

# On a Hierarchical Credibility Model for Quantiles

Georgios Pitselis  
University of Piraeus

June 18-22, 2011, Madrid  
**ASTIN**

## Introduction

In actuarial practice there are cases where the claim distribution is not normal and other cases where the data are contaminated due to large claims (catastrophic events).

On the other hand insurance regulations (Solvency) may require that a risk margin should be established on a basis that is intended to secure the insurance liabilities of the insurer at a given level of sufficiency.

For those cases it is more appropriate to use quantiles instead of the mean for the classical credibility model and quantile regression instead of the ordinary least squares estimation.

The aim of this paper is to present the hierarchical credibility estimation of quantiles. More specifically, we incorporate quantiles into the classical hierarchical model of Jewell (1975) following the the procedure of Bauwelinckx & Goovaerts (1990) and quantile regression into the regression hierarchical model of Sundt (1979). Hierarchical credibility estimators are also established for each model and numerical examples are presented.

## Quantile Functions

The quantile of a distribution is defined as

$$\xi_p = F^{-1}(p) = \inf\{x : F(x) \geq p\}. \quad (1)$$

Let  $X_{(1)}, \dots, X_{(n)}$  denote the order statistics of  $X_1, \dots, X_n$  and let  $\hat{\xi}_p$  denote the sample  $p$ -quantile. The empirical quantile function can be defined as

$$\hat{\xi}_p = n \left( \frac{j}{n} - p \right) X_{(j-1)} + n \left( p - \frac{j-1}{n} \right) X_{(j)}, \quad (2)$$

$$\text{for } \frac{j-1}{n} \leq p \leq \frac{j}{n} \text{ and } j = 1, \dots, n.$$

We call  $n(X_{(j)} - X_{(j-1)})$ ,  $j = 1, \dots, n$  the spacing of the sample.

Estimators of quantiles which may behave better in small samples from symmetric densities can be obtained by a shifted piecewise linear function

$$\hat{\xi}_p = n \left( \frac{2j+1}{n} - p \right) X_{(j)} + n \left( p - \frac{2j-1}{2n} \right) X_{(j+1)} \quad (3)$$

$$\text{for } \frac{2j-1}{n} \leq p \leq \frac{2j+1}{2n} \text{ and } j = 1, \dots, n.$$

The  $\hat{\xi}_p$  is undefined for  $p < \frac{1}{2n}$  or  $p > 1 - \frac{1}{2n}$ .

### Theorem

*Cramer (1946). Let the probability density function  $f_X(\cdot)$  be continuous with continuous derivative in some neighborhood of  $\xi_p$  and let  $f'(\xi_p) \neq 0$ . Then the sample quantile  $\hat{\xi}_p$  is asymptotically distributed according to the normal distribution, with mean  $\xi_p$  and variances*

$$\sigma_{\xi_p}^2 = \frac{p(1-p)}{nf'(\xi_p)^2}.$$

**Remark:**

When  $F$  is of a known location scale type  $F\left(\frac{x-\theta}{\sigma}\right)$  then

$$F_X(\xi_p) = F\left(\frac{\xi_p - \theta}{\sigma}\right) = p$$

and therefore,  $z_p = \left(\frac{\xi_p - \theta}{\sigma}\right)$ , with  $F(z_p) = p$  and  $\xi_p = \sigma z_p + \theta$  can be estimated by the maximum likelihood estimation.

## Two-level Quantile Hierarchical Credibility Model

The model consists of the structural variables  $\Theta_{pj}$  and  $\Theta_p$  and the observable variables  $X_{ijh}$ . Let  $\hat{\xi}_{pjh}$  is the sample  $p - th$  quantile for the  $j(j = 1, \dots, K_h)$  contract and  $h(h = 1, \dots, H)$  subporfolio.

- (i)  $\Xi_{pj}(\Theta_{pjh}, \Theta_{ph}) = E(\hat{\xi}_{pjh} | \Theta_{pjh}, \Theta_{ph})$
- (ii)  $\sigma_{\xi_{pj}}^2(\Theta_{pjh}, \Theta_{ph}) = \text{Var}(\hat{\xi}_{pjh} | \Theta_{pjh}, \Theta_{ph})$
- (iii)  $\Upsilon_p(\Theta_{ph}) = E(\hat{\xi}_{pjh} | \Theta_{ph}) = E[\Xi_{pj}(\Theta_{pjh}, \Theta_{ph}) | \Theta_{ph}]$

The structural parameters are

$$\begin{aligned} \Xi_p &= E[\Upsilon_p(\Theta_{ph})] = E[\Xi_{pj}(\Theta_{pjh}, \Theta_{ph})] = E[\hat{\xi}_{pjh}] \\ s_{\xi_p}^2 &= E[\sigma_{\xi_{pj}}^2(\Theta_{pjh}, \Theta_{ph})] \\ \Phi_{\xi_p} &= E\{\text{Var}[\Xi_{pj}(\Theta_{pjh}, \Theta_{ph}) | \Theta_{ph}]\} \\ U_{\xi_p} &= \text{Var}[\Upsilon_p(\Theta_{ph})]. \end{aligned} \tag{4}$$

The credibility factor for contract level and sector level respectively are:

$$Z_{pjh} = \frac{\Phi_{\xi_p}}{\Phi_{\xi_p} + s_{\xi_p}^2}, \quad (5)$$

$$Z_{ph} = \frac{U_{\xi_p}}{U_{\xi_p} + s_{\xi_p}^2/K_h + \Phi_{\xi_p}/K_h}. \quad (6)$$

### Lemma

Based on the above assumptions we can obtain the expressions for the conditional expectations and covariances:

$$\text{Cov}[\Xi_{pj}(\Theta_{pjh}, \Theta_{ph}), \hat{\xi}_{pj'h'}] = \delta_{hh'}(\delta_{jj'}\Phi_{\xi_p} + U_{\xi_p}) \quad (7)$$

$$\text{Cov}[\Upsilon_p(\Theta_{ph}), \hat{\xi}_{pj'h'}] = \delta_{hh'}U_{\xi_p} \quad (8)$$

$$\text{Cov}(\hat{\xi}_{pjh}, \hat{\xi}_{pj'h'}) = \delta_{hh'}[\delta_{jj'}(\Phi_{\xi_p} + s_{\xi_p}^2) + U_{\xi_p}] \quad (9)$$

$$\text{Cov}(\hat{\xi}_{pjh}, \bar{\xi}_{p.h}) = \text{Cov}(\bar{\xi}_{p.h}, \bar{\xi}_{p.h}) = \frac{1}{K_h}(\Phi_{\xi_p} + s_{\xi_p}^2) + U_{\xi_p} \quad (10)$$

### Theorem

**Credibility estimate on the contract level.** Under the above assumptions the following linearized non-homogeneous quantile estimator for the pure net risk premium for contract  $(j, h)$

$$\widehat{\Xi}_{pj}^{Cred}(\Theta_{pjh}, \Theta_{ph}) = Z_{pjh} \hat{\xi}_{pjh} + (1 - Z_{pjh}) \Xi_p. \quad (11)$$

**Proof:** In order to prove this we have to minimize the following square error

$$E \left[ E \{ [\Xi_{pj}(\Theta_{pjh}, \Theta_{ph}) - c_0 - \sum_{h=1}^H \sum_{j=1}^{k_h} c_{pjh} \hat{\xi}_{pjh}]^2 | (\Theta_{ph}) \} \right]. \quad (12)$$

Differentiating (12) with respect to  $c_0$  we obtain

$$c_0 = E \left[ E[\Xi_{pj}(\Theta_{pjh}, \Theta_{ph}) | \Theta_{ph}] \right] - \sum_{h=1}^H \sum_{j=1}^{K_h} c_{pjh} E[E(\hat{\xi}_{pjh} | \Theta_{ph})]. \quad (13)$$

Inserting the value of  $c_0$  in (12) and differentiating with respect to  $c_{pj'h'}$ , and based on previous Lemma we have the proof.



### Lemma

Consider the two-level quantile hierarchical credibility model. Under the above assumptions and (4) the credibility model may be defined as in (11) and the credibility factor is given by

$$Z_{pjh} = \frac{\Phi_{\xi_p}}{\Phi_{\xi_p} + s_{\xi_p}^2}. \quad (14)$$

**Proof:** The proof is obtained by minimizing the following square error with respect to  $Z_{pjh}$ .

$$\begin{aligned} Q &= E \left[ E \{ [\Xi_{pj}(\Theta_{pjh}, \Theta_{ph}) - \widehat{\Xi}_{pj}^{Cred}(\Theta_{pjh}, \Theta_{ph})]^2 | \Theta_{ph} \} \right] \\ &= E \left[ E \{ [\Xi_{pj}(\Theta_{pjh}, \Theta_{ph}) - \Xi_p - Z_{pjh}(\hat{\xi}_{pjh} - \Xi_p)]^2 | \Theta_{ph} \} \right]. \end{aligned} \quad (15)$$

### Theorem

**Credibility estimate on the sector level.** *Under the above assumptions the following linearized non-homogeneous quantile estimator for the pure net risk premium for sector  $h$  is given*

$$\hat{\Upsilon}_p^{Cred}(\Theta_{ph}) = Z_{ph}\bar{\xi}_{p,h} + (1 - Z_{ph})\Xi_p. \quad (16)$$

**Proof:** The proof is obtained by minimizing the following square error first with respect to  $c_0$  and second with respect the  $c_{pj'h'}$ .

$$E[\Upsilon_p(\Theta_{ph}) - c_0 - \sum_{h=1}^H \sum_{j=1}^{k_h} c_{pj'h} \hat{\xi}_{pj'h}]^2. \quad (17)$$

### Lemma

Consider the two-level quantile hierarchical credibility model. Under the above assumption and (4) the credibility model for the sector level may be defined as in (16) and the credibility factor is given by

$$Z_{ph} = \frac{U_{\xi_p}}{U_{\xi_p} + s_{\xi_p}^2 / K_h + \Phi_{\xi_p} / K_h}. \quad (18)$$

**Proof:** For the proof we have We have to minimize the following square error with respect to  $Z_{ph}$

$$\begin{aligned} Q &= E[\Upsilon_p(\Theta_{ph}) - \widehat{\Upsilon}_p^{Cred}(\Theta_{ph})]^2 \\ &= E[\Upsilon_p(\Theta_{ph}) - \Xi_p - Z_{ph}(\bar{\xi}_{p.h} - \Xi_p)]^2. \end{aligned} \quad (19)$$

## Two-level Hierarchical Credibility Quantile Estimation

We shall present hierarchical credibility estimators based on quantiles:

$$\begin{aligned}
 \bar{\xi}_{p.h} &= \widehat{\Xi}_{pj}(\Theta_{pjh}, \Theta_{ph}) = \frac{1}{K_h} \sum_{j=1}^{K_h} \hat{\xi}_{pjh} \\
 \bar{\xi}_{p..} &= \widehat{\Xi}_p = \frac{1}{\sum_{h=1}^H K_h} \sum_{h=1}^H \sum_{j=1}^{K_h} \hat{\xi}_{pjh} \\
 \widehat{s}_{\xi_{ph}}^2 &= \frac{1}{(\sum_{h=1}^H K_h)} \sum_{h=1}^H \sum_{j=1}^{K_h} \frac{p(1-p)}{n_j [\widehat{f}(\xi_{pjh} | \Theta_{pjh}, \Theta_{ph})]^2} \\
 \widehat{\Phi}_{\xi_p} &= \frac{1}{\sum_{h=1}^H (K_h - 1)} \sum_{h=1}^H \sum_{j=1}^{K_h} (\hat{\xi}_{pjh} - \bar{\xi}_{p.h})^2 - \widehat{s}_{\xi_{ph}}^2 \\
 \widehat{U}_{\xi_p} &= \frac{1}{(H-1)} \sum_{h=1}^H (\bar{\xi}_{p.h} - \bar{\xi}_{p..})^2 - \frac{1}{K_h} \widehat{s}_{\xi_{ph}}^2 - \frac{1}{K_h} \widehat{\Phi}_{\xi_p} \quad (20)
 \end{aligned}$$

The computation of the sample quantiles  $\hat{\xi}_{pjh}$  for contract  $j(j = 1, \dots, K_h)$  and subportfolio  $h(1, \dots, H)$  are based on order statistics, and can be estimated based on (4).

The variance  $\sigma_{\xi_{pj}}^2(\Theta_{pjh}, \Theta_{ph})$  in the assumption (iii) (we will drop the indexes  $j$  and  $h$  for the facility of presentation) can be estimated by an interval estimation from the  $[np]^{th}$  quantile of cumulative distribution function, say  $F(\cdot)$ , denoted by  $\xi_p$  and defined by  $F(\xi_p) = p$  [see Mood et al., (1974), p. 512].

Constructing a symmetric confidence interval of level  $1 - \alpha$  for  $\xi_{pjh}$  we obtain an estimate for  $\sigma_{\xi_p}^2(\Theta_{pjh}, \Theta_{ph})$ :

$$\widehat{\sigma_{\xi_p}^2}(\Theta_{pjh}, \Theta_{ph}) = \frac{n(y_{([np+l])} - y_{([np-l])})^2}{4Z_{1-\alpha/2}^2},$$

where  $l = Z_{1-\alpha/2} \sqrt{np(1-p)}$  and  $Z \sim N(0, 1)$ .

## Remark:

In the case of the weighted credibility analogous to Bühlmann & Straub's (1970) model, the quantiles of a grouped data can be calculated as follows [see Klugman et al. (2008)]

$$\hat{\xi}_p = X_0 + \frac{c}{f_i} \cdot \left( \frac{p \cdot n}{100} - F_{i-1} \right), \quad (21)$$

where

$X_0$  : the lower bound of the cell that contains the 100. $p$  percentile

$f_i$  : is the frequency of the cell  $i$

$c$  : the of the cell

$F_{i-1}$  : is the cumulative frequency of the cell previous of the cell  $i$ .

## Quantile Regression

We assume a sample  $(y_i, \mathbf{x}_i)$ ,  $i = 1, \dots, n$ ,  $y_i$  is the dependent variable and  $\mathbf{x}_i$  denotes a row of a  $n \times k$  the design matrix of explanatory variables and the general model linear model has the form

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i \quad (22)$$

and the quantile regression can be defined as

$$y_{pi} = \mathbf{x}_i' \boldsymbol{\beta}_p + u_{pi}, \quad (23)$$

where  $\boldsymbol{\beta}_p$  is a vector to be estimated and  $u_{pi}$  is the error term. Then the  $p$ -th conditional quantile of  $y_i$  given  $\mathbf{x}_i$  can be written as

$$Q_p(y_i/\mathbf{x}_i) = \mathbf{x}_i' \boldsymbol{\beta}_p \quad (24)$$

where  $\boldsymbol{\beta}_p = \boldsymbol{\beta} + F^{-1}(p)\mathbf{e}_1$  and  $\mathbf{e}_1' = (1, 0, \dots, 0) \in R^k$  (see Bassett & Koenker, 1982).

With quantile regression we can show how various financial characteristics are different at different quantiles.

Thus, the quantile regression method involves allowing the marginal effects to change for claims at different points in the conditional distribution by estimating  $\beta_p$  using several different values of  $p$ ,  $p \in (0, 1)$ .

It is in this way that quantile regression allows for parameter heterogeneity across different types of claims.

In general the  $p$ -th sample quantile ( $0 < p < 1$ ) of  $y$  solves

$$\min_{(\beta)} \left( \sum_{i:y_i \geq b} p|y_i - b| + \sum_{i:y_i < b} (1-p)|y_i - b| \right). \quad (25)$$



In a similar way can be defined the  $p$ -th quantile for the linear model. Let  $\{x_i, i = 1, \dots, n\}$  denote a sequence of (row)  $k$ -vectors of a known design matrix and suppose  $\{y_i, i = 1, \dots, n\}$  is a random sample on the regression process  $u_i = y_i - x_i\beta$  having distribution function  $F$ .

Then the  $p$ -th regression quantile  $0 < p < 1$ , can be defined as any solution to the minimization problem [see Koenker & Bassett (1978), Buchinsky (1998)]

$$\min_{(\beta)} \left( \sum_{i:y_i \geq b} p|y_i - \mathbf{x}'_i\beta| + \sum_{i:y_i < b} (1-p)|y_i - \mathbf{x}'_i\beta| \right) = \min_{\beta} \frac{1}{n} \sum_{i=1}^n \rho_p(u_{p_i}), \quad (26)$$

where  $\rho_p(t) = (p - I(t < 0))t$  is a check function, and  $I(\cdot)$  is the indicator function.

Under certain regulatory conditions can be shown that

$$\sqrt{n}(\widehat{\beta}_p - \beta_p) \rightarrow^L N(0, \mathbf{\Lambda}_p), \quad (27)$$

where  $\mathbf{\Lambda}_p$  is a variance covariance matrix defined as

$$\mathbf{\Lambda}_p = p(1 - p)(E[f_{\xi_p}(0|\mathbf{x}_i)\mathbf{x}_i\mathbf{x}'_i])^{-1}E(\mathbf{x}_i\mathbf{x}'_i)(E(f_{\xi_p}(0|\mathbf{x}_i)\mathbf{x}_i\mathbf{x}'_i))^{-1}. \quad (28)$$

If  $f_{\xi_p}(0|\mathbf{x}_i) = f_{\xi_p}(0)$  with probability 1, i.e., the density of the error term  $u_p$  evaluated at 0 is independent of  $x$ , then  $\mathbf{\Lambda}_p$  in (28) simplifies to

$$\mathbf{\Lambda}_p = \frac{p(1 - p)}{f_{\xi_p}(0)}E(\mathbf{x}_i\mathbf{x}'_i)^{-1} = \sigma_{\xi_p}^2 E(\mathbf{x}_i\mathbf{x}'_i)^{-1}. \quad (29)$$

The last term in (29) can be estimated by  $\hat{E}(\mathbf{x}_i\mathbf{x}'_i) = 1/n \sum_{i=1}^n \mathbf{x}_i\mathbf{x}'_i$ .

## Hierarchical Regression Quantile Credibility Model

In a hierarchical model there is more than one risk factor dividing the portfolio in different subportfolios (sectors, divisions etc.). The model consists of the structural variables  $\Theta_{pjh}$  and  $\Theta_{ph}$ , the observable dependent variables  $Y_{ijh}$ ,  $i = 1, \dots, n_{jh}$  and  $\mathbf{X}_{jh}$  the vector of independent variables in a regression quantile setting.

$\Theta_{ph}$  reflects the distribution of the structure variable, for the  $p - th$  quantile, over the portfolio and  $(\Theta_{pjh}, \Theta_{ph})$ , for the  $p - th$  quantile, over the sector. The sector  $h$  consists of the set of variables  $(\Theta_{pjh}, \Theta_{ph}, Y_{ijh})$ ,  $j = 1, \dots, K_h$ ,  $i = 1, \dots, n_{jh}$  and the contract consists of the variables  $(\Theta_{pjh}, Y_{ijh})$ . It is assumed that

- (i) The subportfolio's  $h = 1, \dots, H$  are independent, i.e. the set  $(\Theta_{pjh}, \Theta_{ph}, \mathbf{Y}_{jh})$  is independent of  $(\Theta_{ph'}, \Theta_{ph'j}, \mathbf{Y}_{jh'})$
- (ii) For each  $h = 1, \dots, H$  and fixed  $\Theta_H = \theta_h$  the contracts, i.e. the pairs  $\Theta_{pjh}, \mathbf{Y}_{jh}$ , are conditionally independent
- (iii) All pairs of variables  $\Theta_{pjh}, \Theta_{ph}$  for  $h = 1, \dots, H$  and  $j = 1, \dots, K_h$  are conditionally equi-distributed
- (iv)  $Q_p(Y_{ijh} | \Theta_{pjh}, \Theta_{ph}) = \mathbf{x}'_{ijh} \beta_p(\Theta_{pjh}, \Theta_{ph})$ , where  $Q_p(Y_{ijh} | \Theta_{pjh}, \Theta_{ph})$  denotes the conditional quantile of  $Y_{ijh}$  conditional on risk parameters  $\Theta_{pjh}$  and  $\Theta_{ph}$ . The vector  $\beta_p(\Theta_{pjh}, \Theta_{ph})$  is an unknown vector of regression parameters and  $\mathbf{x}'_{ijh}$  is the  $i$ th of the design matrix  $\mathbf{X}_{jh}$
- (v) 
$$\text{Cov}(\hat{\beta}_{pjh} | \Theta_{pjh}, \Theta_{ph}) = \frac{p(1-p)}{[f_{\xi_p}(\Theta_{pjh}, \Theta_{ph})]^2} (\mathbf{X}'_{jh} \mathbf{X}_{jh})^{-1} = \sigma_{\xi_p}^2(\Theta_{pjh}, \Theta_{ph}) (\mathbf{X}'_{jh} \mathbf{X}_{jh})^{-1},$$
 where  $\hat{\beta}_{pjh}$  is an estimator of the individual  $p$ -th quantile regression coefficient,  $\beta_p(\Theta_{pjh}, \Theta_{ph})$
- (vi)  $E[\beta_p(\Theta_{pjh}, \Theta_{ph}) | \Theta_{ph}] = \beta_p(\Theta_{ph})$

**Remark:** In what it follows as in Hachemeister's data, we consider the design matrix  $(\mathbf{X}_{jh})$  fixed and the same for each contract.

## Structural parameters

The following are structural parameters that will occur in the credibility premium

- (i)  $\beta_p = E[\beta_p(\Theta_{pjh}, \Theta_{ph})]$
- (ii)  $s_{\xi_p}^2 = E[\sigma_{\xi_p}^2(\Theta_{pjh}, \Theta_{ph})]$
- (iii)  $\Lambda_p = E\{\text{Cov}[\beta(\Theta_{pjh}, \Theta_{ph})|\Theta_{ph}]\}$
- (iv)  $\mathbf{U}_p = \text{Cov}[\beta_p(\Theta_{ph})]$ ,

and the credibility factor for contract level and sector level respectively are:

$$\mathbf{Z}_{pjh} = \Lambda_p[\Lambda_p + s_{\xi_p}^2(\mathbf{X}'_{jh}\mathbf{X}_{jh})^{-1}]^{-1}, \quad (30)$$

$$\mathbf{Z}_{ph} = \mathbf{U}_p[\mathbf{U}_p + \Lambda_p + s_{\xi_p}^2(\mathbf{X}'_{jh}\mathbf{X}_{jh})^{-1}]^{-1} \quad (31)$$

## Regression Quantile Credibility for the contract level

We consider the linear non-homogeneous estimator for the pure net risk premium of the contract  $(p, j)$

$$\mathbf{Y}_{jh}^{Cred} = \mathbf{X}_{jh} \cdot \widehat{\boldsymbol{\beta}}_p^{Cred}(\Theta_{pjh}, \Theta_{ph}), \quad (32)$$

with

$$\widehat{\boldsymbol{\beta}}_p^{Cred}(\Theta_{pjh}, \Theta_{ph}) = [\mathbf{Z}_{pjh} \cdot \widehat{\boldsymbol{\beta}}_{pjh} + (\mathbf{I} - \mathbf{Z}_{pjh})\boldsymbol{\beta}_p(\Theta_{ph})], \quad (33)$$

where  $\widehat{\boldsymbol{\beta}}_{pjh}$  is the individual estimators of  $\boldsymbol{\beta}(\Theta_{pjh}, \Theta_{ph})$ .

### Theorem

Consider the two-level hierarchical Regression model as introduced in the previous subsection. Under the hypothesis (i) - (vi), the optimal non-homogeneous credibility factor for the contract level  $(h, j)$  is given by

$$\mathbf{Z}_{pjh} = \mathbf{\Lambda}_p [\mathbf{\Lambda}_p + s_{\xi_p}^2 (\mathbf{X}'_{jh} \mathbf{X}_{jh})^{-1}]^{-1} \quad (34)$$

where  $\mathbf{\Lambda}$  and  $s_{\xi_p}^2$  are as defined in Section 6.2.

**Proof:** In order the problem to be solved we have to minimize the following expression,

$$Q = E \left\{ E \left[ [\beta_p(\Theta_{pjh}, \Theta_{ph}) - \beta_p(\Theta_{ph}) - \mathbf{Z}_{pjh}(\hat{\beta}_{pjh} - \beta_p(\Theta_{ph}))]' \times \right. \right. \\ \left. \left. \times [\beta_p(\Theta_{pjh}, \Theta_{ph}) - \beta_p(\Theta_{ph}) - \mathbf{Z}_{pjh}(\hat{\beta}_{pjh} - \beta_p(\Theta_{ph}))] | \Theta_{ph} \right] \right\}.$$

Differentiating with respect to  $\mathbf{Z}_{pjh}$  we obtain (34)

## Regression Quantile Credibility on sector level

We consider the linear non-homogeneous estimator for the pure net risk premium of the sector  $p$

$$\mathbf{Y}_h^{Cred} = \mathbf{X}_h \widehat{\boldsymbol{\beta}}_p^{Cred}(\Theta_{ph}) \quad (35)$$

with  $\mathbf{Y}_h = (\mathbf{Y}'_{1h}, \dots, \mathbf{Y}'_{Kh})'$   $\mathbf{X}_h = (\mathbf{X}'_{1h}, \dots, \mathbf{X}'_{Kh})'$  and

$$\widehat{\boldsymbol{\beta}}_p^{Cred}(\Theta_{ph}) = \mathbf{Z}_{ph} \widehat{\boldsymbol{\beta}}_{p,h} + (\mathbf{I} - \mathbf{Z}_{ph}) \boldsymbol{\beta}_p \quad (36)$$

where  $\widehat{\boldsymbol{\beta}}_{p,h}$  and  $\boldsymbol{\beta}_p$  are the individual and collective estimators, respectively, of  $\boldsymbol{\beta}(\Theta_p)$ , with  $\boldsymbol{\beta}_p = E(\widehat{\boldsymbol{\beta}}_{ph})$ .



### Theorem

Under the hypothesis (i) - (iv), the optimal non-homogeneous credibility factor for the factor  $p$  is given by

$$\mathbf{Z}_{ph} = \mathbf{U}_p [\mathbf{U}_p + \mathbf{\Lambda}_p + s_{\xi_p}^2 (\mathbf{X}'_{jh} \mathbf{X}'_{jh})^{-1}]^{-1} \quad (37)$$

where  $\Xi$ ,  $\mathbf{\Lambda}_p$  and  $s_{\xi_p}^2$  are as defined in Subsection 6.2.

**Proof:** The solution of the problem can be obtained by the minimization of the following

$$\begin{aligned} Q &= E \left\{ \left[ \beta_p(\Theta_{ph}) - \widehat{\beta}_p^{Cred}(\Theta_{ph}) \right]' \left[ \beta_p(\Theta_{ph}) - \widehat{\beta}_p^{Cred}(\Theta_{ph}) \right] \right\} \\ &= E \left\{ \left[ \beta_p(\Theta_{ph}) - \mathbf{Z}_{ph} \widehat{\beta}_{p.h} + (\mathbf{I} - \mathbf{Z}_{ph}) \beta_p \right]' \left[ \beta_p(\Theta_{ph}) - \mathbf{Z}_{ph} \widehat{\beta}_{p.h} + (\mathbf{I} - \mathbf{Z}_{ph}) \beta_p \right] \right\} \end{aligned}$$

Differentiating with respect to  $\mathbf{Z}_{ph}$  we obtain (37).

## Hierarchical Quantile Regression Parameter Estimation

For contract  $j$  and subportfolio  $h$  the estimator  $\hat{\beta}_{pjh}$  in (33) can be defined as a solution to the minimization problem

$$\min_{(\beta_{jh})} \left( \sum_{i:y_i \geq b} p |y_{ijh} - \mathbf{x}'_{ijh} \beta_{pjh}| + \sum_{i:y_i < b} (1-p) |y_{ijh} - \mathbf{x}'_{ijh} \beta_{pjh}| \right). \quad (38)$$

An estimator of  $s_{\xi_p}^2 (\mathbf{X}'_{jh} \mathbf{X}_{jh})^{-1}$  is given by

$$\hat{s}_{\xi_p}^2 (\mathbf{X}'_{jh} \mathbf{X}_{jh})^{-1} = \frac{1}{HK_h} \sum_{h=1}^H \sum_{j=1}^{K_h} \frac{p(1-p)}{(\hat{f}_{\xi_p} | \Theta_{pjh}, \Theta_{ph})^2} (\mathbf{X}'_{jh} \mathbf{X}_{jh})^{-1} \quad (39)$$

and estimator of  $\Lambda_p$  is given by

$$\hat{\Lambda}_p = \frac{1}{\sum_{j=1}^H K_h - 1} \sum_{h=1}^H \sum_{j=1}^{K_h} (\hat{\beta}_{pjh} - \hat{\beta}_{p.h}) (\hat{\beta}_{pjh} - \hat{\beta}_{p.h})' - \hat{s}_{\xi_p}^2 (\mathbf{X}'_{jh} \mathbf{X}_{jh})^{-1}, \quad (40)$$

$$\text{where } \hat{\beta}_{p.h} = \frac{1}{K_h} \sum_{j=1}^{K_h} \hat{\beta}_{pjh}. \quad (41)$$

Finally an estimator of  $\mathbf{U}_p$  is given by

$$\hat{\mathbf{U}}_p = \frac{1}{(H-1)} \sum_{h=1}^H (\hat{\beta}_{p,h} - \hat{\beta}_{p..})(\hat{\beta}_{p,h} - \hat{\beta}_{p..})' - \hat{\xi}_{\xi_p}^2 (\mathbf{X}'_{jh} \mathbf{X}_{jh})^{-1} - \hat{\Lambda}_p, \quad (42)$$

$$\text{where } \hat{\beta}_{p..} = \frac{1}{H} \sum_{j=1}^H \hat{\beta}_{p,h}. \quad (43)$$

The term  $\sigma_{\xi_{pj}}^2(\Theta_{pjh}, \Theta_{ph})$ , can be estimated by an interval estimation from the  $[np]^{th}$  quantile of cumulative distribution function, say  $F(\cdot)$ , denoted by  $\xi_{pj}$  and defined by  $F(\xi_{pj}) = p_j$  [see Mood et al., (1974), p. 512].

Two other alternative bootstrap methods can be employed, proposed by Efron (1979). The Design Matrix Bootstrap Estimator, which provides a consistent estimator of the asymptotic matrix under more general conditions, and the Error Bootstrap Estimator, which yields a consistent estimator only under the independence assumption.

For the estimation of covariance matrix  $\Lambda_p$ , Power (1986) considered a kernel estimator by choosing the appropriate bandwidth. For details of the above estimations the reader may be referred to Koenker (2005) and Buchinsky (1998).

## Numerical Example

Here we illustrate how quantile regression performs in a hierarchical credibility framework on a data set that obtained from the largest social security organization in Greece, which covers 5530000 workers and employees, with medical care, medication and hospital care. The 6 contracts correspond to the 6 different classes (contracts) of allowances (expenses) all covered by the social security organization of Greece, for 22 years of claim experiences, from 1980-2001. The 6 different classes (contracts) are:

*Sick<sub>A</sub>*=Sickness allowance,

*Accid<sub>A</sub>*=accident allowance,

*Matern<sub>A</sub>*=maternity benefit,

*Funer<sub>Exp</sub>*=Funeral expenses,

*Other<sub>A</sub>*=Other allowances,

*Manag<sub>Exp</sub>*=Management expenses.

**Table 1: Summary Statistics for Claim Amounts**

<b>Contract</b>	<b>Min</b>	<b>Mean</b>	<b>Median</b>	<b>Max</b>	<b>St.Deviation</b>
<i>Sick<sub>A</sub></i>	3436	17527.50	16782.50	35650	9940.66
<i>Accid<sub>A</sub></i>	464	1970.00	2003.50	3750	969.72
<i>Matern<sub>A</sub></i>	609	6230.96	5332.00	15340	4832.55
<i>Funer<sub>Exp</sub></i>	283	3026.55	2373.50	7281	2340.54
<i>Other<sub>A</sub></i>	232	3027.00	3106.50	8862	1984.07
<i>Manag<sub>Exp</sub></i>	275	2657.96	1884.00	8645	2415.41

Note: Claim amount for contract  $j$ , for the period 1980-2001. The amount of allowances is in million of drachmas (1 Euro=340.75 drachmas).

We suppose that the six classes of allowances (contracts) of our data for some reason belong to two different subportfolios. For this reason we assign contracts  $Sick_A$ ,  $Accid_A$  and  $Matern_A$  to subportfolio 1 and contracts  $Funer_{Exp}$ ,  $Other_A$  and  $Manag_{Exp}$  to subportfolio 2.

As we have mentioned before, with the use of different quantiles we can examine changes at different points of the claim distribution. Figure 1 provides interpretation of the behavior of the six contracts covered by the medical program, each one with respect to year, at different points of the claim distribution. For each contract, seven quantile regression lines for different values of 0.05, 0.1, 0.25, 0.50, 0.75, 0.90, 0.95 are superimposed on the scatter plots in Figure 1. The median  $p = 0.5$  is indicated by the darker solid line and the least squares estimate of the conditional mean function is indicated by the dashed line.

Figure 1 also shows the existence of outlier observations in most of the regression lines, especially in the lines of the 5th contract (other allowances) and the 6th contract (management expenses). The consequence of the nonrobustness is that the least squares estimation provides a rather poor estimate of the conditional mean in comparison with quantile (median) estimation.

Table 2a : Hierarchical Credibility Models						
Contract	Subportfolio 1			Subportfolio 2		
	$Sick_A$	$Accid_A$	$Matern_B$	$Funer_{Exp}$	$Other_A$	$Manag_{Exp}$
<b>Hierarchical Regression Credibility Model</b>						
<b>LS-estimators</b>						
$\hat{\beta}_{0j}^{LS}$	34966.50	3749.12	14585.67	7079.40	4867.30	6562.50
$\hat{\beta}_{1j}^{LS}$	-1516.43	-150.50	-726.50	-352.42	-160.03	-339.52
$R^2$	0.9813	0.9518	0.953	0.956	0.2743	0.8332
<b>Adjusted LS Credibility Estimators</b>						
$B_{0j}^{aLS}$	34959.69	3718.14	14460.27	7049.14	5002.25	6517.30
$B_{1j}^{aLS}$	-1518.12	-153.84	-711.14	-348.11	-176.08	-333.29
	$\hat{\beta}_{p1}^{LS} = (17732.53 \quad -796.21)'$			$\hat{\beta}_{p2}^{LS} = (6169.73 \quad -283.99)'$		
<b>Hierarchical Quantile Regression Credibility Model</b>						
<b>Quantile (Median)-estimators</b>						
$\hat{\beta}_{0j}^{Med}$	34458.80	3680.75	15025.73	7099.73	4517.20	5382.00
$\hat{\beta}_{1j}^{Med}$	-1485.20	-148.25	-762.87	-356.45	-125.36	-269.00
<b>Adjusted Hierarchical Quantile (Median) Credibility Estimators</b>						
$\widehat{\beta}_{p0}^{Cred}(\Theta_{pjh}, \Theta_{ph})$	34473.9640	3284.61	14854.79	7081.13	4647.13	5395.15
$\widehat{\beta}_{p1}^{Cred}(\Theta_{pjh}, \Theta_{ph})$	-1489.54	-106.99	-741.77	-353.20	-190.39	-269.46
	$\hat{\beta}_{p.1}^{Med} = (17721.76 \quad -798.77)'$			$\hat{\beta}_{p.2}^{Med} = (6169.73 \quad -283.99)'$		

Table 2b			
Estimated Structural Parameters			
$\hat{\mathbf{b}}^{LS}$		$\hat{\beta}_{p..}^{Med}$	
11968.41		11694.04	
-540.94		-532.90	
$\hat{\Lambda}^{LS}$		$\hat{\Lambda}_p^{Med}$	
126741757	5492462 8	21733329	-5220233
-5492462	240923.2	-5220233	226452.6
$\hat{\mathbf{U}}^{LS}$		$\hat{\mathbf{U}}_p^{Med}$	
-60159997	2548965	-31167048	1231477
2548965	-111290.9	1231477	-51276.41
$\hat{s}^{2LS}$		$\hat{s}_{\xi_p}^2, p = Median$	
1236546		1372762	
Credibility Factors			
$\hat{\mathbf{Z}}^{LS}$		$\hat{\mathbf{Z}}^{Med}$	
1.08546644	2.0416696	1.08546644	2.0416696
-0.01069404	0.7475193	-0.01069404	0.7475193



Table 3a : Contaminated Hierarchical Credibility Models

Table 3a : Contaminated Hierarchical Credibility Models						
	Subportfolio 1			Subportfolio 2		
Contract	<i>Sick<sub>A</sub></i>	<i>Accid<sub>A</sub></i>	<i>Matern<sub>B</sub></i>	<i>Funer<sub>Exp</sub></i>	<i>Other<sub>A</sub></i>	<i>Manag<sub>Exp</sub></i>
<b>Hierarchical Regression Credibility Model</b>						
<b>Contaminated LS-estimators</b>						
$\hat{\beta}_{0j}^{LS}$	46666.5	3749.12	14585.67	7079.40	4867.30	6562.50
$\hat{\beta}_{1j}^{LS}$	-2279.5	-150.50	-726.50	-352.42	-160.03	-339.52
$R^2$	0.5504	0.9518	0.953	0.956	0.2743	0.8332
<b>Adjusted LS Credibility Estimators</b>						
$B_{0j}^{aLS}$	46058.46	3991.52	14338.32	7014.82	5611.33	6392.32
$B_{1j}^{aLS}$	-2254.52	-176.41	-691.46	-329.49	-251.69	-300.05
	$\hat{\beta}_{p1}^{LS} = (17732.53 \quad -796.21)'$			$\hat{\beta}_{p2}^{LS} = (6169.73 \quad -283.99)'$		
<b>Hierarchical Quantile Regression Credibility Model</b>						
<b>Quantile (Median)-estimators</b>						
$\hat{\beta}_{0j}^{Med}$	34458.80	3680.75	15025.73	7099.73	4517.20	5382.00
$\hat{\beta}_{1j}^{Med}$	-1485.20	-148.25	-762.87	-356.45	-125.36	-269.00
<b>Adjusted Hierarchical Quantile (Median) Credibility Estimators</b>						
$\widehat{\beta}_{p0}^{Cred}(\Theta_{pjh}, \Theta_{ph})$	34473.9640	3284.61	14854.79	7081.13	4647.13	5395.15
$\widehat{\beta}_{p1}^{Cred}(\Theta_{pjh}, \Theta_{ph})$	-1489.54	-106.99	-741.77	-353.20	-190.39	-269.46
	$\hat{\beta}_{p.1}^{Med} = (21632.532 \quad -1050.555)'$			$\hat{\beta}_{p.2}^{Med} = (6169.73 \quad -283.99)'$		

**Table 3b: Estimated Structural Parameters**

$\hat{\mathbf{b}}^{LS}$		$\hat{\boldsymbol{\beta}}_{p..}^{Med}$	
13901.132		11694.04	
-667.273		-532.90	
$\hat{\boldsymbol{\Lambda}}^{LS}$		$\hat{\boldsymbol{\Lambda}}_p^{Med}$	
250401965	-12376396 8	21733329	-5220233
-12376396	612898	-5220233	226452
$\hat{\mathbf{U}}^{LS}$		$\hat{\mathbf{U}}_p^{Med}$	
-131093745	6465844	-31167048	1231477
6465844	-320485	1231477	-51276
$\hat{s}_{\xi}^{2LS}$		$\hat{s}_{\xi_p}^2, p = Median$	
32240024		1372762	
<b>Credibility Factors</b>			
$\hat{\mathbf{z}}^{LS}$		$\hat{\mathbf{z}}^{Med}$	
1.43347706	9.1868387	1.08546644	2.0416696
-0.06572904	-0.3513141	-0.01069404	0.7475193

## References

- Bassett, G., Koenker, R., 1982. An empirical Quantile Function for Linear Models with iid Errors. *Econometrica*, 77(378), 407-415.
- Bauwelinckx, T. & Goovaerts, M. J., 1990. On a multilevel hierarchical credibility algorithm *Insurance: Mathematics and Economics*, 9, 221-228.
- Bauwelinckx, T. & Goovaerts, M. J., 1990. On a multilevel hierarchical credibility algorithm *Insurance: Mathematics and Economics*, 9, 221-228.
- Buchinsky, M., 1998. Recent Advances in Regression Models: A Practical Guideline for Empirical Research. *The Journal of Human Resources*, 33(1), 88-126.
- Goovaerts, M. J, Kaas, R., Van Heerwaarde A. E., Bauwelinckx, T., 1990. *Effective Actuarial Methods*, Amsterdam, The Netherlands.

- Jewell, W. S. 1975. The Use of Collateral data in Credibility Theory: A Hierarchical Model *Giornale dell'Istituto Italiano degli Attuari*, 38, 116.
- Koenker, R., 2005. Quantile Regression. Cambridge University Press, Cambridge.
- Kudryavtsev, A., 2009. Using quantile regression for rate-making. *Insurance: Mathematics and Economics*, 45, 296-304.
- Norberg, N., 1986. Hierarchical Credibility: Analysis of a Random Effect Linear Model with Nested classification *Scandinavian Actuarial Journal*, 204-222.
- Pitt, D. G. W., 2006. Regression quantile analysis of claim termination rates for income protection insurance. *Annals of Actuarial Science* 1 (II), 345-357.
- Sundt, B., 1979. Hierarchical Credibility Regression Model, *Scandinavian Actuarial Journal*, 107-114.
- Sundt, B., 1980. A multi-level hierarchical credibility regression model. *Scandinavian Actuarial Journal*, 1980, 25-32.
- Sundt, B., 1986. Two credibility regression approaches for the classification of passenger cars in a multiplicative tariff, *Astin Bulletin*, 17, 41-70.
- Taylor G. C., 1979. Credibility Analysis of a general Hierarchical Model. *Scandinavian Actuarial Journal*, 1-12.