpha_1 - \alpha_2) \quad \text{and} \quad u = \alpha_1' x_0 = \alpha_2' x_0 > 0.$$

Plugging $x = tx_0$, for all $t \in [-u, u]$, into (9) we obtain

$$g(t) = h(t), \quad \textit{for all } t \in [-u, u]. \tag{10}$$

Suppose that $\theta \neq 0$. Since in this case θ is not parallel to $\alpha_1 - \alpha_2$, there exists $x_1 \in S$ such that $\theta' x_1 \neq 0$ and

$$x_1 \perp (\alpha_1 - \alpha_2), \quad v = \alpha_1' x_1 = \alpha_2' x_1 > 0.$$

Plugging $x = tx_1$, for all $t \in [-1, 1]$, into (9) we obtain

$$t\theta' x_1 = h(tv) - g(tv), \quad \forall\, t \in [-1, 1].$$

Take any $t \in (0, \frac{u}{v})$. By (10) we have that $t\theta' x_1 = h(tv) - g(tv) = 0$, which contradicts the fact that $\theta' x_1 \neq 0$. Thus $\theta = 0$, i.e. $\theta_1 = \theta_2$, and the result follows by Theorem 1.

$\square$

The following corollary extends the above results to domains with certain types of multiple constraints on the covariates.

COROLLARY 1. *The identifiability results in Theorem 1 and Theorem 2 hold when the support S of $m(\cdot)$ can be partitioned into convex sets, each of which contains at least one interior point and $m(\cdot)$ is nonconstant on each convex set.*

Proof. This is a direct consequence of Theorem 1 and Theorem 2. □

2·3. *Additive single-index models*

In this subsection we discuss the identifiability of the model in which

$$m(x) = \sum_{k=1}^{K} g_k(\alpha_k' x_k). \tag{11}$$

Here the covariate x is divided into K parts, i.e., $x' = (x_1', \cdots, x_K')'$. We make the following assumptions.

Assumption 2. Each α_k vector has norm one and the first nonzero component of α_k is positive, $k = 1, \cdots, K$.

Assumption 3. The support of $g_k(\cdot)$ is S_k, $k = 1, \cdots, K$, and the support of $m(\cdot)$ is $S = S_1 \times \cdots \times S_K$.

This model is useful for the situations where the covariate x can be divided into several groups and no interaction effects exist among these groups. Yu & Ruppert (2002) discussed partially linear single-index models of the form

$$m(x) = \alpha' x_1 + g(\beta' x_2),$$

which is a special case of both model (3) and model (11).

Since $m(x)$ is nonconstant on S, there exists at least one component vector, without loss of generality x_K, say, such that there exists a vector a with $x_K = a$ and $m(x)$ is nonconstant on $S_a = \{x \in S \mid x_K = a\}$. Let $z' = (x'_1, \cdots, x'_{K-1})'$ and

$$\tilde{m}(z) = \sum_{i=1}^{K-1} g_i(\alpha'_i x_i).$$

Hence, for $x \in S_a$, $m(x) = m(z, a) = \tilde{m}(z) + g_K(a)$. Since $m(x)$ is nonconstant on S_a, this implies that $\tilde{m}(z)$ is nonconstant on $\tilde{S} = S_1 \times \cdots \times S_{K-1}$. Thus simple induction gives that the first $K-1$ g_k-functions and α_k-vectors are unique. This proves the following theorem.

THEOREM 3. *Suppose that $m(\cdot)$ is a nonconstant continuous function on S and that*

$$m(x) = \sum_{k=1}^{K} g_k(\alpha'_k x_k),$$

for some continuous functions g_k, $k = 1, \cdots, K$. Let Assumptions 1-3 hold. Then the α_k vectors are unique and the $g_k(\cdot)$ functions are unique up to a constant.

2·4. *Additive-index models*

A more general index model has the regression function taking the form

$$m(x) = \sum_{k=1}^{K} g_k(\theta'_k x) + C_0, \tag{12}$$

where $C_0 = m(0)$ and each function $g_k(\cdot)$ is univariate and smooth with $g_k(0) = 0$. These conditions are needed for the identifiability of the model when $K > 1$. These prevent non-identifiability due to shifts. The constant C_0 can be taken as zero in the Single-index model. We now give a necessary condition for the subclass of quadratic models

$$m(x) = \sum_{k=1}^{K} (\theta'_k x)^2 + C_0 \tag{13}$$

followed by its implication for (12). To this end, suppose that we have

$$m(x) = \sum_{k=1}^{K} (\alpha_k' x)^2 = \sum_{k=1}^{K} (\beta_k' x)^2, \tag{14}$$

where $\{\alpha_k, \beta_k\}$ are vectors in $\mathbb{R}^p$. Note that we do not require them to be unit vectors because the functions $g_k(t) = t^2$ are fixed. These vectors can be standardised and the norms can be absorbed into the g_k functions. Since the support of $m(\cdot)$ is non-degenerate, equality holds in (14) if and only if

$$\sum_{k=1}^{K} \alpha_k \alpha_k' = \sum_{k=1}^{K} \beta_k \beta_k'. \tag{15}$$

If all $\{\alpha_k, \beta_k\}$ are free, we may set

$$\sum_{k=1}^{K} \alpha_k \alpha_k' = A, \tag{16}$$

where A is a symmetric matrix. The identifiability of (12) cannot be achieved if (16) has infinite number of solutions. Since A is symmetric, there are $p(p+1)/2$ equations in (16) and there are Kp variables. Thus, to ensure a finite number of solutions, we require that

$$\frac{p(p+1)}{2} \geq Kp,$$

i.e. $p \geq 2K - 1$. However, this is not quite enough since the β_k's could depend on the α_k's. The maximal dependence occurs when

$$\beta_i = \sum_{k=1}^{K} a_{ik} \alpha_k, \quad i = 1, \cdots, K.$$

Therefore, (15) produces $\frac{1}{2}p(p+1)$ equations and $Kp + K^2$ variables. To ensure identifiability we require that

$$\frac{p(p+1)}{2} \geq Kp + K^2,$$

or, equivalently,

$$p \geq \frac{2K-1}{2} + \frac{1}{2}\sqrt{(2K-1)^2 + 8K^2}. \tag{17}$$

For example, when $K = 2$, $p \geq 5$ is necessary for the identifiability of any function in the quadratic subclass (13) unless extra conditions on $m(x)$ are imposed.

The above also indicates that, if one were to determine a sufficient condition for the identifiability of the additive-index models (12), including the quadratic subclass, then one needs (17) to be satisfied for all the functions in (13). Since the quadratic functions in (13) are infinitely differentiable, placing any smoothness condition on $m(x)$ does not help exclude the quadratic subclass. One possible way of relaxing the lower bound on p given by (17) is to assume orthogonality of the index vectors, as in the case of the additive single-index models discussed in § 2.3.

References

Chiou, J. & Müller, H. (2004). Quasi-likelihood regression with multiple indices and smooth link and variance functions. *Scand. J. Statist.* **31**, 367–86.

Duan, N. & Li, K.C. (1991). Slicing regression: A link free regression method. *Ann. Statist.* **19**, 505–30.

Härdle, W., Hall, P. & Ichimura, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.* **21**, 157–78.

Hristache, M., Juditsky, A. & Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-index model. *Ann. Statist.* **29**, 595–23.

Ichimura, H. (1993). Semiparametric least squares (sls) and weighted sls estimation of single-index models. *J. of Economet.* **58**, 71–120.

Naik, P.A. & Tsai, C. (2001). Single-index model selection. *Biometrika* **88**, 821–32.

Powell, J.L., Stock, J.H. & Stoker, T.M. (1989). Semiparametric estimation of index coefficients. *Econometrica* **57**, 1403–30.

Stute, W. & Zhu, L.X. (2005). Nonparametric checks for single-index models. *Ann. Statist.* **33**, 1048–83.

Xia, Y.C., Tong, H. & Li, W.K. (1999). On extended partially linear single-index models. *Biometrika* **86**, 831–42.

Yu, Y. & Ruppert, D. (2002). Penalized spline estimation for partially linear single-index models. *J. Am. Statist. Assoc.* **97**, 1042–54.