

An application of parametric quantile regression to extend the two-stage quantile regression for ratemaking

Fabio Baione & Davide Biancalana

To cite this article: Fabio Baione & Davide Biancalana (2020): An application of parametric quantile regression to extend the two-stage quantile regression for ratemaking, Scandinavian Actuarial Journal, DOI: [10.1080/03461238.2020.1820372](https://doi.org/10.1080/03461238.2020.1820372)

To link to this article: <https://doi.org/10.1080/03461238.2020.1820372>



Published online: 21 Sep 2020.



Submit your article to this journal [↗](#)



Article views: 30




View related articles [↗](#)



View Crossmark data [↗](#)



An application of parametric quantile regression to extend the two-stage quantile regression for ratemaking

Fabio Baione ^a and Davide Biancalana ^b

^aDepartment of Statistics, Sapienza University of Rome, Rome, Italy; ^bDepartment of Methods and Models for Economy Finance and Territory, Sapienza University of Rome, Rome, Italy

ABSTRACT

This paper deals with the use of parametric quantile regression for the calculation of a loaded premium, based on a quantile measure, corresponding to individual insurance risk. Heras et al. have recently introduced a ratemaking process based on a two-stage quantile regression model. In the first stage, a probability to have at least one claim is estimated by a GLM logit, whereas in the second stage several quantile regressions are necessary for the estimate of the severity component. The number of quantile regressions to be performed is equal to the number of risk classes selected for ratemaking. In the actuarial context, when a large number of risk classes are considered (e.g. in Motor Third Party Liability), such approach can imply an over-parameterization and time-consuming. To this aim, in the second stage, we suggest to apply a more parsimonious approach based on Parametric Quantile Regression as introduced by Frumento and Bottai and never used in the actuarial context. This more conservative approach allows you not to lose efficiency in the estimation of premiums compared to the traditional Quantile Regression.

ARTICLE HISTORY

Received 12 January 2020
Accepted 3 September 2020

KEYWORDS

Quantile regression; premium principles; risk margin; ratemaking; quantile regression coefficients modeling

1. Introduction

The interest in the application of quantile regression (QR) in actuarial sciences is growing rapidly. To the best of our knowledge, QR applications in the field of insurance may be found in Pitselis (2009), where it is applied in the measurement of the adequacy of own funds a company requires in order to remain healthy and avoid insolvency in a Solvency 2 context. Dong et al. (2015) use a parametric and a non-parametric QR model in order to derive risk margin for loss reserves and to evaluate solvency capital requirement in non-life insurance applications. Portnoy (2003) applied QR to mortality table rates, whereas Pitt (2006) used censored QR to estimate claim early termination rates for income protection insurance. However, the broader application of QR in the actuarial context refers to insurance ratemaking, which has been firstly discussed by Kudryavtsev (2009) and then used in the working paper of Wolny-Dominiak et al. (2012). Fuzi et al. (2016) apply a Bayesian QR model for claim count data in motor insurance to study the effects of change in the estimates of rating factors on the magnitude of the response variable. More recently, an extension of the original approach of Kudryavtsev has been proposed in this journal by Heras et al. (2018) where an application of a Two-Stage QR (TSQR henceforth) is considered. A similar approach has been introduced by Biancalana (2017) and Baione & Biancalana (2019). Another contribution to QR in insurance is found in Baione et al. (2019), where a study of the diversification effect involved in a portfolio of non-life insurance is computed when

individual loaded premiums are calculated via QR. All the previously mentioned insurance ratemaking papers suggest to estimate the quantiles of aggregate claim amount distributions by decomposing the cost of a claim between a binary variable that indicates whether the policy has claimed, and a continuous variable showing the aggregate claim amount. In actuarial literature such decomposition method is commonly called Two-Part model (Frees 2010, Frees et al. 2013). Anyway, compared to the traditional Two-Part model, the use of QR allows a direct estimate of the loaded premium, i.e. including the safety loading. However, all the previously stated models for insurance ratemaking drive some practical drawbacks:

- Kudryavtsev (2009) does not consider a different claim frequency for each risk class, but it is assumed constant for each insured individual. This is the significant issue because, as well known, the probability of having claims is considerably affected by the features of each insured individual.
- Heras et al. (2018) solve the previous issue by introducing a TSQR model; the first stage involves fitting a logistic regression to estimate, for every policy, the probability of submitting at least one claim, then a QR is used to estimate the individual aggregate claim amount. The approach is used for the application of a Quantile Premium Principle (henceforth QPP), where the premium safety loading is assessed for each risk class using a single probability level of the aggregate claim amount. However, this approach, although if theoretically correct, is questionable from a practical point of view. Its application requires the calibration of a number of QRs on different probability levels equal to the number of risk classes used for the frequency model which, in large portfolios, could be equal to thousands or even millions. Therefore, this approach may be too onerous in terms of model validation, as well as an over parameterization, hence computationally demanding.
- Baione & Biancalana (2019) have suggested an alternative TSQR premium model where the safety loading is calculated on the claim severity. The calibration of the premium safety loading is assessed by means of a quantile approach using a single probability level for each risk class. Thereby, the need to calibrate different QRs is prevented avoiding the subsequent overparameterization. However, by fixing a single probability level on the claim severity it is not possible to obtain a unique probability level of the aggregate claim amount for each risk class as in Heras et al.

The aim of the present work is to introduce a new approach that extends the framework introduced by Heras et al. in order to provide a parsimonious model to calculate the QPP. For this purpose, we suggest to adopt a parametric model of quantile regression (PQR henceforth) recently introduced by Frumento & Bottai (2016), also referred to by the authors as quantile regression coefficients modeling (*qrcm*); we name this approach as Two Stage Parametric Quantile Regression model (TSPQR henceforth). Compared to the traditional QR, where quantiles are estimated one at a time, PQR describes the functional form of the regression coefficients parametrically depending on the order of the quantile. Therefore, this approach introduces a new class of estimators that can be applied to relevant practical problems. As declared by Frumento and Bottai, PQR facilitates estimation and inference, simplifies the interpretation of the results, and generates more efficient estimators. PQR introduces direct modeling of the coefficients function that has never been discussed in literature and enriches the traditional QR theory with a new class of estimators that may have advantages in terms of parsimony, efficiency, and may expand the potential of statistical modeling. This technique is relatively recent and its application is limited in the field of medicine, but as far as we know, it has never been discussed in actuarial and financial fields. The PQR has also been extended to censored and truncated data in Frumento & Bottai (2017). The described estimator has been implemented in the R package *qrcm* (Frumento 2017).

The rest of the paper is arranged as follows. Section 2 provides a brief description of the standard quantile regression model. Section 3 describes the parametric modeling of the quantile regression,

which was introduced by Frumento and Bottai and the mathematical details of TSPQR. A numerical application of TSPQR for Quantile Premium Principle (QPP) is discussed in Section 4. Finally, Section 5 provides the conclusion of this paper.

2. Premium estimation via parametric quantile regression

2.1. Basic elements of quantile regression

Quantile regression techniques (Koenker & Bassett 1978) enable us to define the statistical relationship between the conditional quantile of a response variable Y and a row-vector of independent covariates, $\mathbf{x}_i = (x_{i1}, \dots, x_{im}); 1 \leq i \leq I$. According to Koenker & Bassett (1978), Koenker & Hallock (2001), and Koenker (2005), the relationships between the θ -quantile of the outcome and a set of explanatory variables is given by

$$Q_\theta [Y|\mathbf{x}_i] = Q_\theta [Y_i] = \mathbf{x}_i^T \beta_\theta = \beta_{0,\theta} + \beta_{1,\theta} \cdot x_{i1} + \dots + \beta_{m,\theta} \cdot x_{im}, \quad (1)$$

where β_θ is the corresponding column-vector of regression parameters estimated related to the selected θ th quantile. In statistical practice, the distribution-free approach is often used for estimation (see Koenker & Bassett 1978, Koenker & Hallock 2001, Koenker 2005). In this framework, the estimator $\hat{\beta}_\theta$, of vector β_θ from Equation (1) could be obtained by solving the following minimization problem:

$$\min_{\beta_\theta \in \mathbb{R}^m} \left\{ \sum_{i: y_i \geq \mathbf{x}_i \beta_\theta} \theta |y_i - \mathbf{x}_i \beta_\theta| + \sum_{k: y_k < \mathbf{x}_k \beta_\theta} (1 - \theta) |y_k - \mathbf{x}_k \beta_\theta| \right\}. \quad (2)$$

Alternatively, the minimization of (2) can be performed using an equivalent linear program (Hao & Naiman 2007), whereas the confidence intervals for β_θ can be obtained using the bootstrap method. QR can be easily fitted using the 'R' package `quantreg` (Koenker 2015). It is worth noting that the estimate of the conditional θ th quantile of the outcome variable is based on the estimation of the vector of coefficients β_θ depending on the order of the quantile being estimated. For example, the coefficients for the 50th quantile (median) are generally different from those of the 75th quantile (third quartile). This implies that if we want to investigate the density of a quantile level vector, it is necessary to run several times the QR with reference to each element of a quantile level vector, dependent on the granularity of the analysis required (e.g. using 99 centiles $\theta = 0.01, 0.02, \dots, 0.99$). Such valuation is also relevant in order to discuss the change in the behavior of the regression parameters with respect to a change in the probability level.

To this aim, Figure 1 shows an example taken from Baione & Biancalana (2019) based on a real database from an Australian automobile insurance company (De Jong & Heller 2008), which we will use in the following section for a numerical application. The graph represents the male gender regression parameter which is always positive and shows an increasing trend with the order of the quantile. This means that the estimated conditional quantile for males is greater than that of females; the latter is considered as being base level. Such difference increases as the probability level increases. Anyway, as observable, the underlying trend is not smoothed and it is not uniquely justifiable by random variability.

For example, as observable in Figure 1(a), the estimated difference between females and males is 3.6% at the 62th centile; the value increases to a peak of 9.6% at the 72th centile (orange point) and drops down to 5.4% at the 73th centile.

This means that we have an increasing and significant difference between the two genders at the 72th centile, but not at the 73th with a drop by about 43.8% compared to the previous value, as shown in Figure 1(b).

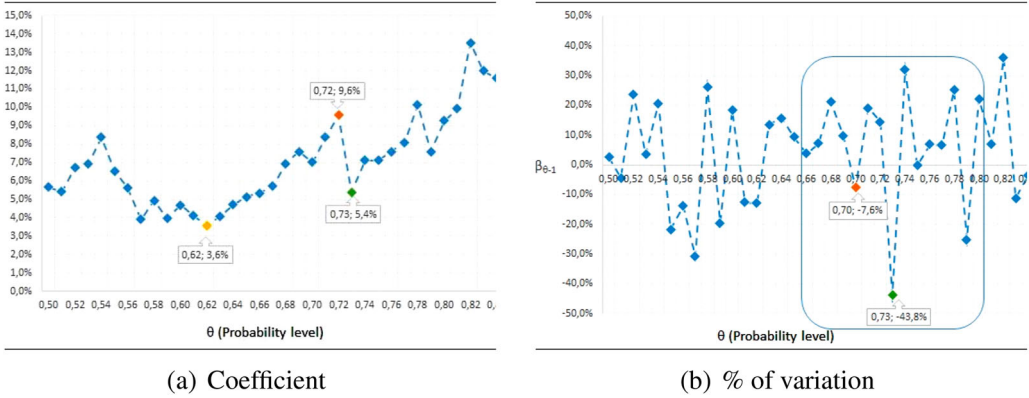


Figure 1. QR parameter estimate β_θ . (a) Coefficient. (b) % of variation.

2.2. An introduction to parametric quantile regression

To overcome the aforementioned drawbacks, it is possible to adopt a parametric approach for the quantile regression coefficients modeling introduced by Frumento & Bottai (2016). The proposed approach allows to model the regression coefficients as parametric functions of the order of the quantile θ , that depends on a finite-dimensional parameter γ of the form:

$$\beta_j(\theta, \gamma_j) = \gamma_{j0} + \gamma_{j1} \cdot b_1(\theta) + \dots + \gamma_{jn} \cdot b_n(\theta), \quad 0 \leq j \leq m, \quad (3)$$

where $b_1(\theta) \dots b_n(\theta)$ are known functions of θ and $b_0(\theta)$ is usually set equal to 1. It is worth noting that the model coefficients $\beta_j(\theta, \gamma_j)$ describe the effect of covariates on the θ th quantile of the response variable. A linear model similar to the Equation (1) is used to express the conditional quantile function as follows:

$$Q_\theta [Y | \mathbf{x}_i, \Gamma] = \mathbf{x}_i^T \Gamma \mathbf{b}(\theta) = \beta_0(\theta, \gamma_0) + \beta_1(\theta, \gamma_1) \cdot x_{i1} + \dots + \beta_m(\theta, \gamma_m) \cdot x_{im}. \quad (4)$$

Therefore, the entire conditional quantile function $Q_\theta [Y | \mathbf{x}_i, \Gamma]$ is described by a $m \times n$ matrix Γ of parameters.

Then, a relevant aspect for ratemaking purposes is that such a technique enables the estimation of the conditional quantile of the response variable, given a set of covariates, for each probability level $\theta \in (0, 1)$ in a single run.

By way of example, assuming the univariate model:

- $Q_\theta [Y | \mathbf{x}_i, \Gamma] = \beta_0(\theta, \gamma_0) + \beta_1(\theta, \gamma_1) \cdot x_{i1}$
 assuming:
 - $\beta_0(\theta, \gamma_0) = \gamma_{00} + \gamma_{01} \cdot 2\theta$,
 - $\beta_1(\theta, \gamma_1) = \gamma_{10} + \gamma_{11} \cdot 2\theta$,

then $b_0(\theta) = 1$, $b_1(\theta) = 2\theta$ and $\Gamma = \begin{bmatrix} \gamma_{00} & \gamma_{01} \\ \gamma_{10} & \gamma_{11} \end{bmatrix}$.

In this example, the conditional quantile of the response variable, for each $\theta \in (0, 1)$, is completely defined by the estimate of 2×2 matrix Γ .

In practice, several sets of $\mathbf{b}(\theta)$ functions could be used (e.g. polynomials, trigonometric, splines), with the only requirement that their combination must induce a θ -monotonically increasing quantile function $Q_\theta [Y | \mathbf{x}_i, \Gamma]$ for some Γ .

For example, $\mathbf{b}(\theta)$ cannot be formed by a single non-monotonic function like $b_1(\theta) = \sin(2\pi\theta)$, because this would cause $Q_\theta [Y | \mathbf{x}_i, \Gamma]$ to be a non-monotonic function of θ at any Γ .

However, the combination $b_0(\theta) = 1, b_1(\theta) = \sin(2\pi\theta), b_2(\theta) = \theta$ is an eligible basis, since

$$Q_\theta [Y | \gamma] = \gamma_0 + \gamma_1 \cdot \sin(2\pi\theta) + \gamma_2 \cdot \theta \quad (5)$$

is a well-defined monotonic quantile function for $\gamma_2 \geq 2\pi|\gamma_1|$. The parameter space will always be a subset of $\mathbb{R}^{m \times n}$, even in very simple models.

The estimate of θ th parametric quantile regression coefficients under model (4) is similar to the standard quantile regression and requires the minimization of the following loss function:

$$L_I(\theta, \Gamma) = I^{-1} \sum_{i=1}^I \left[\theta - \mathbb{I}(y_i \leq \mathbf{x}_i^T \Gamma \mathbf{b}(\theta)) \right] \cdot \left[y_i - \mathbf{x}_i^T \Gamma \mathbf{b}(\theta) \right]. \quad (6)$$

Instead of estimating a discrete set of quantiles, it is possible to estimate the entire quantile process in one time. Following Frumento & Bottai (2016), the estimation of Γ is obtained by the minimization of the following integrated loss function:

$$\bar{L}_I(\Gamma) = \int_0^1 L_I(\theta, \Gamma) d\theta. \quad (7)$$

This approach is defined as integrated loss minimization (ILM). The idea behind ILM is that function (7) can be seen as an average loss function obtained by summing functions (6) by varying θ over the interval $(0, 1)$. The integrated loss (7) can be further generalized by including a nonconstant weight $w(\theta)$. For example, the integral can be calculated over (θ_1, θ_2) rather than in $(0, 1)$ by setting $w(\theta) = \mathbb{I}(\theta_1 \leq \theta \leq \theta_2)$. Note that $\theta_1 = \theta_2 = \theta$ corresponds to ordinary quantile regression of order θ .

The ILM estimator $\hat{\Gamma}_I$ that minimizes the integrated loss function (7), based on a sample of size I , is unbiased. In addition, it has relevant asymptotic properties like consistency and asymptotic normality depending on the structure of $\mathbf{b}(\theta)$. For instance, for the consistency, $\mathbf{b}(\theta)$ should induce a well-defined monotonic quantile function as stated above; and guarantee that the inverse of the quantile function (i.e. the cumulative distribution function) is a continuous function of Γ . Some of the $\mathbf{b}(\theta)$ may be unbounded at $\theta = 0$ or $\theta = 1$, but the functions $B(\theta) = \int_0^\theta \mathbf{b}(u) du$ and $\bar{B} = \int_0^1 B(u) du$ must be finite. Moreover, the first derivative of $\mathbf{b}(\theta)$ must be defined everywhere, although it might be infinite in the extremes.

For instance, valid functions having the above-stated properties are of the form $\theta^\delta, \log(\theta), \log(1 - \theta), \delta^\theta$, the quantile function of any distribution with finite moments, splines, or a combination of the above. For the asymptotic normality of the ILM estimator the reader can refer to Theorem 2 in Frumento & Bottai (2016).

3. Two-stage quantile regression and two-stage parametric quantile regression model for premium estimation

The Two-Stage Quantile Regression model, introduced by Heras et al. (2018), may be considered an extension of the model proposed by Kudryavtsev (2009), who was the first to apply QR to ratemaking in non-life insurance. The starting point is the decomposition of the total claim amount per insured S_i between a binary variable, which indicates whether the policy has at least one claim, and a continuous variable representing the total amount of the claims. The former is an indicator random variable (r.v.) \mathbb{I}_{N_i} indicating whether the i th insured has had at least one claim; it is generally assumed Bernoulli (p_i) distributed and independent from the second strictly positive truncated or conditional r.v. representing the total loss per i th claimant or hereafter simply denoted as severity ($S_i | N_i > 0$) = \hat{S}_i . In addition, the work of Heras et al. also proposes a premium principle based on QR, the so-called

Quantile Premium Principle (QPP), to premium safety loading calculation. The QPP is defined as follows:

$$P_i = \alpha \cdot Q_{\bar{\theta}}[S_i] + (1 - \alpha) \cdot E[S_i], \quad 0 \leq \alpha \leq 1, \quad (8)$$

where the individual premium is a weighted average of the expected value and the quantile to a level $\bar{\theta}$ of S_i . The parameter $\bar{\theta}$ set the unique probability level to define the conditional quantile of the r.v. total claim amount per insured individual. Hence, in order to apply QPP it is necessary to estimate $Q_{\bar{\theta}}[S_i]$.

Then, given an insurance portfolio containing I policyholders ($i = 1, \dots, I$) and a set of independent covariates $\mathbf{x}_i = (x_{i1}, \dots, x_{im})$ for each i , the first stage consists in the estimate of $p_i = E[\mathbb{I}_{N_i} = 0 | \mathbf{x}_i]$, i.e. the probability that policy i makes no claims, by means of a logistic regression model with a logit link function.

For the second stage, starting from the decomposition of the cumulative density function:

$$F_{S_i}(Q_{\bar{\theta}}[S_i | \mathbf{x}_i]) = p_i + (1 - p_i) \cdot F_{\tilde{S}_i}(Q_{\bar{\theta}}[S_i | \mathbf{x}_i]) \quad (9)$$

by setting $\theta_i^* = F_{\tilde{S}_i}(Q_{\bar{\theta}}[S_i | \mathbf{x}_i])$, $i = 1, \dots, I$, as the probability level of the conditional severity \tilde{S}_i , we obtain the following equation:

$$\theta_i^* = \frac{\bar{\theta} - p_i}{1 - p_i}, \quad i = 1, \dots, I. \quad (10)$$

Then from Equations (10) and (9) we have

$$Q_{\bar{\theta}}[S_i | \mathbf{x}_i] = F_{\tilde{S}_i}^{-1}\left(\frac{\bar{\theta} - p_i}{1 - p_i} \middle| \mathbf{x}_i\right) = Q_{\theta_i^*}[\tilde{S}_i | \mathbf{x}_i]. \quad (11)$$

This is a very important relationship, as it permits to estimate the quantile of S_i , working on the continuous r.v. \tilde{S}_i , simply changing the quantile level from $\bar{\theta}$ to θ_i^* , $i = 1, \dots, I$.¹

In a nutshell, the first stage is necessary to determine the I probability levels to be adopted in the second stage for the estimation of the conditional quantiles of the severity.

In TSQR the second stage consists in the application of a traditional QR model. However, if we assume that each insured i is characterized by a specific p_i , to estimate the conditional quantile for each $i = 1, \dots, I$, it is necessary to run a set I of QR models (reduced to H for the allocation in homogenous risk classes), with each one calibrated at a different probability level θ_i^* associated with the relative i th risk profile. Besides, in order to calibrate the TSQR, it is necessary to perform all the diagnostic analyses to validate the model, such as the test for significance of covariates, the test on parameters, the fitting analysis and other diagnostic measures. Furthermore, in the actuarial applications, the ratemaking process involves a large number of rating factors and, as a consequence, a large number of risk classes.

By way of example, assuming a dataset containing 10 dichotomous rating factors, the number of risk classes is $H = 2^{10} = 1,024$; hence, in order to apply the TSQR, it is necessary to calibrate 1,024 quantile regressions, and then analyze the regression diagnostic of $11 \cdot 1,024 = 11,264$ parameters estimated in total. For these reasons, in a real context, the use of the TSQR is questionable from a practical point of view.

So, instead of performing 1,024 QRs, we suggest to substitute the second stage of the TSQR with a PQR approach, which we can define as Two-Stage Parametric Quantile Regression. Hence, we can

¹ It is worth noting that depending on the set of covariates, it is possible to assign the single policyholder a specific risk class generated by the combination of covariates levels.

calculate the conditional quantile of the aggregate claim amount per the i th policy as follows:

$$Q_{\tilde{\theta}}[S_i|\mathbf{x}_i] = Q_{\theta_i^*}[\tilde{S}_i|\mathbf{x}_i] = \mathbf{x}_i^T \Gamma \mathbf{b}(\theta_i^*). \quad (12)$$

In this way, assuming in the previous example $b_0(\theta) = 1$, $b_1(\theta) = 2\theta$, the TSPQR requires a single calibration with an estimate of $11 \cdot 2 = 22$ parameters in one shot, compared to the 11, 264 of the TSQR approach in 1024 separate QRs. It is worth noting that this difference increases as the number of risk classes increases.

4. Numerical investigation

In order to perform a comparison between the TSQR and the TSPQR, we apply the proposed approaches to a real insurer database. Our aim is to estimate the loaded premium by QPP in Equation (8), analyzing the ability of the TSPQR to predict the loaded premium compared to the TSQR model.

4.1. Data set and notation

We set our model on a database from an Australian automobile insurance company (De Jong & Heller 2008) between years 2004 and 2005. The same data set has been used in the numerical applications in Heras et al. (2018) and Baione & Biancalana (2019). The portfolio has $I = 67, 856$ policies, with 4624 policies with at least one claim. The $K = 5$ rating factors in this data set are listed in Table 1, where m_k indicates the number of levels for each rating factor.

Hereafter, we assume the following notation. Let:

- h index the number of the H risk classes;
- $r = \sum_{h=1}^H g_h$ be the total number of policyholders where g_h denotes the number of policyholders belonging to h th risk class;
- i_h index the policyholder belonging to h th risk class;
- \mathbf{x}_h be the row vector of the design matrix, providing information about exposure rating factors of the generic i_h policyholder belonging to the same risk class h .

The loss generated by the i_h th policyholder belonging to h th risk class is modeled by means of the generic non-negative i.i.d. r.v. $S_{i_h} = S_h$. The total or aggregate loss of the portfolio is represented by the r.v.:

$$L = \sum_{h=1}^H L_h = \sum_{h=1}^H \sum_{i=1}^{g_h} S_{i_h} = \sum_{h=1}^H g_h \cdot S_h, \quad (13)$$

where L_h the non-negative r.v. loss of the generic h th sub-portfolio.

Having the goal of a full ratemaking process, the number of potentially identifiable risk classes is $H = \prod_{k=1}^5 m_k = 3744$.

Table 1. Data set rating factors.

Rating Factors	Levels	Number of Levels
Age (band) of the driver	1 (youngest),2,3,4,5,6	6
Gender	M and F	2
Area of residence	A,B,C,D,E,F	6
Vehicle age	1(new),2,3,4	4
Vehicle body type	Bus, convertible, coupe, hatchback, hardtop, motorized caravan-combi, minibus, panel van, roadster, sedan, station wagon, truck, utility	13

Table 2. Risk classes descriptive statistics.

h	Risk class	Number of policies g_h	Number of claims	Total amount
1	1-F	3084.20	514	808,548.20
2	2-F	4740.18	822	1,374,829.90
3	3-F	5504.78	841	1,480,703.60
4	4-F	4625.44	655	1,244,667.40
5	1-M	2254.75	362	76,706.70
6	2-M	3183.50	532	1,111,387.20
7	3-M	4037.33	605	1,237,533.40
8	4-M	4370.64	606	1,310,228.10
	Total	31,800.82	4937	9,314,603.60

Therefore, in order to apply the QR in the second stage one can request to run a QR 3744 times, from which $m = 1 + \sum_{k=1}^5 (m_k - 1) = 27$ parameters are obtained, with a total of 101,088 ($H \times m = 3,744 \times 27$) parameters.

Whereas, using the TSPQR model, assuming e.g. $b_0(\theta) = 1$, $b_1(\theta) = 2\theta$ as in the previous examples (i.e. $n = 2$), we only need to estimate a single PQR in the second stage with 54 ($n \times m = 2 \times 27$) parameters.

Therefore, in the following analysis, to simplify our figures and outcomes, we consider only two rating factors: Vehicle Age (hereafter *Veh_Age*) and *Gender*; in this way, we reduce our analysis on eight rating classes obtained by the combination of four levels (1, 2, 3, 4) and two levels (*M*, *F*) for *Veh_Age* and *Gender*, respectively. In this way, the reduced data set allows us to investigate $H = 8$ risk classes and $m = 5$ parameters for each QR model.

In Table 2, we show some descriptive statistics of the reduced data set.

4.2. First stage: generalized linear models for probability on no claims estimate

The first step of the two proposed models concerns the estimate of the $Pr[\mathbb{I}_{N_h} = 0] = p_h$. The latter is performed by a Generalized Linear Model (GLM) assuming a binomial distribution for the response variable and a Logit link function. The model leads to the following estimate of the conditional mean of the frequency of having at least one claim:

$$p_h = 1 - E[\mathbb{I}_{N_h}] = 1 - \frac{\exp(\mathbf{x}_h^T \cdot \eta)}{1 + \exp(\mathbf{x}_h^T \cdot \eta)}. \quad (14)$$

The parameters' estimate for the Logit GLM used herein are obtained by using the Maximum Likelihood Estimation and are reported in Table 3. As one can observe, regression parameters *Veh_Age*₂ and *Gender*_M show a low significance and levels could be aggregated to the base one. However, for our analysis we leave the levels separated. Anyway, this is not relevant for the next steps requested by the Quantile Premium Principle. For example, the exclusion of the *Gender* covariate from the frequency model (i.e. using the same level of probability p_h for Male and Females), only reduces the number of QRs to be run, but does not exclude the *Gender* from the severity model at all. Indeed, in two-part models the statistical significance of the rating factors is tested independently on each part. The severity component of the QPP is computed by means of a Gamma GLM for the mean estimation and a QR, or a PQR, for the quantiles estimation, respectively. The parameters' estimate of the Gamma GLM model are reported in the Appendix. Hence, to remove completely a covariate, it must not be statistically significant in all models.

Before starting the second stage, we shall firstly fix the probability level for the estimation of the conditional quantile of the total claim amount S_h ; to this aim, we adopt a probability level $\hat{\theta} = 0.95$ in line with the assumption of Heras et al. Then, by means of Equation (10) we define the correspondent θ_h^* probability levels for the r.v. \hat{S}_h whose values are reported in the Table 4 with the corresponding values of the probability of no claim.

4.3. Quantile estimation via quantile regression and parametric quantile regression

Once defined the probability levels θ_h^* for the TSQR model we run eight QRs and we estimate 40 parameters consequently. In Tables 5 and 6, the parameter estimates (β_θ) and the corresponding p -values of the Wald test are reported (for further details, see Koenker 2005), for each run QR.

Table 6 shows that regression parameters $Gender_M$ are not significant (i.e. p -values higher than 0.05) in each risk class h , whereas the regression parameters Veh_Age_2 are not significant for half of the risk classes (bold values). Therefore, the latter risk factor shows an ambiguity on the statistical significance. This ambiguity may arise using of the TSQR approach, as it requires to run H independent QRs each one calibrated at a probability level θ_h^* , $h = 1, \dots, H$. Each QR produces a different set of diagnostic outcomes, then it is necessary to test each run to conduct a full analysis.

It is worth noting that the ambiguity on Veh_Age_2 proves that a not statically significant p -value of the same regression parameter in the GLM Logit (see Table 3), does not imply a no statistical significance in the QRs. Indeed, the frequency and severity regression models refer to different response variables, \mathbb{I}_{N_h} and \tilde{S}_h , respectively. The only relation between GLM Logit and QR models, is that the QR probability level θ_h^* depends on the estimates of p_h by means of Equation (10).

In the following, we implement the PQR procedure. We first shall choose a function for $b(\theta)$. To this aim, we select the class of the Shifted Legendre polynomials (SLP). The SLP are a special class of Legendre polynomials ($P_n(x)$) and are defined as $\tilde{P}_n(x) = P_n(2x - 1)$, implying that the $\tilde{P}_n(x)$ are

Table 3. Estimates of parameters for frequency model.

Parameter	Value	Std. error	p -value
<i>Intercept</i>	-1.689	0.040	
<i>Veh_Age₂</i>	0.026	0.049	0.590
<i>Veh_Age₃</i>	-0.099	0.048	0.039
<i>Veh_Age₄</i>	-0.200	0.049	$<10^{-3}$
<i>Gender_M</i>	-0.027	0.032	0.590

Table 4. Probability of no claim and probability level of the severity conditional quantile.

Risk class	1-F	2-F	3-F	4-F	1-M	2-M	3-M	4-M
p_h	0.844	0.841	0.857	0.869	0.848	0.844	0.860	0.872
θ_h^*	0.679	0.686	0.651	0.619	0.672	0.679	0.643	0.610

Table 5. Estimates of parameters for QRs.

θ_h^*	0.679	0.686	0.651	0.619	0.672	0.679	0.643	0.610
Risk class	1-F	2-F	3-F	4-F	1-M	2-M	3-M	4-M
<i>Intercept</i>	7.116	7.130	7.027	6.920	7.079	7.116	6.998	6.872
<i>Veh_Age₂</i>	0.132	0.137	0.155	0.144	0.157	0.132	0.153	0.166
<i>Veh_Age₃</i>	0.241	0.260	0.178	0.182	0.258	0.240	0.176	0.205
<i>Veh_Age₄</i>	0.382	0.400	0.332	0.320	0.391	0.379	0.307	0.330
<i>Gender_M</i>	0.075	0.064	0.046	0.035	0.063	0.075	0.045	0.041

Table 6. p -values of parameters for QRs.

θ_h^*	0.679	0.686	0.651	0.619	0.672	0.679	0.643	0.610
Risk class	1-F	2-F	3-F	4-F	1-M	2-M	3-M	4-M
<i>Veh_Age₂</i>	0.099	0.086	0.034	0.074	0.036	0.099	0.046	0.046
<i>Veh_Age₃</i>	0.006	0.004	0.025	0.020	0.002	0.007	0.027	0.008
<i>Veh_Age₄</i>	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$	$<10^{-3}$
<i>Gender_M</i>	0.207	0.275	0.398	0.537	0.270	0.207	0.413	0.468

orthogonal on $[0, 1]$. An explicit expression for the SLP of degree n is given by

$$\tilde{P}_n(x) = (-1)^n \sum_{k=0}^n \binom{n}{k} \binom{n+k}{k} (-x)^k. \tag{15}$$

To motivate our choice we observe that (i) in multiple linear regression the adoption of orthogonal polynomial, like the SLP, allows to overcome multicollinearity problem characterizing the polynomial regression; (ii) the estimation algorithm of PQR contained in the 'R' package `iqr` is optimized for objects of the SLP class (see Frumento & Bottai 2016), which means that using SLP of degree k instead of standard polynomials of the same degree will result in a quicker computation, even with $k = 1$, with equivalent results.

Moreover, we have tested different functions based on the same number of parameters to cover a reasonable range of possible choices; the SLP model generates the lowest value of Root Mean Square Error (RMSE).

Therefore, in the following we have assumed a degree $n = 3$ for our analysis and the PQR parameters are defined as follows:

$$\begin{aligned} \beta_j(\theta, \gamma_j) &= \gamma_{j0} + \gamma_{j1} \cdot (2\theta - 1) + \gamma_{j2} \cdot (6\theta^2 - 6\theta + 1) \\ &+ \gamma_{j3} \cdot (20\theta^3 - 30\theta^2 + 12\theta - 1), \quad j = 1, \dots, 5. \end{aligned} \tag{16}$$

Then, in order to estimate the conditional quantile for all the $\theta \in (0, 1)$ by means of Equation (4) we need to solve the minimization problem (7) and estimate the $5 \times 4 = 20$ parameters. Thus, reducing the number of parameters to half the TSQR.

In Tables 7 and 8, we report the parameter estimates and p -values of Wald tests, respectively.

In Table 8, the p -values shown in Wald test for the rating factors can be interpreted as the significance of a test for a null effect of covariates. The second part of Table 8 shows the p -values for the null hypothesis that the corresponding column of Γ is 0 and represents the significance of the components of $b(\theta)$. It is important to note that all the parameters are significant and all tests are completely performed with a single model.

Figure 2 shows the trend of the parameters β_θ and $\beta(\theta, \Gamma)$ over the entire range of probability levels θ calculated by QRs and PQR procedures, respectively. For the β_θ parameters we have run 100

Table 7. $\hat{\Gamma}$ matrix of estimated parameters for PQR.

$\hat{\gamma}_{00} = 5.019$	$\hat{\gamma}_{01} = 1.827$	$\hat{\gamma}_{02} = 0.355$	$\hat{\gamma}_{03} = 0.145$
$\hat{\gamma}_{10} = 0.043$	$\hat{\gamma}_{11} = 0.103$	$\hat{\gamma}_{12} = 0.034$	$\hat{\gamma}_{13} = -0.026$
$\hat{\gamma}_{20} = 0.078$	$\hat{\gamma}_{21} = 0.144$	$\hat{\gamma}_{22} = 0.016$	$\hat{\gamma}_{23} = -0.047$
$\hat{\gamma}_{30} = 0.124$	$\hat{\gamma}_{31} = 0.190$	$\hat{\gamma}_{32} = -0.037$	$\hat{\gamma}_{33} = -0.127$
$\hat{\gamma}_{40} = 0.005$	$\hat{\gamma}_{41} = 0.070$	$\hat{\gamma}_{42} = 0.067$	$\hat{\gamma}_{43} = 0.072$

Table 8. Wald test p -values PQR.

Wald test for rating factors	
Rating factors	p -value
<i>Veh...Age</i>	$<10^{-3}$
<i>Gender</i>	0.008
Wald test for $b(\theta)$	
$b_0(\theta)$	$<10^{-3}$
$b_1(\theta)$	$<10^{-3}$
$b_2(\theta)$	$<10^{-3}$
$b_3(\theta)$	$<10^{-3}$

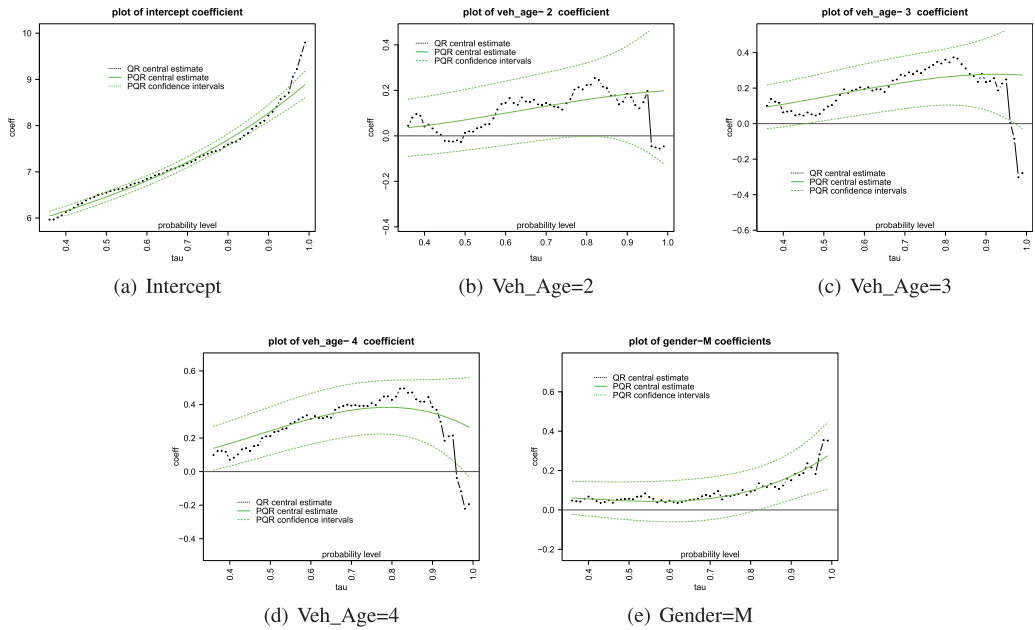


Figure 2. QR and PQR parameter estimates.

times the QR respect to each element of a quantile level vector $\theta = 0.01, 0.02, \dots, 0.99$. Whereas for the PQR it is sufficient to calculate $\beta(\theta, \Gamma)$ simply considering each element of the θ vector as an input of Equation (4).

In each graph, the central continuous line represents the parameter estimates $\hat{\beta}(\theta, \Gamma)$ provided by a single PQR, the external dashed lines represent the upper and lower confidence interval at level 95% estimated by a single PQR, the black dots are the parameter $\hat{\beta}_\theta$ estimates obtained by multiple QRs for different probability levels. For example, in Figure 2(a), Intercept or base level is represented; the green continuous line is obtained letting $j = 0$ in Equation (16). Obviously, the PQR and the QR Intercept parameter increases as the probability level increases.

The second graph, represented in Figure 2(b), shows the *Veh_Age₂* regression parameters. The linear combination with the Intercept parameter allows the estimation of the conditional quantile of the class '2-F'. We can observe that QR parameters show a sort of S-shaped form with some hump and pit with some noise; moreover, it shows an increasing trend until 0.9 and decreasing and more noisy afterwards. A very similar behavior for QR is observed for *Veh_Age₃* (Figure 2(c)) with an incremental trend until 0.8 and a slope afterwards. PQR parameters show a more smoothed trend, as expected, but crossing the QR values. More similar behavior is observable respect to the *Veh_Age₄* and *Gender_M*. In particular, the latter in Figure 2(e) represents the same data shown in Figure 1 for QR. As noticeable, PQR overcomes all the previously commented issues related to the strong variation between neighboring quantiles estimated with QRs.

To better appreciate the differences between the two proposed approaches we show in Table 9 the quantile estimates for each $\theta_h^*, Q_{\hat{\theta}}[S_h]$.

As observable, the two estimates are very close to each other, indeed the differences are in the range $(-1.29\%, 3.52\%)$. In summary, it can be observed that these differences can be considered more or less relevant depending on the purpose of the analysis which they are used in. Indeed, when the analysis requires the estimate of a discrete set of quantiles, or the whole quantile process, PQR approach can result in a correct combination of precision and parsimony.

Table 9. Quantile estimates with $\bar{\theta} = 0.95$.

h	Risk class	Exposure g_h	θ_h^*	$Q_{\bar{\theta}}[S_h]$		$\Delta\%$
				TSQR	TSPQR	
1	1-F	3084.20	0.679	1231.66	1237.68	-0.49%
2	2-F	4740.18	0.686	1432.45	1447.69	-1.05%
3	3-F	5504.78	0.651	1346.64	1364.18	-1.29%
4	4-F	4625.44	0.619	1393.34	1346.01	3.52%
5	1-M	2254.75	0.672	1264.11	1263.20	0.07%
6	2-M	3183.50	0.679	1515.65	1477.00	2.62%
7	3-M	4037.33	0.643	1366.28	1379.23	-0.94%
8	4-M	4370.64	0.610	1398.43	1351.43	3.48%

Table 10. Pure premium estimation.

h	Risk class	Exposure	p_h	$E[\tilde{S}_h]$	$E[S_h]$
1	1-F	3084.20	0.844	1734.85	270.45
2	2-F	4740.18	0.841	1828.16	291.38
3	3-F	5504.78	0.857	1844.88	264.42
4	4-F	4625.44	0.869	1984.13	260.67
5	1-M	2254.75	0.848	2080.32	316.95
6	2-M	3183.50	0.844	2192.21	341.51
7	3-M	4037.33	0.860	2212.26	309.78
8	4-M	4370.64	0.872	2379.24	305.28

4.4. QPP estimate

The differences shown in Table 9 are negligible if we consider that the quantile values are used in a ratemaking process to determine risk loading in the loaded premium in Equation (8). In order to apply the QPP, we first estimate $E[\tilde{S}_h]$ by means of a Gamma GLM with log link function and consequently, we obtain the estimate of the pure (or equivalence) premium $E[S_h] = (1 - p_h) \cdot E[\tilde{S}_h]$. In Table 10, we show the total claim amount (equivalence premium) per risk class. For further details on parameter values and diagnostics the reader can refer to Baione & Biancalana (2019).

Once estimated $E[S_h]$ and $Q_{\bar{\theta}}[S_h]$, following (8) it is important to fix the α level to complete the loaded premium estimate. The choice of the α is strategic and cannot be arbitrary. Following Baione & Biancalana (2019), the basic idea is that the sum of all premiums covers the expected aggregate amount of loss $E[L]$ plus a risk margin value. The overall aggregate amount is generally selected by insurance managers by looking at pricing or solvency conditions, using a functional π that assigns a non-negative real number to random variables L . Hence, it is possible to calibrate the α level so that the following balance equation is satisfied:

$$\pi[L] = \sum_{h=1}^H g_h \cdot P_h = \sum_{h=1}^H g_h [\alpha \cdot Q_{\bar{\theta}}[S_h] + (1 - \alpha) \cdot E[S_h]]. \quad (17)$$

Then

$$\hat{\alpha} = \frac{\pi[L] - E[L]}{\sum_{h=1}^H g_h \cdot [Q_{\bar{\theta}}[S_h] - E[S_h]]}. \quad (18)$$

In the following, according to the Solvency II standard formula for premium risk² in the Motor Third Party Liability insurance (EIOPA 2009), we consider $\pi[L] = E[L] \cdot (1 + 3 \cdot \sigma)$ with $\sigma = 10\%$.

² The standard formula implicitly assumes to measure the distance between the Value at Risk (VaR) at 99.5% confidence level and the mean of the probability distribution of aggregate claims amount by using a fixed multiplier of the volatility equal to 3 for all undertaking.

In our application, this assumption leads to an aggregate amount of $\pi[L] = 12,042,455$ and a mean value $E[L] = 9,263,427$.

Considering the differences between quantiles estimates obtained by means of the TSQR or TSPQR (see Table 9), we obtain two different $\hat{\alpha}$ levels, say $\hat{\alpha}_{QR} = 8.05\%$ and $\hat{\alpha}_{PQR} = 8.13\%$.

In Table 11, the premium estimates are shown for both QRs and PQR procedure:

It is worth noting that the two estimates of QPP are very close, indeed the differences are reduced in the range $(-0.63\%, 0.87\%)$.

In order to demonstrate that these small differences are not occasional and related to our dataset, we have also performed the TSPQR on the same data set used by Heras et al. (2018), where the authors have used the ‘Age of the driver’ (*Age_Cat*) as rating factor instead of ‘Gender’. The *Age_Cat* has six levels, then the number of rating classes as well as the number of parameters increase to $H = 6 \times 4 = 24$ and $m = 9$, respectively. The application of the TSQR requires to run 24 QRs, with a total of 216 ($H \times m = 24 \times 9$) parameters to be tested. While assuming the same polynomial of the

Table 11. Quantile Premium Principle estimation.

<i>h</i>	Risk class	Exposure	$E[S_h]$	P_h		
				TSQRs	TSPQR	$\Delta\%$
1	1-F	3084.20	270.45	347.91	349.10	-0.34%
2	2-F	4740.18	291.38	383.33	385.41	-0.54%
3	3-F	5504.78	264.42	351.63	353.85	-0.63%
4	4-F	4625.44	260.67	351.94	348.92	0.87%
5	1-M	2254.75	316.95	393.27	393.89	-0.16%
6	2-M	3183.50	341.51	436.13	433.84	0.53%
7	3-M	4037.33	309.78	394.92	396.74	-0.46%
8	4-M	4370.64	305.28	393.37	390.35	0.77%

Table 12. TSQR vs. TSPQR: Quantile Premium Principle comparison in the application of Heras et al. (2018).

<i>h</i>	Risk class	P_h		$\Delta\%$
	<i>Veh...Age&Age_Cat</i>	TSQR	TSPQR	
1	(2&1)	603.58	602.48	0.18%
2	(1&1)	546.10	549.97	-0.70%
3	(3&1)	557.49	562.12	-0.82%
4	(2&2)	396.55	396.59	-0.01%
5	(4&1)	564.30	569.59	-0.93%
6	(1&2)	361.42	363.08	-0.46%
7	(2&3)	339.93	339.75	0.05%
8	(1&3)	311.26	310.95	0.10%
9	(2&4)	327.66	329.25	-0.48%
10	(3&2)	369.33	367.74	0.43%
11	(1&4)	300.13	301.63	-0.50%
12	(3&3)	315.34	314.68	0.21%
13	(4&2)	373.96	371.19	0.75%
14	(3&4)	303.49	304.61	-0.37%
15	(4&3)	317.39	317.24	0.05%
16	(4&4)	306.72	306.89	-0.05%
17	(2&5)	239.91	238.85	0.44%
18	(2&6)	261.64	259.39	0.87%
19	(1&5)	218.81	219.19	-0.17%
20	(1&6)	238.55	238.12	0.18%
21	(3&5)	220.02	219.86	0.07%
22	(3&6)	240.95	239.11	0.77%
23	(4&5)	220.54	220.29	0.11%
24	(4&6)	242.18	239.75	1.01%

third degree in (16) for the description parameters, the TSPQR model needs to estimate 36 ($n \times m = 4 \times 9$) parameters at once. In Table 12, we show a comparison of the estimates.

As in the previous case, the distance between the two estimates is very small as it ranges in $(-0.93\%, 1.01\%)$.

5. Conclusion

The paper deals with the assessment of the ratemaking on a portfolio of non-life policies, where the loaded premium is individually calculated according to a quantile premium principle. To this aim, we propose to change a two-part model introduced by Heras et al. (TSQR) (Heras et al. 2018) that, for the estimate of the severity component, requests to perform a number of Quantile Regressions equal to the number of the risk classes selected. Since, in actuarial practice, the ratemaking process can include a remarkable number of risk classes (also in order of billions), this procedure could be onerous and time-consuming; then, following Frumento & Bottai (2016), we introduce the parametric quantile regression (PQR), in order to overcome these drawbacks and define a statistical structure (TSPQR) for ratemaking, which is more parsimonious. The PQR, never applied in the actuarial context, enables the estimate of conditioned quantiles of the severity component, for each probability level in a unique run.

Furthermore, a numerical application to compare the two procedures is produced. TSQR and TSPQR provide similar outcomes that are not too distant. Since this result is obtained by estimating half the number of parameters in the TSPQR case, and this implies a suitable performance of TSPQR in terms of parsimony and efficiency.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Fabio Baione  <http://orcid.org/0000-0002-9926-7869>

Davide Biancalana  <http://orcid.org/0000-0001-6032-2552>

References

- Baione F. & Biancalana D. (2019). An individual risk model for premium calculation based on quantile: a comparison between generalized linear models and quantile regression. *North American Actuarial Journal* **23**, 573–590.
- Baione F., Biancalana D. & De Angelis P. (2019). A quantile regression approach for the analysis of the diversification in non-life premium risk. *Soft Computing* **24**, 8523–8534. Doi: 10.1007/s00500-019-04291-x.
- Biancalana D. (2017). Un approccio quantile regression per la tariffazione danni, basato su un modello a due parti. Ph.D. Thesis. <https://iris.uniroma1.it/retrieve/handle/11573/1209293/944777/Tesi%20dottorato%20Biancalana.pdf>.
- De Jong P. & Heller G. Z. (2008). *Generalized linear models for insurance data*. Cambridge: Cambridge University Press.
- Dong A., Chan J. & Peters G. (2015). Risk margin quantile function via parametric and non-parametric Bayesian approaches. *ASTIN Bulletin* **45**(3), 503–550. doi: doi: 10.1017/asb.2015.8.
- EIOPA (2009). Directive 2009/138/EC of the European parliament and of the council 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II). <https://eiopa.europa.eu>.
- Frees E. W. (2010). *Regression modeling with actuarial and financial applications*. New York: Cambridge University Press.
- Frees E. W., Jin X. & Lin X. (2013). Actuarial applications of multivariate two-part regression models. *Annals of Actuarial Science* **7**, 258–287.
- Frumento P. (2017). QRCM: quantile regression coefficients modeling. R package version 2.1. <https://cran.r-project.org/package=qrcm>.
- Frumento P. & Bottai M. (2016). Parametric modeling of quantile regression coefficient functions. *Biom* **72**, 74–84. doi: doi:10.1111/biom.12410.
- Frumento P. & Bottai M. (2017). Parametric modeling of quantile regression coefficient functions with censored and truncated data. *Biometrics* **73**(4), 1179–1188. doi: doi: 10.1111/biom.12675
- Fuzi M. F. M., Jemain A. A. & Noriszura I. (2016). Bayesian quantile regression model for claim count data. *Insurance, Mathematics & Economics* **66**, 124–137.

- Hao L. & Naiman D. Q. (2007). *Quantile regression*, Vol. 149. Thousand Oaks, CA: Sage.
- Heras A., Moreno I. & Vilar-Zanón J. L. (2018). An application of two-stage quantile regression to insurance ratemaking. *Scandinavian Actuarial Journal* 9, 753–769.
- Koenker R. (2005). *Quantile regression*. Econometric Society Monograph Series, Vol. 38. Cambridge: Cambridge University Press.
- Koenker R. (2015). Quantreg: quantile regression. R package version 5.19. <http://CRAN.R-project.org/package=quantreg>.
- Koenker R. W. & Bassett G. W. (1978). Regression quantiles. *Econometrica* 46, 33–50.
- Koenker R. & Hallock K. (2001). Quantile regression: an introduction. *Journal of Economic Perspectives* 15(4), 43–56.
- Kudryavtsev A. A. (2009). Using quantile regression for rate-making. *Insurance: Mathematics and Economics* 45, 296–304.
- Pitselis G. (2009). Solvency supervision based on a total balance sheet approach. *Journal of Computational and Applied Mathematics* 233(1), 83–96.
- Pitt D. (2006). Regression quantile analysis of claim termination rates for income protection insurance. *Annals of Actuarial Science* 1(2), 345–357.
- Portnoy S. (2003). Censored regression quantiles. *Journal of the American Statistical Association* 98, 1001–1012.
- Wolny-Dominiak A., Ornat-Acedaska A. & Trzpiot G. (2012). Insurance portfolios rate making: quantile regression approach. Proceedings of 30th International Conference Mathematical Methods in Economics, Karviná, Czech Republic.

Appendix

The application of QPP requires the estimate of $E[\tilde{S}(h)]$. To this aim Gamma GLM with log link function is adopted. The model leads to the parameter estimates reported in Table A1.

As observable, compared to the frequency model the $Gender_M$ is statistically significant, whereas all the Veh_Age regression parameters show a low significance and levels should be aggregate to the base one. Anyway, we have leaved the levels separated.

Table A1. Estimates of parameter for claim severity conditional mean.

Parameter	Value	Std. error	p-value
<i>Intercept</i>	7.459	0.065	
<i>Veh_Age₂</i>	0.052	0.078	0.504
<i>Veh_Age₃</i>	0.061	0.077	0.425
<i>Veh_Age₄</i>	0.134	0.079	0.091
<i>Gender_M</i>	0.182	0.052	0.000