T E C H N I C A L
R E P O R T

0560

INDEX COEFFICIENTS ESTIMATION IN SINGLE-INDEX MODELS : THE GENERALIZED MAXIMUM RANK CORRELATION ESTIMATOR

GEENENS G., and L. SIMAR

I A P S T A T I S T I C S
N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

Index coefficients estimation in Single-Index Models : the Generalized Maximum Rank Correlation Estimator

GERY GEENENS*
Institut de Statistique
Université Catholique de Louvain, Belgium
geenens@stat.ucl.ac.be

LÉOPOLD SIMAR
Institut de Statistique
Université Catholique de Louvain, Belgium
simar@stat.ucl.ac.be

November 14, 2005

Abstract

We propose a new estimator, called the Generalized Maximum Rank Correlation Estimator (GMRC), of the index coefficients in the context of the so-called Single-Index Model : $Y_i = g(\beta_0' X_i) + \varepsilon_i$, with g and β_0 unknown. The underlying idea is very simple: given a pair of observations (X_i, Y_i) and (X_j, Y_j), if $g(\beta_0' X_i)$ is greater than $g(\beta_0' X_j)$, it is likely that Y_i be greater than Y_j. In other words, the ranks of the Y_i's and the ranks of the $g(\beta_0' X_i)$'s would be highly positively correlated. The clue is thus to estimate β_0 by the value of β which maximizes an estimated version of the rank correlation. Han (1987) proposed such kind of estimation method, but assuming the strict monotonicity of the link function g. We relax this assumption. The estimator is shown to be root-n consistent and asymptotically normal, and has multiple advantages. In particular, an extensive simulation study shows its very good finite-sample behavior : in most of the situations, it seems that the GMRC estimator represents the best choice in practice.

Key words : semiparametric regression; single-index model; rank correlation; index coefficients.

AMS Classification: 62G05, 62G08, 62G20.

*the authors would like to thank Jean-Marie Rolin for useful discussions. Also, financial support from the "Interuniversity Attraction Pole", Phase V (No. P5/24) from the Belgian Science Policy is gratefully acknowledged.

1 Introduction

Let $Z = (X, Y)$ be an observation from a distribution P, on a set $S \subseteq \mathbf{R}^p \otimes \mathbf{R}$. Consider the regression context, where the link between the response variable Y and the vector of covariates X has to be emphasized. In other words, we are searching for the function $m(x)$, such that

$$Y = m(X) + \varepsilon,$$

where ε is a random disturbance with $E(\varepsilon|X) = 0$. Is is well known that this problem can be tackled different ways, depending on the assumptions made about this function m. One of the most popular model in literature during the last decade is certainly the single-index model. For a comprehensive survey about this topic, see Geenens and Delecroix (2005). This model keeps a great part of the flexibility of the nonparametric models, but avoiding the well-known "curse of dimensionality". The hypothesis made about m is that the effect of the covariates on the response is linear, up to an unknown transformation g. Formally, the model is

$$Y = g(\beta_0' X) + \varepsilon. \tag{1.1}$$

From a sample $\{Z_1, ..., Z_n\}$ of independent observations from P, the estimation problem is thus twofold : the function g and the so-called index coefficients vector β_0 have to be estimated. An important remark is that the scale of $\beta_0' X$ can be arbitrarily chosen. Indeed, any pair (index coefficients vector, link function) from the set $\{(c\beta_0, g_c(.) = g(./c)), c \in \mathbf{R}_0\}$ exactly leads to the same regression function m, so that they cannot be distinguished even if the whole distribution P was known. Hence, for identifiability purpose, it is necessary to fix the scale of β_0. We will choose to fix $\beta_0^{(1)} = 1$, where $v^{(k)}$ denotes the kth component of any vector v.

This paper address the problem of estimating the index coefficients. Note that, because of the above identifiability condition, the parameter space is actually a $p-1$ dimensional subset of $\mathbf{R}^p$. If the set of admissible β is taken to be compact, any β can be represented as $\beta_\theta = (1, \theta)$, where $\theta \in \Theta$, Θ being a compact subset of $\mathbf{R}^{p-1}$. One can also note $\beta_0 = (1, \theta_0)$. Therefore, an estimator of β_0 is directly given by an estimator of θ_0, completed with a leading component equal to 1.

Many estimators of these index coefficients have already been proposed. They can be classified in two families : the M-estimators and the direct estimators. Direct estimators provide an analytic form for the estimator of θ_0. From any data set, the estimation is thus very fast and easy. For example, we can mention the average derivative method (ADE), as in Härdle and Stoker (1989), Powell et al (1989) or Hristache et al (2001), and the Sliced Inverse

Regression (SIR), as in Duan and Li (1991) or Zhu and Fang (1996). The M-estimation approach considers the link function as an infinite-dimensional nuisance parameter. The estimator $\widehat{\theta}$ is the minimizer of a certain criterion, where the unknown link function is replaced by a nonparametric estimator $\widehat{g}$:

$$\widehat{\theta} = \arg\min_{\theta\in\Theta}\sum_{i=1}^{n}\Psi\left(Y_i, \widehat{g}(\beta_\theta' X_i)\right), \tag{1.2}$$

with Ψ a contrast function particularizing the estimator. Typical examples of M-estimators are the semiparametric least squares estimator (Ichimura (1993)) and the semiparametric maximum likelihood estimator (Delecroix et al (2003)). The theoretical properties of the M-estimators are much better of those of the direct ones, nevertheless they require solving an intricate optimization problem (1.2), so that their computation is by far slower.

The estimator proposed in this paper is close to M-estimators, as it is given by the solution of an optimization problem, but is not really as in (1.2). It is actually a generalization of the Maximum Rank Correlation estimator (MRC) of Han (1987) and Sherman (1993). The paper is organized as follows : section 2 describes the estimation scheme. Section 3 derives the asymptotic properties of the estimator. Section 4 gives sufficient conditions to guarantee these asymptotic results. Section 5 provides a general form of the variance-covariance matrix of the estimator in terms of primitives of the model. Practical performances of the estimator are analysed in section 6, before the conclusion in section 7.

2 The Generalized Maximum Rank Correlation Estimator

Considering a slightly different model, namely

$$Y = D \text{ o } F(\beta_0' X, \varepsilon), \tag{2.3}$$

where it is only specified that $D : \mathbf{R} \to \mathbf{R}$ is non-degenerate monotonic and $F : \mathbf{R}^2 \to \mathbf{R}$ is strictly monotonic in each of its variables, Han (1987) proposed an estimator based on very intuitive arguments. Taking $D(x) = x$ and $F(u, v) = g(u) + v$, it appears that the Single-Index Model (1.1) is nothing else but a particular case of (2.3), if the link function g is strictly monotonic. Han's results thus hold in the SIM context and his methodology can be particularized in the following way.

Assume, without loss of generality, that g is strictly increasing. Then, given an inequality $\beta_0'X_i > \beta_0'X_j$ for a pair of samples, it is likely that $Y_i > Y_j$. Indeed, as $g(\beta_0'X_i) > g(\beta_0'X_j)$, only the error term can alter the ranking of the Y_i's. If the variance of this error term is fair, this amounts to say that the ranks of the Y_i's and the ranks of the $\beta_0' X_i$'s would be highly positively correlated. Formally, this rank correlation between the Y_i's and the linear combinations $\beta_\theta' X_i$'s, for any θ, is given by

$$\Gamma_n(\theta) = \binom{n}{2}^{-1} \sum_{\rho} \left\{ 1(Y_i > Y_j) 1(\beta_\theta' X_i > \beta_\theta' X_j) + 1(Y_i < Y_j) 1(\beta_\theta' X_i < \beta_\theta' X_j) \right\}, \qquad (2.4)$$

with $1(.)$ being the indicator function and $\sum\limits_{\rho}$ denoting the summation over the $\binom{n}{2}$ combinations of two distinct elements (i,j) from $\{1,...,n\}$. Note that $\Gamma_n(\theta)$ is the count of the pairs of samples for which the Y's and the $\beta_\theta' X$'s are in concordance. Based on the above idea, the estimator $\widehat{\theta}_n$ of θ_0 is defined as the value which maximizes $\Gamma_n(\theta)$:

$$\widehat{\theta}_n = \arg\max_{\theta \in \Theta} \Gamma_n(\theta).$$

Han (1987) proved the strong consistency of this estimator, while Sherman (1993) showed its root-n consistency and derived its asymptotic normality.

A drawback of the above procedure is obviously the monotonic assumption on the link g. In most of situations, we do not have enough information about it to state that it is the case. Worse, in some phenomena, we know for a fact that it is not. A generalization of the above algorithm is thus needed. The idea is the following : no matter the link function g, if a pair of samples is such that $Y_i > Y_j$, it is likely that $g(\beta_0' X_i) > g(\beta_0' X_j)$. Therefore, the ranks of the Y_i's and the ranks of the $g(\beta_0' X_i)$'s would also be highly positively correlated. Similarly to (2.4), for any θ, the rank correlation between observed and fitted values is given by

$$T_n^*(\theta) = \binom{n}{2}^{-1} \sum_{\rho} \left\{ 1(Y_i > Y_j) 1(g(\beta_\theta' X_i) > g(\beta_\theta' X_j)) + 1(Y_i < Y_j) 1(g(\beta_\theta' X_i) < g(\beta_\theta' X_j)) \right\} \qquad (2.5)$$

and a good estimator of θ_0 should be the value which maximizes $T_n^*(\theta)$. Of course, in the single-index context, such an estimator is not feasible since the link function g is unknown. The way to overcome the problem is to consider it as a nuisance parameter, as it was done in classical M-estimation methods. The link function in (2.5) is thus replaced by any consistent nonparametric estimator, for example the Nadaraya-Watson estimator

$$\widehat{g}^{\theta,h}(u) = \frac{\sum\limits_{i=1}^{n} K\left(\frac{u - \beta_\theta' X_i}{h}\right) Y_i}{\sum\limits_{i=1}^{n} K\left(\frac{u - \beta_\theta' X_i}{h}\right)},$$

with K a kernel function and h a bandwidth. The practical criterion to be maximized is thus

$$T_n(\theta) = \frac{2}{n(n-1)} \sum_{\rho} \{1(Y_i > Y_j) 1(\widehat{g}^{\theta,h}(\beta_\theta' X_i) > \widehat{g}^{\theta,h}(\beta_\theta' X_j)) + 1(Y_i < Y_j) 1(\widehat{g}^{\theta,h}(\beta_\theta' X_i) < \widehat{g}^{\theta,h}(\beta_\theta' X_j))\}$$

which can be written as

$$T_n(\theta) \doteq \frac{2}{n(n-1)} \sum_{\rho} \Phi(Y_i, Y_j, \widehat{g}^{\theta,h}(\beta_\theta' X_i), \widehat{g}^{\theta,h}(\beta_\theta' X_j)). \tag{2.6}$$

The estimator of θ_0, called the Generalized Maximum Rank Correlation Estimator (GMRC estimator), is finally given by

$$\widehat{\theta}_n = \arg\max_{\theta \in \Theta} T_n(\theta).$$

In a sense, the GMRC estimator is a M-estimator as it is the optimizer of the criterion $T_n(\theta)$. However, we observe a difference between (2.6) and the criterion in (1.2) : the function Φ admits two sample values as arguments, while the function Ψ admits only one.

3 Asymptotic theory

In this section we establish the root-n consistency and the asymptotic normality of the Generalized Maximum Rank Correlation Estimator. The proof follows the arguments developped in Han (1987) and Sherman (1993), and is built on the following assumptions :

(1) $\{\varepsilon_i,\ i = 1, ..., n\}$ *are i.i.d. random variables* ;

(2) $\{X_i\}$ *are i.i.d. random p-vectors, independent of* $\{\varepsilon_i\}$, *with distribution function* F_X *such that*

   (i) *the support* S_X *of* F_X *is not contained in any proper linear subspace of* $\mathbf{R}^p$, *and*

   (ii) *there exists at least one regressor, say* $X^{(1)}$, *with everywhere positive Lebesgue density, conditional on the set of the other covariates, say* $X^{(-1)}$, *almost surely* $P_{X^{(-1)}}$;

(3) *the link function* $g : \mathbf{R} \to \mathbf{R}$ *is continuous and does not remain constant on any interval of* $\mathbf{R}$;

(4) θ_0 *is an interior point of* Θ, *a compact subset of* $\mathbf{R}^{p-1}$;

(5) *denoting*

$$\tau(z,\theta) = E\left\{1(y > Y)1(g(\beta_\theta' x) > g(\beta_\theta' X)) + 1(y < Y)1(g(\beta_\theta' x) < g(\beta_\theta' X))\right\}$$

$\forall z = (x, y) \in S$ *and* $\theta \in \Theta$, *and* $\mathcal{N}$ *a neighborhood of* θ_0, *we have :*

(i) *for all* z *in* S, *all mixed second partial derivatives of* $\tau(z, .)$ *exist on* $\mathcal{N}$;

(ii) *there is an integrable function* $M(z)$ *verifying for all* z *in* S *and* θ *in* $\mathcal{N}$

$$\|\nabla_2\tau(z,\theta) - \nabla_2\tau(z,\theta_0)\| \leq M(z)\,\|\theta - \theta_0\| \ ;$$

(iii) $E\,\|\nabla_1\tau(.,\theta_0)\|^2 < \infty$;

(iv) $E\,|\nabla_2|\,\tau(.,\theta_0) < \infty$;

(6) *the nonparametric estimator* $\widehat{g}^{\theta,h}$ *is a consistent estimator of the link function* g.

Note : ∇_k denotes the kth partial derivative operator with respect to θ,

$$|\nabla_k| f(\theta) = \sum_{i_1,\ldots,i_k} \left| \frac{\partial^k}{\partial\theta_{i_1}\ldots\partial\theta_{i_k}} f(\theta) \right|$$

for any function f, and $\|.\|$ denotes the matrix norm.

Assumptions (1) and (4) are classical assumptions in the context of Single-Index models. Assumptions (2) and (3) are designed in order to ensure the identification of the model. Remark that the hypotheses on the link function are very mild. The continuity is needed to allow the use of a classical nonparametric estimator of g in (2.6), while functions g with "flat" parts are to be excluded to avoid ex-aequos in the ranking of the $g(\beta_\theta' X_i)$'s. In the presence of ties, only the error terms ε_i's should determine the ranking of the Y_i's, so that consistency could not be reached. Assumption (5) gives regularity conditions to allow a Taylor expansion of $\tau(z,\theta)$. Practical sufficient conditions for satisfying assumption (5) will be given is section 4. Assumption (6) is obvious. Note that no particular rate of convergence for the bandwidth sequence is needed, except the classical $h \to 0$ and $nh \to \infty$ ensuring the consistency of $\widehat{g}^{\theta,h}$.

theorem 3.1. *Under assumptions (1-6),* $\widehat{\theta}_n$ *is a consistent estimator of* θ_0, *and*

$$\sqrt{n}(\theta - \theta_0) \xrightarrow{\mathcal{L}} N(0, V^{-1}\Delta V^{-1}), \tag{3.7}$$

with $V = \frac{1}{2}E[\nabla_2\tau(Z,\theta_0)]$ *and* $\Delta = E[\nabla_1\tau(Z,\theta_0)(\nabla_1\tau(Z,\theta_0))']$.

Proof. Due to its length, the proof is split into 2 steps : we first establish the consitency of the estimator, and then we derive its rate of convergence and its asymptotic normality.

Step 1 : consistency

Consider the criterion $T_n(\theta)$, the idealized criterion $T_n^*(\theta)$ and define

$$T(\theta) = E[T_n^*(\theta)] = P\left[Y_i > Y_j,\ g(\beta_\theta' X_i) > g(\beta_\theta' X_j)\right].$$

Define also $\delta_n(\theta)$ as the difference between $T_n(\theta)$ and $T_n^*(\theta)$, i.e.

$$\delta_n(\theta) = \frac{1}{n(n-1)} \sum_{i \neq j} 1(Y_i > Y_j) \left[1(\widehat{g}^{\theta,h}(\beta_\theta' X_i) > \widehat{g}^{\theta,h}(\beta_\theta' X_j)) - 1(g(\beta_\theta' X_i) > g(\beta_\theta' X_j)) \right],$$

and note that $\delta_n(\theta) = o_P(1)$ due to the consistency of $\widehat{g}^{\theta,h}$, by assumption (6). Now let (Z_i, Z_j) be a pair of samples such that $g(\beta_0' X_i) > g(\beta_0' X_j)$. Then we have

$$P_{\varepsilon|X}(Y_i > Y_j) > P_{\varepsilon|X}(Y_i < Y_j).$$

Indeed,

$$\begin{aligned} P_{\varepsilon|X}(Y_i > Y_j) &= P_{\varepsilon|X}\left(g(\beta_0' X_i) + \varepsilon_i > g(\beta_0' X_j) + \varepsilon_j\right) \\ &= P_{\varepsilon|X}\left(\varepsilon_i - \varepsilon_j > g(\beta_0' X_j) - g(\beta_0' X_i)\right) \end{aligned} \tag{3.8}$$

which is greater than 1/2 since ε_i and ε_j are two replications of the same random variable. Next, let

$$t_{ij}(\theta) = E_X\left[P_{\varepsilon|X}(Y_i > Y_j) 1(g(\beta_\theta' X_i) > g(\beta_\theta' X_j)) + P_{\varepsilon|X}(Y_i < Y_j) 1(g(\beta_\theta' X_i) < g(\beta_\theta' X_j))\right]$$

and denote $f(x_1, x_2)$ the joint density of (X_i, X_j),

$$A_1(\theta) = \{x = (x_1, x_2) \in \mathbf{R}^p \times \mathbf{R}^p : g(\beta_\theta' x_1) > g(\beta_\theta' x_2)\}$$

and

$$A_2(\theta) = \{x = (x_1, x_2) \in \mathbf{R}^p \times \mathbf{R}^p : g(\beta_\theta' x_1) < g(\beta_\theta' x_2)\}.$$

Then observe that $t_{ij}(\theta)$ can be written as

$$t_{ij}(\theta) = \iint_{\mathbf{R}^p \times \mathbf{R}^p} \alpha(x_1, x_2; \theta) f(x_1, x_2) dx_1 dx_2,$$

where the function α is defined as

$$\alpha(x_1, x_2; \theta) = \begin{cases} P_{\varepsilon|X}(Y_i > Y_j) & \text{if } (x_1, x_2) \in A_1(\theta) \\ P_{\varepsilon|X}(Y_i < Y_j) & \text{if } (x_1, x_2) \in A_2(\theta) \end{cases}.$$

From (3.8), it appears that $\alpha(x_1, x_2; \theta_0)$ is always greater than 1/2, which amounts to say that $t_{ij}(\theta_0)$ is a maximum of the function $t_{ij}(\theta)$. Next it is easy to see that[1]

$$P\left[(X_i, X_j) \in \{(x_1, x_2) \in S_X \times S_X : g(\beta_0' x_1) > g(\beta_0' x_2),\, g(\beta_\theta' x_1) < g(\beta_\theta' x_2)\}\right] > 0. \qquad (3.9)$$

Therefore, the function $\alpha(x_1, x_2; \theta)$ is not maximum for all $(x_1, x_2) \in S_X \times S_X$, and there exists some η such that $t_{ij}(\theta_0) > \eta > t_{ij}(\theta)$. θ_0 is thus the unique maximizer of t_{ij} for each set of (i, j), and therefore is also the unique maximizer of the sum $T(\theta)$. The uniform convergence of the criterion $T_n^*(\theta)$ towards its expectation $T(\theta)$ is direct, due to the U-statistic structure of T_n^*, and by following the step 2 of the proof of Han (1987). We find

$$\sup_{\theta \in \Theta} |T_n^*(\theta) - T(\theta)| \xrightarrow{P} 0.$$

Now, since $\widehat{\theta}_n$ is the maximizer of $T_n(\theta)$, we have :

$$\begin{aligned} T_n(\widehat{\theta}_n) &\geq \max_{\theta \in \Theta} T_n(\theta) \\ T_n^*(\widehat{\theta}_n) + \delta_n(\widehat{\theta}_n) &\geq \max_{\theta \in \Theta} \left[T_n^*(\theta) + \delta_n(\theta)\right] \\ &\geq \max_{\theta \in \Theta} T_n^*(\theta) + \min_{\theta \in \Theta} \delta_n(\theta), \end{aligned}$$

so that

$$T_n^*(\widehat{\theta}_n) \geq \max_{\theta \in \Theta} T_n^*(\theta) - o_p(1).$$

The consistency of $\widehat{\theta}_n$ then follows from the theorem 5.7 of van der Vaart (1998) :

$$\widehat{\theta}_n \xrightarrow{P} \theta_0.$$

Step 2 : $\sqrt{n}$-consistency and asymptotic normality

Consider

$$S_n^*(\theta) = T_n^*(\theta) - T_n^*(\theta_0)$$

and

$$S(\theta) = T(\theta) - T(\theta_0).$$

As T, S is maximized in θ_0 and its maximal value is thus 0. For all $(z_1, z_2) \in S \times S$ and for all $\theta \in \Theta$, define

$$\begin{aligned} f(z_1, z_2, \theta) = {} & 1\{y_1 > y_2\}[1\{g(\beta_\theta' x_1) > g(\beta_\theta' x_1)\} - 1\{g(\beta_0' x_1) > g(\beta_0' x_1)\}] \\ & + 1\{y_1 < y_2\}[1\{g(\beta_\theta' x_1) < g(\beta_\theta' x_1)\} - 1\{g(\beta_0' x_1) < g(\beta_0' x_1)\}]. \end{aligned}$$

[1]a proof is provided in the appendix.

Since $S_n^*(\theta)$ is a U-statistic of order 2 with expectation $S(\theta)$, we may apply the decomposition (5) of Sherman (1993) and write

$$S_n^*(\theta) = S(\theta) + \frac{1}{n}\sum_i a(Z_i,\theta) + \frac{2}{n(n-1)}\sum_\rho b(Z_i,Z_j,\theta) \tag{3.10}$$

with

$$a(z,\theta) = E[f(z,Z,\theta)] + E[f(Z,z,\theta)] - 2S(\theta)$$

and

$$b(z_1,z_2,\theta) = f(z_1,z_2,\theta) - E[f(z_1,Z,\theta)] - E[f(Z,z_2,\theta)] + S(\theta).$$

Assumption (5i) allows to expand $\tau(z,\theta)$ around $\tau(z,\theta_0)$:

$$\tau(z,\theta) = \tau(z,\theta_0) + (\theta-\theta_0)'\nabla_1\tau(z,\theta_0) + \frac{1}{2}(\theta-\theta_0)'\nabla_2\tau(z,\widetilde{\theta})(\theta-\theta_0)$$

with $\widetilde{\theta}$ between θ and θ_0. By assumption (5ii) we have

$$\|(\theta-\theta_0)'[\nabla_2\tau(z,\theta) - \nabla_2\tau(z,\theta_0)](\theta-\theta_0)\| \le M(z)||\theta-\theta_0||^3.$$

Remark also that

$$2S(\theta) = E[\tau(Z,\theta) - \tau(Z,\theta_0)],$$

so that we can write

$$2S(\theta) = (\theta-\theta_0)'E[\nabla_1\tau(Z,\theta_0)] + \frac{1}{2}(\theta-\theta_0)'E[\nabla_2\tau(Z,\theta_0)](\theta-\theta_0) + o(||\theta-\theta_0||^2).$$

Since $S(\theta)$ admits a maximum, the coefficient $E[\nabla_1\tau(z,\theta_0)]$ of the linear term in the above expression must be zero, and V is seen to be negative definite. We finally find

$$S(\theta) = \frac{1}{2}(\theta-\theta_0)'V(\theta-\theta_0) + o(||\theta-\theta_0||^2). \tag{3.11}$$

Now, note that $a(z,\theta) = \tau(z,\theta) - \tau(z,\theta_0) - 2S(\theta)$. Hence, we have

$$\begin{aligned}
\frac{1}{n}\sum_i a(Z_i,\theta) &= \frac{1}{n}\sum_i \tau(Z_i,\theta) - \frac{1}{n}\sum_i \tau(Z_i,\theta_0) - 2S(\theta)\\
&= \frac{1}{n}\sum_i (\theta-\theta_0)'\nabla_1\tau(Z_i,\theta_0) + \frac{1}{2n}\sum_i (\theta-\theta_0)'\nabla_2\tau(Z_i,\theta_0)(\theta-\theta_0)\\
&- (\theta-\theta_0)'V(\theta-\theta_0) + o(||\theta-\theta_0||^2)\\
&= \frac{1}{\sqrt{n}}(\theta-\theta_0)'W_n + \frac{1}{2}(\theta-\theta_0)'D_n(\theta-\theta_0)' + o(||\theta-\theta_0||^2) + R_n(\theta)
\end{aligned}$$

with

$$W_n = \frac{1}{\sqrt{n}}\sum_i \nabla_1\tau(Z_i,\theta_0)$$
$$D_n = \frac{1}{n}\sum_i \nabla_2\tau(Z_i,\theta_0) - 2V$$

and

$$||R_n(\theta)|| \le \frac{1}{n}\sum_i M(Z_i)\ ||\theta-\theta_0||^3.$$

By assumption (5iii), the fact that $E[\nabla_1\tau(z,\theta_0)] = 0$ and the Central Limit Theorem, we have that $W_n \xrightarrow{\mathcal{L}} N(0,\Delta)$ with $\Delta = E[\nabla_1\tau(z,\theta_0)(\nabla_1\tau(z,\theta_0))']$. By assumption (5iv) and a weak law of large numbers, $D_n \xrightarrow{P} 0$. Finally the integrability of M and a weak law of large numbers imply that $R_n(\theta) = o_P(||\theta-\theta_0||^2)$. We then conclude that

$$\frac{1}{n}\sum_i a(Z_i,\theta) = \frac{1}{\sqrt{n}}(\theta-\theta_0)'W_n + o_P(||\theta-\theta_0||^2). \tag{3.12}$$

By assumption (2ii), $\beta_\theta' X$ is absolutely continuous for any θ. Therefore, with assumption (3), we find that $P[g(\beta_0'X_1) = g(\beta_0'X_2)] = 0$. For any $(z_1,z_2) \in S\times S$, $f(z_1,z_2,\theta)$ is thus continuous and bounded in θ_0 almost surely. This remains true for $b(z_1,z_2,\theta)$. Since $f(z_1,z_2,\theta_0) \equiv 0$, one can deduce that $E[b(Z_1,Z_2,\theta)^2] \to 0$ when $\theta \to \theta_0$. The Euclidean properties of the class of functions $\{b(.,.,\theta) : \theta \in \Theta\}$ can be found on the same way as in Sherman (1993) section 5, using the 5-dimensional vector space $\mathcal{H}_\theta = \{h_\theta(.,.,.;\gamma,\gamma_1,\gamma_2,\delta_1,\delta_2) : \gamma,\gamma_1,\gamma_2,\delta_1,\delta_2 \in \mathbf{R}\}$, with

$$h_\theta(z_1,z_2,t;\gamma,\gamma_1,\gamma_2,\delta_1,\delta_2) = \gamma t + \gamma_1 y_1 + \gamma_2 y_2 + \delta_1 g(\beta_\theta' x_1) + \delta_2 g(\beta_\theta' x_2)$$

for any $(z_1,z_2,t) \in S\times S\times\mathbf{R}$. Then, using theorem 3 of Sherman (1993), we have that

$$\frac{2}{n(n-1)}\sum_\rho b(Z_i,Z_j,\theta) = o_P(\frac{1}{n}). \tag{3.13}$$

Using (3.11), (3.12) and (3.13), we can rewrite (3.10) as

$$S_n^*(\theta) = \frac{1}{2}(\theta-\theta_0)'V(\theta-\theta_0) + \frac{1}{\sqrt{n}}(\theta-\theta_0)'W_n + o_P(||\theta-\theta_0||^2) + o_P(\frac{1}{n}).$$

Now write

$$S_n(\theta) = T_n(\theta) - T_n(\theta_0)$$

and see that

$$S_n(\theta) = S_n^*(\theta) + \delta_n(\theta) - \delta_n(\theta_0).$$

For all $\theta \in \Theta$, $\delta_n(\theta)$ can be written as $\Delta_n(\theta)/n$, with

$$\Delta_n(\theta) = \frac{1}{(n-1)} \sum_{i \neq j} 1(Y_i > Y_j) \left[1(\widehat{g}^{\theta,h}(\beta_\theta' X_i) > \widehat{g}^{\theta,h}(\beta_\theta' X_j)) - 1(g(\beta_\theta' X_i) > g(\beta_\theta' X_j)) \right].$$

It follows that $\forall \varepsilon > 0$,

$$P\left[|\Delta_n(\theta)| > \varepsilon \right] \leq P\left[\frac{1}{(n-1)} \sum_{i \neq j} \left| 1(\widehat{g}^{\theta,h}(\beta_\theta' X_i) > \widehat{g}^{\theta,h}(\beta_\theta' X_j)) - 1(g(\beta_\theta' X_i) > g(\beta_\theta' X_j)) \right| > \varepsilon \right].$$

Write $\lambda_n(\theta)$ for $\left| 1(\widehat{g}^{\theta,h}(\beta_\theta' X_i) > \widehat{g}^{\theta,h}(\beta_\theta' X_j)) - 1(g(\beta_\theta' X_i) > g(\beta_\theta' X_j)) \right|$. $\lambda_n(\theta)$ is a random variable such that

$$\lambda_n(\theta) = \begin{cases} 1 & \text{with probability } \pi_n \\ 0 & \text{with probability } 1 - \pi_n \end{cases},$$

with π_n being the probability of discordance between the pairs $(\widehat{g}^{\theta,h}(\beta_\theta' X_i), \widehat{g}^{\theta,h}(\beta_\theta' X_j))$ and $(g(\beta_\theta' X_i), g(\beta_\theta' X_j))$. We find that

$$P\left[n\lambda_n(\theta) > \varepsilon \right] = \begin{cases} \pi_n & \text{if } 0 < \varepsilon < n \\ 0 & \text{if } \varepsilon \geq n \end{cases}$$

and, since $\pi_n \to 0$ due to the consistency of $\widehat{g}^{\theta,h}$,

$$\lim_{n \to \infty} P\left[n\lambda_n(\theta) > \varepsilon \right] = 0 \quad \forall \varepsilon > 0.$$

This means that $\lambda_n(\theta) = o_p(\frac{1}{n})$, and we have that

$$\lim_{n \to \infty} P\left[|\Delta_n(\theta)| > \varepsilon \right] \leq \lim_{n \to \infty} P\left[\frac{1}{(n-1)} \sum_{i \neq j} \lambda_n(\theta) > \varepsilon \right] = 0.$$

This amounts to say that $\forall \theta \in \Theta$, $\delta_n(\theta) = o_P(\frac{1}{n})$, so that

$$S_n(\theta) = \frac{1}{2}(\theta - \theta_0)' V (\theta - \theta_0) + \frac{1}{\sqrt{n}}(\theta - \theta_0)' W_n + o_P(||\theta - \theta_0||^2) + o_P(\frac{1}{n}).$$

Therefore, since V is a negative definite matrix, we can apply to $S_n(\theta)$ theorems 1 and 2 of Sherman (1993) and conclude that

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \xrightarrow{\mathcal{L}} N(0, V^{-1} \Delta V^{-1}).$$

$\square$

4 Sufficient conditions for assumption 5

A key assumption in the theorem is assumption (5). Similarly to what is done in Sherman (1993), we derive here sufficient conditions on the distribution of Z to guarantee this assumption. Remark that, for any $z \in \mathbf{R} \times S_X$ and $\theta \in \Theta$, the function $\tau(z, \theta)$ can be written

$$\tau(z,\theta) = \int_{-\infty}^{y} \int_{-\infty}^{g(\beta_\theta' x)} f^{\theta}_{g(U),Y}(w,t)\ dw\ dt + \int_{y}^{\infty} \int_{g(\beta_\theta' x)}^{\infty} f^{\theta}_{g(U),Y}(w,t)\ dw\ dt.$$

and that

$$\begin{aligned} f^{\theta}_{g(U),Y}(w,t) &= f_{Y|g(U)}(t|w)\ f^{\theta}_{g(U)}(w) \\ &= f_{\varepsilon}(t-w)\ f^{\theta}_{g(U)}(w) \end{aligned}$$

with $U = \beta_\theta' X$, $f^{\theta}_{g(U),Y}$ the joint density of $g(U)$ and Y, $f_{Y|g(U)}$ the density of Y conditional to $g(U)$, f_ε the marginal density of the residual ε and $f^{\theta}_{g(U)}$ the marginal density of $g(U)$. Now suppose that the function g has a continuous derivative at every point and that the derivative is 0 only at the points $a_1, a_2, ..., a_{k-1}$, where $a_1 < a_2 < ... < a_{k-1}$. Let $\mathbf{R}$ be partitioned into k disjoints intervals $I_1, ..., I_k$, such that g is either monotonic increasing or monotonic decreasing (thus invertible) in the ith interval $]a_{i-1}, a_i[$, for each i, and denote $g_i(u)$ the restriction of g to the interval I_i. We have

$$g(u) = \sum_{i=1}^{k} g_i(u)\ 1_{\{u \in I_i\}}. \tag{4.14}$$

Let $u_i(w) = g_i^{-1}(w)$ be the solution of $w = g(u)$ for u in the ith interval. By the above assumption on g, du_i/dw exists for all w, and it is known that the density of $g(U)$ can be written

$$f^{\theta}_{g(U)}(w) = \sum_{\lambda} f^{\theta}_U(u_i(w)) \left| \frac{du_i}{dw}(w) \right|,$$

where the summation λ is taken over the values of i for which $g_i(u) = w$ for some value of u in the ith interval I_i, and f_U is the marginal density of the index U.

We find

$$\begin{aligned} f^{\theta}_{g(U),Y}(w,t) &= f_\varepsilon(t-w) \sum_{\lambda} f^{\theta}_U(u_i(w)) \left| \frac{du_i}{dw}(w) \right| \\ &= f_\varepsilon(t-w) \sum_{\lambda} \int f_{X^{(1)}|X^{(-1)}}(u_i(w) - \theta' r \mid r) f_{X^{(-1)}}(r)\ dr\ \left| \frac{du_i}{dw}(w) \right|, \end{aligned}$$

with $f_{X^{(1)}|X^{(-1)}}$ the density of the first component of vector X conditional to the others components and $f_{X^{(-1)}}$ the density of the vector $(X^{(2)}, ..., X^{(p)})$. Consequently, it is shown that assumption 5 is met if $f_{X^{(1)}|X^{(-1)}}(.|X^{(-1)} = r)$ has bounded derivatives up to order three for each r in the support of $X^{(-1)}$, so that this assumption is not very restrictive.

5 The asymptotic covariance matrix

The theorem should be much more attractive if the covariance matrix in (3.7) was given in terms of model primitives. It is possible to return to them after some arithmetic developments. Note that

$$\tau(Z,\theta) = \int\limits_{g(\beta_\theta' x)<g(\beta_\theta' X)} \xi(Y, g(\beta_0' x)) f_X(x)dx + \int \rho(Y, g(\beta_0' x)) f_X(x)dx, \tag{5.15}$$

with f_X the density of vector X, $\xi(y,w) = E[1(y > Y) - 1(y < Y)|g(\beta_0' X) = w]$ and $\rho(y,w) = E[1(y < Y)|g(\beta_0' X) = w]$. Denote $\nabla^{(i)}$ the ith component of ∇_1, the first partial derivative with respect to θ. We have that

$$\nabla^{(i)}\tau(Z,\theta_0) = \lim_{\varepsilon\to 0}\varepsilon^{-1}[\tau(Z,\theta_0+\varepsilon a_i) - \tau(Z,\theta_0)],$$

where a_i is the unit vector of $\mathbf{R}^{p-1}$ with ith component equal to 1. With (5.15), we have

$$\tau(Z,\theta_0+\varepsilon a_i) - \tau(Z,\theta_0) = \int\limits_{g((\beta_0+\varepsilon b_{i+1})' x)<g((\beta_0+\varepsilon b_{i+1})' X)} \xi(Y, g(\beta_0' x)) f_X(x)dx - \int\limits_{g(\beta_0' x)<g(\beta_0' X)} \xi(Y, g(\beta_0' x)) f_X(x)dx$$

with b_i the unit vector of $\mathbf{R}^p$ with ith component equal to 1. If $\frac{dg}{du}(u)$ exists, we have also that

$$g((\beta_0+\varepsilon b_{i+1})' x) = g(\beta_0' x + \varepsilon x^{(i+1)}) = g(\beta_0' x) + \varepsilon x^{(i+1)}\frac{dg}{du}(\beta_0' x^*),$$

with x^* such that $\beta_0' x^*$ is between $\beta_0' x$ and $\beta_0' x + \varepsilon x^{(i+1)}$. Therefore, we find that

$$\tau(Z,\theta_0+\varepsilon a_i) - \tau(Z,\theta_0) = \int\limits_{g(\beta_0' x)<g(\beta_0' X)+\varepsilon(X^{(i+1)}\frac{dg}{du}(\beta_0' X^*) - x^{(i+1)}\frac{dg}{du}(\beta_0' x^*))} \xi(Y, g(\beta_0' x)) f_X(x)dx - \int\limits_{g(\beta_0' x)<g(\beta_0' X)} \xi(Y, g(\beta_0' x)) f_X(x)dx.$$

Now define the random variable W as $g(\beta_0'X)$. Although g is not injective, it is possible to define a "generalized" inverse g^{-1} of g, based on the decomposition (4.14) of g :

$$g^{-1}(w_u) = \sum_{i=1}^{k} g_i^{-1}(w_u)\ 1(u \in I_i),$$

if it is known that $w_u = g(u)$. Then apply the change of variables from $x = (x^{(1)}, r)$ to $(g(\beta_0'x), r)$. It is found that

$$\tau(Z, \theta_0 + \varepsilon a_i) - \tau(Z, \theta_0) = \int I_Y(r) f_{X^{(-1)}}(r) dr$$

where

$$I_Y(r) = \int_{g(\beta_0'X)}^{g(\beta_0'X)+\varepsilon(X^{(i+1)}\frac{dg}{du}(\beta_0'X^*)-x^{(i+1)}\frac{dg}{du}(\beta_0'x^*))} \xi(Y, w) \left| \frac{dg^{-1}}{dw}(w) \right| f_{W|X^{(-1)}}(w|r) dw.$$

When $\varepsilon \to 0$, $I_Y(r)$ equals

$$\varepsilon(X^{(i+1)} - x^{(i+1)}) \frac{dg}{du}(\beta_0'X) \xi(Y, g(\beta_0'X))) \left| \frac{dg^{-1}}{dw}(g(\beta_0'X)) \right| f_{W|X^{(-1)}}(g(\beta_0'X)|r) + o(\varepsilon),$$

so that, by integration and using the fact that $\frac{dg^{-1}}{dw}(w) = \left(\frac{dg}{du}(g^{-1}(w))\right)^{-1}$,

$$\nabla_1 \tau(Z, \theta_0) = \left(X^{(-1)} - E[X^{(-1)}|g(\beta_0'X)]\right) \xi(Y, g(\beta_0'X)) f_W(g(\beta_0'X)) sign\left(\frac{dg}{du}(\beta_0'X)\right),$$

with $sign(x) = x/|x|$. From the expression of Δ given in the theorem, we find that

$$\Delta = E\left[(X^{(-1)} - E[X^{(-1)}|g(\beta_0'X)])(X^{(-1)} - E[X^{(-1)}|g(\beta_0'X)])' \xi^2(Y, g(\beta_0'X)) f_W^2(g(\beta_0'X))\right].$$

Also, denote $\lambda(y, w) = \xi(y, w) f_W(w) \frac{dg^{-1}}{dw}(w)$. Similar derivations as above lead to

$$2V = E\left[(X^{(-1)} - E[X^{(-1)}|g(\beta_0'X)])(X^{(-1)} - E[X^{(-1)}|g(\beta_0'X)])' \frac{\partial \lambda}{\partial w}(Y, g(\beta_0'X)) \left(\frac{dg}{du}(\beta_0'X)\right)^2\right].$$

Then, remark that, for each w, $\xi(Y, w)$ is bounded and has symmetric distribution around 0 conditional to $g(\beta_0'X) = w$, so that

$$E\left[\xi(Y, w)|g(\beta_0'X) = w\right] = 0.$$

We finally find

$$2V = E[(X^{(-1)} - E[X^{(-1)}|g(\beta_0'X)])(X^{(-1)} - E[X^{(-1)}|g(\beta_0'X)])' \times \frac{\partial \xi}{\partial w}(Y, g(\beta_0'X)) f_W(g(\beta_0'X)) \left|\frac{dg}{du}(\beta_0'X)\right|].$$

6 A simulation study

The aim of this section is to compare the practical performances of the GMRC estimator with other index coefficients estimators. First of all, we have considered the Single-Index Model (1.1) with $p = 2$,

$$\begin{aligned} g(u) &= u^2, \\ \beta_0 &= (1, 2)', \\ X_1 &\sim N(0, 1), \\ X_2 &\sim Bern(0.5), \\ \varepsilon &\sim N(0, 0.05). \end{aligned}$$

The sample size is set to $n = 50$, $n = 100$, $n = 200$ and $n = 500$, and 500 Monte-Carlo replications are drawn in each case, and we estimate θ_0 by the GMRC estimator and by two reliable classical methods : Semiparametric Least Squares (SLS) and Semiparametric Maximum Likelihood (SML)[2]. The mean and the Mean Squared Error of the estimations of θ_0 are given in the table below :

	$n = 50$		$n = 100$		$n = 200$		$n = 500$	
	mean($\widehat{\theta}$)	MSE($\widehat{\theta}$)	mean($\widehat{\theta}$)	MSE($\widehat{\theta}$)	mean($\widehat{\theta}$)	MSE($\widehat{\theta}$)	mean($\widehat{\theta}$)	MSE($\widehat{\theta}$)
GMRC	$1,9095$	$\mathbf{0,0667}$	$1,9282$	$\mathbf{0,0322}$	$1,9800$	$\mathbf{0,0187}$	$2,0112$	$0,0086$
SLS	$1,8047$	$0,2000$	$1,8489$	$0,0820$	$1,8884$	$0,0282$	$1,8887$	$0,0150$
SML	$1,8928$	$0,5104$	$1,7993$	$0,1454$	$1,8509$	$0,0585$	$1,9600$	$\mathbf{0,0082}$

It clearly appears that the GMRC estimator performs better than the other two methods for small to moderately large samples. Actually, the nonparametric estimation of g, what we can expect to be poor when n is small, does not directly arise in criterion (2.6), as it does in criterion (1.2), but only through indicator functions. This fact can explained the better behavior of the GMRC compared with classical M-estimators. Note that for larger n, the Semiparametric Maximum Likelihood seems to become slightly better, what could also be expected as it is known that this estimator is asymptotically efficient. Nevertheless, the GMRC remains very good.

We have also compared the practical performance of the GMRC versus the MRC estimator in the case of an increasing link function. We used the model (1.1) with $p = 2$

[2]The distribution of the covariates is supposed unknown and is nonparametrically estimated : this is what Geenens and Delecroix (2005) called the "Ignorant Semiparametric Maximum Likelihood".

and

$$\begin{aligned}
g(u) &= u^3, \\
\beta_0 &= (1,2)', \\
X_1 &\sim N(0,1), \\
X_2 &\sim N(0,1), \\
\varepsilon &\sim N(0, 0.05).
\end{aligned}$$

We set the sample size to $n = 50$, $n = 100$ and $n = 200$ and again for each case 500 Monte Carlo replications were drawn. We estimated θ_0 by the GMRC and the MRC estimators. We obtained :

	$n = 50$		$n = 100$		$n = 200$	
	mean($\widehat{\theta}$)	MSE($\widehat{\theta}$)	mean($\widehat{\theta}$)	MSE($\widehat{\theta}$)	mean($\widehat{\theta}$)	MSE($\widehat{\theta}$)
GMRC	2, 0081	**0, 0253**	2, 0021	**0, 0078**	1, 9971	**0, 0016**
MRC	2, 0193	0, 0315	2, 0043	0, 0081	1, 9972	0, 0017

An interesting observation is that the GMRC performs as good as the MRC estimator, so that nothing lost by estimating the link function.

A necessary assumption to prove the consistency of the GMRC was to suppose that the link function does not remain constant on any interval of **R**. We studied a third scenario with the same design as the previous one and the logit link function

$$g(u) = \frac{\exp(u)}{1 + \exp(u)},$$

that is a link function almost flat for large absolute values of the index. In this case, we compared the GMRC with the MRC, also probably disadvantaged by the shape of the link, the SLS and the SML. From 500 Monte Carlo replications for sample size set to 50, 100 and 200, we found :

	$n = 50$		$n = 100$		$n = 200$	
	mean($\widehat{\theta}$)	MSE($\widehat{\theta}$)	mean($\widehat{\theta}$)	MSE($\widehat{\theta}$)	mean($\widehat{\theta}$)	MSE($\widehat{\theta}$)
GMRC	1, 4839	**7, 2045**	2, 0329	**6, 1301**	2, 3729	**2, 6000**
MRC	3, 0514	8, 0274	3, 0826	7, 0986	2, 4516	2, 7048
SLS	1, 4408	9, 8741	1, 9661	7, 2968	2, 3653	3, 1638
SML	1, 2308	10, 0190	1, 6675	7, 6794	1, 6582	6, 0794

Due to the shape of the link, the MSE of the estimates are much larger than for the second scenario. However, once again, it is seen that the GMRC estimator remains the best choice, even better than the SLS and the SML estimators.

7 Conclusion

In this paper we have introduced a new estimator of the index coefficients in a Single-Index Model. The underlying idea of the estimation procedure is very intuitive : the ranks of the real unobservable data and the ranks of the observed noisy data should be highly positively correlated. The estimator of the index coefficients is thus given by the vector which maximizes an estimated version of this rank correlation, where the unknown link function g is considered as a infinite-dimensional nuisance parameter and replaced by any consistent nonparametric estimator $\widehat{g}$. Despite the discontinuous nature of the criterion T_n to be maximized, we have developed an asymptotic theory for the estimator, mainly based on the U-statistic structure of an idealized criterion T_n^*, very close to T_n. The estimator is shown to be root-n consistent and asymptotically normal. The advantages of this estimator are multiple : first, the criterion T_n is very simple and fast to compute as it is nothing else but a count of pairs of samples in concordance. Second, the estimated criterion is less influenced by the nonparametric estimation of the link than other classical M-estimators as $\widehat{g}$ arises in T_n only through a rank correlation. This fact can play an important role, especially when the sample size is small. Third, the assumptions made in order to show the asymptotic properties of the estimator are very mild : for the link for example, it is only asked that it is continuous and never flat. Also, the nonparametric estimator $\widehat{g}$ is only supposed to be consistent, without other restriction on the smoothing parameter. Finally, the simulation study has shown the very good finite-sample behavior of the GMRC estimator. In most of the situations, even challenging, is seems that it represents the best choice in practice.

APPENDIX

We first prove the statement (3.9) for the following particular case : $p = 2$, $X^{(2)}$ is a binary regressor and $X^{(1)}$ is a continuous regressor with everywhere positive density conditional to $X^{(2)}$, as required by assumption 2(ii). S_X is therefore equal to $\mathbf{R} \times \{0, 1\}$.

Consider $\theta \neq \theta_0$. Define $\mathcal{U} = \{u \in \mathbf{R} : g(u) \text{ is not a global optimum of } g\}$. By assumption (3), $\mathcal{U}$ is clearly a non-empty open set of $\mathbf{R}$. Let $\mathcal{A}(\theta)$ be the set $\{x \in S_X : \beta_0'x \in \mathcal{U}$ and $\beta_\theta' x \in \mathcal{U}\}$. It is easy to see that $P[X \in \mathcal{A}(\theta)] = 1$. Define

$$\mathcal{B}_1(x;\theta) = \{u \in \mathbf{R} : g(u) < g(\beta_\theta' x)\}$$

and

$$\mathcal{B}_2(x;\theta) = \{u \in \mathbf{R} : g(u) > g(\beta_\theta' x)\}.$$

Let x_1 be any point of $\mathcal{A}(\theta)$ and suppose, without loss of generality, that $g(\beta_0' x_1) > g(\beta_\theta' x_1)$. Remark that $\mathcal{B}_1(x;\theta_0)$ and $\mathcal{B}_2(x_1;\theta)$ are non-empty open sets of $\mathbf{R}$, such that

$$\mathcal{B}_1(x_1;\theta_0) \cap \mathcal{B}_2(x_1;\theta) \neq \emptyset \tag{7.16}$$

and

$$\mathcal{B}_1(x_1;\theta_0) \cup \mathcal{B}_2(x_1;\theta) = \mathbf{R}. \tag{7.17}$$

Now let $a(x_1) \in \mathcal{B}_1(x_1;\theta_0)$ and $b(x_1) \in \mathcal{B}_2(x_1;\theta)$, with $a(x_1) \neq b(x_1)$. The question is : does there exist $x_2 \in S_X$, such that $\beta_0' x_2 = a(x_1)$ and $\beta_\theta' x_2 = b(x_1)$? The answer is affirmative if there is a solution in S_X of the system

$$\begin{cases} x_2^{(1)} + \theta_0 x_2^{(2)} = a(x_1) \\ x_2^{(1)} + \theta x_2^{(2)} = b(x_1) \end{cases} . \tag{7.18}$$

The solution is $x_2^{(2)} = 1$ and $x_2^{(1)} = a(x_1) - \theta_0 = b(x_1) - \theta$, provided $a(x_1) - b(x_1) = \theta_0 - \theta$. Such $a(x_1)$ and $b(x_1)$ exist for all x_1 and whatever $\theta_0 - \theta$, due to (7.16) and (7.17). Inequality (3.9) follows.

Finally, remark that such a solution should naturally exist with greater p or different distribution for $X^{(2)}$, as the number of degrees of freedom of system (7.18) should be greater.

References

[1] Delecroix M., Härdle W. and Hristache M. (2003). Efficient estimation in conditional single-index regression, J. Multivariate Anal., 86, 2, 213-226.

[2] Duan N. and Li K.C. (1991). Slicing Regression : a link free regression method, Ann. Stat., 19, 2, 505-530.

[3] Geenens G. and Delecroix M. (2005). A survey about single-index theory, discussion paper 0508, Institut de Statistique, Université Catholique de Louvain.

[4] Han, A.K. (1987). Nonparametric analysis of a generalized regression model, J. Econometrics, 35, 2-3, 303-316.

[5] Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivatives, J. Amer. Statist. Assoc., 87, 417, 218-226.

[6] Hristache M., Juditsky A. and Spokoiny V. (2001). Direct estimation of the index coefficient in a single-index model, Ann. Statist., 29, 3, 595-623.

[7] Ichimura H. (1993). Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, J. Econometrics 58, 1-2, 71-120.

[8] Powell J.L., Stock J.H. and Stoker T.M. (1989). Semiparametric estimation of index coefficients, Econometrica, 57, 6, 1403-1430.

[9] Sherman R.P. (1993). The limiting distribution of the maximum rank correlation estimator, Econometrica, 61, 1, 123-137.

[10] van der Vaart A.W. (1998). Asymptotic statistics, Cambridge Univ. Press.

[11] Zhu L.X. and Fang K.T. (1996). Asymptotics for kernel estimation of sliced inverse regression, Ann. Statist., 24, 3, 1053-1068.