



HAL
open science

Multiple Pareto index regression with application to large claim costs

Eunjung Kim

► **To cite this version:**

Eunjung Kim. Multiple Pareto index regression with application to large claim costs. Annales de l'ISUP, 2014, 58 (1-2), pp.57-76. hal-03611446

HAL Id: hal-03611446

<https://hal.science/hal-03611446>

Submitted on 17 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

MULTIPLE PARETO INDEX REGRESSION WITH APPLICATION TO LARGE CLAIM COSTS

BY EUN JUNG KIM

In insurance and reinsurance, an important interest is modeling extreme claim costs. For this motivation, Pareto distributions have been successfully applied in extreme value analysis. Under a single-parameter Pareto distribution of a response variable, we propose a new approach to estimate the Pareto index. We assume that the Pareto index is an unspecified function that depends on multiple indices induced by covariates, which constitutes 'a single-parameter Pareto index regression in multiple-index model'. We obtain the parameter estimators by using an approximate maximum likelihood estimation (aMLE) method where the likelihood function for the Exponential distribution approximates the corresponding function of the Pareto distribution through the log-transformation of the response variable. The number of indices is determined by a cross-validation technique. We show our approach by a simulation argument and its application to claim cost data, and discuss its practical performance.

1. Introduction. In insurance and reinsurance, modeling extreme claim costs is an important issue (e.g. Hogg and Klugman (1984)). Such modeling helps actuaries to identify the characteristics of high-risk groups, predict extreme risks, and adjust insurance premium policies of this class for balancing the premiums and the indemnities.

Due to their simplicity, the classical Pareto distributions have been a popular parametric modeling tool in claim cost analysis. Arnold (1983) provides an overview of the variants, properties and inferences of Pareto distributions. Parameter estimation of these distributions is traditionally performed by using maximum likelihood estimation or moment methods. Various estimation techniques, including these traditional ones, are presented and are compared to each other for risk premium pricing in an automobile insurance portfolio by Rytgaard (1990). Other estimation techniques such as robust estimation methods, a comparison between them via measures of goodness-of-fit, as well as their applications to real data are also found in Brazaukas and Serfling (2003). However, in these studies, large claim costs are modeled in the absence of covariates which play a prominent role in identifying the

Keywords and phrases: Extreme claim cost analysis, Single-parameter Pareto, Pareto index regression, Multiple-index model, Approximate maximum likelihood estimation

characteristics of high-risk groups. Because of the heterogeneity of an insurance company's portfolio, it is important to take covariate information into account in claim analysis. A natural approach to measuring the impact of covariates on the response variable is to use regression models, e.g., generalized linear models (McCullagh and Nelder, (1989)). There is a rich literature on actuarial applications. For example, Renshaw (1994) use the generalized linear models in the presence of covariates to model the influence of covariates on the average claim frequency and claim costs for premium pricing. This has been a classical modeling tool for premium pricing. Our interest also lies in using a regression model where the impact of covariates on large claim costs is taken into account. In this paper, we propose a new approach to estimating the parameter of a single-parameter Pareto distribution, where the parameter is called the Pareto index and it is taken to be a function of covariates. The model is a Pareto-index regression model which is based on semiparametric extension of the standard generalized linear regression model.

Recently, statistical techniques for dealing with large claim costs have been developed using extreme value theory (e.g. Beirlant *et al.* (2004)). In particular, an important result in this theory is of interest in large claim analysis. This focuses on the distribution of data exceeding a threshold, which is referred to the Generalized Pareto distribution. Some examples of actuarial applications can be found in Rootzén and Tajvidi (1997) and McNeil (1997). However, in their studies, large claim costs are modeled in the absence of covariates. The approach we propose in this paper would be extended to the Generalized Pareto distribution.

A single-parameter Pareto distribution belongs to the exponential family. To identify the characteristics of high-risk groups, the generalized linear model can be used, where we link the Pareto index to covariates through some specified function. This can be extended so that the Pareto index is an unspecified function of the covariates. However, nonparametric estimations encounter the problem of the so-called 'curse of dimensionality' when the sample size is small but has many covariates. In practice, the sample size which can be used for claim cost analysis is often limited. Hence, it would be interesting if we could decrease the number of the covariates without losing all the information in the covariates. Therefore, as a compromise between the parametric and nonparametric approaches, the semiparametric approach is an appealing alternative. Assuming that the Pareto index function is unspecified, but has a specific structure which is that some multiple indices are induced by covariates and are related to the index function, we introduce a 'single-parameter Pareto index regression in multiple-index model'. The

main idea of the model is not only that all the relevant information provided by covariates is included in some multiple linear combinations of covariates, but also that the indices determine high-risk groups in some nonparametric way. In the literature, multiple-index models are introduced by Ichimura and Li (1991) and investigated by many researchers as one of the dimension reduction techniques (Li (1991), Cook (1994), Hristache *et al.* (2001), Xia *et al.* (2002)) to overcome the disadvantage of nonparametric approaches. One of the dimension techniques proposed by Xia *et al.* (2002) will be used to estimate the Pareto index. This procedure is called ‘refined minimum average variance estimation (rMAVE)’.

Our estimation procedure is based on maximum likelihood estimation. The single-parameter Pareto distribution is related to the Exponential distribution through the log-transformation of the response variable. We can approximate the likelihood function for the Pareto distribution by that of the standard Exponential distribution. For the maximization to work, we need to estimate the Pareto index and use this to replace the unspecified Pareto index function in the approximate likelihood function of the Exponential distribution. This idea is similar to one of Klein and Spady (1993), who propose replacing an unspecified link function with a nonparametric estimator in the likelihood function of the binomial distribution in the case of a binary regression model to define a semiparametric maximum likelihood estimator of the parameter. For the replacement of the unspecified Pareto index function, we first estimate the conditional mean function by using a nonparametric method. Then, from the equation of the relationship between the conditional mean function and the Pareto index under a single-parameter Pareto distribution, we compute the estimator of the Pareto index. The estimator of multiple indices is the solution of this maximization problem. We call this procedure ‘approximate maximum likelihood estimation (aMLE)’.

The paper is organized as follows. In Section 2, we present a Pareto index regression model and explain both rMAVE and aMLE methods. Section 3 describes some issues for our implementation and lists the steps of the algorithm for the aMLE method. In Section 4, a simulation study is presented to show the practical performance of our approach. Section 5 focuses on the application of our approach to extreme claim costs using real data. Section 6 concludes with some suggestions for future work.

2. Pareto index regression model and estimation procedure.

2.1. *The single-parameter Pareto model.* Let Y_1, \dots, Y_n be i.i.d. copies of a random variable $Y \in \mathcal{Y}$ where $\mathcal{Y} \subset \mathbb{R}$ and let X_1, \dots, X_n be i.i.d. copies of a random vector of covariates, X where $X \in \mathcal{X} \subset \mathbb{R}^d$. Y_i represents the

i^{th} response of interest, which is associated with the vector of covariates X_i . In addition, $F_{Y|X}(y | x) = \mathbb{P}(Y \leq y | X = x)$ denotes the cumulative distribution function (c.d.f.) of Y given $X = x$. If Y is Pareto-distributed with the parameter $\alpha(x)$ and a fixed value $y_{\min} = 1$ given $X = x$, we obtain

$$(2.1) \quad 1 - F_{Y|X}(y | x) = y^{-\alpha(x)}$$

where $\alpha(\cdot)$ is a positive parameter. Usually, $\alpha(\cdot)$ is called the Pareto index. It is directly related to the probability of $Y > y$ given $X = x$ as well as the conditional mean with the condition $\alpha(x) > 1$,

$$(2.2) \quad m(x) = \mathbb{E}[Y|X = x] = \frac{\alpha(x)}{\alpha(x) - 1}$$

We are interested in estimating the Pareto index $\alpha(x)$ where $\alpha(x)$ is taken as a function of covariates X . In the parametric GLM framework, we consider the regression model using a log-link function : $\alpha(x) = \exp(\beta'x)$ where $\beta = (c, \beta_1, \dots, \beta_d)'$ is a $(d + 1)$ - dimensional vector of the regression coefficients. Despite the simplicity of the model, there is one known drawback in GLM. The link function is unknown in practice although it is treated as known. An alternative procedure to this is the fully nonparametric procedure: First, we estimate the conditional mean function by using a nonparametric method. Secondly, we replace $m(x)$ by its estimator $\hat{m}(x)$ in (2.2). Finally, we calculate $\hat{\alpha}(x)$ in (2.2). However, the nonparametric procedures suffer from the well-known problem of the curse of dimensionality when $d = \dim(X)$ is high. Thus, our interest lies in adding a dimension reduction assumption to a nonparametric estimation procedure, such that $\alpha(x)$ is an unknown function but has a specific structure. More precisely, assuming that $p > 1$, but much smaller than d and that $\beta_0^1, \dots, \beta_0^p \in \mathbb{R}^d$ exists and imposing that the norm of a parameter vector β_0^i for any $i \in \{1, \dots, p\}$ is equal to one and that any two different vectors β_0^i and β_0^j for $i \neq j$ are orthogonal we propose the following semiparametric multiple-index regression model,

$$\alpha(x) = g(B_0'x)$$

where $g : \mathbb{R}^p \rightarrow \mathbb{R}$. A convenient way to formulate these hypotheses is to introduce the $d \times p$ orthogonal matrix $B_0 = (\beta_0^1, \dots, \beta_0^p)$, i.e. $B_0'B_0 = Id_p$. In fact, the idea behind this model is based on another assumption about the existence of B which satisfies the following equivalence for the conditional mean of Y given $X = x$,

$$(2.3) \quad \mathbb{E}[Y|X = x] = \mathbb{E}[Y|B_0'X = B_0'x]$$

where $B_0' B_0 = Id_p$ is imposed for identifiability reasons (Ichimura and Lee, 1991). Letting $\alpha_{B_0}(\mathbf{u})$ denote the conditional mean of Y given $B_0' X = \mathbf{u}$, we obtain

$$(2.4) \quad \alpha_{B_0}(B_0' x) = \mathbb{E}[Y | B_0' X = B_0' x]$$

where and $\alpha_{B_0} : \mathbb{R}^p \rightarrow \mathbb{R}$. Therefore, from Equation (2.2), the function g becomes explicit, so the Pareto index $\alpha(x)$ can be formulated in the following way,

$$(2.5) \quad \alpha(x) = \frac{\alpha_{B_0}(B_0' x)}{\alpha_{B_0}(B_0' x) - 1}$$

with $\alpha_{B_0}(\cdot) > 1$. These kinds of assumptions have been considered in the literature of the so-called dimension reduction techniques which consider particularly the problem of estimating a space spanned by the columns of B_0 , e.g. see in Li (1991), Cook (1994), Hristache *et al.* (2001) and Xia *et al.* (2002).

2.2. Estimation procedures. In the following, we briefly describe an adaptive dimension reduction technique based on the semiparametric models proposed by Xia *et al.* (2002), which is called 'refined minimum average variance estimation (rMAVE)'. Using this procedure, we can estimate the conditional mean α_{B_0} as well as the matrix of multiple indices B_0 and use them to estimate the Pareto index α . In the next section, we propose a new approach based on the maximum likelihood estimation (MLE) procedure. In the last section, a cross validation technique is proposed to select the number of indices.

2.2.1. Refined Minimum Average Variance Estimation. Let us assume that the number of multiple indices p is known. To estimate B , we note that since $\mathbb{E}[(Y - m(X))^2] = \text{Inf}_h \mathbb{E}[(Y - h(X))^2]$ and (2.4), we obtain

$$(2.6) \quad \mathbb{E}[(Y - \alpha_{B_0}(B_0' X))^2] = \text{Inf}_h \mathbb{E}[(Y - h(X))^2]$$

In addition, let $\alpha_B(\mathbf{u})$ denote the conditional mean of Y given $B' X = \mathbf{u}$ for all matrix B which satisfies $B' B = Id_p$. This allows us to deduce $\alpha_{B_0} = \alpha_B$ and from (2.6), we get

$$\mathbb{E}[(Y - \alpha_{B_0}(B_0' X))^2] = \text{Inf}_B \mathbb{E}[(Y - \alpha_B(B' X))^2]$$

Xia *et al.* (2002) observe that this is also equivalent to $\text{Inf}_B \mathbb{E}[\mathbb{E}\{(Y - \alpha_B(B' X))^2 | B' X\}]$ according to a property of conditional expectation which

is the one relating to double conditional expectations, i.e., $\mathbb{E}[\mathbb{E}(A | M)] = \mathbb{E}[A]$. In fact, $\mathbb{E}\{(Y - \alpha_B(B'X))^2 | B'X\}$ is the conditional variance of Y given $B'X = \mathbf{u}$, which is denoted by $\sigma_B^2(B'X)$. Hence, an estimator of B_0 can be obtained through minimizing the expectation of the conditional variance of Y given $B'X = \mathbf{u}$ in the following way

$$(2.7) \quad \min_B \mathbb{E}\{\sigma_B^2(B'X)\} \quad \text{subject to} \quad B'B = Id_p$$

This procedure is named 'minimum average variance estimation (MAVE)'. If α_B was known, the following empirical criterion based on (x_i, y_i) for $i = 1, \dots, n$, would replace (2.7),

$$(2.8) \quad \min_B \sum_{i=1}^n \left\{ \sum_{j=1}^n (y_j - \alpha_B(B'x_j))^2 w(x_j, x_i) \right\} \quad \text{subject to} \quad B'B = Id_p$$

where $w(x_j, x_i)$ is a weight function which depends on the distance between x_i and x_j . However, we do not know α_B in practice. Therefore, Xia *et al.* (2002) propose using the following approximation by using the first order Talyor expansion of $\alpha_B(B'x_j)$ at $B'x_i$

$$y_j - \alpha_B(B'x_j) \approx y_j - a_i - b_i' B'(x_j - x_i)$$

with $a_i = \alpha_B(B'x_i)$ and b_i being the gradient of α_B calculated at $B'x_i$. The estimators $\{\hat{a}_i, \hat{b}_i\}$ of a_i and b_i for $i = 1, \dots, n$ are obtained by minimizing

$$(2.9) \quad \min_{a \in \mathbb{R}, b \in \mathbb{R}^p} \sum_{j=1}^n [y_j - a - b' B'(x_j - x_i)]^2 w(x_j, x_i)$$

Therefore, plugging the estimators $\{\tilde{a}_i, \tilde{b}_i\}$ for $i = 1, \dots, n$ into (2.8), we obtain an estimator \tilde{B} of B_0 in the following way,

$$(2.10) \quad \tilde{B} = \underset{B: B'B = Id_p}{\operatorname{argmin}} \sum_{i=1}^n \left\{ \sum_{j=1}^n [y_j - \tilde{a}_i - \tilde{b}_i' B'(x_j - x_i)]^2 w(x_j, x_i) \right\}$$

The MAVE method consists of using (2.9) and (2.10). We start by estimating a_i and b_i for $i = 1, \dots, n$ in (2.9) using a preliminary estimator of B_0 , then we plug these estimators of both a_i and b_i into (2.10) to estimate B_0 . Once the estimator of B_0 is obtained, we re-estimate a_i and b_i . Then, we repeat the two estimation processes until the algorithm converges.

In the process of the above estimation, Xia *et al.*(2002) suggest using the two weight functions $w(x_j, x_i)$. Initially, a d -dimensional kernel weight function (a) which does not rely on B is chosen, so that we can easily implement both (2.9) and (2.10),

$$(a) \quad w(x_j, x_i) = \frac{K_h(x_j - x_i)}{\sum_{s=1}^n K_h(x_s - x_i)}, \quad K_h(\mathbf{u}) = \frac{1}{h^d} K\left(\frac{\mathbf{u}}{h}\right)$$

with K being a kernel function on \mathbb{R}^d and its associated bandwidth h going to zero when $n \rightarrow \infty$ and $\sum_{j=1}^n w(x_j, x_i) = 1$. Another choice is a kernel weight function (b) which relies on B . More specifically, using an estimator \tilde{B} , we define a kernel weight function in the reduced space,

$$(b) \quad w(x_j, x_i) = \frac{K_h(\tilde{B}'(x_j - x_i))}{\sum_{s=1}^n K_h(\tilde{B}'(x_s - x_i))}, \quad K_h(\mathbf{u}) = \frac{1}{h^p} K\left(\frac{\mathbf{u}}{h}\right)$$

with K being a kernel function on \mathbb{R}^p . The idea is that when an initial estimator \tilde{B} is available, it is possible to get a more improved estimator through using \tilde{B} to measure the distance between x_j and x_i in the reduced space while iterating (2.9) and (2.10). The last iteration provides the final estimators \hat{B} and $\hat{\alpha}_{\hat{B}}$. Xia *et al.* (2002) call this procedure 'refined minimum average variance estimation (rMAVE)'.

Estimation of α

To estimate the Pareto index α , it suffices to use \hat{B} and $\hat{\alpha}_{\hat{B}}$ obtained from the rMAVE method in (2.5) and calculate $\hat{\alpha}(x)$ in the following way,

$$\hat{\alpha}(x) = \frac{\hat{\alpha}_{\hat{B}}(\hat{B}'x)}{\hat{\alpha}_{\hat{B}}(\hat{B}'x) - 1}$$

We impose that $\hat{\alpha}_{\hat{B}}$ is bigger than 1, since the conditional mean exists if the Pareto index is bigger than 1 when Y given $X = x$ follows a single-parameter Pareto distribution.

2.2.2. Approximate Maximum Likelihood Estimation. Our approach is based on the maximum likelihood estimation (MLE) procedure. The single-parameter Pareto distribution is related to the Exponential distribution through the log-transformation of a response variable: If Y_1, \dots, Y_n follows

the Pareto distribution as (2.1), then $\ln Y_1, \dots, \ln Y_n$ are independent, identically exponentially distributed with the parameter $\alpha(x)$. Let $f(z | x)$ be the probability density function (p.d.f.) of $Z_i = \ln Y_i$ conditional on $X_i = x$. For $z > 0$ and $\alpha(x) > 0$,

$$(2.11) \quad f(z | x) = \alpha(x) \exp^{-\alpha(x)z}$$

Using the definition of the multiple-index regression model in (2.5) for $\alpha(x)$ in the p.d.f. of Z_i in (2.11), let us define the log-likelihood function L_n for the Pareto index $\alpha(x)$ as follows

$$(2.12) \quad L_n(B, \alpha) = \frac{1}{n} \sum_{i=1}^n \left[\ln \alpha(B'x_i) - \alpha(B'x_i)z_i \right]$$

If α was known, the estimator of B_0 would be the solution of maximizing L_n with respect to B , as it works within the parametric framework. However, within our framework it does not work because we assume only the specific structure of the function α in (2.5) but α is unknown. A natural idea is to apply the plug-in method to our log-likelihood function (2.12): we first estimate α , then we substitute $\tilde{\alpha}$ for α , and then finally we maximize the estimated version $\tilde{L}_n(B, \tilde{\alpha})$ with respect to B .

Estimation of α and B

In our regression model, the Pareto index α is unknown, since α relies on the unknown function α_B . To estimate α , we need to estimate α_B . Let $\alpha(\mathbf{u})$ denote the conditional density of Y given $B'X = \mathbf{u}$ for all matrix which satisfies $B'B = I_p$. From Equation (2.5), we deduce $\alpha : \mathbf{u} \rightarrow \alpha_B(\mathbf{u}) / \{\alpha_B(\mathbf{u}) - 1\}$, for all $\mathbf{u} \in \mathbb{R}^p$. Let us define the Pareto index using α_B ,

$$(2.13) \quad \alpha(B'x) = 1 + \frac{1}{\alpha_B(B'x) - 1}$$

Recalling that the conditional mean exists if the Pareto index $\alpha > 1$, under the assumption that Y given $X = x$ follows a single-parameter Pareto distribution, we impose that $\alpha_B > 1$. In the definition (2.13), α_B can be used to define the Pareto index. In addition, note that the Pareto index converges to ∞ when $\alpha_B \rightarrow 1$, while the Pareto index is closer to 1 when $\alpha_B \rightarrow \infty$.

A natural idea about how to define the estimator of α is to use the estimator of α_B in (2.13) in order to compute the estimator of α . Hence, we propose to use kernel smoothing to estimate α_B . Either the Nadaraya-Watson (1964) or local polynomial estimation method (Fan and Gijbels, 1995) can be used.

Let us define the standard Nadaraya-Watson (NW) estimator of $\alpha_B(\mathbf{u})$ at the point $\mathbf{u} = (u_1, \dots, u_p)' \in \mathbb{R}^p$,

$$(2.14) \quad \hat{\alpha}_B(\mathbf{u}) = \frac{\hat{N}_B(\mathbf{u})}{\hat{D}_B(\mathbf{u})}, \quad \hat{D}_B(\mathbf{u}) \neq 0$$

with

$$\begin{aligned} \hat{N}_B(\mathbf{u}) &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_1 \cdots h_p} K \left(\frac{(x'_j \beta^1 - u_1)}{h_1}, \dots, \frac{(x'_j \beta^p - u_p)}{h_p} \right) y_j, \\ \hat{D}_B(\mathbf{u}) &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h_1 \cdots h_p} K \left(\frac{(x'_j \beta^1 - u_1)}{h_1}, \dots, \frac{(x'_j \beta^p - u_p)}{h_p} \right) \end{aligned}$$

where K is a kernel function on \mathbb{R}^p and h_1, \dots, h_p are its associated bandwidth sequence that goes to zero when $n \rightarrow \infty$. For the sake of simplicity, we take a product kernel for K , i.e., $K(\mathbf{u}) = \prod_{i=1}^p K^*(u_i)$ where K^* is an univariate kernel function. Also, we use the same bandwidth sequence h_1, \dots, h_p which satisfies $h_i \propto n^{-\frac{1}{2 \times 2 + p}}$ at the different index number p . The optimal bandwidth obtained by minimizing the asymptotic mean integrated squared error (AMISE) is well-known as $h = Cn^{-\frac{1}{2 \times k + p}}$ where C is a constant and k is the number of derivatives required for α_B . In our setting, we assume that α_B is twice continuously differentiable. Once the estimator $\hat{\alpha}_B$ in (2.14) is obtained, we use this estimator to define an estimator of α ,

$$(2.15) \quad \hat{\alpha}(B'x) = 1 + \frac{1}{\hat{\alpha}_B(B'x) - 1}$$

Next, plugging this estimator into (2.12), we obtain an estimated version of L_n ,

$$(2.16) \quad \hat{L}_n(B, \hat{\alpha}) = \frac{1}{n} \sum_{i=1}^n \left[\ln \hat{\alpha}(B'x_i) - \hat{\alpha}(B'x_i) \ln y_i \right] \tau_n(B'_n x_i, c)$$

where $\tau_n(\mathbf{u}, c)$ is a trimming function, c is a strictly positive constant and B_n is a preliminary estimator of B_0 . This function is commonly used to ensure that the denominators in $\hat{\alpha}_B$ are bounded away from zero. Given that the estimator B_n is obtained, we take $\tau_n(B'_n x_i, c) = \mathbf{1}_{\hat{D}_{B_n}(B'_n x_i) \geq c}$. Such a trimming function is considered an approximation of the ideal trimming function $\mathbf{1}_A(\cdot)$ where $A = \{\mathbf{u} : D_{B_0}(\mathbf{u}) > c, \mathbf{u} \in \mathbb{R}^p\}$ and $D_{B_0}(\cdot)$ is the

density of $B_0'x$ and $c > 0$ (See Delecroix *et al.* (2006)). As a consequence, let us define the approximate maximum likelihood estimator of B_0 ,

$$(2.17) \quad \hat{B} = \operatorname{argmax}_{B: B'B=Id_p} \hat{L}_n(B, \hat{\alpha})$$

The estimation procedure consists of the nonparametric estimation (2.14) and (2.15) as well as the parametric estimation (2.17). The maximization can be accomplished through any numerical optimization procedure.

2.2.3. Determination of the number of indices. An essential question arising from the multiple-index framework is how to select the number of indices. The estimation procedures described in both Section 2.2.1. and Section 2.2.2. can be accomplished only provided a fixed $p = 1, 2, \dots$ (the single-index case can be included), although p is unknown. However, this implies that, for a given p , the performance of each procedure can be measured through the estimation errors such as, for example the sum of residues squared. Since we assume that there is the $d \times p$ orthogonal matrix B_0 satisfying $\mathbb{E}[Y|X] = \mathbb{E}[Y|B_0'X]$, we expect the estimation error to be the lowest when p is chosen correctly. Therefore, we propose determining the p by using the cross-validation technique which restricts that of Xia *et al.* who include a special case $p = 0$ where Y and X are independent (See Xia *et al.* (2002), page 369). Using, for each fixed p , the estimator \hat{B} obtained from each procedure to the computation of the estimators $\hat{\alpha}_{\hat{B}}$ and $\hat{\alpha}$, we then use these estimators to compute the sum of residues squared. The index number p is determined in either of the following ways,

$$(2.18) \quad \hat{p} = \operatorname{argmin}_{1 \leq p \leq d} \sum_{i=1}^n (y_i - \hat{\alpha}_{\hat{B}}(\hat{B}'x_i))^2 \quad \text{or} \quad \hat{p} = \operatorname{argmin}_{1 \leq p \leq d} \sum_{i=1}^n \left(y_i - \frac{\hat{\alpha}(\hat{B}'x_i)}{\hat{\alpha}(\hat{B}'x_i) - 1} \right)^2$$

As we already mentioned, the so-called curse of dimensionality in nonparametric kernel smoothing approaches cannot be avoided, when the number of covariates p is high and the sample size is insufficient for reaching some level of the performance of the estimator. Accordingly, instead of comparing the residual sums for $1 \leq p \leq d$, we compare them for $p = 1, 2, 3$, or up to $p = 4$ at the highest in practice.

3. Implemented algorithm for the aMLE method. For the aMLE method, some technical issues for the implementation as well as the steps of the algorithm will be presented below. However, concerning the implemented algorithm for the rMAVE, see in Xia *et al.* (2002) for details. Moreover, the

rMAVE algorithm in Matlab is available in Xia's website. To focus on the needs of this paper, we rewrote the original program using SAS[®]/IML codes and combined it with the process of estimation of the Pareto index. They were integrated into a new program which includes the algorithms for the aMLE method as well as the maximum likelihood estimation method in the generalized linear model.

3.1. Practical considerations. Initialization of B For our numerical optimization, we need to specify an arbitrary initial matrix of $B = (\beta^1, \dots, \beta^p)$ which satisfies $B'B = Id_p$ for a fixed $p = 1, 2, \dots$ respectively. However, in practice, if the initial matrix selected is outside the neighborhood of the real B , the algorithm can either require many iterations to converge or simply fail to converge. For this reason, we set our initial matrix to a reliable preliminary estimator B_n obtained from a simple procedure, e.g. average derivative estimation (Hardle *et al.*, 1989). Note that this estimator can be also used for the trimming τ_n in (2.16).

Boundary problem of α Let us recall that we impose that $\alpha_B > 1$, since $\alpha > 1$. For this, we introduce a set $M = \{\mathbf{u} : \alpha_{B_0}(\mathbf{u}) > 1\}$ to control the boundary problem of α_B and modify τ_n into $\tau_n(B'_n x_i, c) = \mathbb{1}_{(\widehat{D}_{B_n}(B'_n x_i) \geq c) \cap (\widehat{\alpha}_{B_n}(B'_n x_i) > 1)}$ provided a preliminary estimator B_n . This becomes equivalent to approximating the indicator function of the set $A \cap M$. However, note that since $\mathbb{1}_{\widehat{D}_{B_n}(B'_n x_i) \geq c}$ does not matter in practice but usually it is required to prove some theoretical properties of the nonparametric regression estimator (See Delecroix *et al.* (2006)), one often takes $\mathbb{1}_{\widehat{D}_{B_n}(B'_n x_i) \geq c} \equiv 1$. Accordingly, we use the trimming $\tau_n^*(B'_n x_i) = \mathbb{1}_{\widehat{\alpha}_{B_n}(B'_n x_i) > 1}$ in practice.

Constraints on B While a solution is being searched in (2.17), the maximization is subject to the identification constraints on the $d \times p$ matrix B . For the p columns of B , the number of required constraints is $\frac{p(p-1)}{2} + p$. To reduce some of the computation time in finding a solution satisfying all the constraints, we propose to restrict the number of constraints to $\frac{p(p-1)}{2}$ which is the required number for the orthogonality condition on any two different vectors β_0^i and β_0^j , for $i \neq j$. Therefore, the maximization becomes subject to fewer constraints than in (2.17), that is,

$$\widehat{B}^* = \underset{\beta^i \perp \beta^j, i \neq j}{\operatorname{argmax}} \widehat{L}_n(B, \widehat{\alpha})$$

where $\widehat{B}^* = (\beta_1^*, \dots, \beta_p^*)$. Each vector β_i^* can be normalized by dividing it by its norm. As a consequence, we have $\widehat{B} = (\beta_1, \dots, \beta_p)$ where $\beta_i = \beta_i^* / |\beta_i^*|$.

3.2. *Description of implemented algorithm.* With practical considerations mentioned in Section 3.1, for a fixed p , the algorithm for estimating α and B_0 through an approximate maximum likelihood method is summarized as follows.

Step 0. (Initialization) Obtain a preliminary estimator B_n . For the coherence of notations thereafter, let $\widehat{B}_{(0)}$ be the initial matrix for B . Set $l = 0$ and $\widehat{B}_{(0)} = B_n$.

Step 1. Obtain an estimator $\widehat{\alpha}_{\widehat{B}_{(l)}}^{(l)}$ by the leave-one-out version of the NW method.

Step 2. Compute an estimator $\widehat{\alpha}^{(l)}$ by using the estimator from Step 1.

Step 3. Obtain $\widehat{B}_{(l+1)}^*$ in the following way,

$$\widehat{B}_{(l+1)}^* = \underset{\beta^i \perp \beta^j, i \neq j}{\operatorname{argmax}} \widehat{L}_n^*(B, \widehat{\alpha}^{(l)})$$

where

$$\widehat{L}_n^*(B, \widehat{\alpha}^{(l)}) = \frac{1}{n} \sum_{i=1}^n \left[\ln \widehat{\alpha}^{(l)}(B'x_i) - \widehat{\alpha}^{(l)}(B'x_i) \ln y_i \right] \tau_n^*(B'_l x_i)$$

with respect to B .

Step 4. Normalize the vectors, $\widehat{\beta}_{i(l+1)}^* / |\beta_{i(l+1)}^*|$, $i = 1, \dots, p$ and update $\widehat{B}_{(l+1)}$.

Step 5. Set $l = l + 1$ and go to **Step 1** until convergence is attained.

4. Simulation study.

4.1. *Assessment of the performance of the estimator of multiple indices.* A natural way to evaluate the accuracy of the estimator of multiple indices is to measure the distance between the true matrix B_0 and each estimator \widehat{B} . Let us introduce a simple measurement of the distance denoted by

$$\Delta(\widehat{B}, B_0) = \operatorname{tr}\{(\widehat{B} - B_0)(\widehat{B} - B_0)'\}/d$$

where d is the row number of both B_0 and \widehat{B} . However, there is a discrepancy in the calculation, since the size of \widehat{B} differs from that of B_0 . The column number of \widehat{B} varies from 1 to 3 or up to 4, while the column number of B_0 denoted by p_0 is a fixed constant, i.e., $p_0 = 2$ for the simulated model below. To remedy this problem, we suggest that when $p < p_0$, we add as many zero column vectors to \widehat{B} as necessary to define a new matrix \widehat{B}^* , which includes \widehat{B} as well as the added zero column vectors. The size of \widehat{B}^*

becomes the same as the size of B_0 . The idea is that \widehat{B}^* is regarded as a matrix of a transformation application in \mathbb{R}^{p_0} , but \widehat{B}^* only takes the values in the p -dimensional vector subspace of \mathbb{R}^{p_0} . Hence, we can complement \widehat{B} with zero column vectors to define \widehat{B}^* . When $p > p_0$, we complement B_0 in the same way to define B_0^* of the same size as of \widehat{B} . As a consequence, we propose a new measurement of the distance between B and its estimator \widehat{B} in the following way

$$\begin{aligned}\Delta(\widehat{B}^*, B_0) &= \text{tr}\{(\widehat{B}^* - B_0)(\widehat{B}^* - B_0)'\}/d, \quad \text{if } p < p_0 \\ \Delta(\widehat{B}, B_0^*) &= \text{tr}\{(\widehat{B} - B_0^*)(\widehat{B} - B_0^*)'\}/d, \quad \text{if } p \geq p_0\end{aligned}$$

where \widehat{B}^* and B_0^* are the complemented matrices of \widehat{B} and B_0 , respectively.

4.2. *Simulated example.* We conducted a simulation study to show the performance of our procedure with the finite samples of size $n = 100$ and $n = 200$. For the sake of simplicity, we used a normal product kernel $K(u) = \prod_{i=1}^p K^*(u_i)$ where $K^*(u) = \exp(-x^2/2)/\sqrt{(2\pi)}$ is the classical Gaussian kernel. In addition, the fixed bandwidth was set at $h_n = Cn^{-1/(4+p)}$ where C is a constant, $p = \dim(B'X)$ and the sample size n . In other words, $h_n = 0.75 \times 100^{-1/6}$ was chosen at $C = 0.75$ and the index number $p = 2$ in the sample of size $n = 100$. For an initial matrix $B_{(0)}$, we used an estimator B_n obtained from the refined minimum variance estimation method (Xia *et al.*, 2002).

We considered five independent covariates: Two binary variables X_1 and X_2 were simulated from $\mathbf{1}_{0.2\mathcal{N}(0,1)<0}$ and $\mathbf{1}_{0.5\mathcal{N}(0,1)<0}$, three continuous variables X_3, X_4 and X_5 were simulated from the uniform distribution on the interval $[-1, 1]$ and let $X = (X_1, X_2, X_3, X_4, X_5)'$. Assuming that two indices are needed to capture all the relevant information contained in X , we considered the double-index regression model,

$$(4.1) \quad \alpha(B'_0 X) = g(X'\beta_0^1, X'\beta_0^2)$$

where $\beta_0^1 = (1, 1, 1, 1, 0)'/2$, $\beta_0^2 = (1, -1, 1, -1, 0)'/2$ and $B_0 = (\beta_0^1, \beta_0^2)$. The variable Y was generated from the Pareto distribution with the parameter α which is equal to the regression function $\alpha(B'_0 X)$ in (2.19) and the minimum value of Y which is fixed at one. We specified a double-index regression function $g: \mathbb{R}^2 \rightarrow \mathbb{R}$, that is

$$g(t_1, t_2) = |t_1| + 3t_2^2 + 1$$

Our estimator \widehat{B}^{aMLE} was compared with the other semiparametric estimator \widehat{B}^{rMAVE} which was obtained using the refined minimum variance

estimation method (rMAVE) at the different number of indices. We report the mean of estimation errors over 1,000 simulations from samples of size 100 and 200 in Table 1. The accuracy of each estimator was evaluated by measuring the distance $\Delta(\hat{B}, B_0)$ where $p_0 = 2$ and $1 \leq p \leq 3$, which is defined in Section 4.1.

TABLE 1
Mean (Standard deviation) of estimation errors, $\Delta(\hat{B}, B_0)$

$N = 1,000$	Number of indices (p)	$n = 100$	$n = 200$
\hat{B}^{rMAVE}	$p = 1$	0.43279 (0.14522)	0.42466 (0.14147)
	$p = 2$	0.31351 (0.20133)	0.27909 (0.20661)
	$p = 3$	0.40867 (0.17828)	0.39230 (0.17516)
\hat{B}^{aMLE}	$p = 1$	0.44297 (0.16057)	0.43550 (0.16815)
	$p = 2$	0.11381 (0.06544)	0.09250 (0.05028)
	$p = 3$	0.29791 (0.05060)	0.28654 (0.04521)

When $p = 1$, there is a slight difference in the mean and standard deviation of the estimation errors between the above two estimators, whereas when $p = 2$ or $p = 3$, the mean and standard deviation of our estimator of \hat{B}^{aMLE} are much smaller than \hat{B}^{rMAVE} . Thus, the aMLE estimator performs better when more than one index exist. However, both estimators tend to behave similarly. The mean of estimation errors is the lowest when $p = 2$. It is followed by the mean of estimation errors when $p = 3$ and then when $p = 1$.

We notice that as shown in Table 1, we could not compare the performance of our estimator with that of the estimator obtained from a parametric model such as the generalized linear model (GLM) for single index, i.e., $p = 1$. They are not comparable because under the semiparametric model assumption, we do not estimate the coefficient for the intercept of the vector $(1, X_i)$, which we do have to estimate within the GLM framework.

However, in contrast, our estimator $\hat{\alpha}^{aMLE}$ was compared not only with the other semiparametric estimator $\hat{\alpha}^{rMAVE}$ when $p = 1, 2, 3$, but also with the GLM estimator $\hat{\alpha}^{GLM}$ when $p = 1$. For this, we measured the performance of the estimator of the Pareto index α_{B_0} by computing the average of the squared estimation errors $n^{-1} \sum_{i=1}^n (\hat{\alpha}(\hat{B}'x) - \alpha(B'x))^2$ over 1,000 simulations from samples of size 100 and 200. The computation results are reported in Table 2 and Table 3.

Overall, either semiparametric estimator outperform $\hat{\alpha}^{GLM}$, although the performance of the GLM procedure is improved when the sample size increases. The estimator $\hat{\alpha}^{rMAVE}$ outperforms slightly $\hat{\alpha}^{aMLE}$, when $p = 1$ and $n = 100$. In the other cases, the aMLE method performs much better than the rMAVE method. As expected concerning the choice of the index

TABLE 2
Comparison for single index, $p = 1$

$N = 1,000$,	Number of indices (p)	$n = 100$	$n = 200$
$\hat{\alpha}^{GLM}$	$p = 1$	2.05482 (0.80366)	1.66843 (0.32451)
$\hat{\alpha}^{rMAVE}$	$p = 1$	1.41686 (0.37040)	1.39051 (0.30004)
$\hat{\alpha}^{aMLE}$	$p = 1$	1.43387 (0.45106)	1.38655 (0.39187)

TABLE 3
Comparison for multiple indices, $p \geq 2$

$N = 1,000$	Number of indices (p)	$n = 100$	$n = 200$
$\hat{\alpha}^{rMAVE}$	$p = 2$	1.05040 (0.39238)	0.90359 (0.35394)
	$p = 3$	1.06849 (0.40664)	0.91110 (0.31945)
$\hat{\alpha}^{aMLE}$	$p = 2$	0.99112 (0.36465)	0.82150 (0.27126)
	$p = 3$	1.02147 (0.37667)	0.85893 (0.26171)

number, both rMAVE and aMLE methods show the smallest average of the squared estimation errors when $p = 2$. The number of indices are chosen correctly, i.e. $\hat{p} = 2$.

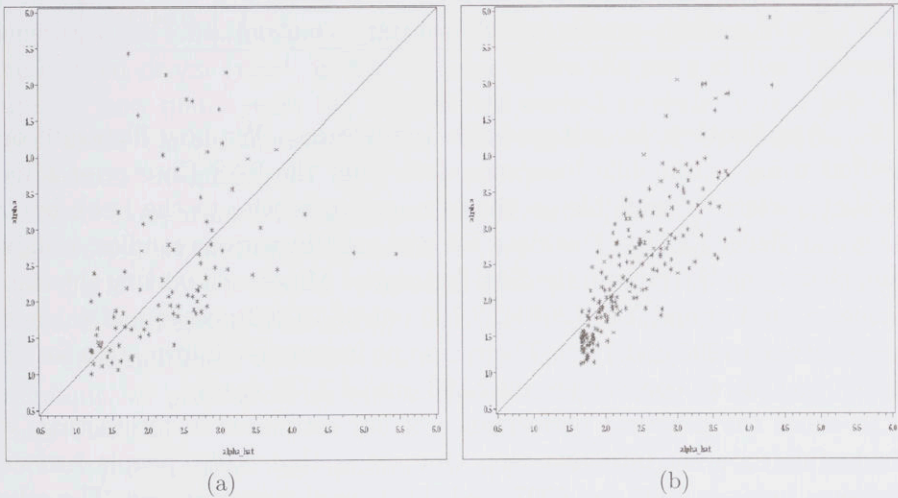


Fig 1: Graphical comparison by using plots of $(\alpha(B'X_i), \hat{\alpha}(\hat{B}'X_i))$ where true Pareto index $\alpha(B'X_i)$ (y -axis) against estimated Pareto index $\hat{\alpha}(\hat{B}'X_i)$ (x -axis) and the 45-degree reference lines at different sample sizes, (a) $n = 100$, (b) $n = 200$

Furthermore, the graphical comparison between the true and estimated estimators is depicted in Figure 1. The left and right panels in Figure 1 are scatter plots of the true Pareto index (x -axis) against the estimated Pareto index (y -axis) at different sample sizes 100 (a) and 200 (b) where the

estimated Pareto index is the estimator $\hat{\alpha}^{aMLE}$ at $\hat{p} = 2$. In the diagram above, we can see that when the true Pareto index is well fitted by the estimated Pareto index, points closely follow a 45-degree reference line. The estimated indices (coefficients of X) used in (a) and (b) is summarized in Table 4.

Figure 1 shows that the estimated Pareto index obtained from the aMLE method fits quite well the true Pareto index, since most of the points are approximately distributed around the 45-degree reference lines except for some points in both panels. Also, it indicates that when sample size increases, the points clearly tend to concentrate along the 45-degree reference line.

TABLE 4
Comparison for two indices used in (a) and (b)

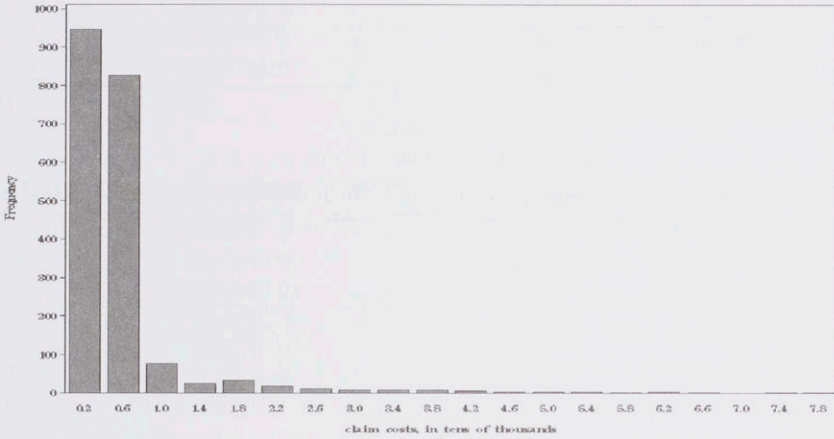
True and estimated indices (coefficients of X)	$\Delta(\hat{B}, B_0)$
$B_0 = \begin{pmatrix} 0.5000 & 0.5000 & 0.5000 & 0.5000 & 0.0000 \\ 0.5000 & -0.5000 & 0.5000 & -0.5000 & 0.0000 \end{pmatrix}'$	0.00000*
$\hat{B}_{(a)} = \begin{pmatrix} 0.2336 & 0.5969 & 0.5104 & 0.5648 & -0.0978 \\ 0.6634 & -0.4191 & 0.3822 & -0.0939 & 0.4789 \end{pmatrix}'$	0.1071189
$\hat{B}_{(b)} = \begin{pmatrix} 0.3224 & 0.5578 & 0.4297 & 0.6147 & 0.1495 \\ 0.4755 & -0.2411 & 0.6843 & -0.4681 & -0.1681 \end{pmatrix}'$	0.041241

5. Application to automobile insurance. We now illustrate our method using automobile insurance data from the SAS Enterprise Miner database, which is available on the companion website to the book by De Jong and Heller (2008). This data set was used in various studies, ranging from classifying the risks in the SAS Enterprise Miner to modeling the claim frequency in Yip and Yau (2004). 2,746 out of 10,303 people had car accidents between 1993 and 1999. There is some incomplete information for 153 people, so that we concentrate on 2,593 out of 2,746 cases.

To check the practical performance of our estimator in forecasting, we divide this data set into two sets. One set is that 1,981 people had car accidents between 1993 and 1997, and this is used for estimation. The other is that 612 people had accidents between 1998 and 1999, and that is used for forecasting. Using the first data set, we display the histogram of claim costs in tens of thousands in Figure 2. There is a big decrease in the frequency of claims between claim costs of 6,000 and 10,000 for claim costs.

The aim of the analysis is to examine how the conditional density of large claim costs is related to a few linear combinations of people's information. For this, we take 0.8 in tens of thousands as a fixed threshold to define large claim costs of our data set. We focus on 208 people whose claim costs are superior to 0.8 in tens of thousands. From this data set, we let the variable

Fig 2: Claims costs from automobile insurance in tens of thousands



Y be claim costs which are more than 8 (in thousands) and the vector X constitute the following five covariates: if there are any children in a car when the insured drives (yes:1, no:0), the area where the insured lives (urban:1, rural:0), how many years the insured has worked at their current job, the value of the car (in thousands), the travel time between home and work of the insured (in minutes).

For the nonparametric part, a product kernel $K(u) = \prod_{i=1}^p K^*(u_i)$ is used where $K^*(u)$ is the classical Gaussian kernel. The fixed bandwidth was set at $h_n = Cn^{-1/(4+p)}$ where $C = 1$, index number $p = \dim(B'X)$ and the sample size n . For an initial matrix $B_{(0)}$, we used an estimator B_n obtained from the refined minimum variance estimation method. Using the cross-validation technique, we present our results at the different number of indices in Table 5 and Table 6.

When $p = 1$, the two semiparametric approaches outperform the parametric approach in Table 5. When we compare the two semiparametric methods, the mean of the errors obtained from rMAVE method is lower than from aMLE method. However, as soon as the number of indices increases ($p \geq 2$) in Table 6, the mean of the errors is improved much more than in the case $p = 1$ in Table 5. Overall, aMLE method outperforms rMAVE method at each number of indices. As regards the choice of the index number, the mean of the errors obtained from both rMAVE and aMLE methods is the lowest when $p = 4$. The four index regression model using aMLE method performs the best among all the proposed models using different methods.

TABLE 5
Comparison for single index, $p = 1$

Method	Number of indices (p)	Mean of estimation errors
GLM	$p = 1$	6.3831594
rMAVE	$p = 1$	2.0758689
aMLE	$p = 1$	2.4478090

TABLE 6
Comparison for multiple indices, $p \geq 2$

Method	Number of indices (p)	Mean of estimation errors
rMAVE	$p = 2$	0.3852293
	$p = 3$	0.0797080
	$p = 4$	0.0546581
aMLE	$p = 2$	0.3505817
	$p = 3$	0.0756347
	$p = 4$	0.0501537

To assess whether our method is useful for prediction, we conducted a supplementary study using a validation data set where claim costs occurred between 1998 and 1999. A natural way is to compare the real claim costs in the validation set with its predicted costs by using the above estimation results in a training data set where car accidents occurred between 1993 and 1997. For this, we examined the results from the GLM and four index regression models using rMAVE and aMLE methods. For example, we first computed \hat{B}^{aMLE} by using aMLE method in a training set, that is

$$\hat{B}^{aMLE} = \begin{pmatrix} -0.081437 & 0.4668623 & -0.249259 & 0.7074869 \\ 0.2109254 & 0.4000173 & -0.335481 & 0.1694367 \\ 0.2175798 & 0.2392426 & 0.8952371 & 0.2732129 \\ 0.9443702 & -0.180302 & -0.154056 & 0.0258089 \\ 0.0985009 & 0.7295751 & 0.0118052 & -0.628842 \end{pmatrix}$$

and we then used it to estimate the claim costs in a validation set. To measure the performance, we computed the prediction errors which are calculated as the mean of residues squares, $n^{-1} \sum_{i=1}^n (y_i - \hat{\alpha}_{\hat{B}}(\hat{B}'x_i))^2$ for each model. In Table 7, the results are reported for the four samples which are obtained using different fixed thresholds s ($s = 2, 4, 6, 8$) in thousands. Here, the thresholds are interpreted as the levels of excessive claim costs, which do not rely on the characteristics of X (or people's information).

Clearly, the prediction errors from the four index regression models are much lower than from GLM. When we compare the two semiparametric approaches, aMLE method tends to outperform rMAVE method more and more as the threshold increases.

TABLE 7
*Prediction errors at different thresholds $s = 2, 4, 6, 8$ in thousands
 (Y = the claim costs in thousand, n = sample size of the validation set)*

Method	$Y > 2$ ($n = 513$)	$Y > 4$ ($n = 316$)	$Y > 6$ ($n = 128$)	$Y > 8$ ($n = 58$)
GLM	25.174654	8.6264527	8.388384	8.9756044
rMAVE	0.0777016	0.0276495	0.0030006	0.0000323
aMLE	0.0748298	0.0267066	0.0020292	0.0000309

6. Conclusion and future work. In this paper, we introduced a new semiparametric procedure for estimating a single-parameter Pareto index model with $\alpha > 1$ under the multiple-index regression assumption. Then, we studied the practical performance of our procedure, although we still do not know the theoretical properties of our estimators. Our approach was well adapted to analyze large claim costs and gave good results compared to the parametric and the other semiparametric (rMAVE) approaches. In practical applications, our model may be a useful alternative to both the classical parametric models or single-index models. The model relies on more flexible assumptions about the Pareto index than parametric models do. Furthermore, multiple indices tend to capture the relevant information carried by covariates better than a single index does throughout our simulation study. We believe that multiple-index regression models become a more efficient tool in the analysis where there are many covariates.

Although the current work concerns the use of semiparametric regression models in practice, it would be of interest to explore the theoretical behavior of our estimators. For practical purposes, it would be useful to try to incorporate a method for automatically selecting bandwidths of kernel estimators into our procedure. To extract only the extreme values from the sample and use the effective sample which is composed of the extracted values for estimation, a method of selecting a threshold which determines the effective sample would also be incorporated. Another possible direction for further work would be to extend our approach to the Generalized Pareto (GP) distribution which is defined as,

$$\begin{aligned}
 F(y) &= 1 - (1 - k\frac{y}{\sigma})^{\frac{1}{k}}, & \text{if } k \neq 0 \\
 &= 1 - \exp(-\frac{y}{\sigma}), & \text{if } k = 0
 \end{aligned}$$

with a shape parameter k and a scale parameter σ ($\sigma > 0$) (See e.g. Davison and Smith (1990)). If $k \leq 0$, $y \geq 0$ and if $k > 0$, $0 < y < \sigma/k$. In the case where $k < 0$, GP distribution is heavy-tailed. We may introduce a positive parameter denoted by $\alpha = -1/k$ and call it the Generalized Pareto (GP) index. This GP index would be linked to covariates via a semiparametric multiple-index regression function.

REFERENCES

- [1] ARNOLD, B. C. (1983). *Pareto distributions*. International cooperative publishing house, Fairland, Maryland.
- [2] BEIRLANT, J., GOEGBEUR, Y., SEGERS, J., AND TEUGELS, J. (2004). *Statistics of extremes: Theory and Application*. John Wiley, New York.
- [3] BEIRLANT, J., GOEGBEUR, Y., SEGERS, J., AND TEUGELS, J. (2004). *Statistics of extremes: Theory and Application*. John Wiley, New York.
- [4] DAVISON, A. C. AND SMITH, R. L. (1990). Models for exceedances over high thresholds (with discussion). *J. Roy, Statist. Soc. Ser. B* (52), 393–442.
- [5] COOK, R. D. (1994). *Using dimension-reduction subspaces to identify important inputs in models of physical systems*. In Proc. Section on Physical and Engineering Sciences 1825. Amer. Statist. Assoc., Alexandria, VA.
- [6] DE JONG, P. AND HELLER, G. J. (2008). *Generalized linear models for insurance data*. International Series on Actuarial Science, Cambridge University Press.
<http://www.acst.mq.edu.au/research/books/GLMsforInsuranceData>.
- [7] DELECROIX, M., HRISTACHE, M. AND PATILEA, V. (2006). On semiparametric M-estimation in single-index regression. *J. Statist. Plann. Inference* 136(3), 730–769.
- [8] FAN, J. AND GJIBELS, I. (1996). *Local polynomial modeling and its applications*. Chapman and Hall, London.
- [9] HÄRDLE, W. AND STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *J. Amer. Statist. Assoc.* 84(408), 986–995.
- [10] HRISTACHE, M., JUDITSKY, A., POLZEHL, J. AND SPOKOINY, V. (2001). Structure adaptive approach for dimension reduction. *Ann. Statist.* 29(6), 1537–1566.
- [11] HOGG, R. AND KLUGMAN, S. (1984). *Loss Distributions*. John Wiley, New York.
- [12] ICHIMURA, H. AND LEE, L.F. (1991). Semiparametric least squares estimation of multiple index models: single equation estimation. In *Nonparametric and Semiparametric Methods in Econometrics and Statistics* 3–50. Cambridge University Press.
- [13] KLEIN, R. AND SPADY, R. (1993). An efficient semiparametric estimator for the binary response model. *Econometrica* 61, 387–421.
- [14] LI, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *J. Amer. Statist. Assoc.* 86, 316–342.
- [15] MCCULLAGH, P. AND NELDER, J. A. (1989). *Generalized linear models.*, 2nd ed. Chapman and Hall, London.
- [16] MCNEIL, A. J. (1997). Estimating the tails of loss severity distributions using extreme value theory. *ASTIN Bulletin* 27, 117–137.
- [17] NADARAYA, E. A. (1964). On estimating regression. *Theory Probab. Appl.* 9, 141–142.
- [18] RENSHAW, A. E. (1994). Modelling the claims process in the presence of covariates. *ASTIN Bulletin* 24, 265–286.
- [19] ROOTZÉN, H. AND TAJVIDI N. (1997). Extreme value statistics and wind storm losses: a case study. *Scandinavian Actuarial Journal* 1, 70–94.
- [20] RYTGAARD, M. (1990). Estimation in the Pareto distribution. *ASTIN bulletin* 20(2), 201–216.
- [21] WATSON, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A* 26, 359–372.
- [22] XIA, Y., LI, W. K., TONG, H. AND ZHANG, D. (2002). An adaptive estimation of dimension reduction space. *J. Roy, Statist. Soc. Ser. B* 64 363–410.

16 AVENUE PIERRE BROSOLETTÉ
92240 MALAKOFF, FRANCE
E-MAIL: eunjungkim.k@gmail.com