THE HANDLING OF CONTINUOUS TARIFF VARIABLES: TIPS AND EXPERIENCES

HANS SCHMITTER
DINO TONIOLO

1998 GENERAL INSURANCE CONVENTION
AND
ASTIN COLLOQUIUM

GLASGOW, SCOTLAND: 7-10 OCTOBER 1998

The handling of continuous tariff variables: Tips and experiences

[Hans Schmitter and Dino Toniolo, Winterthur Insurance, General Guisan-Strasse 40, P.O. Box 357, CH-8401 Winterthur]

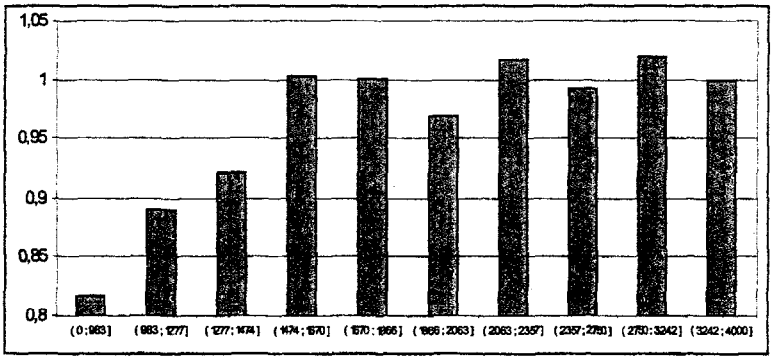
Abstract:

Using continuous variables to describe some data we often incur in some plausibility problems for extreme low and high values. Assuming that the continuous function which describes the relationship between some explanatory variables (covariates) and a response variable lies in a affine linear space (as in the case of the generalised linear models) we can solve these problems with an appropriate choice of the affine linear space.

1. The use of classes

Modern tariffs use several criteria to price a risk. To take an example in Europeia's new motor liability tariff the premium depends on geographical zones, car types, car power, the driver's sex, age and years of driving. Whereas criteria like the zone, the car type or the driver's sex are naturally class variables, the handling of continuous variables like the power of the car or the driver's age is not so obvious. Traditionally, they too have been treated like class variables. This can be done by partitioning the range of the continuous variable into classes. As an example the table below shows cubic capacity classes and the corresponding graph shows the corresponding coefficients resulting from a GLM analysis (taken from an analysis carried out in a company of the Winterthur group):

Class	Ccm range			Coefficient	Class	Ccm range			Coefficient
1	0	to	983	0.8177	6	1867	to	2063	0.9686
2	984	to	1277	0.8906	7	2064	to	2357	1.0171
3	1278	to	1474	0.9213	8	2358	to	2750	0.9927
4	1475	to	1670	1.0026	9	2751	to	3241	1.0196
5	1671	to	866	1.0009	10	3242	and more		1



There are three drawbacks to this approach: The first is the well known problem of setting limits between classes; they should not be too arbitrary. Secondly, the results should lend themselves to a convincing interpretation. In the above example this is not the case: No practitioner would accept the sequence of coefficients of the classes 5 to 9, it just looks too irregular. The third drawback is the large number of parameters: In order to distinguish between 10 cubic capacity classes we need to determine 9 parameters.

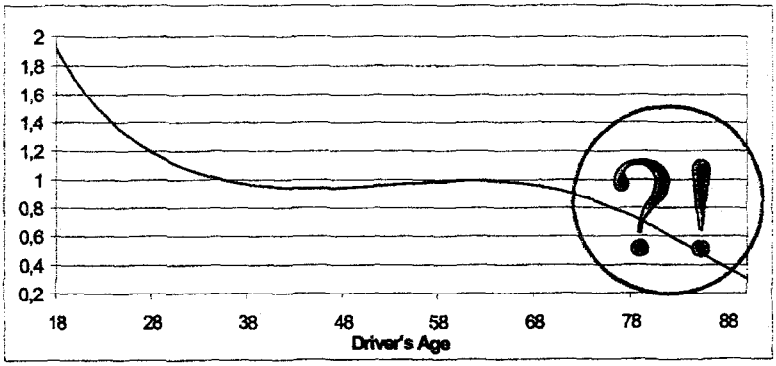
2. Using polynomials

The disadvantages of using classes mentioned in the previous section can be avoided by using polynomials. Thus e. g. within the framework of generalised linear models the influence of the driver's age can be described as follows:

$$\lambda = \exp\left(\ldots + b_1.age + b_2.age^2 + b_3.age^3 \ldots\right)$$

Practitioners assume the age coefficient to be a continuous function of the age. Therefore they accept also the logarithm of the coefficient to be a continuous function of the age, and this continuous function can be approximated by a polynomial.

Usually polynomials of degree 3, 4 or at most 5 lead to good, interpretable results. Therefore models using polynomials normally need fewer parameters than class models. However, anybody who has used polynomials in the construction of tariffs has been faced with the fact that they may lead to useless results in the range where there are no or only few observations. In the example of the age variable this is often the case for the very young and the very old drivers: It happens that the age coefficient increases as a function of the age for very young drivers or drops below every boundary for the very old whereas everybody would expect the contrary – only because the data does not contain enough young and very old drivers.



3. Using piecewise straight functions

There is an alternative to polynomials which has proven very useful in practical applications, namely piecewise straight lines. To take an example, look at the driver's age again: It is known that young drivers have a higher claim frequency than middle-aged drivers, and that for old drivers the frequency increases again. These facts can be modelled using the following covariates in a generalised linear model:

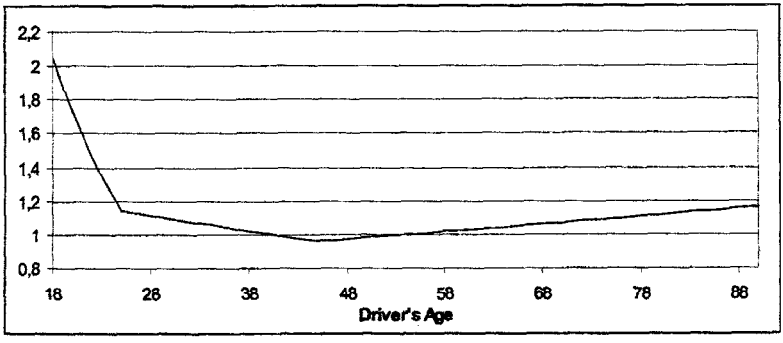
agey =	age	if age <= 25
	0	otherwise
agem =	0	if age<=25
	age-25	if 25 < age <= 45
	20	otherwise
ageo =	0	if age <=45
	age-45	otherwise

Thus the age of the driver is: age= agey + agem + ageo.

Using the model

$$\lambda = \exp(.... + b_1.agey + b_2.agem + b_3.ageo + ...)$$

we determine the coefficients b_1, b_2, b_3. This usually leads to acceptable results which look as follows:



Of course the number of covariates which describe a continuous variable (3 in our example) is arbitrary. On the other hand their maximum values (25 for agey and 20 for agem in the example) can be obtained by maximising the maximum likelihood of the corresponding models. We illustrate this in the simplest case when we have to determine only one value, i. e. when we replace a continuous variable which we call a (for age) by two covariates.

Let

a_i continuous covariate (e.g. age) of observation number i

c limit which must be determined

$$x_{i,1} = a_i - (a_i - c)^+$$

$$x_{i,2} = (a_i - c)^+$$

The derivatives of $x_{i,1}$ and $x_{i,}$ with respect to c are

$$x_{i,1}' = \begin{cases} 0 & \text{if } a_i \le c \\ 1 & \text{otherwise} \end{cases}$$ and

$$x_{i,}' = \begin{cases} 0 & \text{if } a_i \le c \\ -1 & \text{otherwise.} \end{cases}$$

Using the Poisson model, the frequency of the ith observation is

$$\lambda_i = \exp(.... + b_1 x_{i,1} + b\ x_{i,} + ...)$$

the logarithm of the Poisson probability of observing y_i claims, $\log f_i$, is

$$\log f_i = -\exp(.... + b_1 x_{i,1} + b_2 x_{i,2} + ...) + y_i.(.... + b_1 x_{i,1} + b_2 x_{i,2} + ...) - \log(y_i!)$$

and its partial derivative with respect to c is equal to

$$\frac{d}{dc}\log f_i = \begin{cases} -\lambda_i.(b_1 - b_2) + y_i.(b_1 - b_2) & \text{if } a_i > c \\ 0 & \text{otherwise.} \end{cases}$$

Thus the partial derivative of the loglikelihood with respect to c is the same as if there was a *class variable* z_i *defined as follows:*

$$z_i = \begin{cases} 1 & \text{if } a_i < c \\ 0 & \text{otherwise} \end{cases}$$

We can extend our model by introducing such a class variable z_i so that the frequency of the ith observation becomes

$$\lambda_i = \exp(.... + b_1 x_{i,1} + b_2 x_{i,2} + h z_i).$$

If the coefficient h is 0 we have found an optimal c which maximises the loglikelihood of the model. Otherwise we have to try with another value for c. As soon as we find two c values leading to coefficients h of which one is positive and one is negative it takes about 5 iterations of the regula falsi to get the optimal c.

4. Introducing some additional conditions

In a generalised linear model we try to estimate the expected value of a random variable Y, using the inverse of a link function applied to a linear combination of some explanatory variables (covariates):
$E[Y] = link^{-1}(I + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_m x_m)$, where the β_i and the I are estimated using a maximum likelihood approach.

If we want to investigate the relationship between $E[Y]$ and a continuous variable x, we can use for instance a polynomial: $link(E[Y]) = I + \beta_1 x + \beta_2 x^2 + ... + \beta_m x^m$, as in paragraph 2.

The set of real polynomials in one variable up to a given degree m can be considered as a real linear space with basis $(1, x, x^2, ..., x^m)$. So, by trying to estimate the values of β and I, using a maximum likelihood approach, we try to find the element in a linear space, which minimises some kind of error.

By estimating the values β and I, we get a continuous function $F(x) = I + \beta_1 x + \beta_2 x^2 + ... + \beta_m x^m$.
In practice we often observe that these continuous functions give some interpretations problems for high and low values of the variable x. In fact, in the marginal domains of the function F, the quantity of data will not be sufficient to have a reasonable influence on the estimation of the coefficients. Therefore the behaviour of the curve in the marginal domains will only be a "reflex" of the data in the central domain, where the x values of the data are most dense, without considering the "reality" that the small set of data in the marginal domain shows.

In paragraph 3 we saw that in this case an appropriate choice of the functional space can be very helpful (using the functional space of the piecewise straight functions instead of the linear space of the polynomials), but the estimation of an "optimal" functional space (i.e. the estimation of the parameter c) is a non-linear problem.

Another solution is the following: we try to fix some values of the function, so that the function itself will be no longer free in these critical domains. For this solution, is not necessary that the function F is a polynomial, and the function F does not only have to depend on one real variable. The only condition that has to be met (and it is always met by generalised linear models) is that we will search our function F in a linear space V with a finite dimension.
We will present three variations of this kind of solution: in the first we start with a given linear space and we construct an affine linear subspace. In the second and in the third we will define some interesting linear spaces.

4.1. Let $V \subset \{F : \Re^q \to \Re\}$ be a linear space of finite dimension: $\dim(V) = n < \infty$. Let $(f_1,...,f_n)$ be a basis of V.

Let $F = \sum_{i=1}^{n} \beta_i . f_i \in V$. Now we fix k points $a_1,...,a_k$ where the function F has to take some given values $b_1,...,b_k$, i.e. $F(a_j) = b_j \ \forall j = 1,...,k$.

Which conditions must the β_i meet, so that $F(a_j) = b_j \ \forall j = 1,...,k$, where $a_j \in \Re^q$ and $b_j \in \Re$ are given ?

$F(a_j) = b_j$ has the same meaning as $\sum_{i=1}^{n} \beta_i . f_i(a_j) = b_j$ and therefore

$F(a_j) = b_j \ \forall j = 1,...,k$ has the same meaning as $M.\beta = b$, where

$$M := \begin{bmatrix} f_1(a_1) & \cdots & f_n(a_1) \\ \vdots & & \vdots \\ f_1(a_k) & \cdots & f_n(a_k) \end{bmatrix} \in \Re^{k \times n}, \beta := \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \in \Re^{n \times 1} \text{ and } b := \begin{bmatrix} b_1 \\ \vdots \\ b_k \end{bmatrix} \in \Re^{k \times 1}.$$

Without loss of generality: $Rank(M) = k \le n$: If some of the k equations $\sum_{i=1}^{n} \beta_i . f_i(a_j) = b_j$ are linearly depending of the others, then they are superfluous or they lead to a system of equations without solution.

$Rank(M) = k \Rightarrow \exists M^* \in \Re^{k \times k}$ submatrix of M, which can be inverted.

Without loss of generality: let $M^* := \begin{bmatrix} f_1(a_1) & \cdots & f_k(a_1) \\ \vdots & & \vdots \\ f_1(a_k) & \cdots & f_k(a_k) \end{bmatrix} \in GL_k(\Re)$ be this matrix.

We define:

$$M^{**} := \begin{bmatrix} f_{k+1}(a_1) & \cdots & f_n(a_1) \\ \vdots & & \vdots \\ f_{k+1}(a_k) & \cdots & f_n(a_k) \end{bmatrix} \in \Re^{(n-k) \times n}, \ \beta^* := \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \in \Re^{k \times 1}, \ \beta^{**} := \begin{bmatrix} \beta_{k+1} \\ \vdots \\ \beta_n \end{bmatrix} \in \Re^{(n-k) \times 1}.$$

$$M.\beta = b \Leftrightarrow \begin{bmatrix} M^* & M^{**} \end{bmatrix} \begin{bmatrix} \beta^* \\ \beta^{**} \end{bmatrix} = b \Leftrightarrow M^*.\beta^* + M^{**}.\beta^{**} = b \Leftrightarrow$$

$$\beta^* = (M^*)^{-1}.(b - M^{**}.\beta^{**}).$$

On the other hand : $F = \sum_{i=1}^{n} \beta_i . f_i = \begin{bmatrix} f^* & f^{**} \end{bmatrix} \begin{bmatrix} \beta^* \\ \beta^{**} \end{bmatrix}$,

where $f^* = [f_1 \ \cdots \ f_k] \in V^k$ and $f^{**} = [f_{k+1} \ \cdots \ f_n] \in V^{(n-k)}$.

$$F = \begin{bmatrix} f^* & f^{**} \end{bmatrix} \begin{bmatrix} \beta^* \\ \beta^{**} \end{bmatrix} = f^*.\beta^* + f^{**}.\beta^{**} = f^*(M^*)^{-1}.(b - M^{**}.\beta^{**}) + f^{**}.\beta^{**} =$$

$$f^*.(M^*)^{-1}.b + \left[f^{**} - f^*.(M^*)^{-1}.M^{**}\right]\beta^{**}$$

Summary:

Each choice of β^{**} (i.e. of the coefficients $\beta_{k+1},...,\beta_n$) generates a function $F = f^*.(M^*)^{-1}.b + \left[f^{**} - f^*.(M^*)^{-1}.M^{**}\right]\beta^{**}$, which fulfils the k equations $F(a_j) = b_j \ \forall j = 1,...,k$. The set of these functions is an affine linear subspace with origin $f^*.(M^*)^{-1}.b$ and Basis $\left[f^{**} - f^*.(M^*)^{-1}.M^{**}\right]$.

Therefore, in a generalised linear model, we have to replace the old covariates $(f_1,...,f_n)$ with the new covariates $\left[f \quad -f\ .(M\ )^{-1}.M\ \right]$ and we have to add to the previous offset variable the term $f\ .(M\ )^{-1}.b$.

Example:

Let $V = \left\{F : (x,y) \mapsto \sum_{1 \le i+j \le 2} \beta_{ij}.x^i.y^j : \beta_{ij} \in \Re\right\}$. Basis of $V = (y, x^2, xy, x, y^2)$.

We look for functions $F \in V$, which meet the following conditions: $F(0,1) = 1$, $F(1,0) = 1$ and $F(1,1) = 0$

$$\Rightarrow M = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix},\ M^* = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix},\ b = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \Rightarrow$$

$$(M^*)^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & -1 & 1 \end{bmatrix},\ (M^*)^{-1}.b = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix},\ (M^*)^{-1}.M^{**} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \text{ and therefore}$$

$$F = \begin{bmatrix} y & x^2 & xy \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} + \left(\begin{bmatrix} x & y^2 \end{bmatrix} - \begin{bmatrix} y & x^2 & xy \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \right) \begin{bmatrix} \beta_3 \\ \beta_4 \end{bmatrix} =$$

$$(y + x^2 - 2xy) + \beta_3.(x - x^2) + \beta_4.(y^2 - y), \text{ or}$$

$$F = (y + x - 2xy) + \gamma_1.(x - x^2) + \gamma_2.(y - y^2), \text{ where } \gamma_1 := \beta_3 - 1,\ \gamma_2 := -\beta_4$$

So, in a generalised linear model, we will estimate the values γ_1 and γ_2 of the covariates $x - x^2$ and $y - y^2$, which maximise the likelihood function, where the

offset variable in our new model is given by the previous offset variable plus the variable $y+x-2xy$.

Observation: If $b=0$, then the functions F lay in a linear subspace of V.

4.2. It is possible, with an appropriate choice of the linear space V, to define multidimensional domains where the functions F have to be equal to zero (and not only a finite number of points as in 4.1).

Example $(q=2)$:
On $\Re^2$ we take a sequence of n points $(A_1, A_2, ..., A_n)$.
We define $N := \bigcup_{i=1}^{n-1} \overline{A_i A_{i+1}}$ (N is a polygonal way in the plane).
We look for a linear space of continuous functions F, with the characteristic $F : F(x)=0 \ \forall x \in N$.

Idea: We start to find one continuous function G with the suitable property $G : G(x)=0 \ \forall x \in N$.

Construction of G:
For each segment $\overline{AB}$ with $A=(a_1,a_2)$ and $B=(b_1,b_2)$ we define
$g_{\overline{AB}}(x,y) := \max((-x'-|y'|);0;(x'-|y'|-|AB|))+|y'|$, where

$$\begin{pmatrix} x' \\ y' \end{pmatrix} := \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix} \begin{pmatrix} x-a_1 \\ y-a_2 \end{pmatrix} \text{ and}$$

$$\alpha := \arctan\left(\frac{b_2-a_2}{b_1-a_1}\right) \text{ if } b_1 > a_1, \ \alpha := \arctan\left(\frac{b_2-a_2}{b_1-a_1}\right)+\pi \text{ if } b_1 < a_1,$$

$$\alpha := \pi/2 \text{ if } a_1 = b_1 \text{ and } b_2 > a_2, \ \alpha := 3\pi/2 \text{ if } a = b \text{ and } b_2 < a_2.$$

From the construction of $g_{\overline{AB}}$: $g_{\overline{AB}}(x,y)=0 \Leftrightarrow (x,y) \in \overline{AB}$.

There are two natural possibilities to define G:

$$G(x,y) := \prod_{i=1}^{n-1} g_{\overline{A_i A_{i+1}}}(x,y) \text{ and } G(x,y) := \min\left\{g_{\overline{A_i A_{i+1}}}(x,y) : i=1,...,n-1\right\}.$$

Now let W be a linear space of continuous functions mapping $\Re^2$ to $\Re$ with Basis $(w_1, w_2, w_3, ...)$ (for instance $(x, y, x^2, xy, y^2, ...)$ could be our basis).

This basis and the function G generate a new linear space $V = span(G.w_1, G.w_2, G.w_3, ...)$, and its elements are functions that are equal to zero on N. Namely: let $F = \sum_i \beta_i . G.w_i \in V$, then

$$F(x) = \sum_i \beta_i.G(x).w_i(x) = G(x).\sum_i \beta_i.w_i(x) = 0 \ \ \forall x \in N.$$

4.3. In practice, the most used linear space V is the linear space of the polynomials up to a given degree. Unfortunately, in this linear space, is not possible to have some asymptotical properties, like

$$\lim_{x \to 0} F(x) = \infty \text{ or } \lim_{x \to \infty^q} F(x) < \infty.$$

Both properties can be obtained with an easy transformation of the variables:
we replace $x^{old} = (x_1,...,x_q)$ with $x^{new} = (1-1/x_1,...,1-1/x_q)$.

Example: Let $q = 1$. We search functions $F : \lim_{x \to \infty} F(x) = k < \infty$.
Take $V = span(y, y^2, y^3)$, where $y = 1 - 1/x$. Now let $F = \beta_1.y + \beta_2.y^2 + \beta_3.y^3 \in V$
$\Rightarrow \lim_{x \to \infty} F(x) = \beta_1 + \beta_2 + \beta_3$
$\lim_{x \to \infty} F(x) = k \Leftrightarrow k - \beta_1 - \beta_2 = \beta_3$
$F = \beta_1.y + \beta_2.y^2 + (k - \beta_1 - \beta_2)y^3 = k.y^3 + (y - y^3)\beta_1 + (y^2 - y^3)\beta_2.$

Actually, after the transformation of the variable x in the variable y, we applied the same method we showed before in 4.1. In this last example the functions F lies in an affine linear subspace of $V = span(y, y^2, y^3)$ with Basis $(y - y^3, y^2 - y^3)$ and origin $k.y^3$.