Salanti, Ulm:

Modelling Under Order Restrictions

Sonderforschungsbereich 386, Paper 265 (2001)

Online unter: http://epub.ub.uni-muenchen.de/

Modelling under order restrictions

G. Salanti*, K. Ulm
Institute for Medical statistics and Epidemiology
Klinikum rechts der Isar
Technical University of Munich
Munich, Germany

December 3, 2001

Abstract

Order restricted effects of predictors can be represented in models by the monotonic transformation fitted by the pooled adjacent violators algorithm as described by T. Robertson et al. In the context of multivariate modelling, this paper aims to introduce next to additive monotonic models a multidimensional approach. The corresponding permutations test to assess significance for the predictors is described and some feeble points of the approach are discussed. We also introduce a procedure to improve the parsimony of the model by reducing the resulting level sets. The use of monotonic regression in connection with the threshold value estimation problematic is outlined and two similar approaches to assess it are discussed.

*georgia.salanti@imse.med.tu-muenchen.de

1 Introduction

The logit and probit model or nonparametric models are widely used to describe relationships between one or more explanatory variable and a binary outcome. However, an assumption undertaken behind the parametric models is that the relationship is directly linear or after some transformation. This approach turns out to be inadequate in many applications since there can be "a penalty for assuming that a monotone regression is linear". As to non parametric models, as the generalised additive fitted by splines, the result is not always easy to interpret and they are not sometimes convenient when the establishment of a dose-response relationship is of interest. Furthermore, in case of scale predictors, it is often useful - if not indispensable - to assess optimal cutpoints in order to obtain an interpretable and simple model. The monotonic regression retains the monotonicity assumption but relaxes the linearity requirement, and results by fitting step functions. Its use as smoother can extend generalised additive models and multidimensional regression methods. In the present paper, those two methods will be described and a procedure to detect threshold values will be proposed. Throughout this paper we restrict ourselves on cases of binary response. The methodos will be illustrated on an epidemiological study.

Data set used
The data set used to illustrate the methods is taken from the DFG-study "Chronic Bronchitis" [4]. The aim of this study has been to investigate the influence of the occupational total inhalable dust exposure besides the well-known factors smoking and age. The disease rate is suspected to increase while the concentration and the duration of the exposure increase. A further aim of this study has been in case of association between dust and disease, whether a threshold value for the dust concentration could be assessed. The plant from Munich has been analysed.

2 The isotonic procedure

2.1 The isotonic regression as a smoother

Isotonic regression[1] reduces the description of n points from the given population to l homogenous subgroups with respect to the monotonicity restriction.

[1] Isotonic is simply the one-sided restriction of monotonic regression: the trend is considered increasing whereas in monotonic regression there is no assumption about the direction of the trend.

The l constant risk groups in which we summarise the response are called **level sets**. The fact that the variable to be regressed is replaced by constant values under a conditional expectation, refers to a smoothing operation, and therefore many authors call the isotonic transformation the "isotonic smoother". However, the use of the term "smooth" has been criticised since the result is a step function and there is obviously a lack of smoothness. In this paper we retain the will to see in the monotonic process his smoothing characteristics. From this point of view is a smoother having global nature but resulting in locally flat estimations.
The procedure of *"smoothing"* can be described as follows: Starting from the assumption that an isotonic dose-response relationship exists a maximum likelihood estimator under order restrictions of the response is assessed. This estimators can be provided by the *Minimum lower sets algorithm*, the *Maximum upper sets algorithm* or the *"Pool adjacent violators algorithm"* which is used in the present paper considered as the most efficient.
Consider the case of k dose groups where the dose is in increasing order $d_1 \leq d_2 \leq ... \leq d_k$ and the outcome of an experiment is $Y_{ij}, i = 1..k, j = 1..n_i$. The distribution of Y_{ij} can be binary, poisson, continuous or survival probability, belonging to an exponential family having parameter θ_i. We wish to have θ_i in non decreasing order, given that $d_i \leq d_{i+1}, i = 1...k$. That is truth only if all the estimates fulfil the isotonic relationship. If there is somewhere a violator such that $\theta_{i+1} \leq \theta_i$ for some i, then the isotonic estimator of $(\theta_i, \theta_{i+1}) = (\theta^*_{i,i+1})$ is needed to be found.
The PAVA starts with $\theta = (\theta_1, \theta_2, ..., \theta_k)$. If θ is isotonic then $\theta^* = \theta$. Otherwise, both violators θ_{i+1}, θ_i are replaced by their weighted average $\theta^*_i = \theta^*_{i+1} = \frac{w_i\theta_i + w_{i+1}\theta_{i+1}}{w_i + w_{i+1}}$, and w_i, w_{i+1} are both replaced by $w_i + w_{i+1}$. This process is repeated using the new values and weights until an isotonic set of values is obtained.
Within our paper we focus on cases where the response is binary, and therefor $\theta_i = p_i$ (the probability of occurrence of an event) and $w_i = n_i$ the number of observations in each dose group.
Hastie and Tibshirani [7] define the degrees of freedom of a smoother, as the trace of the smoother matrix S. Let B_k be the set of indexes i corresponding to observations p_i that are estimated through p^*_k and $k \in (1, l)$. The smoother matrix will have the following form

$$S_{ij} = \begin{cases} \dfrac{w_i}{\sum\limits_{s \in B_k} w_s} & \text{if } i \text{ and } j \in B_k \text{ for some } k \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Therefore $tr(S_{ij}) = \sum_{ii} \dfrac{w_i}{\sum\limits_{s \in B_k} w_s} = l$, and the degrees of freedom are equal

to the number of level sets. That means that the degrees of freedom of an isotonic smoother are not pre-specified but depend on the data, since the number of level sets as well as their width are automatically selected from PAVA. So, arguments as the number of knots used in other type of smoothing (for example in splines) are not required.
In figure 1 an example of the univariate approach is depicted. The Chronic Bronchitis reaction incidence is considered to increase as long as the dust concentration increases. The dust is summarised in seven constant risk intervals. The isotonic regression is plotted with the equivalent estimator using spline with four degrees of freedom.

2.2 Some important features of the isotonic framework

2.2.1 Test for trend

This approach has some advantages compared to parametric smoothing. No specific assumptions apart from monotonicity is required from the form of the dose-response relationship. An other important advantage of this method is that the test applied to assess significance of the predictor is a test for trend, since the H_1 alternative is the isotonic transformation of the response. In a search of such an adequate test, recall that many commonly used tests for trend give results that depend on the form in which the dose is used. However isotonic regression not only provides one of the most reliable tests for trend [10] [2], but is also expected to increase the power by setting the isotonic transformation of the response as alternative to the constant risk H_0 assumption [11]. This test is known as the isotonic Likelihood Ratio test [6] and follows a weighted X^2 distribution.
We define the following hypothesis:

$$\begin{aligned} H_0 &: \text{equality of } \theta_i \\ H_1 &: \theta_i \text{ are isotonic} \\ H_2 &: \theta_i \text{ is unrestricted} \end{aligned}$$

Then consider T_{01} and T_{12} the statistics that test H_0 against H_1 and H_1 against H_2 respectively. Note that T_{12} tests the adequacy of the isotonic assumption against any other possible shape of the response, and once the T_{01} gives a significant result, it is important to apply it in order to conclude about the isotonic transformation. The test have the following forms:

$$\begin{aligned} T_{01} &= -2(L(\theta_{H_0})) - L(\theta_{H_1})) \\ T_{12} &= -2(L(\theta_{H_1})) - L(\theta_{H_2})) \end{aligned}$$

where $L(\theta_{H_i})$ is the log-likelihood under the hypothesis H_i. Then

$$P(T_{01} \geq c) = \sum_{l=2}^{k} P(l,k,w) P[X^2_{l-1} \geq c] \tag{2}$$

$$P(T_{12} \geq c) = \sum_{l=1}^{k-1} P(l,k,w) P[X^2_{k-l} \geq c] \tag{3}$$

where $P(l,k,w)$ denote the probabilities that under H_0 and given k starting dose-levels the isotonic regression will build l level sets. For a more detailed description of the weights $P(l,k,w)$ see 2.4 in [6].
However, this approximation does not hold in binary response if the number of events per dose group is small i.e. less than 5 events [2]. Furthermore when more than one explanatory variable is taken into account, the likelihood ratio test does not follows any known distribution, and that force us to apply Monte Carlo methods to infer.

2.2.2 Changepoint model

Additionally to its nonparametric nature the isotonic procedure can fit a changepoint model. The monotonicity transformation fitted by pooling adjacent violators is always a step function and therefore automatically detects cutpoints for the explanatory variable without any prior information about the location of the shifts. This feature facilitates potentially the threshold value detection, since it returns a small set of candidate locations.

On analysing dose-response relationships we wish often to model the response with respect to isotonic restrictions regarding more than one explanatory variable. On this aim the isotonic procedure as smoother can be extended to multidimensional isotonic regression [1] and [3] and the additive isotonic models [9] [12]. These two methods are briefly described in the following sections keeping in mind that although we focus on applications were the response is binary, these methods work very well also with continuous and Poisson data.

3 Models under order restrictions

3.1 The multidimensional isotonic model

The isotonic smoothing through PAVA can be extended to more than one dimension, applying an iterative algorithm. For the case of two explanatory

variables for example, imagine the data in a form of a matrix. The algorithm works out by calculating iterative the isotonic-rows matrix and the isotonic-columns matrix. Those two matrices converge to a single matrix which is isotonic with respect to the partial order. That means that for every two elements of the matrix, x_{ij}, x_{kl} having $i \leq k$ and or $j \leq l$ then $x_{ij} \leq x_{kl}$. The result can be visualised as a surface that is non decreasing as long as any of the predictors increases. The algorithm combines both the explanatory variables in l constant risk groups, and therefor each step in the response variable corresponds to a specified bivariate group. Of course this procedure can be extended to more than two variables, but note that if more than three isotonic predictors need to be included in the model, the use of this approach is not recommended due to its great computational complexity. The model for k predictors takes the following form:

$$p_i = \phi^*(x_1, x_2, ..., x_k) \tag{4}$$

where ϕ^* is the isotonic transformation.
A main problem arising from this method is that the convergence is not guarantee in case that the data contain many zero-weighted cells. Therefor the data need to be in pre-selected groups. Even if those groups are very thin and selected objectively (for example using quantiles) that can effect somewhat the results since we reduce the candidate changepoint locations. However, this procedure captures interactions between the explanatory variables, feature that the additive model described in a next section does not provide.
The Likelihood Ratio test has no longer any known large sample approximation. Then, the tests used to assess significance for the predictors are exact. Permutations are used to calculate the p-value of the overall fit and conditional permutations for the effect of each variable included in the model. The the conditional for partial significance procedure can be summarised as follows: In each response $Y_i = 0, 1$ corresponds a vector $x_i = (x_{1i}, x_{2i}, ..., x_{ki})$. To test the effect of the j-th predictor adjusted for the rest $k - 1$ predictors we split the vector x_i and then we combine $(Y_i, x_{1i}, .., x_{j-1,i}, x_{j+1,i}, ..., x_{ki})$ and $d_{j,i}$ randomly. In each combination the isotonic regression and the corresponding *Deviance* - defined here as two time the negative change in the likelihood between isotonic model and constant-response model - are computed. To reject then the non-effect assumption the 95th quantile of the empirical distribution of the deviance is compared to the observed deviance. Of course on can test all predictors at once if so desired by combining Y_i to x_i and then following the same procedure as described above. Following the same idea one can construct confidence surfaces for the estimates, simulating under H_0 or H_1. The width of those intervals can useful informations. In cases where a non significant result is obtained, confidence bands can help us to distinguish between statistical and substantial consistency in the risk.

In case of significant result, one can simulate under the assumption that the isotonic estimates are true to estimate the adequacy of the transformation against any other possible shape (that would be equivalent with a test T_{12}).

3.2 The additive isotonic model

The main advantage of those models compared to multidimensional versions is that the data do not need to be in pre-selected groups, although on the other hand they do not capture interactions. (Note: it is possible to modify the model 5 in order to include interactive terms, but that will slow down the convergence rate). Additive isotonic models start from the assumption that the risk (response) does not decrease as long as the predictors increase, and extend GAMs by letting isotonic transformation act like a "smoother". The local scoring algorithm usually used is here replaced by PAVA and the contribution to the risk of each isotonic variable is a non decreasing step function. Nevertheless it is often the case that some covariates need to enter linearly in the model, leading to a semiparametric model. The additive isotonic model with k isotonic predictors and s linear predictors takes the following form:

$$h(p_i) = \sum_{j=1}^{k} \phi_j^*(x_{ij}) + \sum_{t=1}^{s} \beta x_{it} \tag{5}$$

where h is a link function and ϕ^* denotes the isotonic transformation function. Note that the degrees of freedom of each isotonic term are equal to the number of level sets i.e. the "steps" in to which the isotonic transformation ends up. Once the model is formulated, the isotonic Likelihood Ratio test can be used to infer about the explanatory variables and to test the accuracy of the transformation. The large sample approximation described by Robertson et al. [6] does not hold here, so permutations test need to be applied again in finding the empirical distribution of the test. The procedure would be similar to this described in section 3.1 for the multidimensional regression. Additive isotonic models can prove to be a useful tool for exploratory analysis, since they speed up the checking of variables as possible predictors by rejecting those in whom the best isotonic transformation performs poorly [12].
Isotonic regression automatically detects changepoints for the explanatory variables and combines them in constant risk groups. Therefore additive isotonic models are more than adequate to indicate the location of threshold values [9]. That can be accomplished by pooling adjacent level sets of the variable of interest and examining the change in the likelihood function. A large difference warn for threshold value location, otherwise both groups can be pooled together and the procedure goes on (see section 5).

4 The reduced isotonic regression (RI)

4.1 Description of the method

It has been proved that in many applications the use of the isotonic regression overfits somewhat the data whereas a model with fewer level sets (and therefore df) fits better [8]. Therefor it is of great importance once we fit the isotonic regression, to proceed a backward elimination in order to improve the parsimony of the model and get a nested set of constant response levels. The PAV procedure detects violators of the monotonicity assumption and builds the level sets by amalgamating a neighbourhood until there are no more violators. Some of the resulting level sets could be pooled together, especially those with few elements or those whose estimated values don't differ a lot. In order to compute the reduced isotonic regression two steps have to be considered:

- Which level sets have to be pooled together?
- When to stop the elimination's procedure?

Several methods can be applied to answer those questions. M. Schell [8] proposes a F-test when the response is continuous. When the response is binary P.Baccetti reduces the partial fitted functions in the additive isotonic model 5 by comparing a rather arbitrary count to the change in the likelihood [10]. No matter which backward procedure is used, it is be necessary to compare the reduced isotonic model to its "parent model". There is so far no distribution theory for these models, so simulations must be used to choose between simpler and more complex models. That means that under the assumption that the reduced model is the correct one and a given observed change in the likelihood (between reduced and isotonic model), we can extract simulated data. Then, if the 95% interval of the simulated changes in the likelihood contains the observed one, we conclude to the reduced model.
Focusing on binary response we reduce the multidimensional model by basing the pooling in a corrected Chi-squared test. To make the elimination procedure more clear, we will describe the backward algorithm used to reduce the degrees of freedom in a two-dimensional isotonic regression. Consider that the binary outcome depends now on two isotonic predictors of m and n dimensions respectively and imagine the result as a mn matrix. The entries are the estimated isotonic proportions p_i^* for each bivariate group. Since $|p_i^*| \leq |p_i| = mn$, the level sets l are combinations of the mn groups. Our main is to reduce those combinations to $k < l$ with respect to the outcome. The criteria for elimination will be the following:

1. No level set is allowed to contain less than 1% of the total sample

2. The X^2 test with continuity correction and one degree of freedom is used to determine if the fitted values are significantly different in two adjacent level sets
3. We will start the pooling from the pair that gives the greater p-value

Thus the algorithm can be described as follows:

- Obtain the usual isotonic regression estimator with l level sets
- Compute the X^2 test statistics for every adjacent level set and select to pool those level sets whose combination gives the greater p-value
- Repeat step two until all p-values$< a$

Obviously the reduced isotonic regression depends on the choice of a. For a=1 the reduced isotonic regression is identical to the isotonic level sets whereas for a=0 we get a single level set.
The use of a=0.05 in the backward elimination will not yield a 0.05-level test as usual. We applied the reduced monotonic regression to random noise data (independence of the risk variable from both predictors dust and time) having the same sample size and groups as in our application: the result shows that the procedure will not lead to a singleton level set with a probability of about 64%. This finding is comparable to that of M.Schell - for the univariate case the F-test: using a=0.05 the nominal 5% level is exceeded with probability 50,2%, and in Miller and Siegmund's study (they dichotomised a continuous variable according to a maximal X^2 test) they found an excess of 49%. Not surprising since we use a maximally selected p-value in each pooling. Therefore correction needs to be made to assess the correct significance level a^* that will yield to a 5% test. All we need is to simulate with random noise data:

1. Produce random noise data
2. Assess in each data set the isotonic estimators and their reduced equivalents using $a^* = 0$ and retain the p-value from the last Chi-squared test when only two level sets remain to pool
3. The corrected a^* is the 5% smallest from the set of those "last" p-values

In the case of one explanatory variable M. Schell proposes a empirical formulae for assessing the correct $a^* = 0.012n^{-0.361}$ for sample size between 50 and 800.

5 Threshold value procedure (ThP)

The reduced isotonic regression facilitates even more the threshold value detection by shrinking the possible candidates. An other equivalent approach is recently proposed [1] in order to assess thresholds. In assessing a threshold value the constant risk categories corresponding to the isotonic predictor of interest are lumped together starting now from the two lowest groups and the loss in the fit

$$LR_{Thershold} = -2log(L_k) - (-2log(L_{k-1}))$$

(here L_k denotes the likelihood function that corresponds to k starting groups) is analysed. As long as the fit does not decrease significantly the categories are pooled together. If the fit decreases significantly the cutpoint between the categories is used as threshold. This threshold procedure (ThP) can be combined either with additive models, where the partial isotonic function $\phi_j^*(d_{ij})$ is considered as objects of reduction, or with multidimensional isotonic regression, by lumping adjacent bivariate level sets.
This method can be thought as a variable selection procedure, were one has to choose between two different representations of the same predictor to include in the model. Since the degrees of freedom for each term are the number of the resulting level sets, the change in the Likelihood Ratio test should follow a X^2 distribution with one degree of freedom. This approximation is widely used in the case of fractional polynomials to choose between two possible different degrees of the same explanatory variable.
The main problem rising from this idea is the lack of an appropriate test statistic in order to define more clearly what exactly "a large change in the likelihood" is. Simulations have shown that the Likelihood ratio test is not X_1^2 distributed. Setting

$$H_0\text{:}\textit{No treshold can be assessed}$$

simulated data sets have been drown assuming that $p_i = \alpha + \beta d_i$, (i.e. the proportions are linearly dependent on the predictor) for several α, β, i and number of observations per dose group. The figures show that the X_1^2 approximation is not always consistent. The distribution changes by pooling (figure 2), and depends on the value of α and β (figures 3 and 4). So further investigation remains to be done on this area, in order to estimate if possible, the theoretical distribution of the likelihood test. Note that it is quite complicate to find the theoretical distribution of the test $LR_{Threshold} = T_{01,k} - T_{01,k-1}$: the main problem above the unequal weights is that the distributions (see equation 2) corresponding to $T_{01,k}$ and $T_{01,k-1}$ are not independent. An other criterion that can be used here is AIK criterion, especially in the additive models context. The procedure will be the same as before, but now the AIK

information will indicate a "gap" in the goodness of fit i.e. threshold value location.

6 Simulations study

Although the hypothesis of no threshold value existence is not clearly defined, we allow ourselves to attempt some experiments to estimate the performance of the approaches described above under several circumstances. The reduced isotonic regression and the threshold estimation procedure have been applied in simulated data sets in order to investigate first their behaviour as tests for trend (H_{0a}: constant risk against H_{1a}: increasing trend) and second their capability to detect thresholds (H_{0b}: no threshold against H_{1b}: threshold).
Before trying any simulations, it would be necessary to estimate whether the approximation for a^* proposed [8] holds. The result was not favourable for the approximation: the values resulting from the formulae where far from the values resulting from simulations. Also in our application we have to use an a^* assessed through simulations, as described in previous section.
First data sets have been drawn under the *assumption that a threshold truly exists* (H_{1b}) and the risk is a two-steps function of the dose assuming eight dose groups with equal number of observations in each group (100). The power of reduced isotonic regression as test for trend (H_{0a} against H_{1a}) is not satisfactory only in some situations: in table 1 the probability to get a single level set ($P_{red}(ls = 1)$) is quite large and becomes even larger if the "step" in the true function is less than 5%. The same probabilities corresponding to threshold procedure are about zero. However, simulating under H_{0a} the error of type I is very large using Th.P (for example using the same sample size and 5% risk in each group, the probability to get more than one level set is 90,6%), whereas using RI the error of type I is controlled to remains to 5%. So, Th.P is inconvenient as test for trend, and should be used only in cases that the increase in the risk is proved.
The probabilities to assess the correct cutpoint are comparable using both methods. As expected, it decreases as long as the "step" in the true function becomes shorter, having success of about 50% when the increase is of 5% ($P_{red}(cut)$ and $P_{thres}(cut)$in table 1). The location of the cutpoint does not effect the capability of the procedures to detect it (table 2).
To simulate under *no threshold existence* H_{0b} we assumed linear regression using different slopes. The results are depicted in table 3. The probability for each group limit to be assessed as a threshold is approximately normal distributed around the level set that corresponds in a increase in the risk from the first dose group of about 10% for RI or 5% for ThP.

Assumed regression	$P_{red}(ls=1)$	$P_{red}(cut)$	$P_{thres}(cut)$
0.05 0.05 0.15 0.15 0.15 0.15 0.15 0.15	0.032	0.851	0.831
0.05 0.05 0.10 0.10 0.10 0.10 0.10 0.10	0.411	0.387	0.408
0.10 0.10 0.20 0.20 0.20 0.20 0.20 0.20	0.108	0.698	0.711
0.10 0.10 0.40 0.40 0.40 0.40 0.40 0.40	0.000	0.992	0.961
0.15 0.15 0.30 0.30 0.30 0.30 0.30 0.30	0.007	0.895	0.851
0.05 0.05 0.05 0.05 0.15 0.15 0.15 0.15	0.004	0.888	0.828
0.05 0.05 0.05 0.05 0.10 0.10 0.10 0.10	0.275	0.443	0.543
0.10 0.10 0.10 0.10 0.20 0.20 0.20 0.20	0.028	0.762	0.747
0.15 0.15 0.15 0.15 0.30 0.30 0.30 0.30	0.000	0.898	0.871
0.20 0.20 0.20 0.30 0.30 0.30 0.30 0.30	0.157	0.596	0.597
0.20 0.20 0.20 0.35 0.35 0.35 0.35 0.35	0.009	0.857	0.835
0.25 0.25 0.25 0.30 0.30 0.30 0.30 0.30	0.672	0.313	0.262

Table 1: Simulations under the existence of a threshold value. As $P_{red}(ls=1)$ we denote the probability to get a single level set and $P(cut)$ is the probability to assess the correct cutpoint as threshold.

Assumed regression	$P_{red}(cut)$	$P_{thres}(cut)$
0.15 0.15 0.30 0.30 0.30 0.30 0.30 0.30	0.852	0.828
0.15 0.15 0.15 0.30 0.30 0.30 0.30 0.30	0.864	0.829
0.15 0.15 0.15 0.15 0.30 0.30 0.30 0.30	0.867	0.815
0.15 0.15 0.15 0.15 0.15 0.30 0.30 0.30	0.872	0.818
0.15 0.15 0.15 0.15 0.15 0.15 0.30 0.30	0.867	0.789

Table 2: Simulations under the existence of a threshold value-different locations of the threshold

7 Application

7.1 Additive Models

The data are summarized in table 4. A preliminary exploration of the data has been performed by applying generalised additive models using natural cubic splines. The independent variables "time since first exposure" and "dust concentration" enter the model as smooth terms and smoking enters linearly. The overall examination of the fitted functions indicated a rough but in general increasing trend for dust effect. The variable "time since first exposure" presents also an increasing trend. In total eight outliers have been

α, β	Threshold group (RI)	Threshold group (ThP)
0.05, 0.05	4	2
0.10, 0.05	4	2
0.05, 0.10	2	1
0.05 0.025	5	3

Table 3: Simulations under the absence of a threshold value (linear regression $\alpha + \beta d_i$)

	total	without CBR	with CBR
Nonsmokers and ex-smokers			
Dust (mg/cm)	1.43(0-15.05)	1.43(0-15.4)	1.45(0.36-8)
Time (in years)	24(1-66)	23(1-55)	33(8-66)
Smokers			
Dust (mg/cm)	1.40(0.20-15)	1.07(0.20-15)	4.62(0.25-12.07)
Time (in years)	25(3-51)	24(3-51)	28(6-49)

Table 4: Descriptive statistics table

detected and dropped out. Using the Likelihood Ratio test [7], all three predictors have been significant. The reader should keep in mind that the judgement about the trend regarding the partial fit, does not need to be very strict, since GAMs are often objects of over-interpretation and pure local effects can confuse. In any case, the isotonic Likelihood Ratio test applied in next parts will provide the answer we need.

Using the sample and informations resulting from the GAM analysis we fitted additive isotonic models of the form 5. The fitted isotonic function for dust concentration is presented in figure 5. The next step in our analysis is to establish if the isotonic assumption for dust concentration holds, by comparing the full model to the model without dust, resulting in the isotonic Likelihood Ratio test. Although the change in the likelihood function and the equivalent change in the degrees of freedom can guide us to some preliminary conclusions, p-values need to be assessed applying permutations test. With D and D* denoting the original sample deviance and the deviance in the permuted data set respectively, the test can be described as below: In every worker having disease status c_i we assign a vector (d_i, t_i, s_i, c_i) having elements the dust, time and smoking habits observations and the endpoint. To test for constant dust effect, against the increasing risk alternative and adjusting for time and smoking, we break up the vector to (d_i) and (t_i, s_i, c_i) and then we combine them randomly. For every combination an additive isotonic model is fitted and the deviance D* of the model is assessed. Then the 95th quantile of the empirical distribution of D* gives the critical value that if exceeded by D, H_0 is to be rejected. Table 5 summarises the results

Model	-2log(Lik)	df	p-value
$\phi^*(time) + 0.68 smoking$	1318.68	5	
$\phi_1^*(time) + \phi_2^*(dust) + 0.71 smoking$	1260.22	9	0.001

Table 5: Additive isotonic fit. The p-value is assessed using 1000 permutations

Group's limit	-2log(Lik)	df
4.94	1260.22	9
6.23	1275.47	8
7.72	1281.74	7
9.27	1298.98	6

Table 6: Threshold value estimation using additive isotonic models

from fitting the additive isotonic model without and with dust effect, as well as the result from the permutations test for each city sample. Obviously, the dust presents a significant isotonic influence to the risk.

In the estimation of a maximal allowed dust concentration we proceed as described in the previous section. We start pooling dust groups and we examine the change in likelihood. In order to conclude which difference is large enough to indicate a threshold value limit, we compare it to $2.71(=X^2_{1,90})$, since we delete one level set and therefore one degree of freedom. Following this idea a threshold value at 4.94 mg/cm is found (table 6). We conclude at the same threshold even using AIC criterion in the elimination procedure.

7.1.1 Multidimensional regression

A three-dimensional model 4 is applyied using time, dust and smoking as possible predictors. We depicte the results in figure 6. The same permutations procedure as described in section 7.1 is applied to test the conditional influence of dust separate for the sample of smokers and nonsmokers (figure 7). The test results in p-value = 0.01 for smokers and p-value = 0.23 for non smokers. Simulated data sets have been used also to construct confidence surfaces for the isotonic estimators. For non smokers we simulated under the constant risk assumption, whereas for smokers the surfaces correspond to H_1 i.e. the isotonic estimates. Their width is not remarkably large, except perhaps the third and last dust groups, so the consistency of the risk for non smokers and the isotonic transformation for smokers are rather likely to occur.
In estimating the reduced model, we used simulations to assess the corrected a-level. Let us call a^* the value of a that yields an 0.05-level test of H_0 versus

H_1 in the backward procedure. We produced 1.000 random permutations of the response variable as if it was independent on the explanatory variables. We fitted in each permutation the isotonic and reduced isotonic regression. Then to get the a^* that leads to 0.05-significance test we picked the 50th smallest p-value when only two level set remain. Moreover, the 49th and 51th lowest p-values give the 95% confidence limits for a^*. Figure 8present the reduced model corresponding to smokers. For this application we used in the backward elimination procedure a^*=0.00038 (0.00036-0.00039) to get a 5% level test.
The change in the -2logLikelihood between isotonic and reduced model is 14.30. We proceed a last step in order to compare the reduced model to its equivalent isotonic: simulating (1.000 simulations) under the assumption that the reduced model is the correct one, we conclude that such a large change in the Likelihood as the observed could have occur with probability p-value = 0.533. Thus we conclude to the more parsimonious model, and we have found a useful stratification for both variables, that splits them in two groups of high and lower risk. Moreover, it is obvious that a threshold value between 4 and 5 mg/cm can be obtained.

7.1.2 Classification

The additive isotonic model, the multidimensional isotonic model and the generalised additive model have been applied to Chronic Bronchitis data and the ROC curves have been estimated as an index of their classification's capability. All three of them have been found to be significant under the null hypothesis "true area = 0.5" ($p<0.001$). The areas and their standard error were 0.630 (0.008), 0.635 (0.008) and 0.649 (0.008) respectively. The generalised additive model seems to classify slightly better than the other two methods.

8 Discussion

Although much debated from many statisticians, the use of step functions in modelling can prove to be very useful in lots of cases. Step functions are easy to interpret and by fitting changepoint models they summarise the predictors in such a way to define groups of constant response. This feature can be desired in many studies. As an example we refer the case of car insurance studies where the main is to define groups of drivers having common risk to produce accident regarding some prognostic factors in order to assess the convenient fee for each group. Recently methods elegant and simple, as the classification and regression trees, are getting always more popular to fit such changepoint models.

Even more than splitting the explanatory variables in constant response groups, it is sometimes of great importance to establish dose-response relationship. This requirement is important in cases that the aim of a study is to prove causality between factors and outcome. On this direction, we wish that the regression function has a monotonic shape. That is crucial in medical studies, where the researcher want to test whether the increasing exposure in a risk factor is associated to an increasing disease rate.

Monotonic regression combines both desired characteristics described above resulting in a monotonic step function. The more efficient algorithm to fit the model is PAVA. The test for trend that is provided is very powerful, however his theoretical distribution should be used with caution in cases of small event rate.

In modelling with more than one predictor there are two ways to fit the data under the monotonicity assumption. The additive isotonic model does not contain any interactive terms so far, but its advantage is that the predictors can be used as continuous if so required. Multidimensional regression is useful to model the relationship when the predictors are suspected to interact. The disadvantage of this approach is that the explanatory variables can not be more than three, and they need to be in pre-selected quantiles. The main problem arising from both methods is the luck of an appropriate approximation for the distribution of the test statistic. Conditional permutations need to be used in order to test the effect of each predictor that makes the use of the model cumbersome. However, since the test is a test for trend (the monotonic transformations is the alternative to the constant risk assumption) it is sometimes important to assess despite the computational complexity.

The reduced isotonic regression has also been introduced. The procedure focuses on finding a subset of the level sets resulting form the isotonic regression. The model becomes more parsimonious but the selection of the best model should be based again on simulations. The constant risk assumption can be rejected if the reduced isotonic regression results to more than one

level set. Given that the elimination's procedure has be designed so that eliminates the error of type I, the power of the test is satisfactory.
Monotonic regression can be used in the threshold value assessment context. Defining as threshold the first cutpoint of the reduced level sets, reduced isotonic regression can also provide a threshold value detection procedure, presenting good results for thresholds that correspond to more than 5% increase.

An other similar approach to the threshold problem is by pooling adjacent level sets and then testing the homogeneity of the pooled levels by a Likelihood Ratio test. Although the procedure appears to yield satisfactory results, similar to those taken from the reduced isotonic regression, it is unlikely that the X_1^2 approximation for the Likelihood Ratio test works correctly. However both approaches can be used in combination with other threshold value detection procedures, as additional tool.
Further developments of isotonic models are also possible, as the introduction of random effects, their combination with smoothing splines. It would be interesting also to investigate the introduction of the isotonic smoothing to dynamic Cox model.

References

[1] Ulm, K.(1999), "Nonparametric Analysis of Dose-Response Relationships" ,Annals of the New York academy of Sciences 895, 223-231

[2] Ulm K.,Dannegger F.,Becker U.,(1998), Tests for trends in binary response, Discussion paper 115, SFB 386

[3] Salanti G., Ulm K., (2001), Multidimensional Isotonic Regression and Estimation of the Threshold Value. Discussion Paper No. 234, Sonderforschungsbereich 386 der Ludwigs-Maximilian-University Munich

[4] Ulm K., Salanti G.,(2001), Estimation of the general thresholdlimit values for dust, Int Arch Occ Env Health (to be appear)

[5] Chambers J., Programming with data, USA ,1998

[6] Robertson T., Wright F. T. and Dykstra R. L. Order Restricted Statistical Inference, Wiley, Neu York, USA, 1988

[7] Hastie TJ., Tibshirani RJ.,(1995), Generalized Additive Models for Medical Research. Stat Methods Med Res 4: 187-196

[8] Schell M.J., Singh B.,(1997), The reduced Monotonic Regression Method, J.Am.Stat.Assoc. 92:128-135

[9] Morton-Jones T., Diggle P., Parker L.,(2000), Additive Isotonic Regression Models in Epidemiology, Stat Med 19: 849-860

[10] Chuang-Stein C, Agresti A (1997), Tutorial in Biostatistics: A review of tests for detecting a monotone dose-response relationship with ordinal response data, Stat Med 16: 2599-2618

[11] Mancuso J.,Hongshik A.,Chen J.,(2001),Order-restricted dose-related trend tests, Stat Med 20: 2305-2318

[12] Baccetti P.,(1989), Additive Isotonic Models, JASA 84: 289-294