Profile Least Squares Estimators in the Monotone Single Index Model

Fadoua Balabdaoui and Piet Groeneboom

Abstract We consider least squares estimators of the finite regression parameter $\boldsymbol{\alpha}$ in the single index regression model $Y = \psi(\boldsymbol{\alpha}^T \boldsymbol{X}) + \varepsilon$, where $\boldsymbol{X}$ is a d-dimensional random vector, $\mathbb{E}(Y|\boldsymbol{X}) = \psi(\boldsymbol{\alpha}^T \boldsymbol{X})$, and ψ is a monotone. It has been suggested to estimate $\boldsymbol{\alpha}$ by a profile least squares estimator, minimizing $\sum_{i=1}^n (Y_i - \psi(\boldsymbol{\alpha}^T \boldsymbol{X}_i))^2$ over monotone ψ and $\boldsymbol{\alpha}$ on the boundary $\mathcal{S}_{d-1}$ of the unit ball. Although this suggestion has been around for a long time, it is still unknown whether the estimate is $\sqrt{n}$-convergent. We show that a profile least squares estimator, using the same pointwise least squares estimator for fixed $\boldsymbol{\alpha}$, but using a different global sum of squares, is $\sqrt{n}$-convergent and asymptotically normal. The difference between the corresponding loss functions is studied and also a comparison with other methods is given.

1 Introduction

The monotone single index model tries to predict a response from the linear combination of a finite number of parameters and a function linking this linear combination to the response via a monotone *link function* ψ_0 which is unknown. So, more formally, we have the model

$$Y = \psi_0(\boldsymbol{\alpha}_0^T \boldsymbol{X}) + \varepsilon,$$

where Y is a one-dimensional random variable, $\boldsymbol{X} = (X_1, \ldots, X_d)^T$ is a d-dimensional random vector with distribution function G, ψ_0 is monotone, and ε is a one-dimensional random variable such that $\mathbb{E}[\varepsilon|\boldsymbol{X}] = 0$ G almost surely. For identifiability, the regression parameter $\boldsymbol{\alpha}_0$ is a vector of norm $\|\boldsymbol{\alpha}_0\|_2 = 1$, where

F. Balabdaoui (✉)
Seminar für Statistik ETH, Zürich, Switzerland
e-mail: fadoua.balabdaoui@stat.math.ethz.ch

P. Groeneboom
Delft University of Technology, DIAM, Mekelweg 4, 2628 CD Delft, The Netherlands
e-mail: P.Groeneboom@tudelft.nl

A. Daouia and A. Ruiz-Gazen (eds.), *Advances in Contemporary Statistics and Econometrics*, https://doi.org/10.1007/978-3-030-73249-3_1

$\|\cdot\|_2$ denotes the Euclidean norm in $\mathbb{R}^d$, so $\boldsymbol{\alpha}_0 \in \mathcal{S}_{d-1}$, the unit $(d-1)$-dimensional sphere.

The ordinary profile least squares estimate of $\boldsymbol{\alpha}_0$ is an M-estimate in two senses: for fixed $\boldsymbol{\alpha}$, the least squares criterion

$$\psi \mapsto n^{-1}\sum_{i=1}^{n}\left\{Y_i - \psi(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\right\}^2 \tag{1}$$

is minimized for all monotone functions ψ (either decreasing or increasing) which gives an $\boldsymbol{\alpha}$-dependent function $\hat{\psi}_{n,\boldsymbol{\alpha}}$, and the function

$$\boldsymbol{\alpha} \mapsto n^{-1}\sum_{i=1}^{n}\left\{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\right\}^2 \tag{2}$$

is then minimized over $\boldsymbol{\alpha}$. This gives a profile least squares estimator $\hat{\boldsymbol{\alpha}}_n$ of $\boldsymbol{\alpha}_0$, which we will call LSE in the sequel. Although this estimate of $\boldsymbol{\alpha}_0$ has been known now for a very long time (more than 30 years probably), it is not known whether it is $\sqrt{n}$-convergent (under appropriate regularity conditions), let alone that we know its asymptotic distribution. Also, simulation studies are rather inconclusive. For example, it is conjectured in Tanaka (2008) on the basis of simulations that the rate of convergence of $\hat{\boldsymbol{\alpha}}_n$ is $n^{9/20}$. Other simulation studies, presented in Balabdaoui et al. (2019a), are also inconclusive. In that paper, it was also proved that an ordinary least squares estimator (which ignores that the link function could be non-linear) is $\sqrt{n}$-convergent and asymptotically normal under elliptic symmetry of the distribution of the covariate $\boldsymbol{X}$. Another linear least squares estimator of this type, where the restriction on $\boldsymbol{\alpha}$ is $\boldsymbol{\alpha}^T\boldsymbol{S}_n\boldsymbol{\alpha} = 1$, S_n is the usual estimate of the covariance matrix of the covariates, and a renormalization at the end is not needed (as it is in the just mentioned linear least squares estimator) was studied in Balabdaoui et al. (2019b) and was shown to have similar behavior. If this suggests that the profile LSE should also be $\sqrt{n}$-consistent, the extended simulation study in Balabdaoui et al. (2019b) shows that it is possible to find other estimates which exhibit better performance in these circumstances.

An alternative way to estimate the regression vector is to minimize the criterion

$$\boldsymbol{\alpha} \mapsto \left\| n^{-1}\sum_{i=1}^{n}\left\{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\right\}\boldsymbol{X}_i \right\|^2 \tag{3}$$

over $\boldsymbol{\alpha} \in \mathcal{S}_{d-1}$, where $\|\cdot\|$ is the Euclidean norm. Note that this is the sum of d squares. The rational behind minimizing (3) is the fact that the true index vector, $\boldsymbol{\alpha}_0$, satisfies the (population) score equation

$$\mathbb{E}\left\{(Y - \psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X}))\boldsymbol{X}\theta(\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\} = \boldsymbol{0}, \tag{4}$$

where θ is any measurable and bounded function. This clearly follows from the iterative law of expectations and the fact that $\mathbb{E}\{Y|\boldsymbol{\alpha}_0^T\boldsymbol{X}\} = \psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})$. If the function θ is taken to be the constant 1, then the goal is to find the minimizer of the Euclidean norm of the empirical counterpart of the above score equation, after replacing the unknown link function, ψ_0, by its estimator $\hat{\psi}_{n,\boldsymbol{\alpha}}$.

We prove in Sect. 3 that this minimization procedure leads to a $\sqrt{n}$-consistent and asymptotically normal estimator, which is a more precise and informative result compared to what we know now about the LSE.. Using the well-known properties of isotonic estimators, it is easily seen that the function (3) is piecewise constant as a function of $\boldsymbol{\alpha}$, with finitely many values, so the minimum exists and is equal to the infimum over $\boldsymbol{\alpha} \in \mathcal{S}_{d-1}$. Notice that this estimator does not use any tuning parameters, just like the LSE.

In Balabdaoui et al. (2019b), a similar Simple Score Estimator (SSE) $\hat{\boldsymbol{\alpha}}_n$ was defined as a point $\boldsymbol{\alpha} \in \mathcal{S}_{d-1}$ where all components of the function

$$\boldsymbol{\alpha} \mapsto n^{-1}\sum_{i=1}^{n}\left\{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\right\}\boldsymbol{X}_i$$

cross zero. If the criterion function were continuous in $\boldsymbol{\alpha}$, this estimator would have been the same as the least squares estimator, minimizing (3), with a minimum equal to zero, but in the present case we cannot assume this because of the discontinuities of the criterion function.

The definition of an estimator as a crossing of the d-dimensional vector $\mathbf{0}$ makes it necessary to prove the existence of such an estimator, which we found to be a rather non-trivial task. Defining our estimator directly as the minimizer of (3), so as a least squares estimator, relieves us from the duty to prove its existence. Since our estimator has the same limit distribution as the SSE, we refer to it here under the same name.

A fundamental function in our treatment is the function $\psi_{\boldsymbol{\alpha}}$, defined as follows.

Definition 1 Let $\mathcal{S}_{d-1}$ denote again the boundary of the unit ball in $\mathbb{R}^d$. Then, for each $\boldsymbol{\alpha} \in \mathcal{S}_{d-1}$, the function $\psi_{\boldsymbol{\alpha}} : \mathbb{R} \to \mathbb{R}$ is defined as the nondecreasing function which minimizes

$$\psi \mapsto \mathbb{E}\{Y - \psi(\boldsymbol{\alpha}^T\boldsymbol{X})\}^2$$

over all nondecreasing functions $\psi : \mathbb{R} \to \mathbb{R}$. The existence and uniqueness of the function $\psi_{\boldsymbol{\alpha}}$ follows, for example, from the results in Landers and Rogge (1981).

The function $\psi_{\boldsymbol{\alpha}}$ coincides in a neighborhood of $\boldsymbol{\alpha}_0$ with the ordinary conditional expectation function $\tilde{\psi}_{\boldsymbol{\alpha}}$

$$\tilde{\psi}_{\boldsymbol{\alpha}}(u) = \mathbb{E}\left\{\psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})|\boldsymbol{\alpha}^T\boldsymbol{X} = u\right\}, \qquad u \in \mathbb{R}; \tag{5}$$

see Balabdaoui et al. (2019b), Proposition 1. The general definition of $\psi_{\boldsymbol{\alpha}}$ uses conditioning on a σ-lattice, and $\psi_{\boldsymbol{\alpha}}$ is also called a *conditional* 2-*mean* (see Landers and Rogge 1981).

The importance of the function $\psi_{\boldsymbol{\alpha}}$ arises from the fact that we can differentiate this function w.r.t. $\boldsymbol{\alpha}$, in contrast with the least squares estimate $\hat{\psi}_{n,\boldsymbol{\alpha}}$, and that $\psi_{\boldsymbol{\alpha}}$ represents the least squares estimate of ψ_0 in the underlying model for fixed $\boldsymbol{\alpha}$, if we use $\boldsymbol{\alpha}^T\boldsymbol{x}$ as the argument of the monotone link function.

It is also possible to introduce a tuning parameter and use an estimate of $\frac{d}{du}\psi_{\boldsymbol{\alpha}}(u)\big|_{u=\boldsymbol{\alpha}^T\boldsymbol{X}}$. This estimate is defined by

$$\tilde{\psi}'_{n,h,\boldsymbol{\alpha}}(u) = \frac{1}{h}\int K\left(\frac{u-x}{h}\right)\,d\hat{\psi}_{n,\boldsymbol{\alpha}}(x), \tag{6}$$

where K is one of the usual kernels, symmetric around zero and with support $[-1, 1]$, and h is a bandwidth of order $n^{-1/7}$ for sample size n. For fixed $\boldsymbol{\alpha}$, the least squares estimate $\hat{\psi}_{n,\boldsymbol{\alpha}}$ is defined in the same way as above. Note that this estimate is rather different from the derivative of a Nadaraya-Watson estimate which is also used in this context and is in fact the derivative of a ratio of two kernel estimates. If we use the Nadaraya-Watson estimate, we need in principle two tuning parameters, one for the estimation of ψ_0 and another one for the estimation of the derivative ψ_0'.

Using the estimate (6) of the derivative, we now minimize

$$\boldsymbol{\alpha} \mapsto \left\| n^{-1}\sum_{i=1}^{n}\left\{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\right\}\boldsymbol{X}_i\,\tilde{\psi}'_{n,h,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\right\|^2 \tag{7}$$

instead of (3), where $\|\cdot\|$ is again the Euclidean norm. The motivation for considering such a minimization problem is very similar to the one given above for the SSE. The only difference now is that the current approach allows us to take the function θ to be equal to the derivative of ψ_0', which is replaced in the empirical version of the population score in (4) by its estimator $\tilde{\psi}'_{n,h,\boldsymbol{\alpha}}$. A variant of this estimator was defined in Balabdaoui et al. (2019b) and called the Efficient Score Estimator (ESE) there, since, if the conditional variance $\text{var}(Y|\boldsymbol{X}=\boldsymbol{x}) = \sigma^2$, where σ^2 is independent of the covariate $\boldsymbol{X}$ (the homoscedastic model), the estimate is efficient. As in the case of the simple score estimator (SSE), the estimate was defined as a crossing of zero estimate in Balabdaoui et al. (2019b) and not as a minimizer of (7). But the definition as a minimizer of (7) produces an estimator that has the same limit distribution.

The qualification "efficient" is somewhat dubious, since the estimator is no longer efficient if we do not have homoscedasticity. We give an example of that situation in Sect. 5, where, in fact, the SSE has a smaller asymptotic variance than the ESE. Nevertheless, to be consistent with our treatment in Balabdaoui et al. (2019b) we will call the estimate, $\hat{\boldsymbol{\alpha}}_n$, minimizing (7), again the ESE.

Dropping the monotonicity constraint, we can also use as our estimator of the link function a cubic spline $\hat{\psi}_{n,\boldsymbol{\alpha}}$, which is defined as the function minimizing

$$\sum_{i=1}^{n} \left\{\psi(\boldsymbol{\alpha}^T \boldsymbol{X}_i) - Y_i\right\}^2 + \mu \int_a^b \psi''(x)^2 \, dx, \tag{8}$$

over the class of functions $\mathcal{S}_2[a, b]$ of differentiable functions ψ with an absolutely continuous first derivative, where

$$a = \min_i \boldsymbol{\alpha}^T \boldsymbol{X}_i, \qquad b = \max_i \boldsymbol{\alpha}^T \boldsymbol{X}_i,$$

see Green and Silverman (1994), pp. 18 and 19, where $\mu > 0$ is the penalty parameter. Using these estimators of the link function, the estimate $\hat{\boldsymbol{\alpha}}_n$ of $\boldsymbol{\alpha}_0$ is then found in Kuchibhotla and Patra (2020) by using a $(d-1)$-dimensional parameterization β and a transformation $S : \boldsymbol{\beta} \mapsto S(\boldsymbol{\beta}) = \boldsymbol{\alpha}$, where $S(\boldsymbol{\beta})$ belongs to the surface of the unit sphere in $\mathbb{R}^d$, and minimizing the criterion

$$\boldsymbol{\beta} \mapsto \sum_{i=1}^{n} \{Y_i - \hat{\psi}_{S(\boldsymbol{\beta}),\mu}(S(\boldsymbol{\beta})^T \boldsymbol{X}_i)\}^2,$$

over $\boldsymbol{\beta}$, where $\hat{\psi}_{S(\boldsymbol{\beta}),\mu}$ minimizes (8) for fixed $\boldsymbol{\alpha} = S(\boldsymbol{\beta})$.

Analogous to our approach above, we can skip the reparameterization and minimize instead

$$\left\| \frac{1}{n} \sum_{i=1}^{n} \left\{\hat{\psi}_{n,\boldsymbol{\alpha},\mu}(\boldsymbol{\alpha}^T \boldsymbol{X}_i) - Y_i\right\} \boldsymbol{X}_i \, \tilde{\psi}'_{n,\boldsymbol{\alpha},\mu}(u)\big|_{u=\boldsymbol{\alpha}^T \boldsymbol{X}_i} \right\| \tag{9}$$

where $\tilde{\psi}_{n,\boldsymbol{\alpha},\mu}$ minimizes (8) for fixed $\boldsymbol{\alpha}$ and $\tilde{\psi}'_{n,\boldsymbol{\alpha},\mu}$ is its derivative. We call this estimator the spline estimator.

We finally give simulation results for these different methods in Sect. 5, where, apart from the comparison with the spline estimator, we make a comparison with other estimators of $\boldsymbol{\alpha}_0$ not using the monotonicity constraint: the Effective Dimension Reduction (EDR) method, proposed in Hristache et al. (2001) and implemented in the `R` package `edr`, the (refined) Mean Average conditional Variance Estimator (MAVE) method, discussed in Xia (2006), and implemented in the `R` package `MAVE`, and Estimation Function Method (EFM), discussed in Cui et al. (2011).

For reasons of space, the proofs of the statements of our paper are given in Balabdaoui and Groeneboom (2020).

2 General Conditions and the Functions $\hat{\psi}_{n,\hat{\alpha}}$ and $\psi_{\hat{\alpha}}$

We give general conditions that we assume to hold in the remainder of the paper here and give graphical comparisons of the functions $\hat{\psi}_{n,\boldsymbol{\alpha}}$ and $\psi_{\boldsymbol{\alpha}}$, where $\psi_{\boldsymbol{\alpha}}$ is defined in Definition 1.

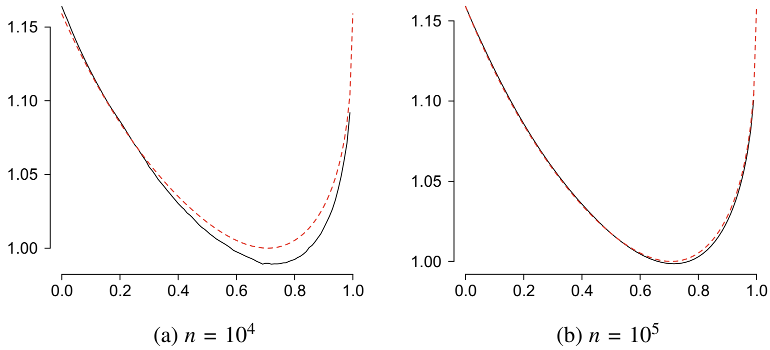


Fig. 1 The loss functions L^{LSE} (red, dashed) and $\widehat{L}_n^{\mathrm{LSE}}$ (solid), where $n = 10^4$ and $n = 10^5$

Example 1 As an illustrative example, we take $d = 2$, $\psi_0(x) = x^3$, $\boldsymbol{\alpha}_0 = (1/\sqrt{2}, 1/\sqrt{2})^T$, $Y_i = \psi_0(\boldsymbol{\alpha}_0^T X_i) + \varepsilon_i$, where the ε_i are i.i.d. standard normal random variables, independent of the $\boldsymbol{X}_i$, which are i.i.d. random vectors, consisting of two independent Uniform(0, 1) random variables. In this case, the conditional expectation function (5) is a rather complicated function of $\boldsymbol{\alpha}$ which we shall not give here but can be computed by a computer package such as Mathematica or Maple. The loss functions:

$$L^{\mathrm{LSE}} : \alpha_1 \mapsto \mathbb{E}\{Y - \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{X})\}^2 \quad \text{and} \quad \widehat{L}_n^{\mathrm{LSE}} : \alpha_1 \mapsto n^{-1} \sum_{i=1}^n \left\{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{X}_i)\right\}^2 \tag{10}$$

where the loss function $\widehat{L}_n^{\mathrm{LSE}}$ is for sample sizes $n = 10,000$ and $n = 100,000$, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$. For $\alpha_1 \in [0, 1]$ and α_2 equal to the positive root $\{1 - \alpha_1^2\}^{1/2}$, we get Fig. 1. The function L^{LSE} has a minimum equal to 1 at $\alpha_1 = 1/\sqrt{2}$, and $\widehat{L}_n^{\mathrm{LSE}}$ has a minimum at a value very close to $1/\sqrt{2}$ (furnishing the profile LSE $\hat{\boldsymbol{\alpha}}_n$), which gives a visual evidence for consistency of the profile LSE.

In order to show the $\sqrt{n}$-consistency and asymptotic normality of the estimators in the next sections, we now introduce some conditions, which correspond to those in Balabdaoui et al. (2019b). We note that we do not need conditions on reparameterization.

(A1) $\boldsymbol{X}$ has a density w.r.t. Lebesgue measure on its support $\mathcal{X}$, which is a convex set $\mathcal{X}$ with a nonempty interior, and satisfies $\mathcal{X} \subset \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\| \le R\}$ for some $R > 0$.
(A2) The function ψ_0 is bounded on the set $\{u \in \mathbb{R} : u = \boldsymbol{\alpha}_0^T \boldsymbol{x},\ \boldsymbol{x} \in \mathcal{X}\}$.
(A3) There exists $\delta > 0$ such that the conditional expectation $\tilde{\psi}_{\boldsymbol{\alpha}}$, defined by (5), is nondecreasing on $I_{\boldsymbol{\alpha}} = \{u \in \mathbb{R} : u = \boldsymbol{\alpha}^T \boldsymbol{x},\ x \in \mathcal{X}\}$ and satisfies $\tilde{\psi}_{\boldsymbol{\alpha}} = \psi_{\boldsymbol{\alpha}}$, so minimizes

$$\left\|\mathbb{E}\left\{Y - \psi(\boldsymbol{\alpha}^T \boldsymbol{X})\right\} \boldsymbol{X}\right\|^2,$$

over nondecreasing functions ψ, if $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| \le \delta$.

(A4) Let a_0 and b_0 be the (finite) infimum and supremum of the interval $\{\boldsymbol{\alpha}_0^T \boldsymbol{x},\ \boldsymbol{x} \in \mathcal{X}\}$. Then ψ_0 is continuously differentiable on $(a_0 - \delta R, a_0 + \delta R)$, where R and δ are as in Assumption A1 and A3.

(A5) The density g of $\boldsymbol{X}$ is differentiable and there exist strictly positive constants c_1 to c_4 such that $c_1 \le g(\boldsymbol{x}) \le c_2$ and $c_3 \le \frac{\partial}{\partial x_i} g(\boldsymbol{x}) \le c_4$ for $\boldsymbol{x}$ in the interior of $\mathcal{X}$.

(A6) There exists a $c_0 > 0$ and $M > 0$ such that $\mathbb{E}\{|Y|^m | \boldsymbol{X} = \boldsymbol{x}\} \le m! M_0^{m-2} c_0$ for all integers $m \ge 2$ and $\boldsymbol{x} \in \mathcal{X}$ almost surely w.r.t. dG.

These conditions are rather natural, and are discussed in Balabdaoui et al. (2019b). The following lemma shows that, for the asymptotic distribution of $\hat{\boldsymbol{\alpha}}_n$, we can reduce the derivation to the analysis of $\psi_{\hat{\boldsymbol{\alpha}}_n}$. We have the following result (Proposition 4 in Balabdaoui et al. 2019b) on the distance between $\hat{\psi}_{n,\hat{\boldsymbol{\alpha}}}$ and $\psi_{\hat{\boldsymbol{\alpha}}}$.

Lemma 1 *Let conditions (A1)–(A6) be satisfied and let G be the distribution function of $\boldsymbol{X}$. Then we have, for $\boldsymbol{\alpha}$ in a neighborhood $\mathcal{B}(\boldsymbol{\alpha}_0, \delta)$ of $\boldsymbol{\alpha}_0$*

$$\sup_{\boldsymbol{\alpha} \in \mathcal{B}(\boldsymbol{\alpha}_0, \delta)} \int \left\{\hat{\psi}_{n\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{x}) - \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{x})\right\}^2 dG(\boldsymbol{x}) = O_p\left((\log n)^2 n^{-2/3}\right).$$

3 The Limit Theory for the SSE

In this section, we derive the limit distribution of the SSE introduced above. In our derivation, the function $\psi_{\boldsymbol{\alpha}}$ of Definition 1 plays a crucial role. Below, we will use the following assumptions, additionally to (A1)–(A6).

(A7) There exists a $\delta > 0$ such that for all $\boldsymbol{\alpha} \in (\mathcal{B}(\boldsymbol{\alpha}_0, \delta) \cap \mathcal{S}_{d-1}) \setminus \{\boldsymbol{\alpha}_0\}$, the random variable

$$\mathrm{cov}\left((\boldsymbol{\alpha}_0 - \boldsymbol{\alpha})^T \boldsymbol{X}, \psi_0(\boldsymbol{\alpha}_0^T \boldsymbol{X}) \;\middle|\; \boldsymbol{\alpha}^T \boldsymbol{X}\right)$$

is not equal to 0 almost surely.

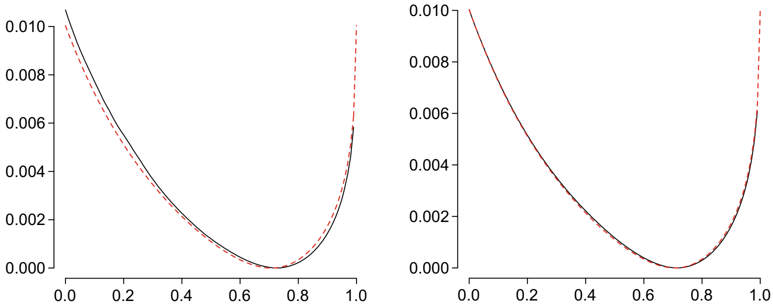
(A8) The matrix

$$\mathbb{E}\left[\psi_0'(\boldsymbol{\alpha}_0^T \boldsymbol{X})\, \mathrm{cov}(\boldsymbol{X} | \boldsymbol{\alpha}_0^T \boldsymbol{X})\right]$$

has rank $d - 1$.

We start by comparing (3) with the function

$$\boldsymbol{\alpha} \mapsto \left\|\mathbb{E}\left\{Y - \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{X})\right\} \boldsymbol{X}\right\|^2. \tag{11}$$



(a) $n = 10^4$ (b) $n = 10^5$

Fig. 2 The loss functions L^{SSE} (red, dashed) and $\widehat{L}_n^{\text{SSE}}$ (solid), where $n = 10^4$ and $n = 10^5$

As in Sect. 1, the function $\hat{\psi}_{n,\boldsymbol{\alpha}}$ is just the (isotonic) least squares estimate for fixed $\boldsymbol{\alpha}$.

Example 2 (*Continuation of Example* 1) We consider the loss function given by

$$L^{\text{SSE}} : \alpha_1 \mapsto \left\| \mathbb{E}\left\{Y - \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{X})\right\} \boldsymbol{X} \right\|^2, \tag{12}$$

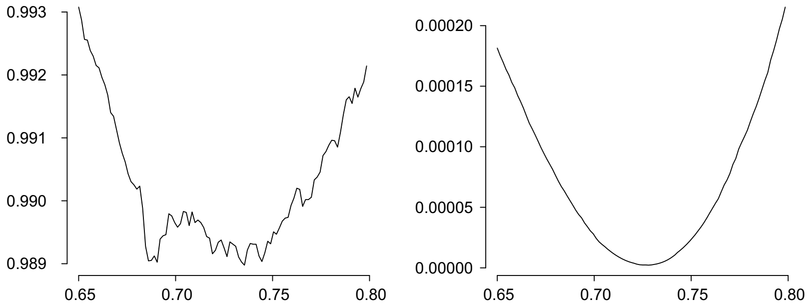
and compare this with the loss function

$$\widehat{L}_n^{\text{SSE}} : \alpha_1 \mapsto \left\| n^{-1} \sum_{i=1}^{n} \left\{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T \boldsymbol{X}_i)\right\} \boldsymbol{X}_i \right\|^2, \tag{13}$$

for the same data as in Example 1 in Sect. 2. If we plot the loss functions L^{SSE} and $\widehat{L}_n^{\text{SSE}}$ for the model of Example 1, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$, for $\alpha_1 \in [0, 1]$ and α_2 the positive root $\sqrt{1 - \alpha_1^2}$, we get Fig. 2. The function L^{LSE} has a minimum equal to 0 at $\alpha_1 = 1/\sqrt{2}$ while $\widehat{L}_n^{\text{SSE}}$ attains its minimum at a value that is very close to $1/\sqrt{2}$.

In general, the curve $\widehat{L}_n^{\text{SSE}}$ will be smoother than the curve $\widehat{L}_n^{\text{LSE}}$. The rather striking difference in smoothness of the loss functions $\widehat{L}_n^{\text{LSE}}$ and $\widehat{L}_n^{\text{SSE}}$ can be seen in Fig. 3, where we zoom in on the interval $[0.65, 0.80]$ for $n = 10,000$ and the examples of Figs. 1 and 2. The question is whether this difference in smoothness explains why the SSE is $\sqrt{n}$-consistent while this might not be the case for the profile LSE.

In the computation of the SSE, we have to take a starting point. For this, we use the LSE, which is proved to be consistent in Balabdaoui et al. (2019a). The proof of the consistency of the SSE is a variation on the proof for corresponding crossing of the zero estimator in Balabdaoui et al. (2019b) in (D.2) of the supplementary material. We use the following lemma, which is a corollary to Proposition 2 in the material of Balabdaoui et al. (2019b).



(a) $\widehat{L}_n^{\text{LSE}}$ (b) $\widehat{L}_n^{\text{SSE}}$

Fig. 3 The loss functions $\widehat{L}_n^{\text{LSE}}$ and $\widehat{L}_n^{\text{SSE}}$ on $[0.65, 0.80]$, for $n = 10^4$

Lemma 2 *Let ϕ_n and ϕ be defined by*

$$\phi_n(\boldsymbol{\alpha}) = \int \boldsymbol{x}\left\{y - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\right\} d\mathbb{P}_n(\boldsymbol{x}, y),$$

and

$$\phi(\boldsymbol{\alpha}) = \int \boldsymbol{x}\left\{y - \psi_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{x})\right\} dP(\boldsymbol{x}, y).$$

Then, uniformly for $\boldsymbol{\alpha}$ in a neighborhood $\mathcal{B}(\boldsymbol{\alpha}_0, \delta) \cap \mathcal{S}_{d-1}$ of $\boldsymbol{\alpha}_0$

$$\phi_n(\boldsymbol{\alpha}) = \phi(\boldsymbol{\alpha}) + o_p(1).$$

Remark 1 The proof in Balabdaoui et al. (2019b) used reparameterization, but this is actually not needed in the proof.

Theorem 1 (Consistency of the SSE) *Let $\hat{\boldsymbol{\alpha}}_n \in \mathcal{S}_{d-1}$ be the SSE of $\boldsymbol{\alpha}_0$ and let conditions (A1)–(A8) be satisfied. Then*

$$\hat{\boldsymbol{\alpha}}_n \xrightarrow{p} \boldsymbol{\alpha}_0.$$

Lemma 3 *Let $\hat{\boldsymbol{\alpha}}_n \in \mathcal{S}_{d-1}$ be a minimizer of*

$$\left\| n^{-1}\sum_{i=1}^{n}\left\{Y_i - \hat{\psi}_{n,\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\right\}\boldsymbol{X}_i\right\|^2, \tag{14}$$

for $\boldsymbol{\alpha} \in \mathcal{S}_{d-1}$, where $\|\cdot\|$ denotes the Euclidean norm. Then, under conditions (A1)–(A8) we have

$$n^{-1}\sum_{i=1}^{n}\left\{Y_i-\hat\psi_{n,\hat{\boldsymbol\alpha}_n}(\hat{\boldsymbol\alpha}_n^T\boldsymbol X_i)\right\}\boldsymbol X_i=n^{-1}\sum_{i=1}^{n}\left\{Y_i-\psi_{\hat{\boldsymbol\alpha}_n}(\hat{\boldsymbol\alpha}_n^T\boldsymbol X_i)\right\}\left\{\boldsymbol X_i-\mathbb{E}\left(\boldsymbol X|\hat{\boldsymbol\alpha}_n^T\boldsymbol X_i\right)\right\}+o_p\left(n^{-1/2}\right). \tag{15}$$

We now have the following limit result.

Theorem 2 (Asymptotic normality of the SSE) *Let $\hat{\boldsymbol\alpha}_n$ be the minimizer of*

$$\left\|n^{-1}\sum_{i=1}^{n}\left\{Y_i-\hat\psi_{n,\boldsymbol\alpha}(\boldsymbol\alpha^T\boldsymbol X_i)\right\}\boldsymbol X_i\right\|^2, \tag{16}$$

for $\boldsymbol\alpha\in\mathcal{S}_{d-1}$, where $\|\cdot\|$ denotes the Euclidean norm. Let the matrices $\boldsymbol A$ and $\boldsymbol\Sigma$ be defined by

$$\boldsymbol A=\mathbb{E}\Big[\psi_0'(\boldsymbol\alpha_0^T\boldsymbol X)\,Cov(\boldsymbol X|\boldsymbol\alpha_0^T\boldsymbol X)\Big], \tag{17}$$

and

$$\boldsymbol\Sigma=\mathbb{E}\left[\left\{Y-\psi_0(\boldsymbol\alpha_0^T\boldsymbol X)\right\}^2\left\{\boldsymbol X-\mathbb{E}(\boldsymbol X|\boldsymbol\alpha_0^T\boldsymbol X)\right\}\left\{\boldsymbol X-\mathbb{E}(\boldsymbol X|\boldsymbol\alpha_0^T\boldsymbol X)\right\}^T\right]. \tag{18}$$

Then, under conditions $(A1)$–$(A8)$, we have

$$\sqrt{n}(\hat{\boldsymbol\alpha}_n-\boldsymbol\alpha_0)\to_d N\left(\boldsymbol 0,\boldsymbol A^-\boldsymbol\Sigma\boldsymbol A^-\right),$$

where $\boldsymbol A^-$ is the Moore-Penrose inverse of $\boldsymbol A$.

Example 3 (*Continuation of Example 2*) We compute the asymptotic covariance matrix for Example 2. In this case, we get for matrix $\boldsymbol A$ in part (ii) of Theorem 2

$$\begin{aligned}\boldsymbol A&=\mathbb{E}\Big[\psi_0'(\boldsymbol\alpha_0^T\boldsymbol X)\,\mathrm{Cov}(\boldsymbol X|\boldsymbol\alpha_0^T\boldsymbol X)\Big]\\&=\frac{3}{4}\mathbb{E}\left[\left(\frac{X_1+X_2}{\sqrt2}\right)^2\left(\boldsymbol X-\mathbb{E}(\boldsymbol X|\boldsymbol\alpha_0^T\boldsymbol X)\right)\left(\boldsymbol X-\mathbb{E}(\boldsymbol X|\boldsymbol\alpha_0^T\boldsymbol X)\right)^T\right]\\&=\begin{pmatrix}1/15&-1/15\\-1/15&1/15\end{pmatrix}.\end{aligned}$$

The Moore-Penrose inverse of $\boldsymbol A$ is given by

$$\boldsymbol A^-=\begin{pmatrix}15/4&-15/4\\-15/4&15/4\end{pmatrix}.$$

Furthermore, we get

$$\begin{aligned}\boldsymbol{\Sigma} &= \mathbb{E}\left[\left\{Y - \psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}^2 \left\{\boldsymbol{X} - \mathbb{E}(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}\left\{\boldsymbol{X} - \mathbb{E}(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}^T\right] \\ &= \mathbb{E}\left\{\boldsymbol{X} - \mathbb{E}(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}\left\{\boldsymbol{X} - \mathbb{E}(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}^T \\ &= \begin{pmatrix} 1/24 & -1/24 \\ -1/24 & 1/24 \end{pmatrix}.\end{aligned}$$

So the asymptotic covariance matrix is given by

$$\boldsymbol{A}^-\boldsymbol{\Sigma}\boldsymbol{A}^- = \begin{pmatrix} 75/32 & -75/32 \\ -75/32 & 75/32 \end{pmatrix} \approx \begin{pmatrix} 2.34375 & -2.34375 \\ -2.34375 & 2.34375 \end{pmatrix}.$$

Remark 2 Theorem 2 corresponds to Theorem 3 in Balabdaoui et al. (2019b), but note that the estimator has a different definition. Reparameterization is also avoided.

4 The Limit Theory for ESE and Cubic Spline Estimator

The proofs of the consistency and asymptotic normality of the ESE and spline estimator are highly similar to the proofs of these facts for the SSE in the preceding section. The only extra ingredient is the occurrence of the estimate of the derivative of the link function. We only discuss the asymptotic normality.

In addition to the assumptions (A1)–(A7), we now assume the following:

(A8') $\psi_{\boldsymbol{\alpha}}$ is twice differentiable on $(\inf_{x\in\mathcal{X}}(\boldsymbol{\alpha}^T\boldsymbol{x}), \sup_{x\in\mathcal{X}}(\boldsymbol{\alpha}^T\boldsymbol{x}))$.

(A9) The matrix

$$\mathbb{E}\left[\psi_0'(\boldsymbol{\alpha}_0^T\boldsymbol{X})^2 \operatorname{cov}(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\right]$$

has rank $d-1$.

An essential step is again to show that

$$\begin{aligned}&\int \boldsymbol{x}\left\{y - \hat\psi_{n,\hat{\boldsymbol{\alpha}}_n}(\hat{\boldsymbol{\alpha}}_n^T\boldsymbol{x})\right\}\hat\psi'_{n\hat{\boldsymbol{\alpha}}_n}(\hat{\boldsymbol{\alpha}}_n^T\boldsymbol{x})\,d\mathbb{P}_n(\boldsymbol{x},y) \\ &= \int \left\{\boldsymbol{x} - \mathbb{E}(X|\hat{\boldsymbol{\alpha}}_n^T\boldsymbol{X})\right\}\left\{y - \hat\psi_{n,\hat{\boldsymbol{\alpha}}_n}(\hat{\boldsymbol{\alpha}}_n^T\boldsymbol{x})\right\}\hat\psi'_{n\hat{\boldsymbol{\alpha}}_n}(\hat{\boldsymbol{\alpha}}_n^T\boldsymbol{x})\,d\mathbb{P}_n(\boldsymbol{x},y) \\ &\qquad + o_p(n^{-1/2}) + o_p(\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0).\end{aligned}$$

For the ESE, this is done by defining the piecewise constant function $\bar\rho_{n,\boldsymbol{\alpha}}$ for u in the interval between successive jumps τ_i and τ_{i+1}) of $\hat\psi_{n\boldsymbol{\alpha}}$ by

$$\bar\rho_{n,\boldsymbol{\alpha}}(u) = \begin{cases} \mathbb{E}[\boldsymbol{X}|\boldsymbol{\alpha}^T\boldsymbol{X} = \tau_i]\psi'_{\boldsymbol{\alpha}}(\tau_i) & \text{if } \psi_{\boldsymbol{\alpha}}(u) > \hat\psi_{n\boldsymbol{\alpha}}(\tau_i) \text{ for all } u \in (\tau_i, \tau_{i+1}), \\ \mathbb{E}[\boldsymbol{X}|\boldsymbol{\alpha}^T\boldsymbol{X} = s]\psi'_{\boldsymbol{\alpha}}(s) & \text{if } \psi_{\boldsymbol{\alpha}}(s) = \hat\psi_{n\boldsymbol{\alpha}}(s) \text{ for some } s \in (\tau_i, \tau_{i+1}), \\ \mathbb{E}[\boldsymbol{X}|\boldsymbol{\alpha}^T\boldsymbol{X} = \tau_{i+1}]\psi'_{\boldsymbol{\alpha}}(\tau_{i+1}) & \text{if } \psi_{\boldsymbol{\alpha}}(u) < \hat\psi_{n\boldsymbol{\alpha}}(\tau_i) \text{ for all } u \in (\tau_i, \tau_{i+1}); \end{cases}$$

see Appendix E in the supplement of Balabdaoui et al. (2019b). The remaining part of the proof runs along the same lines as the proof for the SSE. For additional details, see Appendix E in the supplement of Balabdaoui et al. (2019b).

The corresponding step in the proof for the spline estimator is given by the following lemma.

Lemma 4 *Let the conditions of Theorem 5 in Kuchibhotla and Patra (2020) be satisfied. In particular, let the penalty parameter* μ_n *satisfy* $\mu_n = o_p(n^{-1/2})$. *Then we have for all* $\boldsymbol{\alpha}$ *in a neighborhood of* $\boldsymbol{\alpha}_0$ *and for the corresponding natural cubic spline* $\hat\psi_{n\boldsymbol{\alpha}}$

$$\int \mathbb{E}(\boldsymbol{X}|\boldsymbol{\alpha}^T\boldsymbol{X}) \left\{y - \hat\psi_{n\boldsymbol{\alpha}}\left(\boldsymbol{\alpha}^T\boldsymbol{x}\right)\right\} \hat\psi'_{n\boldsymbol{\alpha}}\left(\boldsymbol{\alpha}^T\boldsymbol{x}\right)\, d\mathbb{P}_n(\boldsymbol{x}, y) = O_p(\mu_n) = o_p\left(n^{-1/2}\right).$$

Remark 3 The result shows that we have as our basic equation in $\boldsymbol{\alpha}$

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\left\{\hat\psi_{n\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i) - Y_i\right\}\hat\psi'_{n\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\boldsymbol{X}_i\\
&= \frac{1}{n}\sum_{i=1}^n\left\{\hat\psi_{n\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i) - Y_i\right\}\hat\psi'_{n\boldsymbol{\alpha}}(\boldsymbol{\alpha}^T\boldsymbol{X}_i)\left\{\boldsymbol{X}_i - \mathbb{E}(\boldsymbol{X}_i|\boldsymbol{\alpha}^T\boldsymbol{X}_i)\right\} + o_p\left(n^{-1/2}\right)\\
&= o_p\left(n^{-1/2}\right).
\end{aligned}$$

The remaining part of the proof of the asymptotic normality can either run along the same lines as the proof for the corresponding fact for the SSE, using the function $u \mapsto \psi_{\boldsymbol{\alpha}}(u) = \mathbb{E}\{\psi_0(\boldsymbol{\alpha}^T\boldsymbol{x})|\boldsymbol{\alpha}^T\boldsymbol{X} = u\}$, or directly use the convergence of $\hat\psi_{n\hat{\boldsymbol{\alpha}}_n}$ to ψ_0 and of $\hat\psi'_{n\hat{\boldsymbol{\alpha}}_n}$ to ψ_0' (see Theorem 3 in Kuchibhotla and Patra 2020). For the SSE and ESE, we were forced to introduce the intermediate function $\psi_{\boldsymbol{\alpha}}$ to get to the derivatives, because for these estimators the derivative of $\hat\psi_{n\hat{\boldsymbol{\alpha}}_n}$ did not exist.

We get the following result.

Theorem 3 *Let either* $\hat{\boldsymbol{\alpha}}_n$ *be the ESE of* $\boldsymbol{\alpha}_0$ *and let Assumptions (A1)–(A7) and (A8') and (A9) of the present section be satisfied, or let* $\hat{\boldsymbol{\alpha}}_n$ *be the spline estimator of* $\boldsymbol{\alpha}_0$ *and let Assumptions (A0)-(A6) and (B1)–(B3) of Kuchibhotla and Patra 2020) be satisfied. Moreover, let the bandwidth* $h \asymp n^{-1/7}$ *in the estimate of the derivative of* $\psi_{\boldsymbol{\alpha}}$ *for the ESE. Define the matrices*

$$\tilde{\boldsymbol{A}} := \mathbb{E}\Big[\psi_0'(\boldsymbol{\alpha}_0^T\boldsymbol{X})^2\, Cov(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\Big], \tag{19}$$

and

$$\tilde{\boldsymbol{\Sigma}} := \mathbb{E}\left[\left\{Y - \psi_0(\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}^2 \psi_0'(\boldsymbol{\alpha}_0^T\boldsymbol{X})^2 \left\{\boldsymbol{X} - \mathbb{E}(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}\left\{\boldsymbol{X} - \mathbb{E}(\boldsymbol{X}|\boldsymbol{\alpha}_0^T\boldsymbol{X})\right\}^T\right]. \tag{20}$$

Then

$$\sqrt{n}(\tilde{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) \to_d N_d\left(\mathbf{0}, \tilde{\boldsymbol{A}}^- \tilde{\boldsymbol{\Sigma}} \tilde{\boldsymbol{A}}^-\right),$$

where $\tilde{\boldsymbol{A}}^-$ *is the Moore-Penrose inverse of* $\tilde{\boldsymbol{A}}$.

This corresponds to Theorem 6 in Balabdaoui et al. (2019b) and Theorem 5 in Kuchibhotla and Patra (2020), but note that the formulation of Theorem 5 in Kuchibhotla and Patra (2020) still contains the Jacobian connected with the lower dimensional parameterization. Consequently, the ESE and the cubic spline estimator admit the same weak limit under the conditions stated above.

5 Simulation and Comparisons with Other Estimators

In this section, we compare the LSE with the Simple Score Estimator (SSE), the Efficient Score Estimator (ESE), the Effective Dimension Reduction (EDR) estimate, the spline estimate, the MAVE estimate, and the EFM estimate. We take part in the simulation settings in Balabdaoui et al. (2019a), which means that we take the dimension d equal to 2. Since the parameter belongs to the boundary of a circle in this case, we only have to determine a one-dimensional parameter. Using this fact, we use the parameterization $\boldsymbol{\alpha} = (\alpha_1, \alpha_2) = (cos(\beta), \sin(\beta))$ and determine the angle β by a golden section search for the SSE, ESE, and spline estimate. For EDR, we used the `R` package `edr`; the method is discussed in Hristache et al. (2001). The spline method is described in Kuchibhotla and Patra (2020), and there exists an `R` package `simest` for it, but we used our own implementation. For the MAVE method, we used the `R` package `MAVE`; for theory, see Xia (2006). For the EFM estimate (see Cui et al. 2011), we used an `R` script, due to Xia Cui and kindly provided to us by her and Rohit Patra. All runs of our simulations can be reproduced by running the `R` scripts in Groeneboom 2018.

In simulation model 1, we take $\boldsymbol{\alpha}_0 = (1/\sqrt{2}, 1/\sqrt{2})^T$ and $\boldsymbol{X} = (X_1, X_2)^T$, where X_1 and X_2 are independent Uniform(0, 1) variables. The model is now

$$Y = \psi_0(\boldsymbol{\alpha}_0^T \boldsymbol{X}) + \varepsilon,$$

where $\psi_0(u) = u^3$ and ε is a standard normal random variable, independent of $\boldsymbol{X}$.

In simulation model 2, we also take $\boldsymbol{\alpha}_0 = (1/\sqrt{2}, 1/\sqrt{2})^T$ and $\boldsymbol{X} = (X_1, X_2)^T$, where X_1 and X_2 are independent Uniform(0, 1) variables. This time, however, the model is (Table 1)

$$Y = \text{Bin}\left(10, \exp(\boldsymbol{\alpha}_0^T \boldsymbol{X}) / \left\{1 + \exp(\boldsymbol{\alpha}_0^T \boldsymbol{X})\right\}\right);$$

see also Table 2 in Balabdaoui et al. (2019a). This means

Table 1 Simulation, model 1; ε_i is standard normal and independent of $\boldsymbol{X}_i$, consisting of two independent Uniform(0, 1) random variables. The mean value $\hat{\mu}_i = \text{mean}(\hat{\alpha}_{in}),\ i = 1, 2$ and n times the variance-covariance $\hat{\sigma}_{ij} = n \cdot \text{cov}(\hat{\alpha}_{in}, \hat{\alpha}_{jn}),\ i,\ j = 1, 2$, of the Efficient Dimension Reduction Estimate (EDR), computed by the `R` package `edr`, the Least Squares Estimate (LSE), the Simple Score Estimate (SSE), the Efficient Score Estimate (ESE), the spline estimate, the MAVE estimate, and the EFM estimate for different sample sizes n. The line, preceded by ∞, gives the asymptotic values (unknown for EDR and LSE). The values are based on 1000 replications

Method	n	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_{11}$	$\hat{\sigma}_{22}$	$\hat{\sigma}_{12}$
EDR	100	0.621877	0.361894	11.409222	36.869184	9.152389
	500	0.701217	0.686094	7.334756	11.468453	–3.881349
	1000	0.701669	0.702244	6.437653	8.090771	–3.552562
	5000	0.706021	0.706798	7.344431	7.276717	–7.288047
	∞	0.707107	0.707107	?	?	?
LSE	100	0.672698	0.697350	3.148912	2.975246	–2.915427
	500	0.702163	0.701718	3.620960	3.665710	–3.588491
	1000	0.704706	0.704320	3.665561	3.664711	–3.637541
	5000	0.707262	0.705690	4.435842	4.485168	–4.453713
	∞	0.707107	0.707107	?	?	?
SSE	100	0.673997	0.6919403	3.338637	3.362656	–3.141408
	500	0.699986	0.706198	2.849647	2.807978	–2.793798
	1000	0.706477	0.704191	2.501106	2.510047	–2.494237
	5000	0.707090	0.706423	2.473765	2.485884	–2.477371
	∞	0.707107	0.707107	2.343750	2.343750	–2.343750
ESE	100	0.682781	0.687949	3.067802	2.991976	–2.855176
	500	0.702940	0.702462	3.100843	3.116337	–3.064151
	1000	0.704055	0.706387	2.676388	2.653164	–2.650667
	5000	0.707130	0.706444	2.257541	2.265547	–2.259443
	∞	0.707107	0.707107	1.885522	1.885522	–1.885522
Spline	100	0.690741	0.705485	1.801235	1.762567	–1.711552
	500	0.703670	0.702640	1.795384	1.778454	–1.773560
	1000	0.705684	0.706007	1.786589	1.781797	–1.777691
	5000	0.706404	0.707193	2.180466	2.181544	–2.179269
	∞	0.707107	0.707165	1.885522	1.885522	–1.885522
MAVE	100	0.686503	0.684887	2.423618	3.546768	–2.245708
	500	0.703333	0.705537	1.897806	1.876220	–2.040677
	1000	0.705840	0.705660	1.929966	1.907128	–1.911452
	5000	0.707328	0.706299	2.071168	2.082169	–2.074914
	∞	0.707107	0.707107	1.885522	1.885522	–1.885522
EFM	100	0.686292	0.684274	2.802308	3.280956	–2.312445
	500	0.703236	0.705133	2.082162	2.045150	–2.044960
	1000	0.705629	0.705950	1.866486	1.860184	–1.856340
	5000	0.707269	0.707251	1.953800	1.964081	–1.957351
	∞	0.707107	0.707107	1.885522	1.885522	–1.885522

Table 2 Simulation, model 2; $Y_i \sim \text{Bin}\left(10, \exp(\boldsymbol{\alpha}_0^T \boldsymbol{X}_i)/\left\{1+\exp(\boldsymbol{\alpha}_0^T \boldsymbol{X}_i)\right\}\right)$, where $\boldsymbol{X}_i$ consists of two independent Uniform(0, 1) random variables. The mean value $\hat{\mu}_i = \text{mean}(\hat{\alpha}_{in}),\ i = 1, 2$ and n times the variance-covariance $n\text{cov}(\hat{\alpha}_{in}, \hat{\alpha}_{jn}),\ i,\ j = 1, 2$, of the Efficient Dimension Reduction Estimate (EDR), computed by the `R` package `edr`, the Least Squares Estimate (LSE), the Simple Score Estimate (SSE), the Efficient Score Estimate (ESE), the spline estimate, the MAVE estimate, and the EFM estimate for different sample sizes n. The line, preceded by ∞, gives the asymptotic values (unknown for EDR and LSE). The values are based on 1000 replications

Method	n	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_{11}$	$\hat{\sigma}_{22}$	$\hat{\sigma}_{12}$
EDR	100	0.587264	0.202005	13.33724	48.15572	11.87625
	500	0.670702	0.602469	26.76111	66.92737	14.09701
	1000	0.696075	0.666591	21.89080	49.31544	9.345753
	5000	0.704424	0.706604	11.39598	11.11493	–11.17376
	∞	0.707107	0.707107	?	?	?
LSE	100	0.658631	0.699725	4.069966	3.596783	–3.609490
	500	0.695541	0.703007	5.650618	5.362877	–5.358190
	1000	0.704497	0.701243	5.909494	6.043808	–5.911246
	5000	0.704805	0.707621	6.303320	6.321866	–6.298515
	∞	0.707107	0.707107	?	?	?
SSE	100	0.667908	0.694376	3.760921	3.420387	–3.356968
	500	0.698498	0.706423	3.358458	3.182044	–3.223734
	1000	0.707276	0.702390	3.179623	3.236283	–3.184724
	5000	0.706162	0.707286	2.718742	2.707549	–2.709870
	∞	0.707107	0.707107	2.727482	2.727482	–2.727482
ESE	100	0.684804	0.688063	2.892165	2.874755	–2.744223
	500	0.698078	0.706159	3.562625	3.457337	–3.446605
	1000	0.707879	0.701445	3.420159	3.470217	–3.418606
	5000	0.706321	0.707110	2.775092	2.760287	–2.764230
	∞	0.707107	0.707107	2.737200	2.737200	–2.737200
Spline	100	0.677287	0.695301	3.009781	2.779876	–2.714928
	500	0.699117	0.706946	2.952928	2.784383	–2.830415
	1000	0.707890	0.702001	3.027712	3.064772	–3.026082
	5000	0.706200	0.707312	2.764447	2.762986	–2.760530
	∞	0.707107	0.707232	2.737200	2.737200	–2.737200
MAVE	100	0.667849	0.654361	3.891510	8.700093	–2.325804
	500	0.699108	0.706377	3.155191	2.990569	–3.031249
	1000	0.707520	0.702341	3.040201	3.097965	–3.049075
	5000	0.707657	0.705827	2.572343	2.573418	–2.570275
	∞	0.707107	0.707107	2.737200	2.737200	–2.737200
EFM	100	0.663227	0.666070	5.681573	5.978194	–2.503058
	500	0.698920	0.706295	3.279110	3.055940	–3.118757
	1000	0.707878	0.706275	3.102414	3.157143	–3.108516
	5000	0.706043	0.701894	2.669352	2.650343	–2.656742
	∞	0.707107	0.707107	2.737200	2.737200	–2.737200

$$Y = \psi_0(\boldsymbol{\alpha}_0^T \boldsymbol{X}) + \varepsilon,$$

where

$$\psi_0(\boldsymbol{\alpha}_0^T \boldsymbol{X}) = 10 \exp(\boldsymbol{\alpha}_0^T \boldsymbol{X})/\{1 + \exp(\boldsymbol{\alpha}_0^T \boldsymbol{X})\}, \qquad \varepsilon = N_n - \psi_0(\boldsymbol{\alpha}_0^T \boldsymbol{X}),$$

and

$$N_n = \text{Bin}\left(10, \frac{\exp(\boldsymbol{\alpha}_0^T \boldsymbol{X})}{1 + \exp(\boldsymbol{\alpha}_0^T \boldsymbol{X})}\right).$$

Note that indeed $\mathbb{E}\{\varepsilon|\boldsymbol{X}) = 0$, but that we do not have independence of ε and $\boldsymbol{X}$, as in the previous example.

It was noticed in Xia (2006), p. 1113, that, although it was shown in Hristache et al. (2001) that the $\sqrt{n}$ rate of convergence for the estimation of $\boldsymbol{\alpha}_0$ can be achieved, the asymptotic distribution of the method proposed in Hristache et al. (2001) was not derived, which makes it difficult to compare the limiting efficiency of the estimation method with other methods. In Xia (2006), the asymptotic distribution of the rMAVE estimate is derived (see Theorem 4.2 of Xia 2006), which shows that this limit distribution is actually the same as that of the ESE and the spline estimate. Since Xia is one of the authors of the recent `MAVE` `R` package, we assume that the rMAVE method has been implemented in this package, so we will identify MAVE with rMAVE in the sequel.

The proof of the asymptotic normality result for the MAVE method uses the fact that the iteration steps, described on p.1117 of Xia (2006), start in a neighborhood $\{\boldsymbol{\alpha} : \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\| \le Cn^{-1/2+c_0}\}$ of $\boldsymbol{\alpha}_0$, where $C > 0$ and $c_0 < 1/20$, and indeed our

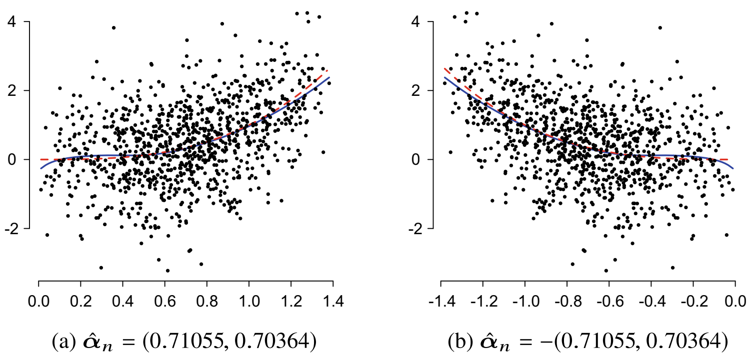


Fig. 4 Two MAVE estimates of $\boldsymbol{\alpha}_0 = 2^{-1/2}(1, 1)^T$ for model 1 with sample size $n = 1000$: **a** from starting the iterations at $\boldsymbol{\alpha}_0$, **b** from starting the iterations at $-\boldsymbol{\alpha}_0$; the blue solid curve is the estimate of the link function, based on $\hat{\boldsymbol{\alpha}}_n$; the blue dashed function is $t \mapsto t^3$ in **a** and $t \mapsto -t^3$ in **b**. Note that in **b** also the sign of the first coordinates of the points $(\hat{\boldsymbol{\alpha}}_n^T \boldsymbol{X}_i, Y_i)$ in the scatterplot is reversed. Under the restriction that the link function is nondecreasing **b** cannot be a solution

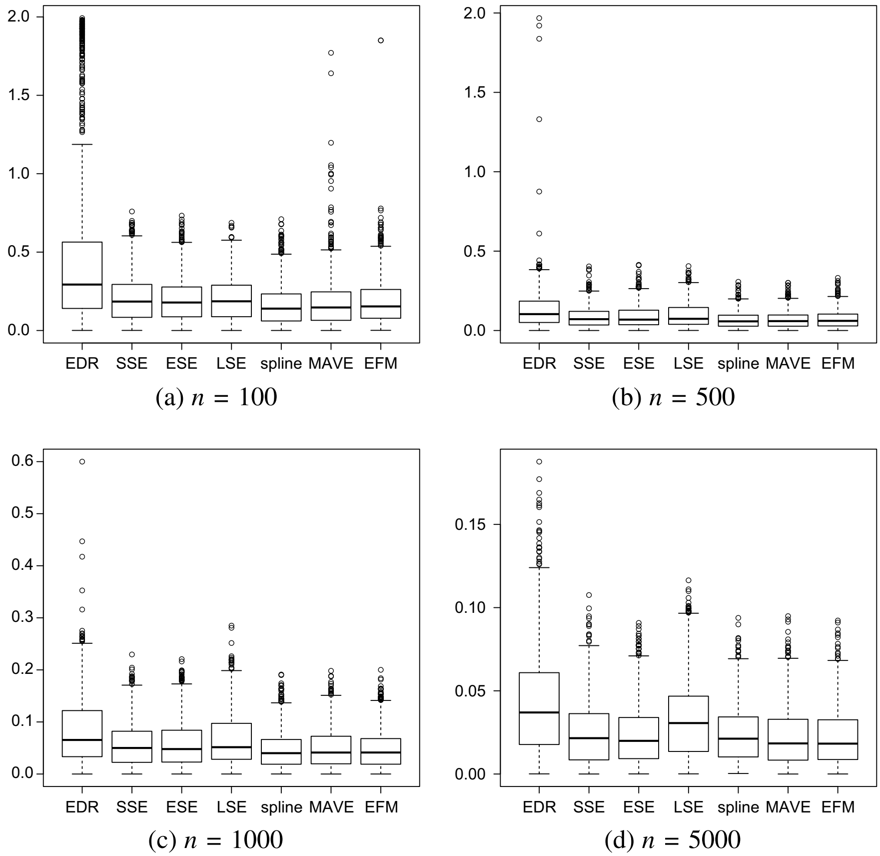
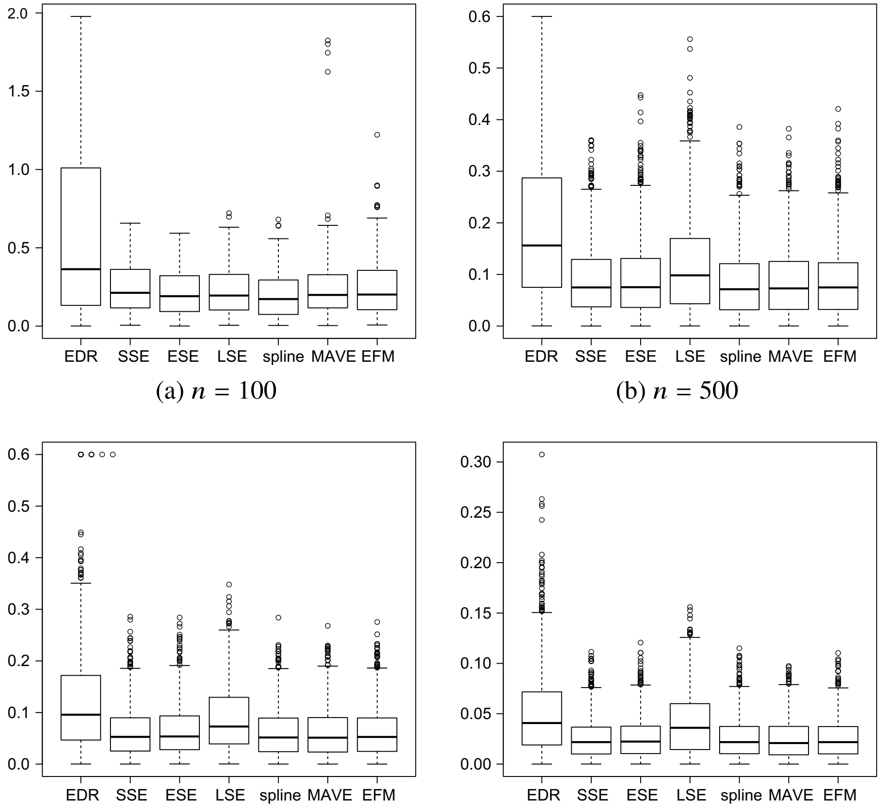


Fig. 5 Boxplots of $\sqrt{n/2}\,\|\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\|_2$ for model 1. In **b** and **c**, the values of EDR were truncated at 0.6 to show more clearly the differences between the other estimates

original experiments with the `R` package showed many outliers, probably due to starting values not sufficiently close to $\boldsymbol{\alpha}_0$. A further investigation revealed that there were many solutions in the neighborhood of the points $-\boldsymbol{\alpha}_0$. This phenomenon is illustrated in Fig. 4, generated by our own implementation of the algorithm in Xia (2006). The link function is constructed from the values $a_j^{\hat{\boldsymbol{\alpha}}_n}$ in the algorithm in Xia (2006), p. 1117, where the ordered values of $\hat{\boldsymbol{\alpha}}_n^T \boldsymbol{X}_j$ are the first coordinates.

Because of the difficulty we just discussed, we reversed in the results of the `MAVE R` package the sign of the solutions in the neighborhood of $-\boldsymbol{\alpha}_0$. By the parameterization with a positive first coordinate in Cui et al. (2011), situation (b) in Fig. 4 cannot occur for the EFM algorithm. We also tried a modification of the same type as our modification of the MAVE algorithm for the EDR algorithm, but this did not lead to a similar improvement of the results.



(a) $n = 100$ (b) $n = 500$ (c) $n = 1000$ (d) $n = 5000$

Fig. 6 Boxplots of $\sqrt{n/2}\,\|\hat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\|_2$ for model 2. In **b** and **c**, the values of EDR were truncated at 0.6 to show more clearly the differences between the other estimates

It follows from Theorem 2 that the variance of the asymptotic normal distribution for the SSE is equal to 2.727482 and from Theorem 3 that the variance of the asymptotic normal distribution for the ESE and spline estimator equals 2.737200. We already noticed in Sect. 4 that the present model is not homoscedastic. In this case, the asymptotic covariance matrix for the SSE of Theorem 2 is in fact given by $\boldsymbol{A}^{-} = \boldsymbol{A}^{-}\boldsymbol{\Sigma}\boldsymbol{A}^{-}$.

It is clear that the estimate EDR is inferior to the other methods for these models; even the LSE for which we do not know the rate of convergence has a better performance, see Figs. 5 and 6. In Hristache et al. (2001), not only it is assumed that the errors have a normal distribution, but also in model 1, where this condition is satisfied, the behavior is clearly inferior, in particular for the lower sample sizes.

6 Concluding Remarks

We replaced the "crossing of zero" estimators in Balabdaoui et al. (2019b) with profile least squares estimators. The asymptotic distribution of the estimators was determined and its behavior illustrated by a simulation study, using the same models as in Balabdaoui et al. (2019a).

In the first model, the error is independent of the covariate and homoscedastic and in this case, four of the estimators were efficient. In the other (binomial-logistic) model, the error was dependent on the covariates and not homoscedastic. It was shown that the Simple Score Estimate (SSE) had in fact a smaller asymptotic variance in this model than the other estimators for which the asymptotic variance is known, although the difference is very small and does not really show up in the simulations.

There is no uniformly best estimate in our simulation, but the EDR estimate is clearly inferior to the other estimates, including the LSE, in particular for the lower sample sizes. On the other hand, the LSE is inferior to the other estimators except for the EDR. All simulation results can be reproduced by running the `R` scripts in Groeneboom (2018).

Acknowledgements We thank Vladimir Spokoiny for helpful discussions during the Oberwolfach meeting "Statistics meets Machine Learning", January 26–February 1, 2020. We also feel very honored to make this contribution to Christine's Festschrift and wish her all the best in her future endeavors.

References

Balabdaoui, F., & Groeneboom, P. (2020). Profile least squares estimators in the monotone single index model. Version with proofs. https://arxiv.org/abs/2001.05454.

Balabdaoui, F., Durot, C., & Jankowski, H. (2019a). Least squares estimation in the monotone single index model. *Bernoulli*, *25*(4), 3276–3310.

Balabdaoui, F., Groeneboom, P., & Hendrickx, K. (2019b). Score estimation in the monotone single-index model. *Scandinavian Journal of Statistics*, *46*(2), 517–544. ISSN 0303-6898.

Cui, X., Härdle, W. K., & Zhu, L. (2011). The efm approach for single-index models. *Annals of Statistics*, *39*(3), 1658–1688, 06. https://doi.org/10.1214/10-AOS871.

Green, P. J., & Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*.Monographs on statistics and applied probability (Vol. 58). London: Chapman & Hall. ISBN 0-412-30040-0. https://doi.org/10.1007/978-1-4899-4473-3. A roughness penalty approach.

Groeneboom, P. (2018). Algorithms for computing estimates in the single index model. https://github.com/pietg/single_index.

Hristache, M., Juditsky, A., & Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-index model. *Annals of Statistics*, *29*(3), 595–623. ISSN 0090-5364. 10.1214/aos/1009210681. https://doi.org/10.1214/aos/1009210681.

Kuchibhotla, A. K., & Patra, R. K. (2020). Efficient estimation in single index models through smoothing splines. *Bernoulli*, *26*(2), 1587–1618. ISSN 1350-7265. https://doi.org/10.3150/19-BEJ1183.

Landers, D., & Rogge, L. (1981). Isotonic approximation in L_s. *Journal of Approximation Theory*, *31*(3), 199–223. ISSN 0021-9045. https://doi.org/10.1016/0021-9045(81)90091-5.

Tanaka, H. (2008). Semiparametric least squares estimation of monotone single index models and its application to the iterative least squares estimation of binary choice models.

Xia, Y. (2006). Asymptotic distributions for two estimators of the single-index model. *Econometric Theory*, *22*(6), 1112–1137. ISSN 0266-4666. https://doi.org/10.1017/S0266466606060531.