Estimating Insurance Premiums using Credibility Theory and Quantiles

Li Yang*

Abstract

This thesis is based on the paper *'Quantile credibility models'* by Georgios Pitselis, which was published in 2013. It introduces credibility theory and shows how quantiles can be incorporated in the Bühlmann-Straub model and the Hachemeister's regression model. For each model, a numeric example is presented.

1 Introduction

Insurance contracts must be familiar to most people, but it is little known how premiums are estimated. The modern credibility[1] theory estimates an insured's pure risk premium by striking a balance between the individual's average claims and the overall mean of the claim data. Below is an illustration of this premium estimation technique.
It is well known that an insurance company insures many kinds of risks. These risks are grouped into 'similar risks', called the collective, on the basis of 'objective' risk characteristics [1]. Examples of such risk characteristics are age and sex. Based on the observed data and statistical information, the structure of the collectives could be determined. When setting a premium for a new risk, about which there is no pre-existing claim experience, one could estimate it by the overall mean of the claim data $\overline{X}$, i.e. the average claim amount over all risk collectives. After a period of n years, the aggregate claim amounts observed could contribute to the estimation process. Then the premium for this risk could be estimated by a combination of $\overline{X}$ and the individual average claims, i.e. $\overline{X_j}$, the structure is as follows:

$$z_j \bar{X}_j + (1 - z_j)\bar{X}.$$

A premium such as this is called *a credibility premium*, z_j is *the credibility factor*, expressing how much faith you can have in the individual average claims. [2].
Bühlmann introduced *the balanced Bühlmann model* in 1967 and together with Straub

*Li Yang received a bachelor degree in Econometrics & Operations Research at Maastricht University in 2017, where she currently takes the Research Master in the same field. Contact: li.yang@student.maastrichtuniversity.nl

[1] A premium estimation technique.

introduced *the Bühlmann-Straub model* in 1970 which established the theoretical foundation of modern credibility theory. Hachemeister extended these models in *Hachemeister's regression model* by using linear trend model. In 2013, Georgios Pitselis published a paper illustrating links between credibility theory and quantiles. This improves the premium estimation technique as claim distributions are in general heavy-tailed, and quantile credibility regression detects information about the tail behaviour of claim distributions.
This paper aims to introduce *the Bühlmann-Straub model*, *the Quantile credibility model*, *Hachemeister's regression model* and *the Quantile regression credibility model.* The rest of the paper is organized as follows. Section 2 covers the concept of quantiles and conditional quantiles. Section 3 discusses quantile regression and its estimation method. *The Bühlmann-Straub model* will be introduced in section 4, as well as how quantiles are incorporated into this model. *Hachemeisters's regression model* will be introduced in section 5, as well as how quantiles are incorporated in this model. Section 6 discusses the effect of outliers on these models. Finally, Section 7 draws conclusions.

2 Quantiles and Conditional Quantiles

2.1 Definition

Let Y be a random variable with the distribution F_Y and p be a real number between 0 and 1, i.e. $0 < p < 1$. Then the p-quantile ξ_p is defined as follows:

$$\xi_p = F_Y^{-1}(p) = inf\{y : F_Y(y) \geq p\}.$$

As the payment could take any value of some entire multiple of the monetary unit, which results in a very large set of possible values, each of them with a very small probability, a continuous CDF for F_Y is more appropriate here. Then the p-quantile of F_Y can be obtained by minimizing the following objective function with respect to ξ_p:

$$\begin{aligned} &p\int_{y\geq\xi_p} |y-\xi_p|\,\mathrm{d}F_Y(y) + (1-p)\int_{y<\xi_p} |y-\xi_p|\,\mathrm{d}F_Y(y) \\ &= p\int_{y\geq\xi_p} (y-\xi_p)\,\mathrm{d}F_Y(y) - (1-p)\int_{y<\xi_p} (y-\xi_p)\,\mathrm{d}F_Y(y). \end{aligned} \tag{1}$$

Take the derivative of equation (1) w.r.t. ξ_p:

$$\begin{aligned} &-p\int_{y\geq\xi_p} dF_Y(y) + (1-p)\int_{y<\xi_p} dF_Y(y) \\ &= -p[1-F_Y(\xi_p)] + (1-p)F_Y(\xi_p) \\ &= -p + F_Y(\xi_p) \\ &\overset{set}{=} 0. \end{aligned}$$

The second derivative $= f_Y(y) \geq 0$, which implies that $\xi_p = F_Y^{-1}(p)$ indeed minimizes the objective function mentioned above.

Similarly, in case that Y has a conditional distribution $F_{Y|X}$, the p-quantile is defined as follows:

$$\xi_p = F_{Y|X}^{-1}(p) = inf\{y : F_{Y|X}(y) \geq p\}.$$

ξ_p is a function of X and minimizes the following objective function:

$$p \int_{y \geq \xi_p} |y - \xi_p| \, \mathrm{d}F_{Y|X}(y) + (1 - p) \int_{y < \xi_p} |y - \xi_p| \, \mathrm{d}F_{Y|X}(y). \tag{2}$$

2.2 Empirical Quantile Function

Let $Y_1, Y_2, ..., Y_n$ denote the order statistics of $X_1, X_2, ..., X_n$, and let $\hat{\xi}_p$ denote the sample p-quantile. Assume that the order statistics $Y_1, Y_2, ..., Y_n$ partition the support of X into n parts and thereby create n equalling areas under $f(x)$ and above the X-axis, see *figure 1*, then each area is on average: $\dfrac{1}{n}$.

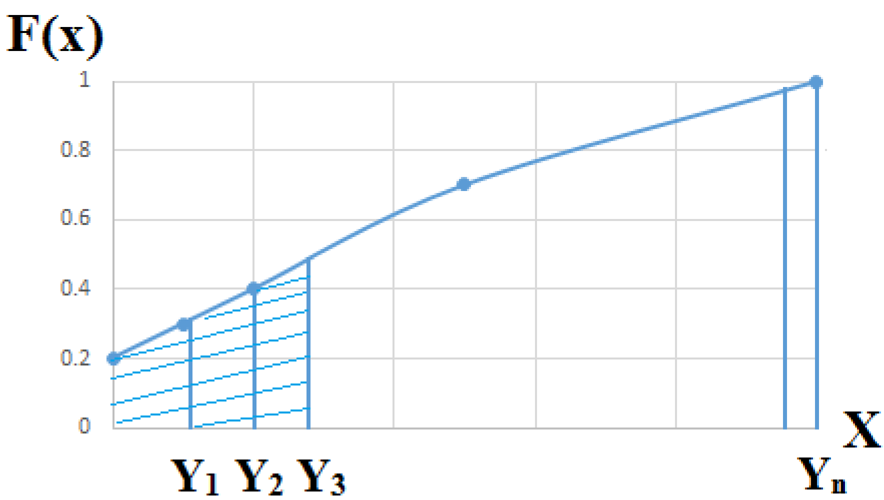


FIGURE 1: Sample Quantiles

According to the definition, p is the area under $f(x)$ to the left of ξ_p. If np is an integer, $Y_j (1 \leq j \leq n)$ serves as an estimator of ξ_p, namely:

$$\hat{\xi}_p = Y_j, if\, p = \frac{j}{n}.$$

In case $j - 1 < np < j$, the empirical quantile function can be defined as:

$$\begin{aligned} \hat{\xi}_p &= Y_{j-1} + \{np - (j-1)\}(Y_j - Y_{j-1}) \\ &= (j - np)Y_{j-1} + \{np - (j-1)Y_j\}. \end{aligned}$$

In summary, $\forall$ integer-valued j $\in [1, n]$,

$$\hat{\xi}_p = \begin{cases} Y_j & \text{if } np = j; \\ (j - np)Y_{j-1} + \{np - (j-1)Y_j\} & \text{if } j - 1 < np < j. \end{cases}$$

2.3 Confidence Interval

This subsection shows how confident that ξ_p is contained in the interval, say (Y_j, Y_k). Simply, it is to calculate $P(Y_j < \xi_p < Y_k)$. Let N denotes the number of X_i which is smaller than ξ_p, then N is a binomial random variable with n mutually independent trials and with probability of success $p = P(X_i < \xi_p)$. To ensure that ξ_p is sandwiched by Y_j and Y_k, the order of Y_j must be less than ξ_p, and the order of Y_k greater than ξ_p, meaning there are at least j X_is to the left of ξ_j and at most (k-1) X_is to the right of ξ_j, then $j \leq N \leq k-1$. It immediately follows that:

$$P(Y_j < \xi_p < Y_k) = \sum_{i=j}^{k-1} P(N = i) = \sum_{i=j}^{k-1} \binom{n}{i} (p)^i (1-p)^{n-i}.$$

3 Quantile Regression

3.1 Classical Multiple Linear Regression Model

To analyse the behaviour of a dependent variable, given a set of explanatory variables, a standard approach is to use multiple linear regression and estimate the parameters by minimizing the sum of squared residuals, which leads to an approximation to the mean function of the conditional distribution of the dependent variable. Below is an illustration of this method.
The form of this regression model for a single observation is as follows:

$$y_i = \beta_1 + x_{i2}\beta_2 + \cdots + x_{iK}\beta_K + \epsilon_i \ .$$

If stack all n observations:

$$\underbrace{Y}_{\text{n} \times 1} = \underbrace{X}_{\text{n}\times\text{K}} \underbrace{\beta}_{\text{K}\times 1} + \underbrace{\epsilon}_{\text{n}\times 1} \ ,$$

where

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1' \\ 1 & x_2' \\ \vdots & \vdots \\ 1 & x_n' \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & x_{13} & \cdots & x_{1K} \\ 1 & x_{22} & x_{23} & \cdots & x_{2K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n2} & x_{n3} & \cdots & x_{nK} \end{pmatrix}$$

$Y: vector\ of\ dependent\ variables, n\ observations;$
$X: matrix\ of\ k\ explanatory\ variables\ of\ n\ observations, with\ k \leq n;$
$\beta: vector\ of\ k\ unknown\ parameters;$
$\epsilon_i: assumed\ to\ be\ iid, with\ E[\epsilon_i|X] = 0, and\ Cov[\epsilon_i, \epsilon_j|X] = 0, Var[\epsilon_i|X] = \sigma^2,\ let$

$$\Sigma = \begin{pmatrix} \epsilon_1^2|X & \epsilon_1\epsilon_2|X & \cdots & \epsilon_1\epsilon_n|X \\ \epsilon_2\epsilon_1|X & \epsilon_2^2|X & \cdots & \epsilon_2\epsilon_n|X \\ \epsilon_n\epsilon_1|X & \epsilon_n\epsilon_2|X & \cdots & \epsilon_n^2|X \end{pmatrix},\ then\ \Sigma = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix};$$

$the\ population\ regression\ is: E[Y|X] = X\beta;$
$the\ estimate\ of\ E[Y|X]: \hat{Y} = X\hat{\beta};$
$the\ residual: e = Y - \hat{Y} = Y - X\hat{\beta};$
$$the\ sum\ of\ squared\ residuals: \sum_{i=1}^{n} e'e = \sum_{i=1}^{n}(Y - X\hat{\beta})'(Y - X\hat{\beta}).$$

The estimated coefficients are calculated as follows:

$$\hat{\beta} = (X'X)^{-1}(X'Y),$$
$$with\ \Sigma_{\hat{\beta}} = \sigma^2(X'X)^{-1}.$$

When the assumption $Var[\epsilon_i|X] = \sigma^2$ is violated, meaning the value of σ_i^2 may differ from each other, namely:

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}.$$

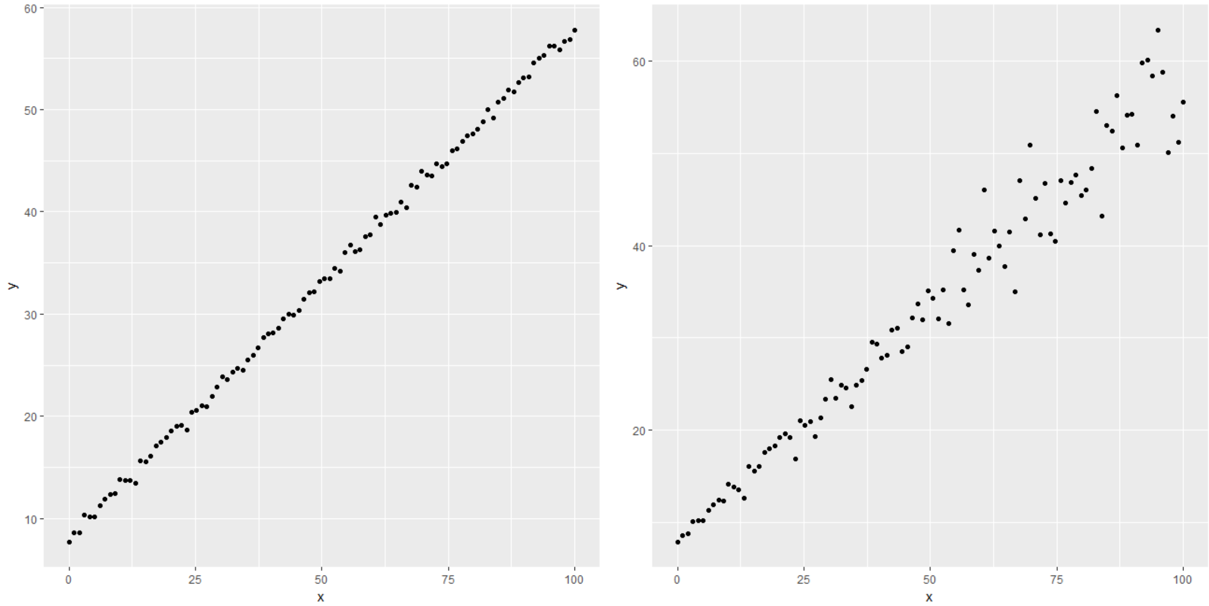
the method of **weighted least squares** is used, then

$$\hat{\beta} = (X'\Sigma^{-1}X)^{-1}(X'\Sigma^{-1}Y),$$
$$with\ \Sigma_{\hat{\beta}} = \sigma^2(X'\Sigma^{-1}X)^{-1}.$$

Weighted least squares estimation is used in the Hachemeister's regression model which will be covered in later section.[4]

3.2 OLS vs Quantile

The mean measure, obtained from the classical multiple linear regression, represents the average behaviour of a distribution, but provides little information about the tail behaviour of that distribution. In 1918, Koenker and Bassett proposed the quantile regression, which enables to estimate various quantile functions of a conditional distribution. This approach provides a more comprehensive picture of the effect of dependent variables on independent variable by putting different quantile regressions together in one graph. To see the difference two simple examples are presented. Both disturbances are normally distributed, data of *figure 2a* is with constant variance, data of *figure 2b* is with non-constant variance. As can be seen from *figure 2b* that the dependent variable becomes more variable when the independent variable increases.



(A) Data (constant variance) (B) Data (non-constant variance)

FIGURE 2: Two simple data sets

These two data sets are generated by employing the following R-codes :

Data (constant variance):

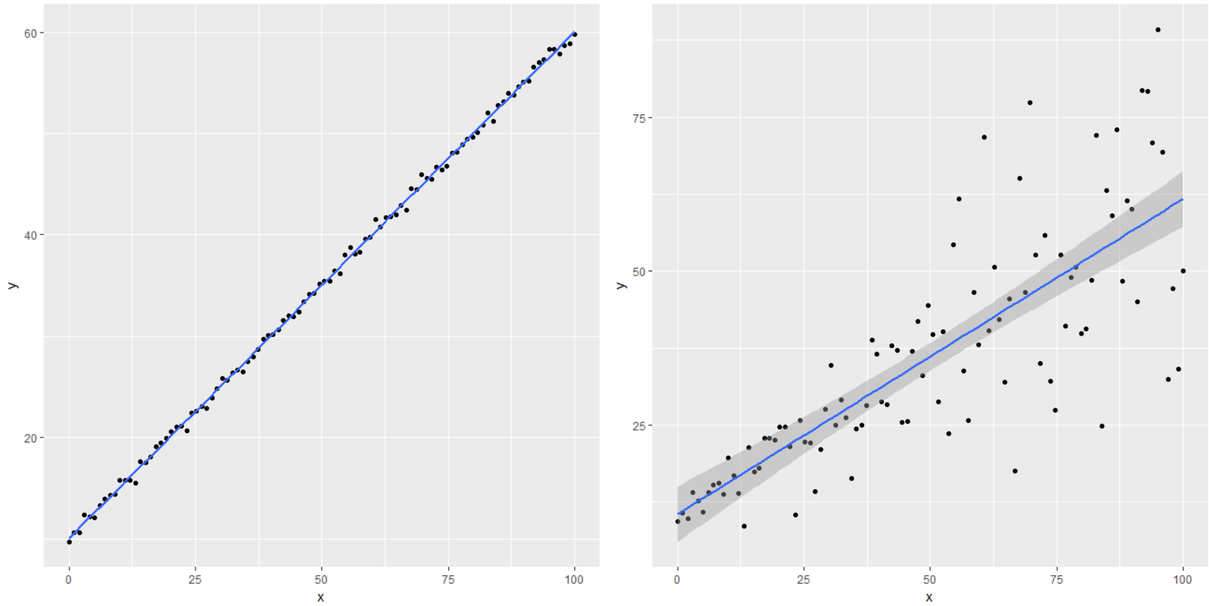
```
set.seed(1)
x <- seq(0, 100, length.out=100)
b_0 <- 10
b_1 <- 0.5
e <- rnorm(100, mean=0, sd=0.5)
y <- b_0 + b_1*x + e
```

Data (non-constant variance):

```
set.seed(1)
x <- seq(0, 100, length.out=100)
b_0 <- 10
b_1 <- 0.5
sig <- 1 + 0.2*x
e <- rnorm(100,mean=0,sd=sig)
y <- b_0 + b_1*x + e
```

FIGURE 3: Two artificial data

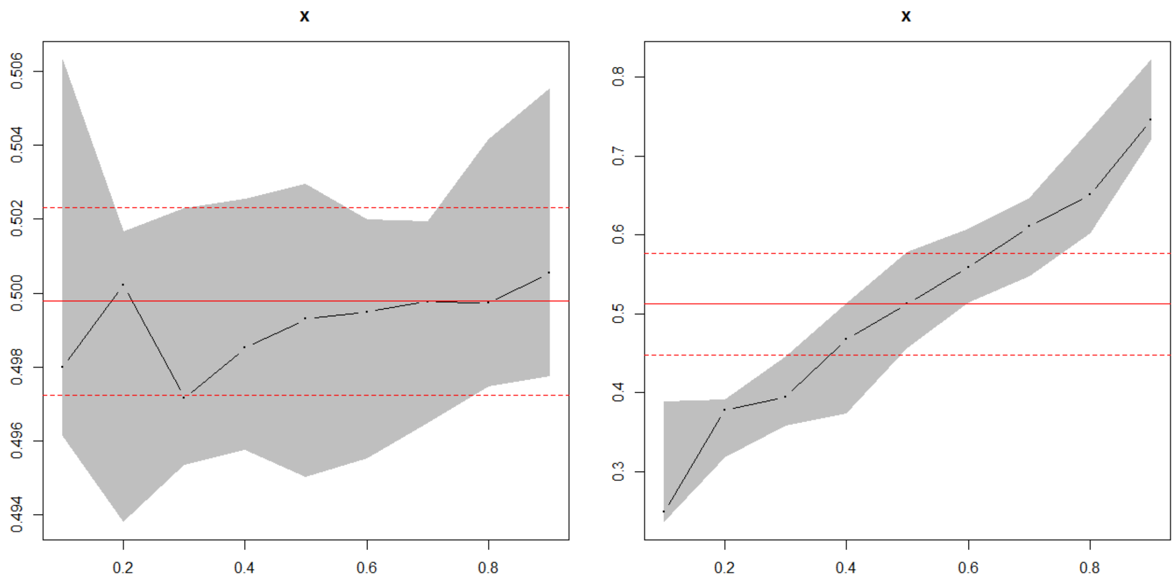
First, apply OLS regression to both data sets. *Figure 4a* and *figure 4b* show the results separately. When the variance is constant, the conditional mean from OLS regression provides a good estimate of dependent variable. However, it is not the case when the variance is not constant, as the conditional mean becomes less meaningful when the explanatory variable increases, seen from *figure 4b*.



(A) OLS (constant variance) (B) OLS (non-constant variance)

Figure 4: OLS regression

Second, apply quantile regression to both data sets. The slope coefficient for the quantile indicated on the x-axis for both data sets are shown separately in *figure 5a* and *figure 5b*. The horizontal solid lines are the OLS coefficient estimates, the horizontal dotted lines are its confidence interval; the black curved lines are the quantile coefficient estimates, the shaded area is its confidence interval.



(A) Quantile coefficients (constant variance) (B) Quantile coefficients (non-constant variance)

Figure 5: Quantile regression

Figure 5a shows that the quantile coefficient estimates fall well inside the confidence interval when the variance is constant, meaning the quantile estimates are not significantly different from the OLS coefficient estimates. On the other hand, when the variance is not constant, *figure 5b* shows that the majority of quantile coefficient estimates fall outside of the OLS confidence interval, meaning quantile coefficient estimates are significantly different from OLS coefficient estimates. This implies that, under the assumption that the disturbance is normally distributed and with constant variance, there is no efficiency gain by using quantile regression; but quantile estimations perform better when the variance is not constant, as the coefficients can vary across quantiles. To point out, the estimated coefficients are significantly different from 0, when 0 is not in the confidence interval. In case the error term is not normally distributed, quantile regression may be more efficient than the least squares estimation, mentioned in Pitselis' paper.

3.3 The Method of Quantile Regression

Similar to general linear regression model, quantile regression model has the following regression equation:

$$y_{pi} = x_i'\beta_p + u_{pi} \ .$$

Here, y_{pi} is the p-quantile of the observed risks $y_{i1}, ..., y_{in}$; β_p is the corresponding regression coefficients. As p can take any value between 0 and 1, β_p may have different value for each choice of p, which is different from the linear regression model where β_p is a fixed parameter. The p-conditional quantile of y_i given x_i is:

$$Q_p(y_i|x_i) = x_i'\beta_p$$

In view of (1), Q_p of quantile regression can be written in the following form:

$$Q_p(y_i|x_i) = p\int_{y_i \geq x_i'\beta_p} |y_i - x_i'\beta_p| \, \mathrm{d}F_{Y|X}(y) + (1-p)\int_{y_i < x_i'\beta_p} |y_i - x_i'\beta_p| \, \mathrm{d}F_{Y|X}(y) \qquad (3)$$

The estimator of β_p can be obtained by minimizing its sample counterpart:

$$\begin{aligned} \hat{Q}_p(y_i|x_i) &= \frac{1}{N}\left[p\sum_{i:y_i \geq x_i'\beta_p}^{N} |y_i - x_i'\beta_p| + (1-p)\sum_{i:y_i < x_i'\beta_p}^{N} |y_i - x_i'\beta_p|\right] \\ &= \frac{1}{N}\sum_{i=1}^{N} g(y_i - x_i'\beta_p|p) \end{aligned}$$

where

$$g(y_i - x_i'\beta_p|p) = \begin{cases} p(y_i - x_i'\beta_p) & \text{if } y_i - x_i'\beta_p \geq 0 \\ (1-p)(y_i - x_i'\beta_p) & \text{if } y_i - x_i'\beta_p < 0 \end{cases}$$

For $p = 0.5$,

$$2\hat{Q}_{0.5}(y_i|x_i) = \frac{1}{N}\sum_{i=1}^{n} |y_i - x_i'\beta_p|$$

This is a regression estimated via the method of LAD: Least Absolute Deviations Estimation, referred to as a "median regression". This minimization problem can be set up as

a linear programming problem, and quantile regression can be implemented in software such as R.
The function $g(y_i - x_i'\beta_p|p)$ is known as the "check function": piecewise linear and not differentiable at $y_i = x_i'\beta_p$. One way to minimize $\hat{Q}_p(y_i|x_i)$ is to use *the directional derivatives* of $\hat{Q}_p(y_i|x_i)$. *The directional derivatives* works as follows:
let ω be an arbitrary direction,

$$\frac{d}{d\delta}\hat{Q}_p\Big(\beta_p + \delta\omega; p\Big)\Big|_{\delta=0} = \frac{1}{N}\Big(\frac{d}{d\delta}\sum_{i=1}^{N}\big(\xi_p - x_i'\beta_p - \delta x_i'\omega\big)\big(p - I_{\{\xi_p - x_i'\beta_p - \delta x_i'\omega<0\}}\big)\Big)\Big|_{\delta=0}$$
$$= -\frac{1}{N}\sum_{i=1}^{N}\psi_p\big(\xi_p - x_i'\beta_p, -x_i'\omega\big)x_i'\omega$$

where

$$\psi_p(\xi_p - x_i'\beta_p, -x_i'\omega) = \begin{cases} p - I_{\{\xi_p - x_i'\beta_p<0\}} & \text{if } \xi_p - x_i'\beta_p \neq 0 \\ p - I_{\{-x_i'\omega<0\}} & \text{if } \xi_p - x_i'\beta_p = 0 \end{cases}$$

To determine the quantile regression estimator of β_p is to find a point that minimizes the function $Q_p(\beta_p; p)$, which is equal to find a point where $\frac{d}{d\delta}Q_p\Big(\beta_p + \delta\omega; p\Big)$ is non-negative at all directions. The p-quantile regression estimator of β_p is denoted as $\hat{\beta}_p$, quantile regression residuals are $\hat{e}_i(p) = \xi_p - x_i'\hat{\beta}_p$.

If $plim\frac{1}{n}X'X$ equals a finite and positive definite matrix, meaning data is well-behaved, $\hat{\beta}_p$ is consistent and asymptotically normally distributed with asymptotic covariance matrix

$$Asy.Var[\hat{\beta}_p] = \frac{1}{n}H'GH',$$

where

$$H = plim\frac{1}{n}\sum_{i=1}^{n} f_p(0|x_i)x_ix_i' ,$$
$$G = plim\frac{p(-p)}{n}\sum_{i=1}^{n} x_ix_i' .$$

It is worth mentioning that computation of $f_p(0|x_i)$ could be complicated.[4]

4 Credibility Models

4.1 The Bühlmann-Straub model

Let X_{jt} be a claim amount for contract j at time t, then X_{jt} can be decomposed into three parts:

$$X_{jt} = m + \Xi_j + \Xi_{jt}, \quad j = 1, ..., J,\ t = 1, ..., T+1$$

m: the overall mean claim, i.e. the expected value of the claim amount for an arbitrary policyholder in the portfolio;

Ξ_j : denotes a random deviation from m, iid, with $E[\Xi_j] = 0, Var[\Xi_j] = a$;
Ξ_{jt}: denotes the deviation for year t from the long-term average. Ξ_{j1}, Ξ_{j2},... are iid, with $E[\Xi_{jt}] = 0$, $Var[\Xi_{jt}] = s^2/w_{jt}$, w_{jt} is the weight attached to observation X_{jt}.
This model assumes equal numbers of policies in X_{jt}.
Then the unbiased predictor of $X_{j,n+1}$ which minimizes the following objective function :

$$E\left[\{m + \Xi_j - \sum_{j=1}^{K}\sum_{i=1}^{n} h_{ji}X_{ji}\}^2\right], \; subject\ to\ E[m + \Xi_j] = \sum_{i,t} h_{ji}X_{ji},$$

equals the credibility premium:

$$z_j X_{jw} + (1 - z_j)X_{zw},$$

where

$$z_j = \frac{aw_{j\Sigma}}{s^2 + aw_{j\Sigma}} \; ; \; z_\Sigma = \sum_{j=1}^{J} z_j \; ;$$
$$w_{j\Sigma} = \sum_{t=1}^{T} w_{jt} \; ; \; w_{\Sigma\Sigma} = \sum_{j=1}^{J} w_{j\Sigma} \; ;$$
$$X_{jw} = \sum_{t=1}^{T} \frac{w_{jt}}{w_{j\Sigma}} X_{jt}; \; X_{ww} = \sum_{j=1}^{J} \frac{w_{j\Sigma}}{w_{\Sigma\Sigma}} X_{jw} \; ; \; X_{zw} = \sum_{j=1}^{J} \frac{z_j}{z_\Sigma} X_{jw} \; .$$

Unbiased parameter estimates:

$$\widetilde{m} = X_{ww};$$
$$\widetilde{s^2} = \frac{1}{J(T-1)} \sum_{j,t} w_{jt}(X_{jt} - X_{jw})^2 \; ;$$
$$\widetilde{a} = \frac{\sum_j w_{j\Sigma}(X_{jw} - X_{ww})^2 - (J-1)\widetilde{s^2}}{w_{\Sigma\Sigma} - \sum_j w_{j\Sigma}^2 / w_{\Sigma\Sigma}} \; , in\ case\ \tilde{a} = 0, set\ \tilde{a} = (J-1) \times \tilde{s}^2 \; .$$

It is worth noting that all the equations hold if given Ξ_j, Ξ_{jt} is iid and with $E[\Xi_{jt}|\Xi_j] = 0$. Then $Cov[\Xi_{jt}, \Xi_{ju}] = 0 \; (t \neq u), Cov[\Xi_j, \Xi_{jt}] = 0$, which means that Ξ_{jt} and Ξ_{ju} are uncorrelated, Ξ_j and Ξ_{jt} are uncorrelated, but the random variables X_{jt} are not marginally uncorrelated.
In the Quantile credibility model and Quantile regression credibility model Georgios Pitselis uses Θ_j instead of Ξ_j, and assumes that given Θ_j, $X_{1j,}, ..., X_{nj}$ are conditionally independent and with the same distribution function.[2]

Numeric Example

Apply the Bühlmann-Straub model to Hachemeister's claims data set (the severity average loss per claim per state in 12 periods). The data set is shown below.

P	Average loss per claim					Number of claims per period				
	State1	State2	State3	State4	State5	State1	State2	State13	State4	State5
1	1738	1364	1759	1223	1456	7861	1622	1147	407	2902
2	1642	1408	1685	1146	1499	9251	1742	1357	396	3172
3	1794	1597	1479	1010	1609	8706	1523	1329	348	3046
4	2051	1444	1763	1257	1741	8575	1515	1204	341	3068
5	2079	1342	1674	1426	1482	7917	1622	998	315	2693
6	2234	1675	2103	1532	1572	8263	1602	1077	328	2901
7	2032	1470	1502	1953	1606	9456	1964	1277	352	3275
8	2035	1448	1622	1123	1735	8003	1515	1218	331	2697
9	2115	1464	1828	1343	1607	7365	1527	896	287	2663
10	2262	1831	2155	1243	1573	7832	1748	1003	384	3017
11	2267	1612	2233	1762	1613	7849	1654	1108	321	3242
12	2517	1471	2059	1306	1690	9077	1861	1121	342	3425

TABLE 1: Hachemeister's claims data set.

Calculations:

$w_{1\Sigma} = 100155,\ w_{2\Sigma} = 19895,\ w_{3\Sigma} = 13735,\ w_{4\Sigma} = 4152,\ w_{5\Sigma} = 36110;$
$w_{\Sigma\Sigma} = 174047;$
$X_{1w} = 2060.921,\ X_{2w} = 1511.224,\ X_{3w} = 1805.843\ ,\ X_{4w} = 1352.976,\ X_{5w} = 1599.829;$
$X_{ww} = 1865.404;$
$\widetilde{s^2} = 139120026,\ \widetilde{a} = 89638.73;$
$Z_1 = 0.9847404,\ Z_2 = 0.9276352,\ Z_3 = 0.8984754,\ Z_4 = 0.7279092,\ Z_5 = 0.9587911.$
$Z_\Sigma = 4.497551,\ X_{zw} = 1683.713.$

Premiums estimated:

	state1	state2	state3	state4	state5
premium	2055.165	1523.706	1793.444	1442.967	1603.285

TABLE 2: premiums Bühlmann-Straub model

4.2 Quantile Credibility Model

Based on Bülmann's classical model assumptions, Georgios Pitselis developed *the quantile credibility model.* Let $X_{j1}, X_{j2}, ..., X_{jn_j}$ be the observed total claim amounts (or the total number of claims) in period i=1,2,.., n_j for contract j = 1,2,...,K, let Θ_j be an unobservable risk parameter that describes for contract j, and $\hat{\xi}_{pj}$ be the estimator of ξ_{pj}.

The model assumptions:
(i) Given $\Theta_j = \theta_j$, the observations $X_{j1}, ..., X_{jn_j}$ are conditionally independent with the same distribution function;
(ii) Θ_j is a random variable with distribution U;
(iii) $\Xi_p(\Theta_j) = E(\hat{\xi}_{pj}|\Theta_j)$;
(iv) $\nu_p = Var(\hat{\xi}_{pj}|\Theta_j)$.
The structural parameters are defined as follows:
$\Xi_p = E[\Xi_p(\Theta_j)]$, $s^2_{\xi_p} = E[\nu_p(\Theta_j)]$, $\psi_p = Var[\Xi_p(\Theta_j)]$.
Then the linear quantile credibility estimation with K contracts can be defined as:

$$\Xi_p^{Cred} = Z_{pj}\hat{\xi}_p + (1 - Z_{pj})\Xi_p \ ,$$

where

$$Z_{pj} = \frac{\psi_p}{E[\nu_p(\Theta_j)] + \psi_p} \ .$$

Proof Define a linear Bayes estimator of ξ_p as $g_j(c_0^j, c_{pl}^j, \hat{\xi}_{lp}) = c_0^j + \sum_{l=1}^{K} c_{lp}^j \hat{\xi}_{lp}$. The best estimator $g(X_j)$ minimizes the following objective function:

$$Q = E\left[\Xi_p(\Theta_j) - c_0^j - \sum_{l=1}^{K} c_{lp}^j \hat{\xi}_{lp}\right]^2 . \tag{4}$$

Taking the derivative of (4) with respect to c_0^j, c_{lp}^j[2], respectively

$$\begin{cases} \dfrac{dQ}{dc_0^j} = 2E\left[\left(\Xi_p(\Theta_j) - c_0^j - \sum\limits_{l=1}^{K} c_{lp}^j \hat{\xi}_{lp}\right)(-1)\right] \overset{set}{=} 0 \ , \\ \dfrac{dQ}{dc_{l'p}^j} = 2E\left[\left(\Xi_p(\Theta_j) - c_0^j - \sum\limits_{l=1}^{K} c_{lp}^j \hat{\xi}_{lp}\right)(-\hat{\xi}_{l'p})\right] \overset{set}{=} 0 \ . \end{cases}$$

Then

$$\begin{cases} E[\Xi_p(\Theta_j)] - c_0^j - \sum\limits_{l=1}^{K} c_{lp}^j E(\hat{\xi}_{lp}) = 0 \ , \quad (\star) \\ E[\Xi_p(\Theta_j)\hat{\xi}_{l'p}] - c_0^j E(\hat{\xi}_{l'p}) - E(\hat{\xi}_{l'p}) \sum\limits_{l=1}^{K} c_{lp}^j E(\hat{\xi}_{lp}) = 0 \ . \end{cases}$$

Multiply the equation $(\star)$ by $E(\hat{\xi}_{l'p})$:

$$\begin{cases} E[\Xi_p(\Theta_j)]E[\hat{\xi}_{l'p}] - c_0^j E(\hat{\xi}_{lp}) - \sum\limits_{l=1}^{K} c_{lp}^j E(\hat{\xi}_{pl}) E[\hat{\xi}_{l'p}] = 0 \ , \\ E[\Xi_p(\Theta_j)\hat{\xi}_{l'p}] - c_0^j E(\hat{\xi}_{l'p}) - E[\hat{\xi}_{l'p}] \sum\limits_{l=1}^{K} c_{lp}^j E(\hat{\xi}_{lp}) = 0 \ . \end{cases}$$

[2] In Pitselis' paper, he uses p' insteas of p, which could be a typo.

It immediately follows that:

$$E[\Xi_p(\Theta_j)]E[\hat{\xi}_{l'p}] - E[\Xi_p(\Theta_j)\hat{\xi}_{l'p}] = E[\hat{\xi}_{l'p}]\sum_{l=1}^{K} c_{lp}^{j} E(\hat{\xi}_{lp}) - \sum_{l=1}^{K} c_{lp}^{j} E(\hat{\xi_{pl}})E[\hat{\xi}_{l'p}]$$

$$Cov[\Xi_p(\Theta_j), \hat{\xi}_{l'p}] = \sum_{l=1}^{K} c_{lp}^{j} Cov[\hat{\xi}_{lp}, \hat{\xi}_{l'p}\ ] \ .$$

Therefore, if $l' = j$:

$$\begin{aligned}
Cov[\Xi_p(\Theta_j), \hat{\xi}_{jp}] &= c_{jp} Var[\hat{\xi}_{jp}] \\
c_{jp} &= \frac{Cov[\Xi_p(\Theta_j), \hat{\xi}_{jp}]}{Var[\hat{\xi}_{jp}]} \\
&= \frac{E\big[Cov[\Xi_p(\Theta_j), \hat{\xi}_{jp}|\Theta_j]\big] + Cov\big[E[\Xi_p(\Theta_j)|\Theta_j], E[\hat{\xi}_{jp}(\Theta_j)|\Theta_j]\big]}{E\big[Var(\hat{\xi}_{jp}|\Theta_j)\big] + Var\big[E(\hat{\xi}_{jp}|\Theta_j)\big]} \\
&= \frac{Var[\Xi_p(\Theta_j)]}{E[\nu_p(\Theta_j)] + Var[\Xi_p(\Theta_j)]} \\
&= \frac{\psi_p}{E[\nu_p(\Theta_j)] + \psi_p} \quad .
\end{aligned}$$

Parameter estimation:

$$\widehat{\Xi_p} = \bar{\xi}_{p.} = \frac{1}{K}\sum_{j=1}^{K} \hat{\xi}_{jp} \ ,$$

$$E\big[\widehat{\nu_p(\Theta_j)}\big] = \frac{1}{K}\sum_{j=1}^{K} \frac{\hat{\omega}_p(\Theta_j)}{n_j} \ ,$$

$$\hat{\psi}_p = \frac{1}{K-1}\sum_{j=1}^{K}(\hat{\xi}_{jp} - \bar{\xi}_{p.})^2 - \frac{1}{K}\sum_{j=1}^{K} \frac{\hat{\omega}_p(\Theta_j)}{n_j} \ .$$

Then the linear quantile credibility with K contracts can be estimated as:

$$\widehat{\Xi_{jp}}^{Cred} = \widehat{Z}_{jp}\hat{\xi}_{jp} + (1 - \widehat{Z}_{jp})\widehat{\Xi_p} \ ,$$

where

$$\begin{aligned}
Z_{jp} &= \frac{\widehat{\psi_p}}{E[\widehat{\nu_p(\Theta_j)}] + \widehat{\psi_p}} \ , \\
\widehat{\omega_p}(\Theta) &= \frac{n^2(y_{\lfloor np+l \rfloor} - y_{\lfloor np-l \rfloor})^2}{4Z_{1-\alpha/2}^2} \ , \\
l &= Z_{1-\alpha/2}\sqrt{np(1-p)}.
\end{aligned}$$

Note that $\lfloor * \rfloor$ is the integer part of *, α is the significance level.

Numeric Example
Apply this method to Hachemeister's claims data set. Here only premiums for the median (p=0.5) are presented. α is chosen to be 0.05, then $l = 3.394757$. A shortcoming of this approach, however, if p is below 0.366 (approximately), $y_{\lfloor np-l \rfloor}$ will be invalid, as $0 < np - l < 1$ because of small sample size, whereas order statistic starts at y_1.

Premiums estimated:

	state1	state2	state3	state4	state5
premium	1521.66	1681.396	1602.864	1730.946	1644.133

Table 3: Premiums Quantile credibility model

5 Regression Credibility Models

5.1 Hachemeister's Regression Model

Hachemeister forecasted average claim amounts for bodily injury claims classified by state in the USA. The claim data consists 12 periods, ranges from the third quarter of 1970 to the second quarter of 1973. Due to inflation these data is affected by time trends. In his model he regressed the claims averages for each state on explanatory variables: constant and trend(1:12), constructed a diagonal weighting matrix by putting claim frequencies in the diagonal, computed the regression coefficients using weighted least squares estimation. He estimated the credibility factor for the individual regression coefficients, then set premium as the predicted claims average for the next period.
The model is as follows:
let $Y_j = (Y_{j1}, Y_{j2}, \cdots, Y_{jn})'$ be an observation vector of the jth risk, in Hachemeister's regression model the entries are the claims averages of state j in quarter t; let $w_j = (\omega_{j1}, \omega_{j2}, \cdots, \omega_{in})'$ be the associated known weights, in this model ω_{jt} are the corresponding number of claims; let Θ_j be the set of all potential and possible values of the risk profile θ_j in the portfolio.
This model assumes that given Θ_j, X_j satisfies the regression equation:

$$Y_j = X_j \beta_j + \epsilon_j \ .$$

The model assumptions:
(i) $\Theta_1, \Theta_2, \cdots$ are identically distributed, the pairs $(\Theta_1, Y_1), \cdots$ are independent;
(ii)Given Θ_j, $Y_{j1}, Y_{j2}, \cdots$, are independent, and

$$\underbrace{E[Y_j|\Theta_j]}_{\text{n} \times 1} = \underbrace{X_j}_{\text{n}\times\text{K}} \underbrace{\beta(\Theta_j)}_{\text{K}\times 1} \ ,$$

where

$\beta(\Theta_j) = vector\ of\ unknown\ coeffients\ of\ length\ K(K \leq n),\ is\ considered\ as\ a\ random\ variable,\ its\ distribution\ is\ determined\ by\ the\ sturcture\ of\ the\ collective,$
$X_j = known\ fixed\ matrix\ of\ rank\ K;$

(iii)$Cov(X_j, X_j'|\Theta_j) = \sigma^2(\Theta_j)W_j^{-1}$.

The structural parameters:
$s^2 = E[\sigma^2(\Theta_j)],\ A = Cov(\beta(\Theta_j), \beta(\Theta_j)'),\ b = E[\beta(\Theta_j)]$.

The credibility estimator for $\beta(\Theta_j)$:
$\beta(\Theta_j)^{Cred} = Z_j\widehat{\beta}_j + (I - Z_j)b$, where $Z_j = A[A + s^2(X_j'W_jX_j)^{-1}]^{-1},\ \widehat{\beta}_j = (X_j'(W_j)^{-1}X_j)^{-1}X_j'W_jY_j$.

Parameter estimation:

$$\widehat{b} = (\sum_{j=1}^{J} Z_j)^{-1} \sum_{j=1}^{J} Z_j\widehat{\beta}_j;$$

$$\widehat{s^2} = \frac{1}{J}\sum_{j=1}^{J}\widehat{s_j^2},\ with\ \widehat{s_j^2} = \frac{1}{n-K}(Y_j - X_j\widehat{\beta}_j)'(Y_j - X_j\widehat{\beta}_j);$$

$$\widehat{A} = \frac{1}{J-1}\sum_{j=1}^{J} Z_j(\widehat{\beta}_j - \widehat{b})(\beta_j - \widehat{b})'.$$

Numeric Example
Apply this method to Hachmeister's claims data.

State Indiv.coef.: $\widehat{\beta}_j = (X_j'(W_j)^{-1}X_j)^{-1}X_j'W_jY_j$,

where $X = X_j = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 12 \end{pmatrix}$, $W_j = \begin{pmatrix} w_{j1} & 0 & \cdots & 0 \\ 0 & w_{j2} & \cdots & 0 \\ 0 & 0 & \cdots & w_{j,12} \end{pmatrix}$, $Y_j = \begin{pmatrix} L_{j1} \\ \vdots \\ L_{j,12} \end{pmatrix}$.

The results are shown in the table below:

	state1	state2	state3	state4	state5
intercept	1658.47243	1398.30252	1532.99872	1176.70407	1521.89933
slope coefficient	62.39246	17.13975	43.30732	27.80702	11.87448

Table 4: Regression coefficients Hachemeister's regression model

All the slope coefficients are positive, then there indeed exists a long-term increase in the data, which can also be seen from the regression lines (upward sloping) shown in *figure 6*.

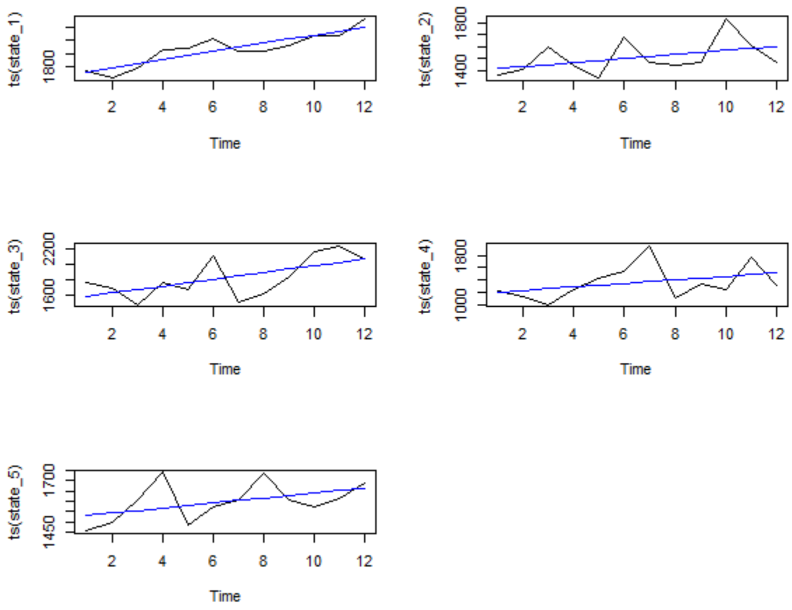


FIGURE 6: Weighted least square regressions

Within state variance: $\widehat{s^2} = \dfrac{1}{5}\sum_{j=1}^{5}\widehat{s_j^2} = 49870187,$

$where\ \widehat{s_1^2} = 121262869, \widehat{s_2^2} = 30174010, \widehat{s_3^2} = 52483869, \widehat{s_4^2} = 24359005, \widehat{s_5^2} = 21071182;$

State Adj.coef.:

	state1	state2	state3	state4	state5
intercept	1693.52313	1373.02958	1545.36429	1314.54855	1417.40928
slope coef.	57.17147	21.34641	40.61014	14.80935	26.30721

TABLE 5: Regression coefficients Hachemeister's regression model

Finally, Premium $= Adj.\ intercept + 13 \times slope\ coef.$:

	state1	state2	state3	state4	state5
premium	2436.75	1650.53	2073.30	1507.07	1759.40

TABLE 6: Premiums Hachemeister's regression model

5.2 Quantile Regression Credibility Model

This section shows how quantiles could be incorporated into Hachemeister's model. Here only unweighted case will be considered. Assume that there are J contracts, for each contract n years of claims experience, or other characteristics. Let $Q_P(y_{jt}|\Theta_j)(0 < p < 1)$ be the conditional quantile of y_{jt} corresponding to the unobservable random risk parameter Θ_j and $Y_j = (y_{j1}, ..., y_{jn})'$ an observable vector of risks.

Further assumptions:
(i) $\Theta_1, \Theta_2, \cdots$ are identically distributed, the pairs $(\Theta_1, Y_1), \cdots$ are independent;
(ii) Given Θ_j,

$$\underbrace{Q_p(Y_j|\Theta_j)}_{\text{n} \times 1} = \underbrace{X_j}_{\text{n}\times\text{K}} \underbrace{\beta_p(\Theta_j)}_{\text{K}\times 1},$$

where

$\beta_p(\Theta_j) = vector\ of\ unknown\ coefficients\ of\ length\ K(K \leq n)\ for\ pth\ quantile,\ is\ considered\ as\ a\ r$
$variable,\ its\ distribution\ is\ determined\ by\ the\ structure\ of\ the\ collective;$
$X_j = known\ fixed\ matrix\ of\ rank\ K;$
$Cov(\widehat{\beta_{pj}}, \widehat{\beta_{pj}}'|\Theta_j) = \sigma^2_{\xi_p}(\Theta_j)(X_j'X_j)^{-1}.$

The structural parameters:
$s^2_{\xi_p} = E[\sigma^2_{\xi_p}(\Theta_j)],\ A_p = Cov(\beta_p(\Theta_j), \beta_p(\Theta_j)'),\ \beta_p = E[\beta_p(\Theta_j)].$

The credibility estimator for $\beta(\Theta_j)$:

$$\beta_{pj}(\Theta_j)^{Cred} = Z_{pj}\widehat{\beta_{pj}} + (I - Z_{pj})\beta_p,$$

where
$Z_{pj} = A_p[A_p + s^2_{\xi_p}(X_j'X_j)^{-1}]^{-1}, \quad \widehat{\beta_{pj}} = (X_j'X_j)^{-1}X_j'Y_{pj}$.

Z_{pj} is solution to the following minimization problem:

$$\begin{aligned} Q &= E\Bigg([\beta_p(\Theta_j) - \widehat{B_{pj}}^{Cred}]'[\beta_p(\Theta_j) - \widehat{B_{pj}}^{Cred}]\Bigg) \\ &= E\Bigg([\beta_p(\Theta_j) - \beta_p - Z_{pj}(\widehat{\beta_{pj}} - \beta_p)]'[\beta_p(\Theta_j) - \beta_p - Z_{pj}(\widehat{\beta_{pj}} - \beta_p)]\Bigg) \end{aligned}$$

Let $\beta^0_p(\Theta_j) = \beta_p(\Theta_j) - \beta_p,\ \beta^0_{pj} = \widehat{\beta}_{pj} - \beta_p$, then the above equation becomes:

$$\begin{aligned} Q &= E\Bigg([\beta^0_p(\Theta_j) - Z_{pj}\beta^0_{pj}]'[\beta^0_p(\Theta_j) - Z_{pj}\beta^0_{pj}]\Bigg) \\ &= E\Bigg(\beta^0_p(\Theta_j)'\beta^0_p(\Theta_j) + (\beta^0_{pj})'(Z_{pj})'Z_{pj}\beta^0_{pj} - (\beta^0_p(\Theta_j))'Z_{pj}\beta^0_{pj} - (Z_{pj}\beta^0_{pj})'\beta^0_p(\Theta_j)\Bigg) \end{aligned}$$

Using the product rule and differentiating with respect to the matrix Z_{pj},

$$\frac{dQ}{dZ_{pj}} = -2E[\beta^0_p(\Theta_j)(\beta^0_{pj})' - Z_{pj}\beta^0_{pj}(\beta^0_{pj})'] \overset{set}{=} 0$$

as

$$\begin{aligned}
&E[\beta_p^0(\Theta_j)(\beta_{pj}^0)' - Z_{pj}\beta_{pj}^0(\beta_{pj}^0)'] \\
&= E\bigg([\beta_p(\Theta_j) - \beta_p][\beta_p(\Theta_j) - \beta_p]' - Z_{pj}[\widehat{\beta}_{pj} - \beta_p][\widehat{\beta}_{pj} - \beta_p]'\bigg) \\
&= E\bigg([\beta_p(\Theta_j) - \beta_p][\beta_p(\Theta_j) - \beta_p]'\bigg) - E\bigg(Z_{pj}[\widehat{\beta}_{pj} - \beta_p][\widehat{\beta}_{pj} - \beta_p]'\bigg) \\
&= Cov(\beta_p(\Theta_j)) - Z_{pj}Cov(\widehat{\beta}_{pj}) \\
&= Cov(\beta_p(\Theta_j)) - Z_{pj}\bigg(E(Cov[\widehat{\beta}_{pj}|\Theta_j]) + cov(E[\widehat{\beta}_{pj}|\Theta_j])\bigg)
\end{aligned}$$

then $Z_{pj} = Cov(\beta_p(\Theta_j))\bigg(E(Cov[\widehat{\beta}_{pj}|\Theta_j]) + cov(E[\widehat{\beta}_{pj}|\Theta_j])\bigg)^{-1}$

Parameter estimation:

$$\begin{aligned}
&\widehat{\beta_{pj}} = \underset{\beta_p}{\text{argmin}}\left(\sum_{i:y_i \geq b} p|y_{ji} - x_{ji}'\beta_j| + \sum_{i:y_j < b}(1-p)|y_{ji} - x_{ji}'\beta_j|\right); \\
&\widehat{\beta}_p = \frac{1}{J}\sum_{j=1}^{J}\widehat{\beta_{pj}}; \\
&\widehat{A_p} = \frac{1}{J-1}\sum_{j=1}^{J}(\widehat{\beta}_{pj} - \widehat{\beta}_p)(\widehat{\beta}_{pj} - \widehat{\beta}_p)' - \frac{1}{J}\sum_{j=1}^{J}\widehat{\sigma_{\xi_p}^2}(\Theta_j)(X_j'X_j)^{-1}; \\
&Cov(\widehat{\beta_{pj}}|\Theta_j) = \sigma_{\xi_p}^2(\Theta_j)(X_j'X_j)^{-1} = \frac{p(1-p)}{n(f_{\xi_p})^2}(X_j'X_j)^{-1}.
\end{aligned}$$

$\sigma_{\xi_p}^2(\Theta_j)$ is estimated using order statistic estimation:

$$\begin{aligned}
&\widehat{\sigma_{\xi_p}^2} = \frac{1}{J}\sum_{j=1}^{J}\frac{\widehat{\omega_p(\Theta_j)}}{n}, \\
&\textit{where} \\
&\widehat{\omega_p(\Theta_j)} = \frac{n^2(y_{\lfloor np+l\rfloor} - y_{\lfloor np-l\rfloor})^2}{4Z_{1-\alpha/2}^2}, l = Z_{1-\alpha/2}\sqrt{np(1-p)}.
\end{aligned}$$

Numeric Example

Apply this method to Hachmeister's claims data.
Premiums estimated:

	state1	state2	state3	state4	state5
premium	2098.915	1684.715	1960.715	1618.415	1762.39

Table 7: Premiums Quantile regression credibility model

6 Effect of outliers on the models

The premiums estimated in previous sections are summarized in *Table 8*. It is noticeable from the summary statistics in the table that for each state premium estimated using Hachemeister's regression model is greater than using Bühlman-Straub model, which is due to the fact that Hachemeister's regression model considers long-term increas in next period, while Bühlman-Straub model doesnot.

	premium1	premium2	premium3	premium4	premium5
Bühlman-Straub	2055.17	1523.71	1793.44	1442.97	1603.29
Hachemeister regression	2436.75	1650.53	2073.30	1507.07	1759.40
Quantile(0.5) credi.	1521.66	1681.40	1602.87	1730.95	1644.13
Quantile regression credi.(0.5)	2098.92	1684.72	1960.72	1618.42	1762.39

TABLE 8: Summary premiums.

In order to see the effect of outliers on these four models, Hachemeister's claims data is revised in two steps. The results are presented in *Table 9*.

(A) Increase the last observation of state1 to 100000

	premium1	premium2	premium3	premium4	premium5
Bühlman-Straub	10895.74	1511.25	1805.87	1353.10	1599.84
Hachemeister regression	-	-	-	-	-

(B) Increase the last observation of state1, state2 and state3 to 100000

	premium1	premium2	premium3	premium4	premium5
Bühlman-Straub	10895.75	10727.69	9799.37	1353.31	1599.87
Hachemeister regression	-	-	-	-	-
Quantile(0.5)	1477.74	1698.73	1590.08	1767.28	1647.18

TABLE 9: Summary premiums for revised data.

First, increase the last observation of state 1 to 100000. The results indicate that the most significant change is in the premium estimated for state1 using the Bühlmann-Straub model. To point out, $\tilde{a}$ in the Bühlmann-Straub model is negative in this case, so set $\tilde{a} = (J-1) \times \tilde{s}^2$. Furthermore, the Hachemeister's regression model is invalid under this circumstance, because the design matrix is not invertible, hence it cannot be used to develop a regression model. Note that the premiums estimated using Quantile credibility model and Quantile regression credibility model are omitted in the table, as there is no changes to these premiums.

Next, also increase the last observation of state 2 and state 3 to 100000. Similarly, the premiums estimated for state 1, state2 and state3 are largely affected using the Bühlmann-Straub model. Moreover, $\tilde{a}$ in the Bühlmann-Straub model is negative and the Hachemeister's regression model remains invalid. What is particlarly interersting is that there is just a small increase to premiums estimated using Quantile credibility model, but no changes to the premiums estimated using Quantile regression credibility model. It can be clear from the results that outliers have less impact on Quantile credibility model and Quantile credibility regression model than on the Bühlmann-Straub model and the Hachemeister's regression model.

7 Conclusion

Despite the fact that the models are applied to Hachemeister's data set which is different from Pitselis' paper, similar conclusions can be drawn. They are summarized as follows. Firstly, incorporating with quantiles enables to estimate premiums as well as changes in these premiums at different points of the claims distributions. Secondly, Quantile credibility model and Quantile credibility regression model are less sensitive to outlying data. Finally, as the claims distributions are in general heavy-tailed and skewed to the right, quantile estimation is more desirable than least squares estimation in the context of the insurance industry. However, it is important to note that there is a shortcoming in the Hachemeister's data set. That is, the size of Hachemeister's data set is relatively small, which could influence the impact of the outliers on the models. Therefore, further researches such as applying these models to data with a larger sample size are recommended.

References

[1] Bühlmann, H., Gisler, A. (2005). *A course in Credibility Theory and its Applications.* Heidelberg: Springer.

[2] Kaas, R. and et al (2001). *Modern Actuarial Risk Theory.* Dordrecht: Kluwer Academic publishers.

[3] Koenker, R. and Bassett, G. (1978). *Regression Quantiles.* Econometrica, 46, 33-50.

[4] Greene, W.H. (2012). *Econometric Analysis.* 7th ed. Harlow: Pearson Education Limited.

[5] Pitselis, G. (2013). Quantile credibility models. *Insurance: Mathematics and Economics*, 52, 477-489.