# Nonlinear Regression Models with Applications in Insurance

Rastislav Potocký and Milan Stehlík*

*Department of Applied Mathematics and Statistics, Faculty of Mathematics, Physics and Informatics, Comenius University, Mlynská dolina, 842 48 Bratislava 4, Slovak Republic*

**Abstract:** Two possible applications of nonlinear regression models in insurance are discussed. The first part deals with modelling IBNR reserves when a cubic approximation to the solution locus is used instead of linear or quadratic ones. A formula is given for construction of improved confidence regions for parameters in such models.Using this approach IBNR reserves for a data set are computed.In the second part a method is proposed of how to measure the influence of additive perturbations on nonlinear regression model parameters. An example is given which shows how this method can be used to preserve privacy of sensitive data in insurance business.

**Keywords:** Nonlinear regression models, confidence regions, additive perturbations, IBNS reserves, privacy of data.

## 1. INTRODUCTION

Due to the latest developments in insurance and reinsurance industries nonlinear regression models became more popular with steadily increasing importance. Let us mention two up-to date applications: nonlinear credibility estimation and nonlinear modelling of IBNR reserves (i.e. reserves for incurred but not reported losses).

De Vylder [1] extended Hachemeister's linear regression credibility model to a nonlinear regression model by assuming that observations are an arbitrary function f(β(**θ**)) of the unknown vector β(**θ**). This model lacks of robustness of credibility estimators. Pitselis [2] applied robust statistics to De Vylder's nonlinear credibility estimation and presented an application to Hachemeister's data.

The process of IBNR modelling and estimation has been studied by actuaries for many years since both RBNS (reported but not settled) and IBNR reserves are the largest liabilities of insurance companies. Recall that incurred but not reported loss or IBNR is the difference between ultimate loss and reported loss. Thus quantifying the uncertainty in estimation of IBNR plays the important role in insurance business. The classical approach makes use of run-off triangles (e.g. the chain-ladder method and its modifications). There exist also direct methods for computing or modelling of IBNR, for instance those using copulas and indirect methods based on estimation of loss development factors. For instance, Stelltjes [3] presents a model for predicting losses as a function of exposures, calendar period and development age. Typically, then a nonlinear regression model is used for estimating the 95% confidence interval of IBNR loss for an accident period.However using the quadratic approximation of the model function may lead to inaccurate confidence regions for parameters of the model.

Consider the usual nonlinear regression model

$$y_a = f(x_a, \boldsymbol{\theta}) + e_a \qquad a = 1,...n \tag{1}$$

$\boldsymbol{Y} = (y_1,...y_n)'$ denotes the vector of observations, the function $\boldsymbol{f}(\boldsymbol{\theta}) = (f(x_1, \boldsymbol{\theta}),..., f(x_n, \boldsymbol{\theta}))$ has a known form dependent on p unknown parameters $\boldsymbol{\theta} = (\theta_1,...\theta_p)'$, $x_a$ are known vector-valued variables, $e_a$ are independent and normally distributed random errors with zero mean and variance $\sigma^2$.

The problem of finding acceptable confidence regions for $\boldsymbol{\theta}$ has been discussed by many authors, see e.g. [4-6]. Cook and Goldberg [7] showed examples of models for which the Bates-Watts methodology based on quadratic approximations did not work. On the other hand Clarke [8] presented a method of constructing regions with higher precision. However, his investigations deal with a single parameter $\theta_i$, not with the whole vector $\boldsymbol{\theta}$**.** The aim of our paper is to construct confidence regions based on cubic approximation which are more accurate than those currently used.

In what follows pre-and post-, square bracket and $\otimes$– multiplications of a three-dimensional array $U^{\bullet}_{\bullet\bullet} = \left(U^a_{ij}\right)$ or a four-dimensional array $U^{\bullet}_{\bullet\bullet\bullet} = \left(U^a_{kij}\right)$ by a matrix *E* mean summation over the indexes i, j, a, k, respectively. The reader is recommended to consult Table **1** for better understanding of these operations. Recall that if $U^{\bullet}_{\bullet\bullet} = \left(U^a_{ij}\right)$, a=1,…,n, i=1,…,p, j=1,…,q is an nxpxq array, then its a-th face is pxq matrix $\left(U^a_{ij}\right)$ and its ij-th column $\left(U^1_{ij},...,U^n_{ij}\right)^T$ is *n*-vector.

*Address correspondence to this author at the Institut für angewandte Statistik, Johannes Kepler University in Linz, Freistädter Strasse 315, 2. Stock, A-4040 Linz a. D., Austria; Tel:+43 732 2468 5881; Fax: +43 732 2468 9846; E-mails: mlnstehlik@gmail.com, Milan.Stehlik@jku.at

1876-5270/10 2010 Bentham Open

**Table 1. Rules for Multiplication**

| | | |
|---|---|---|
| $U^{\cdot}_{\cdot\cdot}$ - n x p x q ; $E$- s x p | $EU^{\cdot}_{\cdot\cdot}$ -n x s x q | $\left(EU^{\cdot}_{\cdot\cdot}\right)^{a}_{ij} = \sum_t E_{it}U^{a}_{tj}$ |
| $U^{\cdot}_{\cdot\cdot}$ - n x p x q ; $E$-q x s | $U^{\cdot}_{\cdot\cdot}E$-n x p x s | $\left(U^{\cdot}_{\cdot\cdot}E\right)^{a}_{ij} = \sum_t U^{a}_{it}E_{tj}$ |
| $U^{\cdot}_{\cdot\cdot}$ - n x p x q ; $E$- s x n | $[E]\ [U^{\cdot}_{\cdot\cdot}]$ - s x p x q | $\left([E][U^{\cdot}_{\cdot\cdot}]\right)^{a}_{ij} = \sum_t E_{at}U^{t}_{ij}$ |
| $U^{\cdot}_{\cdot\cdot\cdot}$ -n x r x p x q;$E$ -s x r | $E \otimes U^{\cdot}_{\cdot\cdot\cdot}$ - n x s x p x q | $\left(E \otimes U^{\cdot}_{\cdot\cdot\cdot}\right)^{a}_{ijk} = \sum_t E_{it}U^{a}_{tjk}$ |

Put

$$V = \left(\frac{\partial f_a}{\partial \theta_i}\right)_{\theta=\hat{\theta}}, \qquad V^{\cdot}_{\cdot\cdot} = \left(\frac{\partial^2 f_a}{\partial \theta_i \partial \theta_j}\right)_{\theta=\hat{\theta}} \quad \text{and} \quad V^{\cdot}_{\cdot\cdot\cdot} = \left(\frac{\partial^3 f_a}{\partial \theta_i \partial \theta_j \partial \theta_k}\right)_{\theta=\hat{\theta}}$$

where a=1,…n, i, j, k=1,…p and $\hat{\theta}$ is the least –squares estimate of $\theta$. Let $V = UR$ be the unique QR-decomposition of V where $R$ is an upper triangular matrix and the colums of $U$ form an orthogonal basis for the column space of $V$.

Recall that by the Bates-Watts parameter-effects array we understand the $pxpxp$ array

$$A^{\cdot}_{\cdot\cdot} = \left[U^T\right]\left[\left(R^{-1}\right)^T V^{\cdot}_{\cdot\cdot} R^{-1}\right] \quad \text{and similarly} \quad A^{\cdot}_{\cdot\cdot\cdot} = \left[U^T\right]\left[R^{-1} \otimes \left(\left(R^{-1}\right)^T V^{\cdot}_{\cdot\cdot\cdot} R^{-1}\right)\right]$$

is the four dimensional parameter-effects array (for details see [9]). Bates and Watts have used $A^{\cdot}_{\cdot\cdot}$ to compute the maximum parameter-effects curvature which, together with the maximum intrinsic curvature, give answer to the question whether the model may be considered sufficiently linear or not for the set of values of parameters we are interested in. As mentioned above the quadratic approximation of the model function does not give satisfactory results in some cases, so a cubic approximation is needed..

Consider a partition $\theta = \left(\theta^T_{(1)}, \theta^T_{(2)}\right)^T$, $\theta_{(1)} = \left(\theta_1, \ldots \theta_q\right)^T$, $\theta_{(2)} = \left(\theta_{q+1}, \ldots \theta_p\right)^T$, where $\theta_{(2)}$ is the parameter of interest, $\theta_{(1)}$ being a nuisance parameter. Let $S(\theta)$ be the usual sum of squares for the model (1) and $g\left(\theta_{(2)}\right) = \left(g_1\left(\theta_{(2)}\right), \ldots, g_q\left(\theta_{(2)}\right)\right)^T$ be the value of $\theta_{(1)}$ which minimizes $S\left(\theta_{(1)}\right) = S\left(\theta_{(1)}, \theta_{(2)}\right)$ for given $\theta_{(2)}$. Put $\alpha\left(\theta_{(2)}\right) = \left(g\left(\theta_{(2)}\right), \theta_{(2)}\right)$ and $h\left(\theta_{(2)}\right) = f\left(\alpha\left(\theta_{(2)}\right)\right)$.

The approximate confidence region based on the likelihood ratio (the likelihood region for short) is such a set of values $\theta_{(2)}$ for which

$$\left(Y - h\left(\theta_{(2)}\right)\right)^T \left(Y - h\left(\theta_{(2)}\right)\right) - S\left(\hat{\theta}\right) \le c^2\rho^2, \qquad (2)$$

where $\rho = \sigma$ and $c^2 = \chi^2\left(p-q, \delta\right)$ if $\sigma$ is known or $\rho^2 = \left(p-q\right)s^2$ and $c^2 = F\left(p-q, n-p, \delta\right)$ if $\sigma^2$ is estimated by $s^2 = S\left(\hat{\theta}\right)/\left(n-p\right)$.

In the first part the solution of the above-mentioned problem for parameter vector is given which is then compared to a result of Clarke [8]. Next we discuss the effect of perturbation of the response vector on the values of parameters in a nonlinear model. In the final part our results are applied to the problem of finding acceptable reserves for claims in insurance business and the problem of preserving privacy of stored data.

## 2. APPROXIMATE CONFIDENCE REGIONS

We assume that

1) $f(\theta)$ is a continuous function of $\theta$ with finite derivatives up to and including degree 4

2) the vector $\frac{\partial S(\theta)}{\partial \theta}$ vanishes at one point $\hat{\theta}$ where $S$ is the corresponding sum of squares

3) the Bates-Watts intrinsic curvature as well as $\left(\hat{e}\right)^T (\varphi)^T \otimes \varphi^T V^{\cdot}_{\cdot\cdot\cdot} \varphi$ ) and $\hat{e}^T\left(\varphi^T \otimes \left(\varphi^T V^{\cdot}_{\cdot\cdot\cdot} \varphi\right)\right)$ can be neglected where $\varphi = \theta - \hat{\theta}$ and $\hat{e} = Y - f\left(\hat{\theta}\right)$.

Following the argument in [9] and using the cubic approximation of $h\left(\theta_{(2)}\right)$ we rewrite (2) as

$$\eta_2^T\eta_2 + \eta_2^T\left(\eta_2^T A^2_{22}\eta_2\right) + \frac{1}{3}\eta_2^T\left(\eta_2^T \otimes \left(\eta_2^T A^2_{222}\eta_2\right)\right) - \eta_2^T\left[\eta_2^T\right]\left[A^2_{21}\right]\left(\eta_2^T A^1_{22}\eta_2\right) - \| \left[\eta_2^T\right]\left[A^2_{12}\right]\eta_2 \|^2 + \frac{1}{4}\| \eta_2^T A^2_{22}\eta_2 \|^2 + \ldots \le c^2\rho^2, \qquad (3)$$

where $\eta_2 = R_{22}\left(\theta_{(2)} - \hat{\theta}_{(2)}\right)$ and the neglected terms are of degree 4 and higher in $\eta$.

Put $\Psi_2 = \lambda b, \Psi_2^* = \lambda^* b,$ where $b$ is a unit vector of $R^q$, $\lambda, \lambda^* \geq 0$, $\Psi_2$ and $\Psi_2^*$ are boundary points of the region (3) and the *L-region* $\eta_2^T \eta_2 \leq c^2\rho^2$, respectively. Inserting $\Psi_2 = \lambda b$ into (3)

and expressing $\lambda$ as a power series in $c\rho$ we obtain

**Theorem 1.** Under the above assumptions

$$\lambda(b) = \lambda^*(b)\left\{1 - \frac{c\rho}{2}\Gamma(b) + \frac{c^2\rho^2}{2}\mathrm{B}(b) + \ldots\right\} \tag{4}$$

where $b$ is a unit vector of $R^q$ and

$$\Gamma(b) = b^T\left[b^T\right]\left[A_{22}^2\right]b,$$

$$\mathrm{B}(b) = \frac{5}{4}\left(b^T\left[b^T\right]\left[A_{22}^2\right]b\right)^2 - \frac{1}{4}\left\|b^T A_{22}^2 b\right\|^2 + \left\|\left[b^T\right]\left[A_{12}^2\right]b\right\|^2 + b^T\left[b^T\right]\left[A_{21}^2\right]\left(b^T A_{22}^1 b\right) - \frac{1}{3}b^T \otimes \left(b^T\left[b^T\right]\left[A_{222}^2\right]b\right)$$

$$\lambda^*(b) = c\rho \,\|\, R_{22}^{-1}b \,\|$$

The expression (4) provides a method for construction of the likelihood region (2): by substituting different values for $b$ we obtain the bounds of this region.

If there are no nuisance parameters, we obtain the likelihood region for the whole parameter vector $\theta$.

**Theorem 2**.Let $d$ be a unit vector in $R^p$. Then

$$\lambda(d) = \lambda^*(d)\left\{1 - (c\rho/2)\Gamma(d) + \left(c^2\rho^2/2\right)\beta(d) + \ldots\right\} \tag{5}$$

where the neglected terms are of degree 4 and higher in $c\rho$ and

$$\lambda^*(d) = c\rho\left\|R^{-1}d\right\|$$

$$\Gamma(d) = d^T\left[d^T\right]\left[A_{\cdot\cdot}^{\cdot}\right]d$$

$$\beta(d) = \frac{5}{4}\left(d^T\left[d^T\right]\left[A_{\cdot\cdot}^{\cdot}\right]d\right)^2 - \frac{1}{4}\left\|d^T A_{\cdot\cdot}^{\cdot} d\right\|^2 - \frac{1}{3}d^T \otimes \left(d^T\left[d^T\right]\left[A_{\cdots}^{\cdot}\right]d\right) \tag{6}$$

For every array $U, U_{\cdot\cdot}^{\cdot}, U_{\cdots}^{\cdot}$ let us denote $U_{\cdot r} = \left(U_{1r,\ldots}U_{nr}\right)^T$, $U_r = \left(U_{r1},\ldots U_{rp}\right)$, $U_{\cdot r}^a =$

$\left(U_{1r}^a,\ldots U_{pr}^a\right)^T$, $U_{rs}^{\cdot} = \left(U_{rs}^1,\ldots U_{rs}^n\right)^T$, and so on. Put $G = V^T V = \left(g_{ij}\right)$, $G^{-1} = \left(V^T V\right)^{-1} = \left(g^{ij}\right)$.

For one-dimensional parameter vector (2) becomes the likelihood interval with the bound points defined from the equation

$$\theta_p - \hat{\theta}_p = \left(g^{pp}\rho^2\right)^{1/2}\bar{c}\left(1 - \left(\bar{c}\rho/2\right)\Gamma + \left(\bar{c}\rho\right)^2 \mathrm{B}/2\right) \tag{7}$$

where

$$\bar{c} = \pm c$$

$$\Gamma = A_{pp}^p$$

$$\mathrm{B} = -\left(A_{pp}^p\right)^2 + \| A_{\cdot p}^p \|^2 + A_{p\cdot}^p A_{pp}^{\cdot} - 1/3 A_{ppp}^p \tag{8}$$

It follows from (7) that there will be one value of $\theta_p$ for +c and another for –c (c is always taken positive). Two values of $\theta_p$ so determined will not be symmetric with respect to $\hat{\theta}_p$.

Clarke [8] proved the similar expression

$$\theta_p - \hat{\theta}_p = \left(g^{pp}\rho^2\right)^{1/2}\bar{c}\left(1 - \left(c\rho/2\right)\Gamma + \left(c^2\rho^2/2\right)\Psi + \ldots\right) \tag{9}$$

where

$$\Psi = -\Gamma^2 + g^{pp}\left(\left(\Gamma^s\Gamma^s\right) + \left(\Gamma^s\Gamma_s\right) - \frac{1}{3}k\right) \tag{10}$$

with

$$\Gamma^s\Gamma^s = r_{pp}^6\left(G^{-1}\left[G^{-1}V^T\right]\left[V_{\cdot\cdot}^{\cdot}\right]G^{-1}\right)_{p\cdot}^p\left(\left[G^{-1}V^T\right]\left[V_{\cdot\cdot}^{\cdot}\right]G^{-1}\right)_{\cdot p}^p$$

$$\Gamma^s\Gamma_s = r_{pp}^6\left(G^{-1}\left[G^{-1}V^T\right]\left[V_{\cdot\cdot}^{\cdot}\right]\right)_{p\cdot}^p\left(G^{-1}\left[G^{-1}V^T\right]\left[V_{\cdot\cdot}^{\cdot}\right]G^{-1}\right)_{pp}^{\cdot}$$

$$k = r_{pp}^6\left(G^{-1} \otimes \left(G^{-1}\left[G^{-1}\right]V^T\left[V_{\cdots}^{\cdot}\right]G^{-1}\right)\right)_{ppp}^p \tag{11}$$

where $r_{pp}$ means the (p,p) -element of the matrix $R$.

It is shown in [10] that $\mathrm{B} = \Psi$ i.e. (7) and (9) are the same.

## 3.THE PROBLEM OF PERTURBATIONS

In the second part the problem of additive perturbations of the observation vector **Y** is discussed. Suppose modifying **Y** by addition of a vector **u**. Let L(**θ** ) denote the log-likelihood corresponding to the postulated model and L( **θ** / **u** ) the log-likelihood corresponding to the perturbed model, i.e.

$$\mathrm{L}(\boldsymbol{\theta}/\mathbf{u}) = -n/2\ln(2\pi\sigma^2) - 1/2\sigma^2 \left|\left|\mathbf{Y}+\mathbf{u} - \mathbf{f}(\boldsymbol{\theta})\right|\right|^2$$

St. Laurent and Cook [11, 12] have shown that the so-called generalized leverage vector and Jacobian leverage vector can be used to assess the influence of perturbations on

fitted values in nonlinear regression. Our aim is to give a tool for assessment of the influence of additive perturbations on the values of parameters themselves. Then the relationship among the new measure, leverages introduced by Laurent and Cook and the intrinsic curvature is explored.

Let $\boldsymbol{\theta}_{\mathbf{u}}^{*}$ denote the least squares estimator of $\boldsymbol{\theta}$ under $L(\boldsymbol{\theta}/\mathbf{u})$ .We have $\boldsymbol{\theta}^{*} = \boldsymbol{\theta}_{\mathbf{0}}^{*}$ and $L(\boldsymbol{\theta}) = L(\boldsymbol{\theta}/\mathbf{0})$. Let **V** and **W** be the first and second derivatives of $f(\boldsymbol{\theta})$ with elements

$$V_j^a=(\partial f_a/\partial \theta_j)_{\theta*} \quad \text{and} \quad W_{jk}^a = (\partial^2 f_a / \partial \theta_j \partial \theta k)_{\theta*}$$

respectively. For each a=1,...n denote by $\mathbf{T}_a$ the px px p - array with elements

$$(\partial^3 f_a/\partial \theta_j \partial\theta_k \partial \theta_l)_{\theta*} .$$ Put $\mathbf{e}^{*}= \mathbf{Y}-\mathbf{f}(\boldsymbol{\theta}^{*})$.

Consider a perturbation of **Y** in the form $\mathbf{u} = c\,\mathbf{b}$, where $c \in R$ and $||\mathbf{b}||=1$.

We define the parameter leverage vector due to a perturbation of the data by c in the direction **b** as

$$P(c, \mathbf{b}) = (\boldsymbol{\theta}_{c\mathbf{b}}^{*} - \boldsymbol{\theta}^{*}) / c$$

and the Jacobian parameter leverage vector as

$$P(\mathbf{b}) = \lim_{c\to 0} P(c, \mathbf{b})$$

Finally we define the Jacobian parameter leverage matrix as

$$\mathbf{P} = (\mathbf{V}'\mathbf{V} - [\mathbf{e}^{*'}][\mathbf{W}])^{-1}\mathbf{V}'$$

where the square bracket multiplication is as above.

**Theorem 3.** *The parameter leverage vector due to a perturbation of the data by c in the direction* ***b*** *is, up to terms of order c* [13],

$$P(c,\mathbf{b})= \mathbf{A}^{-1}\mathbf{V}'\mathbf{b} + 1/2\,\mathbf{b}'(\mathbf{P}_1+\mathbf{P}_2+\mathbf{P}_3+\mathbf{P}_4)\mathbf{b}c \qquad (12)$$

where

$$\mathbf{A}= \mathbf{V}'\mathbf{V}- [\mathbf{e}^{*'}][\mathbf{W}]$$
$$\mathbf{P_1} = -[\mathbf{A}^{-1}][\mathbf{V}\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{V}']$$
$$(\mathbf{P_2})_{tu}^{s} = ([\mathbf{I}- \mathbf{V}\mathbf{A}^{-1}\mathbf{V}'][\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{V}'])_{su}^{t}$$
$$(\mathbf{P}_3)_{tu}^{s} = ([\mathbf{I}- \mathbf{V}\mathbf{A}^{-1}\mathbf{V}'][\mathbf{A}^{-1}\mathbf{W}\mathbf{A}^{-1}\mathbf{V}']_{st})^{u}$$
$$\mathbf{P}_4 = [\mathbf{A}^{-1}][\mathbf{V}\mathbf{A}^{-1}(\Sigma e_i \mathbf{T}_i)\mathbf{A}^{-1}\mathbf{V}']$$

**Corollary 1.** *The Jacobian parameter leverage vector due to a perturbation of the data by c in the direction* ***b*** *is* $P(\mathbf{b}) = \mathbf{P}\mathbf{b}$.

In order to get deeper insight into the matrix **P** we express it in a standard form. Let the QR decomposition of **V** be given by $\mathbf{V}= \mathbf{U}\mathbf{R}$. Consider now the reparameterization

$$\boldsymbol{\tau} = \mathbf{U}'(\mathbf{f}(\boldsymbol{\theta}) - \mathbf{f}(\boldsymbol{\theta}^{*}))$$

where $\boldsymbol{\tau}$ is called the normal parameter. It can be shown that the Jacobian parameter leverage matrix with respect to the normal parameter is

$$\mathbf{P} = (\mathbf{I} - \mathbf{B})^{-1}\mathbf{U}' \qquad (13)$$

where $\mathbf{B} = [\mathbf{e}^{*'}][\mathbf{L}'\mathbf{W}\mathbf{L}]$ , $\mathbf{L} = \mathbf{R}^{-1}$

The relationship between the Jacobian parameter leverage matrix and the intrinsic curvature of Bates and Watts [4] follows from the fact that

$$\mathbf{B} = [\mathbf{e}^{*'}\mathbf{N}][\mathbf{C}]$$

where $\mathbf{N}'\mathbf{e}^{*}$ is the rotated residual vector and **C** is the intrinsic curvature array.

The norm $||\mathbf{P}\mathbf{b}||$ measures the influence of a perturbation of data in the direction **b** on values of parameters. In the case of the derived linear model $\mathbf{f}(\boldsymbol{\theta}^{*}) + \mathbf{V}(\boldsymbol{\theta} - \boldsymbol{\theta}^{*})$ ,the Jacobian parameter leverage matrix for the normal parameter is $\mathbf{H} = \mathbf{U}'$**.** The matrix **P** takes into account the normal curvatures of the expectation surface at $\mathbf{f}(\boldsymbol{\theta}^{*})$, while **H** does not.

For the normal parameter

$$(\mathbf{P})_{max} = \max_{||\mathbf{b}||=1} ||(\mathbf{I} - \mathbf{B})^{-1}\mathbf{U}'\mathbf{b}|| = (1-\lambda_p)^{-1} \qquad (14)$$

where $\lambda_p$ is the largest eigenvalue of **B**. Denoting by $\mathbf{b_i}$ the i-th standard basis vector in $R^n$,

we have

$$P_i = ||\mathbf{P}\mathbf{b_i}|| = ||(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{U}')_i|| \qquad (15)$$

where $(\mathbf{U}')_i$ means the i-th column of $\mathbf{U}'$ . Evidently

$$P_i^2 = \mathbf{b}_i'\mathbf{U}(\mathbf{I} - \mathbf{B})^{-2}\mathbf{U}'\mathbf{b_i} \qquad (16)$$

We recall that the Jacobian leverage matrix introduced by St.Laurent and Cook [11, 12] is $\mathbf{J} = \mathbf{U}(\mathbf{I} - \mathbf{B})^{-1}\mathbf{U}'$. Its i-th diagonal element $j_{ii}$ measures the rate of change in the fitted value $y_i^{*}$ with respect to the rate of change in $y_i$. It follows from (16) that if $(\mathbf{I} - \mathbf{B})^{-1}$ differs from $(\mathbf{I} - \mathbf{B})^{-2}$ , the case maximazing $j_{ii}$ may be different from that with the largest Jacobian parameter leverage. If $(\mathbf{I}- \mathbf{B})^{-1} = (\mathbf{I}- \mathbf{B})^{-2}$, then $\mathbf{B}= \mathbf{O}$. Two special cases in which this occurs are when the model is intrinsically linear (the elements of $\mathbf{L}'\mathbf{W}\mathbf{L}$ are zeros) and when the model provides an exact fit to the data (the residual vector $\mathbf{e}^{*}$ is zero).

## 4. APPLICATIONS

### 4.1. Application of Approximate Confidence Regions to IBNR Reserves

Our results will be applied to the case of computing IBNR reserves by the method described above which is completely different from chain-ladder methods. The results obtained show that asymmetric (with respect to the least squares estimators of parameters) confidence intervals are more accurate than intervals based on an approximation of lower order.

The model used by Stelltjes [3] $f(x, \boldsymbol{\theta})=\alpha\exp(\beta x)+\gamma\exp(\omega x)$ , i.e. 4-parameter bi-exponential model is by computation ill-conditioned. More precisely, the practical im-

plementation in computational software S-plus leads to the error message "singular gradient matrix" (which is more or less expectable if one exponential fits data better) and "step factor reduced below minimum". In the present paper we do not publish the original data of [3] since paper is online and insurance data are considered as privacy data.

If correlations among incremental pure premiums are not negligible, what is also case of the data given by [3], usage of the correct model produces different values. For example, the PROC NLIN based method in [3] gave resulting estimators 3.1994, -0.0754, 29.4446, -0.5480 and 95% confidence intervals for estimators (2.0596, 4.3392), (-0.0942, -0.0566), (18.5334, 40.3557) and (-0.6986,-0.3974) of parameters $\alpha$, $\beta$, $\gamma$, $\omega$, respectively. However, using the correlation, as addressed in [3], we obtain the estimators 2.36488, -0.07767, 21.61184, -0.566532 and 95% confidence intervals for estimators (1.2746, 3.4551), (-0.0959, -0.05937), (10.664, 32.5596) and (-0.72473,-0.4083) of parameters $\alpha$, $\beta$, $\gamma, \omega$, respectively.

One of the main practical and theoretical problems of the IBNR confidence intervals implementation is the efficient estimation of the covariance structure and, consequently, sums of squares $S\left(\hat{\theta}\right)$. This problem evidently goes behind the framework of this contribution and will be addressed in some future work. For the related topic see designs of experiments literature addressing the covariance estimation, for recent results in the parametrized covariance see e.g. [3].

Applying the results of the second part of this paper reveals the important fact, namely, that we can neglect the effect of $\mathbf{B}$ in (7). Here we have the following quantities needed for calculation of confidence intervals by means of (7) :

$g^{11} = 0.1129 \times 10^{-6}$, $g^{22} = 0.31 \times 10^{-10}$, $g^{33} = 0.1033 \times 10^{-4}$, $g^{44} = 0.1968 \times 10^{-8}$,

$A_{11}^1 = 0, A_{22}^2 = 0.2134 \times 10^{-3}, A_{33}^3 = 0.1231 \times 10^{-3}, A_{44}^4 = 0.3603 \times 10^{-3}$,

$c = \sqrt{F(1,36,0.95)} \approx \sqrt{4.13}$.

Using them we obtain

$\alpha$ : upper bound 3.1994 + 0.5807=3.7801 ; lower bound 3.1994 – 0.5807= 2.6187

$\beta$ upper bound –0.0754 +0.013= -0.0624; lower bound -0.0754 –0.027= -0.0781

$\gamma$ upper bound 29.4446 +8.85= 38.2946 lower bound 29.4446-13.73= 15.7146

$\omega$ upper bound –0.5480 +0.057=-0.4903 lower bound –0.5480-0.157= -0.7050.

Comparing them to the original results of Stelltjes shows that our method gives narrower intervals than the classical approach (the only exception being $\gamma$). Recall that the intervals are not symmetric with respect to the least-squres estimates thus reflecting the nonlinearity of the model.

Admitting for a moment that the hypothesis $\beta = 0$ holds, we get the famous Mitcherlitz model f(x, $\theta$)=$\alpha$ +$\gamma$ exp($\omega$ x).A thorough analysis of this model can be found in [7, 8, 14] to illustrate that Bates- Watts method is not always reliable . It follows that for this model our approach outperforms the classical one based on quadratic approximation as it produces confidence regions whose true coverage is closer to the nominal level than for the classical ones.

The outlook of the research in this direction brings a possibility to compute 2-parameter confidence interval, for instance for $(\alpha, \gamma)$ or $(\gamma, \omega)$. Notice that Stelltjes [3] has computed $S\left(\hat{\theta}\right)$ as the sum of weighted least squares and not as reported in (2.3.1), p. 359, it has been a typo. Thus, we are using the same approach as Stehlík [15] and employ a weight function that is inversely proportional to the variance of the data.

Probably, the main objective of the actuarial work is to construct the confidence interval for IBNR reserves for various accident quarters. A classical approach how to construct approximate confidence intervals is given by Bates and Watts [4], p. 58. For instance, 95% confidence interval for total IBNR for data in [3] is (27459851, 32751218) and (25254267, 40083031) when the parameter variance - covariance matrix is used (see [3]). The outlook is to construct a confidence interval based on cubic approximation of the model function (1). However, one should take care about the fact, that traditional nonlinear regression assumes that the error terms are normal which is a symmetric distribution with a range of whole real line. Incremental pure premium data may actually be skewed and can hardly ever be highly negative, therefore, using the normal distribution is approximation at best.

### 4.2. Application of Additive Perturbation Models for Privacy

#### *Preserving Data Mining in Insurance*

In insurance companies, many data sets have to be stored. A large fraction of them uses randomized data distortion techniques to mask the data for preserving the privacy of sensitive data. The additive perturbation attempts to hide the sensitive data by randomly modifying the data values often using additive noise. Random matrices have ‘predictable’ structures in the spectral domain and it develops a random matrix-based ‘Spectral Filtering Technique’ (SPF) to retrieve original data from the dataset distorted by adding random values (see [16]). In the present example we consider the same model as in section 4.1. As mentioned above the PROC NLIN based method gave estimators 3.1994, -0.0754, 29.4446, -0.5480 of model parameters. Then for x=10 we have f=1.628027.

Assuming that the error has a Gaussian N(0,1) form we get an observed value of 3.97282 and additive perturbation by N(0,0.001) will give us -0.0001432867, while the additive

perturbation by N(0,0.01) will give 3.949242. Thus tuning of the standard deviation can tune the necessary level of privacy of the data. The control over this procedure is guaranteed by the theoretical results in the section 3, and may be properly analysed for every insurance model.

## ACKNOWLEDGEMENTS

Authors are thankful to the referee comments which improved the quality of the paper. Work was partially supported by AKTION 54p13 and Vega grant 1/0077/09.

## REFERENCES

[1] F. De Vylder, "Non-linear regression in credibility theory", Insurance: *Mathematics and Economics*, vol. 4, No. 3, pp. 163-172, 1985.

[2] G. Pitselis, "De Vylder's robust nonlinear regression credibility", *Belgian Actuarial Bulletin*, vol. 4, no. 1, pp. 1-4, 2004.

[3] S. Stelljes, "*Nonlinear Regression Model of Incurred but Not Reported Losses*", Casualty Actuarial Society Forum Casualty Actuarial Society – Arlington: Virginia, Fall, 2006, pp. 353-377.

[4] D. M. Bates, and D. G. Watts, *Nonlinear Regression Analysis and its Applications*, Wiley Series in Probability and Statistics: New York, 2007.

[5] D. Hamilton, Confidence regions for parameter subsets in nonlinear regression, *Biometrika*, vol. 73, pp. 57-64, 1986.

[6] L. M. Haines, T. E. O'Brien, and G. P. Y. Clarke, "Kurtosis and curvature measures for nonlinear regression models", *Statistica Sinica*, vol. 14, pp. 547-570, 2004.

[7] R. D. Cook, and M. L. Goldberg, "Curvatures for parameter subsets in nonlinear regression", *Annals Statistics*, vol. 14, pp. 1399-1418, 1986.

[8] G. P. Y. Clarke, "Marginal curvatures and their usefulness in the analysis of nonlinear regression models", *Journal of the American Statistical Association*, vol. 82, pp. 844-850, 1987.

[9] R. Potocký, and T. Van Ban, "Confidence regions in nonlinear regression models", *Application of Mathematics*, vol. 37, pp. 29-39, 1992.

[10] R. Potocký, and M. Stehlík, "Improved non-linear regression modelling of parameters of IBNR reserves", *Communications in Dependability and Quality Managements*, vol. 11, no. 4, pp. 54-60, 2008.

[11] R. J. St. Laurent, and R. D. Cook, "Leverage and superleverage in nonlinear regression", *Journal of the American Statistical Association*, vol. 87, pp. 985-990, 1992.

[12] R. J. St.Laurent, and R. D. Cook, "Leverage, local influence and curvature in nonlinear regression", *Biometrika*, vol. 80, pp. 99-106, 1993.

[13] R. Potocký, and T. Van Ban, "On additive perturbations in nonlinear regression models", *Tatra Mountains Mathematical Publications*, vol. 17, pp. 65-71, 1999.

[14] R. Potocký, "Approximate confidence regions for parameter subsets in a non-linear model", *Tatra Mountains Mathematical Publications*, vol. 26, pp. 227-236, 2003.

[15] M. Stehlík, "Covariance related properties of D-optimal correlated designs", in the *5th St. Petersburg Workshop on Simulation*, 2005, pp. 645-652.

[16] S. Datta, "*On Random Additive Perturbation for Privacy Preserving Data Mining*", M.S. thesis, University of Maryland, MD, USA, 2004.

Received: June 18, 2010 Revised: August 17, 2010 Accepted: August 29, 2010