Insurance: Mathematics and Economics 111 (2023) 242–256

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

# Parametric expectile regression and its application for premium calculation

Suhao Gao, Zhen Yu *

*School of Statistics, Renmin University of China, Beijing, China*

A R T I C L E I N F O

*Article history:*
Received February 2022
Received in revised form January 2023
Accepted 9 May 2023
Available online 18 May 2023

*JEL classification:*
C13
G22

*Keywords:*
Aggregate claims model
Parametric expectile regression
Premium principle
Two-stage model
Gini index

A B S T R A C T

Premium calculation has been a popular topic in actuarial sciences over the decades. Generally, a two-stage model is used to develop the premium calculation process. It can be decomposed into estimating the probability of having at least one claim by the logistic regression in the first stage and calibrating the severity in the second stage. Existing methods for the second stage include generalized linear models (GLMs) and quantile regression. However, the GLMs fail to provide information associated with the extreme claim amount and the quantile is not sensitive to the size of the extreme losses. Given that the magnitude of the extreme claim amount may lead to a huge loss for the insurer, we introduce the expectile risk measure into premium calibration and propose an expectile-based risk premium, which outperforms other methods for heavy-tailed distributions. Naively adopting the conventional expectile regression in the second stage is not preferred because it would be over-parameterized and time-consuming if the portfolio contains a large number of risk classes. Thus, we put forward a two-stage parametric expectile regression (TSPER) with parametric expectile regression (PER) method used in the second stage. The consistency and asymptotic normality of the proposed PER estimator are established under mild conditions. We also propose a novel Expectile Premium Principle to allocate the total premium for each policy. In the analysis of the automobile insurance data set, the proposed TSPER method outperforms other two-stage methods in terms of the ordered Lorenz curve and the Gini index.



## 1. Introduction

Calibrating the risk premiums to cover the costs of policyholders is a crucial task in the assessment of underwriting risk. One of the popularly adopted methods is the Two-Part Model (Frees, 2010; Frees et al., 2013). It decomposes the cost of the policy into a binary variable indicating whether the policyholder has claimed and a continuous variable representing the aggregate claim amount. Traditionally, GLMs are used to model both parts. Although GLMs can identify risk factors that affect the expectations of the aggregate claim amount, they fail to provide information associated with the extreme claim amount. The ignorance of such information may lead to an inaccurate assessment of the solvency, inflicting losses or even financial distress on the insurance company. In addition, GLMs are also limited by the strong assumptions of the underlying distribution and risk homogeneity (Kudryavtsev, 2009), which often deviates from the real-world situations.

Recently, quantile regression (QR) developed by Koenker and Bassett (1978) has gained a lot of attention. Kudryavtsev (2009) is the first to apply the QR method in insurance ratemaking. The method provides some knowledge of the characteristics of policyholders who are prone to make large claims. As an extension of Kudryavtsev's approach, Heras et al. (2018) present a Two-Stage Quantile Regression (TSQR) model to estimate the aggregate claim amount. However, applying the TSQR method necessitates conducting a number of QRs at different quantile levels. The number of these QRs equals the number of risk classes, which results in heavy computation when there are plenty of risk classes. To alleviate the problem of over-parameterization, Baione and Biancalana (2019) fix the quantile level for the severity and introduce a premium principle based on a quantile risk measure to calculate the risk margin. As an alternative method for parsimony and efficiency, Baione and Biancalana (2021) introduce parametric quantile regression (PQR) (Frumento and Bottai, 2016) into premium calculations. However, quantile is not a coherent risk measure (Artzner et al., 1999) and is not sensitive to the size of the extreme losses

* Corresponding author.
*E-mail address:* amy.yuzhen@ruc.edu.cn (Z. Yu).

https://doi.org/10.1016/j.insmatheco.2023.05.004

(Kuan et al., 2009) that generally exist in large portfolios either. Thus, it cannot reveal the riskiness of the policies sufficiently and may lead to great losses for the insurers.

In this paper, we introduce an expectile risk measure into premium calculation. Expectile (Newey and Powell, 1987) is elicitable, law-invariant, and coherent (Ziegel, 2014). Unlike the quantile risk measure, the expectile risk measure depends on the tail expectation rather than the tail probability (Daouia et al., 2021). Thus, it cares about not only the probabilities of the extreme events but also their magnitude (Kim and Lee, 2016; Bellini and Di Bernardino, 2017), offering crucial information on the riskiness of the policies for the insurer. As far as we know, expectile regression (ER), which is a distribution-free methodology and allows for heteroscedasticity (Gu and Zou, 2016; Zhao et al., 2018), has not been applied to premium calculation, though related work does exist in fields like actuarial and financial risk management (Cai and Weng, 2016; Kim and Lee, 2016; Girard et al., 2020; Barigou et al., 2022; Pitera and Schmidt, 2022). Additionally, the asymmetric squared error loss for the positive and negative deviations used by the expectile regression provides a good way to describe gain-loss asymmetry characterized by the insurer with loss aversion (preferring avoiding great losses to making an equivalent profit). In this paper, we adopt and extend the expectile regression for the calculation of the risk premium.

The main contributions of this paper are as follows. First, we propose a parametric expectile regression (PER) method to conduct multiple expectile regressions in a single run, which is more parsimonious and efficient than the ER method. We introduce a PER estimator that generalizes that of the ordinary ER method and establish its consistency and asymptotic normality under mild conditions. By parametrically modeling the regression coefficient functions, the PER method provides more stable results on the extreme tail compared with the ER method and can alleviate or even limit the crossing problem. Second, given that the extreme claim amount is crucial for the evaluation of the riskiness of the policies, we propose an expectile-based individual risk premium and develop the Two-Stage Expectile Regression (TSER) and Two-Stage Parametric Expectile Regression (TSPER) methods for its calculation. The actuarial meaning of the expectile-based risk premium is discussed in detail. In the analysis of a popular automobile insurance data set, we show that in terms of the ordered Lorenz curve and the Gini index, the TSPER method outperforms other two-stage methods, which include the Expected Value Premium Principle (EVPP) (Szynal and Teugels, 1993; de Lourdes Centeno, 2002; Asmussen and Møller, 2003; Young, 2004), the Standard Deviation Premium Principle (SDPP) (Szynal and Teugels, 1993; Kaluszka, 2001; Asmussen and Møller, 2003; Young, 2004), the TSQR, the two-stage parametric quantile regression (TSPQR), and the TSER. Third, we put forward an Expectile Premium Principle (EPP) to allocate the total risk premium calculated by the bootstrap method to each policy using a single quantile level for the aggregate claim amount. We choose the unique quantile level such that the sum of the individual risk premiums for all the policies covers the total risk premium. Unlike the existing methods, the unique quantile level used in the EPP is not subjectively assigned but chosen to ensure all the policies have the same probability that the actual loss exceeds the risk premium.

The rest of the article is organized as follows. After reviewing the basics of the ER method in Section 2, we propose the PER method and establish its asymptotic properties in Section 3. Section 4 proposes the TSPER method and the EPP for premium calculation, provides the actuarial meaning of the expectile-based risk premium, and discusses the relationship between the proposed EPP and the expected value premium principle. Section 5 applies the proposed TSPER method to an Australian automobile insurance data set and compares it with other two-stage methods. We conclude this paper in Section 6.

## 2. Expectile regression

Compared with the QR, the ER method is a more effective tool to take into account the magnitude of the extreme losses in the actuarial context (Newey and Powell, 1987). Let $\mathrm{ER}_\tau(Y_i|\mathbf{x}_i)$ be the conditional $\tau$-th expectile of the response variable $Y_i$ given covariates $\mathbf{x}_i \in \mathcal{R}^{m+1}$ for $i = 1, \ldots, n$, where $n$ is the sample size. For any expectile level $\tau \in (0, 1)$, the ER method is given by

$$\mathrm{ER}_\tau\left(Y_i|\mathbf{x}_i\right) = \mathbf{x}_i^\top \boldsymbol{\varphi}(\tau), \quad i = 1, \ldots, n, \tag{1}$$

where $\boldsymbol{\varphi}(\tau) = \left(\varphi_0(\tau), \varphi_1(\tau), \ldots, \varphi_m(\tau)\right)^\top$ is an unknown $(m+1)$-dimensional expectile-specific coefficient vector. The estimator $\hat{\boldsymbol{\varphi}}(\tau)$ can be obtained by minimizing the asymmetric least squares (ALS) loss function

$$L_n(\boldsymbol{\varphi}(\tau)) = \frac{1}{n}\sum_{i=1}^{n}\left|\tau - \omega_{\tau,i}\right|\left(y_i - \mathbf{x}_i^\top \boldsymbol{\varphi}(\tau)\right)^2, \tag{2}$$

where $\omega_{\tau,i} = \mathbb{I}\left(y_i \le \mathbf{x}_i^\top \boldsymbol{\varphi}(\tau)\right)$ and $\mathbb{I}(\cdot)$ denotes the indicator function. The ER method can be easily implemented using the R package *expectreg*. When $\tau = 0.5$, the expectile estimator reduces to the least squares estimator, which assigns equal weights to all observations in the data set. The asymmetric loss function in (2) assigns weights $\tau$ and $1-\tau$ to positive and negative deviations, respectively.

Kuan et al. (2009) propose expectile-based Value at Risk (EVaR) as a risk measure related to the expectiles. Ziegel (2014) and Bellini et al. (2014) analyze its theoretical properties: (a) $\mathrm{EVaR}_\tau(Y_i)$ is strictly monotonic in $\tau$ for $\tau \in (0,1)$; (b) $\mathrm{EVaR}_\tau(Y_i)$ is strictly monotonic in $Y_i$, that is, if $Y_i^* \ge Y_i$ a.s. and $P\left(Y_i^* > Y_i\right) > 0$, then $\mathrm{EVaR}_\tau\left(Y_i^*\right) > \mathrm{EVaR}_\tau\left(Y_i\right)$; (c) $\mathrm{EVaR}_\tau(-Y_i) = -\mathrm{EVaR}_\tau(Y_i)$; (d) if $Y_i$ is symmetric with respect to $y$, then $\mathrm{EVaR}_\tau(Y_i) + \mathrm{EVaR}_{1-\tau}(-Y_i) = 2y$; (e) $\mathrm{EVaR}_\tau(Y_i)$ is the only risk measure that is coherent and elicitable.

However, the ER method has some drawbacks. Note that the unknown parameter $\boldsymbol{\varphi}(\tau)$ is expectile-specific. When we are interested in the covariate effects on the response at a number of expectile levels, we have to fit the ERs for these expectile levels one by one, which is onerous in terms of model specification and validation and computationally demanding. In addition, the ER method also suffers from the well-known crossing problem, making it hard for interpretation. As a solution to the aforementioned problems, we suggest adopting a parametric model for the expectile regression coefficient functions.

## 3. Parametric expectile regression

Following the idea of the parametric modeling in Frumento and Bottai (2016), we propose the PER method for parsimony and derive its asymptotic properties under some mild conditions in this section. This theoretically sound estimator will be adopted for individual premium calculation in Section 4.

We model the coefficients in (1) hierarchically as parametric functions of $\tau$:

$$\boldsymbol{\varphi}(\tau,\boldsymbol{\gamma})=\boldsymbol{\gamma}\mathbf{b}(\tau)=\begin{pmatrix}\gamma_{00} & \gamma_{01} & \dots & \gamma_{0h}\\ \gamma_{10} & \gamma_{11} & \dots & \gamma_{1h}\\ \vdots & \vdots & \ddots & \vdots\\ \gamma_{m0} & \gamma_{m1} & \dots & \gamma_{mh}\end{pmatrix}\begin{pmatrix}b_0(\tau)\\ b_1(\tau)\\ \vdots\\ b_h(\tau)\end{pmatrix}, \tag{3}$$

where the unknown $\boldsymbol{\gamma}=(\boldsymbol{\gamma}_0,\dots,\boldsymbol{\gamma}_m)^\top$ is a $(m+1)\times(h+1)$ matrix, $\mathbf{b}(\tau)=[b_0(\tau),b_1(\tau),\dots,b_h(\tau)]^\top$ is a set of $h+1$ known functions of $\tau\in(0,1)$, and $b_0(\tau)$ is generally set as 1. Given an expectile level $\tau\in(0,1)$, we define the PER method for a predefined $\mathbf{b}(\tau)$ by

$$\mathrm{PER}_\tau(Y_i|\mathbf{x}_i,\boldsymbol{\gamma})=\varphi_0(\tau,\boldsymbol{\gamma}_0)+\varphi_1(\tau,\boldsymbol{\gamma}_1)x_{i1}+\dots+\varphi_m(\tau,\boldsymbol{\gamma}_m)x_{im}=\mathbf{x}_i^\top\boldsymbol{\gamma}\mathbf{b}(\tau). \tag{4}$$

A specific $\mathbf{b}(\tau)$ can be used if the corresponding model is believed to be true, or we can define it as a collection of flexible functions of $\tau$, such as trigonometric functions, logarithmic functions, and polynomials, as long as it induces a well-defined expectile function of the response for some $\boldsymbol{\gamma}$.

The advantage of parametric modeling is that after obtaining the estimates of the parameter $\boldsymbol{\gamma}$ in a single run, we are able to get the estimated expectile-specific coefficients $\boldsymbol{\varphi}(\tau,\boldsymbol{\gamma})$ in (4) and the estimated expectile of the response $\widehat{\mathrm{PER}}_\tau(Y_i|\mathbf{x}_i,\boldsymbol{\gamma}), i=1,\dots,n$ at an arbitrary expectile level $\tau\in(0,1)$ simply by taking $\tau$ as an input of (4).

**Remark 1.** In expectile regression (1), for varying $\tau$, $\boldsymbol{\varphi}(\tau)$ can be viewed as an infinite-dimensional parameter. Generally, the estimated coefficients are non-smooth functions of $\tau$. As for the PER, we further model the expectile regression coefficient functions, $\boldsymbol{\varphi}(\tau)$, as parametric functions of the order of the expectile. Then, the ER can be viewed as the special case of the PER where $\boldsymbol{\varphi}(\tau,\boldsymbol{\gamma})$ is described by a "huge" number of parameters. A similar relationship between the QR and the PQR can also be found in the existing literature (Frumento and Bottai, 2016; Frumento and Bottai, 2017; Bottai and Orsini, 2019; Sottile et al., 2020; Sottile and Frumento, 2021; Frumento and Salvati, 2021).

### 3.1. The estimator

This section proposes a PER estimator that generalizes the ordinary ER estimator.

Note that modeling of the expectile coefficients $\boldsymbol{\varphi}(\tau)$ in (3) across different expectile levels shares the same parameter $\boldsymbol{\gamma}$. To combine the information across the entire expectile process, we adopt the integrated loss minimization method (Frumento and Bottai, 2016) to estimate $\boldsymbol{\gamma}$ with the integrated ALS loss function, i.e.,

$$\bar{L}_n(\boldsymbol{\gamma})=\int_0^1 L_n(\boldsymbol{\varphi}(\tau,\boldsymbol{\gamma}))d\tau. \tag{5}$$

A more general form of (5) is $\bar{L}_n(\boldsymbol{\gamma})=\int_0^1 w(\tau)L_n(\boldsymbol{\varphi}(\tau,\boldsymbol{\gamma}))d\tau$, where $w(\tau)$ is a nonconstant weight. When $w(\tau)=\mathbb{I}(\tau_1\le\tau\le\tau_2)$ and $\tau_1=\tau_2=\tau$, the PER method degenerates to the ordinary $\tau$-th ER method. By applying the chain rule, the gradient of (5) with respect to $\boldsymbol{\gamma}$ can be written as

$$\bar{\mathbf{S}}_n(\boldsymbol{\gamma})=\int_0^1 S_n(\boldsymbol{\varphi}(\tau,\boldsymbol{\gamma}))\mathbf{b}(\tau)^\top d\tau, \tag{6}$$

where $S_n(\boldsymbol{\varphi}(\tau,\boldsymbol{\gamma}))=-2n^{-1}\sum_{i=1}^n\left|\tau-\omega_{\tau,i}\right|\mathbf{x}_i\left(y_i-\mathbf{x}_i^\top\boldsymbol{\varphi}(\tau,\boldsymbol{\gamma})\right)$ is the $(m+1)$-dimensional gradient vector of the loss function (2) with respect to $\boldsymbol{\varphi}$.

Denote that $b_k'(\tau)=\frac{\mathrm{d}b_k(\tau)}{\mathrm{d}\tau}$, $B_k(\tau)=\int_0^\tau b_k(u)\mathrm{d}u$, $\bar{B}_k(\tau)=\int_0^\tau B_k(u)\mathrm{d}u$, $B_{k_1k_2}(\tau)=\int_0^\tau b_{k_1}(u)b_{k_2}(u)\mathrm{d}u$, $\bar{B}_{k_1k_2}(\tau)=\int_0^\tau B_{k_1k_2}(u)\mathrm{d}u$, and $a_{k_1k_2}^{(i)}=\left(\sum_{j=0}^m x_{ij}\gamma_{jk_1}\right)\left(\sum_{j=0}^m x_{ij}\gamma_{jk_2}\right)$ for $k,k_1,k_2=0,1,\dots,h$. Let $\mathbf{A}(\tau)$ be a $(h+1)\times(h+1)$ matrix whose $k_1,k_2$-th element is $B_{k_1k_2}(\tau)$, $\overline{\mathbf{A}}(\tau)=\int_0^\tau\mathbf{A}(u)du$, and $\mathbf{B}(\tau),\overline{\mathbf{B}}(\tau),\mathbf{b}_k'(\tau)$ be the $h$-dimensional vectors composed of $B_k(\tau),\overline{B}_k(\tau),b_k'(\tau)$, respectively.

A direct decomposition and computation of the integrated loss (5) and the integrated gradient (6) (a $(m+1)\times(h+1)$ matrix) yield

$$\begin{aligned}\bar{L}_n(\boldsymbol{\gamma})=&\,n^{-1}\sum_{i=1}^n y_i^2(1.5-3\tau_i+\tau_i^2)+2n^{-1}\sum_{i=1}^n y_i\mathbf{x}_i^\top\boldsymbol{\gamma}\left[-\overline{\mathbf{B}}(1)+(1-2\tau_i)\mathbf{B}(\tau_i)+2\overline{\mathbf{B}}(\tau_i)\right]\\ &+n^{-1}\sum_{i=1}^n\sum_{k_1}^h\sum_{k_2}^h a_{k_1k_2}^{(i)}\left[-\mathbf{B}_{k_1k_2}(\tau_i)+\overline{\mathbf{B}}_{k_1k_2}(1)-2\overline{\mathbf{B}}_{k_1k_2}(\tau_i)\right]\end{aligned} \tag{7}$$

and

$$\begin{aligned}\bar{\mathbf{S}}_n(\boldsymbol{\gamma})=&\,2n^{-1}\sum_{i=1}^n y_i\mathbf{x}_i\left[\overline{\mathbf{B}}(1)+(2\tau_i-1)\mathbf{B}(\tau_i)-2\overline{\mathbf{B}}(\tau_i)\right]\\ &+2n^{-1}\sum_{i=1}^n\mathbf{x}_i\mathbf{x}_i^\top\boldsymbol{\gamma}\left[(1-2\tau_i)\mathbf{A}(\tau_i)-\overline{\mathbf{A}}(1)+2\overline{\mathbf{A}}(\tau_i)\right],\end{aligned} \tag{8}$$

where $\tau_i(\boldsymbol{\gamma})$ (we use $\tau_i$ for abbreviation) denotes the order of the expectile that satisfies $\mathbf{x}_i^\top \boldsymbol{\varphi}(\tau_i, \boldsymbol{\gamma}) = y_i$ for $i = 1, \ldots, n$. When $\tau_i$ equals to the expectile level of $y_i$ for all $i$, $\bar{\mathbf{S}}_n(\boldsymbol{\gamma})$ approaches zero. $\bar{\mathbf{S}}_n(\boldsymbol{\gamma})$ is a smooth function of $\boldsymbol{\gamma}$ when $\tau_i$'s are near the true values.

The unknown parameters $\boldsymbol{\gamma}$ can be estimated by the Newton-Raphson method or the quasi-Newton method (Li and Fukushima, 1999; Bottai et al., 2015). As will be illustrated in Algorithm 1 for expectile-based premium calculation, the loss function (7) and the gradient (8) are crucial for the iterative updates of the quasi-Newton method.

### 3.2. Asymptotic properties

In the following, we establish, under mild conditions, the consistency and asymptotic normality of the proposed estimator by using the theory of extremum estimators (Amemiya, 1985; Newey and Mcfadden, 1994).

Let $\hat{\boldsymbol{\gamma}}_n$ be the estimator of the true parameter $\boldsymbol{\gamma}_0 \in \Gamma$ obtained by minimizing the integrated loss function $\bar{L}_n(\boldsymbol{\gamma})$. As $\mathrm{E}\left[\bar{\mathbf{S}}_n(\boldsymbol{\gamma}_0)\right] = 0$, unbiasedness follows.

**Theorem 3.1.** *(Consistency) Assume that $\Gamma$ is a compact set. If there exists a function $\bar{L}_0(\boldsymbol{\gamma})$ such that (i) $\bar{L}_0(\boldsymbol{\gamma})$ is continuous; (ii) $\bar{L}_0(\boldsymbol{\gamma})$ is uniquely minimized at $\boldsymbol{\gamma}_0$; (iii) the integrated ALS loss function $\bar{L}_n(\boldsymbol{\gamma})$ converges uniformly in probability to $\bar{L}_0(\boldsymbol{\gamma})$, then $\hat{\boldsymbol{\gamma}}_n \xrightarrow{p} \boldsymbol{\gamma}_0$.*

The validity of the conditions in Theorem 3.1 relies on the structure of $\mathbf{b}(\tau)$, which should (i) induce a well-defined expectile regression such that $\Gamma$ is not empty; and (ii) ensure that $\tau_i$ is a continuous function of $\boldsymbol{\gamma}$. Although $\mathbf{b}(\tau)$ may not be bounded at $\tau = 0$ or 1, $\mathbf{B}(\tau)$, $\bar{\mathbf{B}}(\tau)$, $\mathbf{A}$ and $\bar{\mathbf{A}}(\tau)$ must be finite. The possible choices of the predefined $\mathbf{b}(\tau)$ could be trigonometric functions, logarithmic functions, polynomials, splines, piecewise linear functions, and combinations of the above. Additionally, $\mathbf{b}'(\tau)$ should be defined in the interior of [0, 1], but it could be infinite in the extremes. For the proof of Theorem 3.1, we refer to Theorem 2.1 in Newey and Mcfadden (1994).

For simplification in the analysis of the asymptotic normality, we convert the $(m+1) \times (h+1)$ parameter matrices $\boldsymbol{\gamma}_0$ and $\hat{\boldsymbol{\gamma}}_n$ to vectors of length $(m+1) \times (h+1)$ of the form $\boldsymbol{\gamma} = (\gamma_{00}, \ldots, \gamma_{jk}, \ldots, \gamma_{mh})$, and redefine $\bar{L}_n(\boldsymbol{\gamma})$ and $\bar{\mathbf{S}}_n(\boldsymbol{\gamma})$ accordingly.

**Theorem 3.2.** *(Asymptotic normality) Suppose that the conditions of Theorem 3.1 are satisfied. If (i) $\boldsymbol{\gamma}_0$ lies in the interior of $\Gamma$; (ii) $\bar{L}_n(\boldsymbol{\gamma})$ is twice continuously differentiable in a neighborhood $\mathcal{N}$ of $\boldsymbol{\gamma}_0$; (iii) $\sqrt{n}\nabla_{\boldsymbol{\gamma}} \bar{L}_n\left(\boldsymbol{\gamma}_0\right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Omega})$; (iv) there exists a $\mathbf{H}(\boldsymbol{\gamma})$ that is continuous at $\boldsymbol{\gamma}_0$ and satisfies $\sup_{\boldsymbol{\gamma} \in \mathcal{N}} \left\| \nabla_{\boldsymbol{\gamma}\boldsymbol{\gamma}} \bar{L}_n(\boldsymbol{\gamma}) - \mathbf{H}(\boldsymbol{\gamma}) \right\| \xrightarrow{p} \mathbf{0}$, and (v) $\mathbf{H} = \mathbf{H}\left(\boldsymbol{\gamma}_0\right)$ is nonsingular, then $\sqrt{n}\left(\hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{H}^{-1}\boldsymbol{\Omega}\mathbf{H}^{-1}\right)$.*

For more details of the proof of Theorem 3.2, we refer readers to Theorem 3.1 in Newey and Mcfadden (1994). When $\bar{\mathbf{S}}_n(\boldsymbol{\gamma})$ is continuous and continuously differentiable, condition (ii) is satisfied. By applying a central limit theorem to $\bar{\mathbf{S}}_n(\boldsymbol{\gamma})$, we have condition (iii). Let $\bar{\mathbf{S}}_i(\boldsymbol{\gamma})$ be the $i$-th summand of $\bar{\mathbf{S}}_n(\boldsymbol{\gamma})$, $\boldsymbol{\Omega} = \mathrm{E}\left[\bar{\mathbf{S}}_i\left(\boldsymbol{\gamma}_0\right)\bar{\mathbf{S}}_i\left(\boldsymbol{\gamma}_0\right)^\top\right]$, and $\mathbf{H} = \mathrm{E}\left[\nabla_{\boldsymbol{\gamma}}\bar{\mathbf{S}}_i(\boldsymbol{\gamma})\right]|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}_0}$. $\boldsymbol{\Omega}$'s and $\mathbf{H}$'s rows and columns are indexed according to $\boldsymbol{\gamma}$, i.e., $(00, \ldots, jk, \ldots, mh)$. Then, the $st, s't'$ element of $\boldsymbol{\Omega}$ and $\mathbf{H}$ $(s, s' = 0, 1, \ldots, m;\ t, t' = 0, 1, \ldots, h)$ is expressed by

$$\begin{aligned}\mathbf{H}_{st,s't'} = \frac{\partial \bar{S}_n\left(\gamma_{st}\right)}{\partial \gamma_{s't'}} &= 2n^{-1}\sum_{i=1}^{n}\left(\mathbf{x}_i^\top \boldsymbol{\gamma}\mathbf{b}'(\tau_i)\right)^{-1} x_{is} b_{t'}\left(\tau_i\right)\{Y_i x_{is}\left[\left(1-2\tau_i\right) b_t\left(\tau_i\right)\right] \\ &\quad -\left[\sum_{j=0}^{m}\sum_{k=0}^{h} x_{is}x_{ij}\gamma_{jk}\left(1-2\tau_i\right) b_k(\tau_i) b_t(\tau_i)\right]\bigg\}\end{aligned} \tag{9}$$

and

$$\begin{aligned}\boldsymbol{\Omega}_{st,s't'} &= n^{-1}\sum_{i=1}^{n}\bar{S}_i(\gamma_{st})\bar{S}_i(\gamma_{s't'}) \\ &= 4n^{-1}\sum_{i=1}^{n}\left\{Y_i x_{is}\left[\bar{B}_t(1)+(2\tau_i-1)B_t(\tau_i)-2\bar{B}_t(\tau_i)\right]+\sum_{j=0}^{m}\sum_{k=0}^{h}x_{is}x_{ij}\gamma_{jk}\left[(1-2\tau_i)A_{kt}-\bar{A}_{kt}(1)+2\bar{A}_{kt}(\tau_i)\right]\right\} \\ &\quad \left\{Y_i x_{is'}\left[\bar{B}_{t'}(1)+(2\tau_i-1)B_{t'}(\tau_i)-2\bar{B}_{t'}(\tau_i)\right]+\sum_{j=0}^{m}\sum_{k=0}^{h}x_{is'}x_{ij}\gamma_{jk}\left[(1-2\tau_i)A_{k'}-\bar{A}_{kt'}(1)+2\bar{A}_{kt'}(\tau_i)\right]\right\},\end{aligned} \tag{10}$$

where $\mathbf{x}^\top\boldsymbol{\gamma}\mathbf{b}'(\tau)$ is the derivative of the conditional expectile function with respect to $\tau$. The problem of expectile crossing can be limited by choosing specific forms of $\mathbf{b}'(\tau)$ such that $\mathbf{x}^\top\boldsymbol{\gamma}\mathbf{b}'(\tau)$ is positive.

By estimating the asymptotic covariance matrix in Theorem 3.2, we can obtain the asymptotic standard errors and asymptotic confidence intervals (CIs) for the unknown parameters, as shown in our empirical analysis in Section 5.

## 4. Premium estimation

This section introduces the expectile risk measure into the premium calculation. Section 4.1 proposes an expectile-based individual risk premium and then uses the PER method as the second stage of the two-stage approach (TSPER) for expectile-based risk premium calculation. Section 4.2 introduces the bootstrap method to obtain a total risk premium and then proposes the EPP as a top-down allocation rule to distribute the total risk premium to each policy. We discuss the meaning of the expectile-based risk premium in the actuarial context in Section 4.3 and describe in Section 4.4 the relationship between the EPP and the EVPP for a better understanding of the proposed method.

#### 4.1. Two-stage parametric expectile regression for premium estimation

We first define some essential notations. Suppose that an insurance portfolio contains $n$ independent policyholders with the $i$-th ($i = 1, \ldots, n$) insured belonging to one of the $L$ risk classes. Let $n_l$ ($l = 1, \ldots, L$) denote the number of policies belonging to the $l$-th risk class, $n_l'$ of whom have positive aggregate claim amount. Denote the aggregate claim amount for the $i$-th policy by $Y_i$ and the corresponding number of claims by $N_i$ for $i = 1, \ldots, n$. Then, we have $n = \sum_{l=1}^{L} n_l$ and the total aggregate claim amount of the portfolio given by

$$Y = \sum_{l=1}^{L} \sum_{i=1}^{n_l} \mathbb{I}_{N_i} Y_i = \sum_{l=1}^{L} \sum_{i=1}^{n_l'} \tilde{Y}_i, \tag{11}$$

where the indicator random variable $\mathbb{I}_{N_i}$ denotes the event of whether the $i$-th policyholder makes at least one claim and $\tilde{Y}_i \triangleq Y_i | (N_i > 0)$ represents the positive truncated aggregate claim amount (severity) of the $i$-th policy. Generally, $\mathbb{I}_{N_i}$ is assumed Bernoulli distributed with the probability of submitting no claim over a one-year policy period being $p_i$ and independent from the severity $\tilde{Y}_i$.

Given that the size of the extreme claim amount is crucial for the evaluation of the riskiness of the policies, we now propose an expectile-based individual risk premium which can be calibrated with the help of the PER method in Section 3. Recall that the expectation of the aggregate claim amount over a one-year policy period for the $i$-th policyholder can be factorized into the product of the probability of having claims and the expected severity by using the double expectation formula, i.e.,

$$\mathrm{E}(Y_i) = (1 - p_i)\, \mathrm{E}(\tilde{Y}_i), \quad i = 1, \ldots, n. \tag{12}$$

Following the idea of Baione and Biancalana (2019) who propose a quantile-based method for premium calculation, we extend (12) to the framework of the expectile risk measure and propose an expectile-based risk premium for the $i$-th policy for a given expectile level $\tau$,

$$\mathrm{P}(Y_i) = (1 - p_i)\, \mathrm{PER}_\tau(\tilde{Y}_i), \quad i = 1, \ldots, n, \tag{13}$$

where $\mathrm{PER}_\tau(\tilde{Y}_i)$ represents the $\tau$-th expectile of the $i$-th severity $\tilde{Y}_i$.

As shown in (13), the individual risk premium is a product of the probability of having at least one claim and the expectiles of the severity for a given expectile level. In the following, we propose a TSPER method for its calculation, in which we can estimate $1 - p_i$ by the logistic regression in the first stage and obtain $\mathrm{PER}_\tau(\tilde{Y}_i)$ for any $\tau \in (0, 1)$ via a single PER in the second stage. We will discuss the actuarial meaning of the expectile-based risk premium (13) in Section 4.3.

##### 4.1.1. Logistic regression for the probability of having at least one claim

Given that the significance of claims varies among policies with different exposures, the exposure difference should be adjusted in statistical modeling. As the probability of having at least one claim is proportional to the exposure (De Jong and Heller, 2008), it is widely assumed that the probability of having at least one claim over the exposure $e_i$ is given by $1 - p_i^* = e_i \times (1 - p_i)$ (Heras et al., 2018; Baione and Biancalana, 2019; Kang et al., 2020; Kang et al., 2021; Baione and Biancalana, 2021; Hou, 2022), where $1 - p_i$ is the probability of having at least one claim over one year. This assumption will be further explained in Appendix A. In the first stage of the TSPER method, an exposure-adjusted logistic regression is used to model the probabilities of having claims given a set of covariates $\mathbf{x}_i$, which is expressed by

$$\mathrm{logit}\left((1 - p_i^*)/e_i\right) = \mathbf{x}_i^\top \boldsymbol{\beta}, \quad i = 1, \ldots, n, \tag{14}$$

where $\mathrm{logit}(x) = \ln \frac{x}{1-x}$ is the logit function and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_m)^\top$ is an unknown $(m+1)$-dimensional coefficient vector. By fitting (14), we can obtain the estimated coefficients $\hat{\boldsymbol{\beta}}$ and the probabilities of submitting at least one claim over a one-year policy period, i.e., $1 - \hat{p}_i = (1 - \hat{p}_i^*)/e_i$ for $i = 1, \ldots, n$. Policyholders belonging to the same risk class share the same $\hat{p}_i$.

##### 4.1.2. Parametric expectile regression for risk premium

In the second stage of the TSPER, we apply the PER method in Section 3 to regress the expectile of the severities on the covariates and model the regression coefficients hierarchically as parametric functions of the expectile level $\tau \in (0, 1)$, i.e.,

$$\mathrm{PER}_\tau(\tilde{Y}_i | \mathbf{x}_i, \boldsymbol{\gamma}) = exp\{\mathbf{x}_i^\top \boldsymbol{\varphi}(\tau, \boldsymbol{\gamma})\} = exp\{\mathbf{x}_i^\top \boldsymbol{\gamma}\, \mathbf{b}(\tau)\}, \quad i = 1, \ldots, n, \tag{15}$$

where the unknown parameters $\boldsymbol{\gamma}$ can be estimated by the quasi-Newton method as shown in Algorithm 1. The response variable $\tilde{Y}_i$ is transformed by logarithm in our paper. The choices of the "basis" $\mathbf{b}(\tau)$ must be defined in advance. As stated in Section 3, $\mathbf{b}(\tau)$ can be specified if a particular model is believed to be true; otherwise, it can be defined as a collection of known functions (such as polynomials and splines) such that $\mathbf{b}(\tau)$ induces a well-defined expectile function of the response for some $\boldsymbol{\gamma}$.

After obtaining the estimated parameters $\hat{\boldsymbol{\gamma}}$ in a single run, we are able to get the estimated expectile of the severity at an arbitrary expectile level $\tau \in (0, 1)$ simply by taking $\tau$ as an input of (4).

As shown in the following Section 4.2, the individual premium calculation requires estimating several conditional expectiles of the response, and the number of expectiles equals the number of risk classes. Though we can also use multiple times of ERs for estimating these expectiles in the second stage, it is more efficient to adopt a single PER for an estimated $\hat{\boldsymbol{\gamma}}$ and obtain the expectiles for the obtained expectile levels $\tilde{\tau}_i$'s. The PER reduces the crossing problem and also yields a parsimonious and efficient estimation for the coefficients, making it easier to interpret compared with standard expectile regression (Eilers, 2013; Schnabel and Eilers, 2013; Waltrup et al., 2015; Frumento and Bottai, 2016; Frumento and Salvati, 2021; Frumento et al., 2021; Fusco et al., 2022). The TSQR, TSER, and TSPQR can be derived by substituting the PER in the second stage of the TSPER with QR, ER, and PQR, respectively. The premium principle of the EVPP and SDPP can be found in Appendix A (Kaluszka, 2001; Young, 2004).

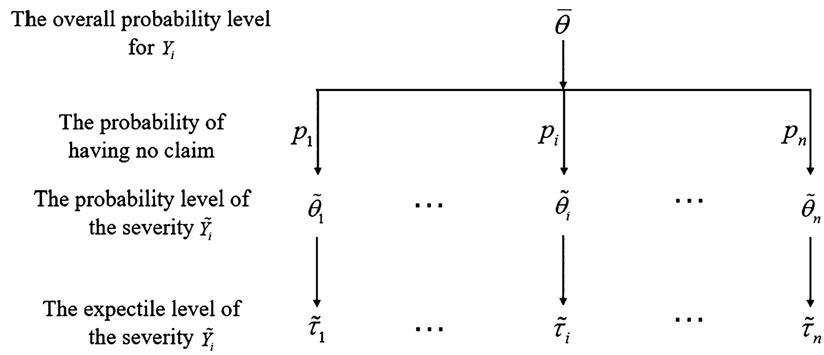


**Fig. 1.** The process to derive the expectile levels $\tilde{\tau}_1, \ldots, \tilde{\tau}_n$ for the severities $\tilde{Y}_1, \ldots, \tilde{Y}_n$ in (16) from the overall quantile level $\bar{\theta}$ for all $Y_i$ and the probabilities of having no claims $p_1, \ldots, p_n$.

### *4.2. Expectile premium principle*

This section introduces the method for obtaining the total risk premium and the premium principle to allocate the total to each policy.

To ensure collected risk premiums will be sufficient, to a specific probability, to cover the claim amount made by the insured, we first obtain a total risk premium $\pi(Y)$ for the portfolio by using the quantile of the bootstrap distribution of the total aggregate claim amount. Its quantile level $\kappa$ is generally chosen to be 99.5% according to Solvency II (EIOPA, 2009). Such a way of premium calculation has been widely applied in existing actuarial studies (see England and Verrall, 1999; England, 2002; Björkwall et al., 2009; Björkwall et al., 2010; Peremans et al., 2017).

To allocate a given total premium to each policy, we propose the EPP for a chosen quantile level $\bar{\theta}$ of all $Y_i$ by using the expectile-based individual risk premiums, which is expressed by

$$\pi(Y) = \sum_{l=1}^{L}\sum_{i=1}^{n_l} e_i\,\mathrm{P}(Y_i) = \sum_{l=1}^{L}\sum_{i=1}^{n_l} e_i(1-p_i)\,\mathrm{PER}_{\tilde{\tau}_i}(\tilde{Y}_i), \tag{16}$$

where $\pi(Y)$ is the total risk premium for the portfolio obtained by the bootstrap method, $\mathrm{P}(Y_1), \ldots, \mathrm{P}(Y_n)$ are the TSPER-based individual risk premiums, and $e_1, \ldots, e_n$ are the exposures of the policies. In (16), all of the $\tilde{\tau}_i$-th expectile of the severity $\tilde{Y}_i$ for $i = 1, \ldots, n$ are determined by the given $\bar{\theta}$-th quantile of the individual aggregate claim amounts $Y_i$ after adjusting the probability of having no claim for each risk class. This single quantile level $\bar{\theta}$ is chosen to balance (16).

Fig. 1 illustrates how to derive the expectile levels $\tilde{\tau}_i$'s of $\tilde{Y}_i$ in (16) from the quantile level $\bar{\theta}$ of all $Y_i$ and the probabilities of having no claims, $p_1, \ldots, p_n$, (which can be obtained by the first stage of the TSPER method) with the help of the intermediate parameters $\tilde{\theta}_i$'s, i.e., the quantile level of $\tilde{Y}_i$. Let $\mathrm{Q}_{\bar{\theta}}[Y_i|\mathbf{x}_i]$ be the $\bar{\theta}$-th conditional quantile of $Y_i$ given $\mathbf{x}_i$ and $F(\cdot)$ be the cumulative distribution function of the corresponding random variable. By applying the double expectation formula, the relationship between the quantile levels of $Y_i$ and $\tilde{Y}_i$ can be derived by (Heras et al., 2018; Baione and Biancalana, 2019; Baione and Biancalana, 2021)

$$F_{Y_i}(\mathrm{Q}_{\bar{\theta}}[Y_i|\mathbf{x}_i]) = p_i + (1-p_i)\,F_{\tilde{Y}_i}(\mathrm{Q}_{\bar{\theta}}[Y_i|\mathbf{x}_i]), \quad i = 1, \ldots, n. \tag{17}$$

Let $\tilde{\theta}_i \triangleq F_{\tilde{Y}_i}\left(\mathrm{Q}_{\bar{\theta}}[Y_i|\mathbf{x}_i]\right)$ for $i = 1, \ldots, n$. Then, for a given quantile level $\bar{\theta}$ of $Y_i$'s (i.e., $F_{Y_i}\left(\mathrm{Q}_{\bar{\theta}}[Y_i|\mathbf{x}_i]\right) = \bar{\theta}$ for all $i$), the quantile level of $\tilde{Y}_i$ can be obtained by

$$\tilde{\theta}_i = \frac{\bar{\theta} - p_i}{1 - p_i}, \quad i = 1, \ldots, n. \tag{18}$$

From (17) and (18), it is easy to see that

$$\mathrm{Q}_{\tilde{\theta}_i}(\tilde{Y}_i|\mathbf{x}_i) = F_{\tilde{Y}_i}^{-1}\left(\frac{\bar{\theta}-p_i}{1-p_i}\middle|\mathbf{x}_i\right) = \mathrm{Q}_{\bar{\theta}}(Y_i|\mathbf{x}_i), \quad i = 1, \ldots, n. \tag{19}$$

Since the quantile levels and the expectile levels are one-to-one corresponding to each other (Efron, 1991; Jones, 1994; Yao and Tong, 1996; Kuan et al., 2009), we can then derive the expectile level $\tilde{\tau}_i$ of $\tilde{Y}_i$ that corresponds to the $\tilde{\theta}_i$-th quantile of $\tilde{Y}_i$ for $i = 1, \ldots, n$ (i.e., $\mathrm{PER}_{\tilde{\tau}_i}(\tilde{Y}_i) = \mathrm{Q}_{\tilde{\theta}_i}(\tilde{Y}_i|\mathbf{x}_i)$ for all $i$). In sum, for a given quantile level $\bar{\theta}$ for all $\tilde{Y}_i$, we can find the corresponding expectile level $\tilde{\tau}_i$ of $\tilde{Y}_i$ for $i = 1, \ldots, n$ in (16).

An important difference between the proposed EPP and the other existing principles (Heras et al., 2018; Baione and Biancalana, 2021) is that the EPP offers a way to select the quantile level $\bar{\theta}$ based on the balance equation in (16) instead of using an arbitrary quantile level for all the policies. It can be easily seen from (18) and the one-to-one correspondence between $\tilde{\tau}_i$ and $\tilde{\theta}_i$ that $\tilde{\tau}_i$'s are positively proportional to $\bar{\theta}$ given $p_i$'s. Then, as $\bar{\theta}$ increases, all the expectile levels $\tilde{\tau}_i$'s raise and so do the individual risk premiums. Thus, we can find a unique quantile level $\bar{\theta}$ for all the policies by gradually increasing the candidate quantile level until the weighted sum of the TSPER-based individual risk premiums evaluated at the corresponding $\tilde{\tau}_i$'s in (16) covers the total risk premium for the portfolio. By assigning this single quantile level $\bar{\theta}$ for all $Y_i$, it guarantees that all the policies have the same probability $(1-\bar{\theta})$ that the claim amount exceeds the risk premium.

Algorithm 1 summarizes the main steps for premium calculation via the EPP: After obtaining a total risk premium for the portfolio by the bootstrap method, we obtain the building blocks ($\hat{p}_1, \ldots, \hat{p}_n$, and $\hat{\boldsymbol{\gamma}}$) for individual premium calculation in (13) via the TSPER method and then choose the quantile level $\bar{\theta}$ that balances (16), from which we can derive the expectile levels $\tilde{\tau}_i$'s and use them as inputs of the fitted PER model for individual risk premium calculation.

**Algorithm 1** Premium calculation via the Expectile Premium Principle.

***Step 1 – Total risk premium calculation via the bootstrap method***

Generate $B$ bootstrap samples of size $n$ for the aggregate claim amounts $\{Y_1^b, \ldots, Y_n^b\}_{b=1}^B$ and compute the sum of aggregate claim amounts for each bootstrap sample by $\pi^b(Y) = \sum_{i=1}^n Y_i^b$ for $b = 1, \ldots, B$.
Let the total risk premium $\pi(Y)$ be the $\kappa$-th empirical quantile of $\pi^1(Y), \ldots, \pi^B(Y)$.

***Step 2 – Individual risk premium allocation via the* TSPER *method***

Fit exposure-adjusted logistic model (14) to obtain the estimated probability of having claims $1 - \hat{p}_i$ for $i = 1, \ldots, n$.
Let $j = 0$. Initialize the starting point $\boldsymbol{\gamma}^0$, the step length $\alpha^0$, convergence tolerance $\varepsilon = 10^{-6}$, and the inverse of the approximate Hessian matrix $\Psi^0 = \boldsymbol{I}$, where $\boldsymbol{I}$ is an $m \times h$ identity matrix.
**while** $\|\bar{\mathbf{S}}_n(\boldsymbol{\gamma}^j)\| > \varepsilon$ **do** ▷ PER estimation via the quasi-Newton method
  Update $\boldsymbol{\gamma}^{j+1} \leftarrow \boldsymbol{\gamma}^j - \alpha^j \boldsymbol{\Psi}^j \bar{\mathbf{S}}_n(\boldsymbol{\gamma}^j)$ and let $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^{j+1}$;
  Update $\mathbf{s}^j \leftarrow \boldsymbol{\gamma}^{j+1} - \boldsymbol{\gamma}^j$ and $\Delta\bar{\mathbf{S}}_n^j \leftarrow \bar{\mathbf{S}}_n(\boldsymbol{\gamma}^{j+1}) - \bar{\mathbf{S}}_n(\boldsymbol{\gamma}^j)$;
  Update $\boldsymbol{\Psi}^{j+1}$ by using the Davidon-Fletcher-Powell formula
  $\boldsymbol{\Psi}^{j+1} \leftarrow \boldsymbol{\Psi}^j - \frac{\boldsymbol{\Psi}^j \Delta\bar{\mathbf{S}}_n^j \Delta\bar{\mathbf{S}}_n^{j\,T} \boldsymbol{\Psi}^j}{\Delta\bar{\mathbf{S}}_n^{j\,T} \boldsymbol{\Psi}^j \Delta\bar{\mathbf{S}}_n^j} + \frac{\mathbf{s}^j \mathbf{s}^{j\,T}}{\mathbf{s}^{j\,T} \Delta\bar{\mathbf{S}}_n^j}$;
  $\alpha^{j+1} \leftarrow \arg\min_{\alpha \geq 0} \bar{L}_n(\boldsymbol{\gamma}^{j+1} - \alpha \boldsymbol{\Psi}^{j+1} \bar{\mathbf{S}}_n(\boldsymbol{\gamma}^{j+1}))$;
  Set $j \leftarrow j + 1$.
**end while**
Choose the overall quantile level $\bar{\theta}$ for all the policies such that (16) holds and find the corresponding expectile levels $\tilde{\tau}_1, \ldots, \tilde{\tau}_n$ for severities $\tilde{Y}_1, \ldots, \tilde{Y}_n$ based on (18).
Calculate individual risk premiums $\mathrm{P}(Y_i) = (1 - \hat{p}_i)\widehat{\mathrm{PER}}_{\tilde{\tau}_i}(\tilde{Y}_i|\mathbf{x}_i) = (1 - \hat{p}_i) exp\{\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}} \mathbf{b}(\tilde{\tau}_i)\}$ for all $i$ by taking $\tilde{\tau}_1, \ldots, \tilde{\tau}_n$ as inputs of (15).

**Remark 2.** Although two-stage models are extensively adopted to calculate risk premiums (Kudryavtsev, 2009; Heras et al., 2018; Baione and Biancalana, 2019; Baione and Biancalana, 2021; Hou, 2022), all the existing methods, to the best of our knowledge, are based on the quantile risk measure, which does not account for the magnitude of severe losses. Additionally, the proposed EPP also differs from the existing premium calculation principles in the way of selecting the quantile level $\bar{\theta}$. In the existing literature, there are primarily two approaches for setting the quantile levels for premium calculation. One approach is to choose a single quantile level for each severity such that the sum of the individual risk premiums matches the total risk premium (see, e.g., Baione and Biancalana (2019)). However, since $p_i$ varies among risk classes, it is impossible to ensure that policies belonging to different risk classes have the same probability that the aggregate claim amount exceeds the risk premium. The other approach is to fix a single quantile level $\bar{\theta}$ of $Y_i$ for each policy. In the same fashion of (18), after adjusting for the probabilities of having at least one claim, a set of $\tilde{\theta}_i$'s for severities can be obtained (see, e.g., Heras et al. (2018); Baione and Biancalana (2021)) and then according to the Quantile Premium Principle, individual risk premiums are calculated by a weighted sum of the pure premium $\mathrm{E}(Y_i)$ and the quantile of the total claim amount $Q_{\bar{\theta}}(Y_i)$ (or the claim's variance $\mathrm{Var}(Y_i)$) with weights being called loading factors. However, unlike Baione and Biancalana (2019), the unique quantile level $\bar{\theta}$ in both Heras et al. (2018) and Baione and Biancalana (2021) are set with no basis. In this paper, the proposed EPP combines the advantages of these two approaches and chooses $\bar{\theta}$ for all $Y_i$ by using the balance equation in the premium principle, which not only simplifies the allocation procedure but also avoids the possible subjectivity in the quantitative analysis.

**Remark 3.** By using the proposed EPP framework, we calibrate the total risk premium and the individual risk premiums for a portfolio based on the existing claims in the portfolio. Similar to the existing premium calculation methods in the literature (Kudryavtsev, 2009; Heras et al., 2018; Baione and Biancalana, 2019; Kang et al., 2020; Baione and Biancalana, 2021; Kang et al., 2021; Hou, 2022; Hou et al., 2022), we do not consider new customers in this framework.

**Remark 4.** When the expectile-based methods are applied to insurance data, claims that incurred greater losses are given larger weights as shown in (2) and their characteristics are better modeled than smaller losses. Intuitively, the expectile-based methods outperform other methods with equal weights given to all observations (like the conventional mean regression) when data are generated from heavy-tailed distributions. The quantile regression (QR) method that is less affected by the outliers cannot achieve the same goal.

### *4.3. The actuarial meaning of the expectile-based premium*

To provide a better understanding of the expectile-based premium defined in (13), this section explains its actuarial meaning and shows the advantage of adopting the proposed expectile-based risk premium over the quantile-based counterparts.

Recall that under the EPP framework, we estimate the $\tilde{\tau}_i$-th expectile of the severity $\tilde{Y}_i$, i.e., $\mathrm{EVaR}_{\tilde{\tau}_i}(\tilde{Y}_i)$, as a building block of the expectile-based individual risk premium via the PER method in (15) for $i = 1, \ldots, n$. By the definition of expectile (Newey and Powell, 1987), the expectile risk measure $\mathrm{EVaR}_{\tilde{\tau}_i}(\tilde{Y}_i)$ for a given expectile level $\tilde{\tau}_i$ satisfies

$$\frac{\int_{-\infty}^{\mathrm{EVaR}_{\tilde{\tau}_i}(\tilde{Y}_i)} \left|\tilde{Y}_i - \mathrm{EVaR}_{\tilde{\tau}_i}(\tilde{Y}_i)\right| \mathrm{d}F(\tilde{Y}_i)}{\int_{-\infty}^{\infty} \left|\tilde{Y}_i - \mathrm{EVaR}_{\tilde{\tau}_i}(\tilde{Y}_i)\right| \mathrm{d}F(\tilde{Y}_i)} = \tilde{\tau}_i, \quad i = 1, \ldots, n. \tag{20}$$

Let $\tilde{Y}_{i+} = \max\{\tilde{Y}_i - \mathrm{EVaR}_{\tilde{\tau}_i}(\tilde{Y}_i), 0\}$ and $\tilde{Y}_{i-} = \max\{\mathrm{EVaR}_{\tilde{\tau}_i}(\tilde{Y}_i) - \tilde{Y}_i, 0\}$ for the $i$-th severity. Denote $\tilde{Y}_i' = (1 - p_i)\tilde{Y}_i - P(Y_i)$, where the first term can be viewed as the claim amount of the $i$-th policy after adjusting for the probability of having claims of the $i$-th policyholder and $P(Y_i)$ is the expectile-based risk premium defined in (13). Then, $P(Y_i)_- = \max\{\tilde{Y}_i', 0\}$ represents how much the risk premium $P(Y_i)$

is lower than the $p_i$-adjusted aggregate claim amount, suggesting the loss of the insurer for the $i$-th policy, and $P(Y_i)_+ = \max\{-\tilde{Y}_i', 0\}$ denotes the amount of the $i$-th risk premium that is greater than the $p_i$-adjusted aggregate claim amount, representing the "redundant" premium the $i$-th policyholder pays or the profit of the insurer for the $i$-th policy. If the expectile-based risk premium (13) is collected for the $i$-th policy, then the "gain-loss" ratio for the insurer can be calculated as

$$\frac{\mathrm{E}[P(Y_i)_+]}{\mathrm{E}[P(Y_i)_-]} = \frac{\mathrm{E}[\max\{-\tilde{Y}_i', 0\}]}{\mathrm{E}[\max\{\tilde{Y}_i', 0\}]} = \frac{\mathrm{E}[\max\{(1-p_i)(\mathrm{EVaR}_{\tilde{\tau}_i}(\tilde{Y}_i) - \tilde{Y}_i), 0\}]}{\mathrm{E}[\max\{(1-p_i)(\tilde{Y}_i - \mathrm{EVaR}_{\tilde{\tau}_i}(\tilde{Y}_i)), 0\}]} = \frac{\mathrm{E}[\tilde{Y}_{i-}]}{\mathrm{E}[\tilde{Y}_{i+}]} = \frac{\tilde{\tau}_i}{1-\tilde{\tau}_i}, \quad i = 1, \ldots, n, \tag{21}$$

where the last equality follows from (20). It is easy to see from (21) that as the expectile level $\tilde{\tau}_i$ increases, the "gain-loss" ratio raises and the loss caused by high-risk policyholders is relatively lower. Unfortunately, in practice, $\tilde{\tau}_i$ cannot be set too high because a relatively high "redundant" premium $\mathrm{E}[Y_{i-}]$ may lead to customer churn. Thus, there is a trade-off between the customer churn risk and the insurer's shortfall risk. By collecting the expectile-based risk premium for a higher expectile level $\tilde{\tau}_i$, the insurer chooses to face a relatively lower loss with increased potential for customer churn.

This paper adopts the expectile-based risk premium for the premium principle and thus differs from the quantile-based risk premium applied in the Quantile Premium Principle (QPP, Heras et al., 2018; Baione and Biancalana, 2019; Baione and Biancalana, 2021). For the quantile-based risk measure Value at Risk (VaR) (Khorshidi and Ghezavati, 2019) with level $\tilde{\theta}_i \in (0, 1)$, i.e., the possible maximum aggregate claim amount over the exposure with probability $\tilde{\theta}_i$ (Kuan et al., 2009), we can derive from its definition that

$$\frac{\int_{-\infty}^{\mathrm{VaR}_{\tilde{\theta}_i}(\tilde{Y}_i)} \mathrm{d}F(\tilde{Y}_i)}{\int_{-\infty}^{\infty} \mathrm{d}F(\tilde{Y}_i)} = \tilde{\theta}_i, \quad i = 1, \ldots, n. \tag{22}$$

If the insurer collects the quantile-based risk premium $P(Y_i) = (1-p_i)\,\mathrm{VaR}_{\tilde{\theta}_i}(\tilde{Y}_i)$ (Baione and Biancalana, 2019) for the $i$-th policy obtained via the Quantile Premium Principle (Heras et al., 2018), then we can obtain that

$$\frac{P(\tilde{Y}_i' \le 0)}{P(\tilde{Y}_i' > 0)} = \frac{P((1-p_i)\tilde{Y}_i \le P(Y_i))}{P((1-p_i)\tilde{Y}_i > P(Y_i))} = \frac{P(\tilde{Y}_i \le \mathrm{VaR}_{\tilde{\theta}_i}(\tilde{Y}_i))}{P(\tilde{Y}_i > \mathrm{VaR}_{\tilde{\theta}_i}(\tilde{Y}_i))} = \frac{\tilde{\theta}_i}{1-\tilde{\theta}_i}, \quad i = 1, \ldots, n, \tag{23}$$

where the notations remain the same as those in (21) except for the quantile-based risk premium $P(Y_i)$ and the second equality of (23) is derived from (22). $P(\tilde{Y}_i' > 0)$ can be viewed as the probability of the risk premium being lower than the $p_i$-adjusted aggregate claim amount, representing the shortfall risk; and $P(\tilde{Y}_i' \le 0)$ denotes the probability of the risk premium being greater, which relates to the customer churn risk. Then, (23) suggests that there also exists a trade-off between the shortfall risk and the customer churn risk for the quantile-based risk premium. However, unlike the expectile-based risk premium, the quantile-based risk premium at a particular quantile level can only reflect a balance between the probabilities of loss and gain but not their magnitude. This suggests that the quantile-based risk premium cannot sufficiently reveal the riskiness of the policies and may lead to great losses for the insurers. Compared with the above quantile-based risk premium, the expectile-based one used in the EPP takes a step further by taking account of both the probability of the extreme events and their magnitudes (Kuan et al., 2009) simultaneously. The EPP may be favored by insurers who are more apprehensive about catastrophic events that could cause great losses.

### 4.4. The relationship between EPP and EVPP

For a better understanding of the proposed EPP, we introduce the loading factor in this subsection to show that the proposed EPP is actually equivalent to the EVPP applied for each risk class with a set of chosen risk class-specific loading factors. It is worth pointing out that the loading factors are determined after balancing (16) and are not involved in the process of calculating the expectile-based risk premium.

Recall that the application of the expected value premium principle generally involves choosing a loading factor to balance $\pi(Y) = \sum_{l=1}^{L}\sum_{i=1}^{n_l} e_i\,\mathrm{P}_{\mathrm{EVPP}}(Y_i) = \sum_{l=1}^{L}\sum_{i=1}^{n_l} e_i(1+\alpha)\,\mathrm{E}(Y_i)$, where the loading factor $\alpha$ quantifies the level of the risk premium $\mathrm{P}_{\mathrm{EVPP}}(Y_i)$ higher than the pure premium $\mathrm{E}(Y_i) = (1-p_i)\,\mathrm{E}(\tilde{Y}_i)$ for all $i$. $\alpha = (\mathrm{P}(Y_i) - \mathrm{E}(Y_i))/\,\mathrm{E}(Y_i)$ is usually used for evaluation of the profit made by the insurer for risk bearing and regulatory purposes (IAA, 2009; EIOPA, 2009; IFRS, 2017). Then, for given expectile-based risk premiums, the loading factor for all policies (indexed by $i$) belonging to the $l$-th risk class, $\alpha_l$, $l = 1, \ldots, L$, can be obtained by

$$1 + \alpha_l = \frac{\mathrm{P}(Y_i)}{\mathrm{E}(Y_i)} = \frac{(1-p_i)\,\mathrm{PER}_{\tilde{\tau}_i}(\tilde{Y}_i)}{(1-p_i)\,\mathrm{E}(\tilde{Y}_i)} = \frac{\mathrm{PER}_{\tilde{\tau}_i}(\tilde{Y}_i)}{\mathrm{E}(\tilde{Y}_i)} = \frac{e^{\mathbf{x}_i^\top \boldsymbol{\varphi}(\tau,\boldsymbol{\gamma})}}{e^{\mathbf{x}_i^\top \boldsymbol{\beta}}} = e^{\mathbf{x}_i^\top(\boldsymbol{\varphi}(\tau,\boldsymbol{\gamma}) - \boldsymbol{\beta})}. \tag{24}$$

This implies that the expectile-based loading factors vary for different risk classes and are determined by the riskiness of the corresponding risk classes. Then, the expectile-based risk premiums calculated by the EPP can be viewed as applying the EVPP method for each risk class with the loading factor being $\alpha_l$ for the $l$-th risk class in (24). Rather than fixing a unique loading factor for all risk classes as the EVPP does, the EPP allows for a more precise description of the distribution of loss for each risk class by varying the loading factor across risk classes.

In addition, as $\alpha_l$ is unit-free, we also interpret the EPP-based loading factor for each risk class as a measure of relative "risk" with respect to the pure premium for policies belonging to the risk class (Risk Margin Working Group, 2009). It enables us to compare across different risk classes, offering some knowledge of the volatility of loss across the risk class and reflecting the market price an insurer would offer for bearing the risk (Olivieri and Pitacco, 2011). The higher the loading factor is, the higher the risk an insurer will bear for the extreme loss incurred by the policies belonging to the corresponding risk class.

As in (24), the loading factors for the TSER, TSPQR, and TSQR methods can also be calculated with the corresponding risk premiums estimated via the ER, PQR, and QR methods, respectively, and interpreted in the same fashion. However, the loading factors for EVPP and SDPP methods are constant across different risk classes and thus cannot contribute to the comparison of the riskiness across risk classes.

**Table 1**
The notations in the automobile insurance dataset.

| Notations | Type | Description |
|---|---|---|
| $Y_i$ | Continuous | The claim amount of the $i$-th policy ($) |
| $e_i$ | Continuous | The exposure of the $i$-th policy |
| $A$ | Categorical | The levels of drivers' age: 1 (youngest), 2, 3, 4, 5, 6 |
| $V$ | Categorical | The levels of the age of the vehicle: 1 (youngest), 2, 3, 4 |

**Table 2**
The parameter estimates of logistic regression in the first stage of the TSQR, TSPQR, TSER, and TSPER methods.

| Probability model | Parameter | Estimation | S.E. | $z$ value | $p$ value | |
|---|---|---|---|---|---|---|
| | Intercept | -1.406 | 0.062 | -22.619 | 0.000 | *** |
| | $V = 2$ | 0.031 | 0.051 | 0.620 | 0.535 | |
| | $V = 3$ | -0.095 | 0.050 | -1.927 | 0.054 | . |
| | $V = 4$ | -0.190 | 0.051 | -3.734 | 0.000 | *** |
| logistic regression | $A = 2$ | -0.199 | 0.064 | -3.111 | 0.002 | ** |
| | $A = 3$ | -0.260 | 0.062 | -4.193 | 0.000 | *** |
| | $A = 4$ | -0.303 | 0.062 | -4.880 | 0.000 | *** |
| | $A = 5$ | -0.533 | 0.068 | -7.800 | 0.000 | *** |
| | $A = 6$ | -0.536 | 0.077 | -6.925 | 0.000 | *** |

Notes: Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1, ' ' 1.

**Table 3**
MSE, Bias, and the Gini index for different choices of $\mathbf{b}(\tau)$ of the TSPER method. The standard errors of the Gini indices are given in parentheses. The boldface denotes the best performance.

| Choice of $b(\tau)$ | $b_1(\tau)$ | $b_2(\tau)$ | $b_3(\tau)$ | $b_4(\tau)$ | $b_5(\tau)$ |
|---|---|---|---|---|---|
| MSE | 0.30 | **0.21** | 0.27 | 0.43 | 0.30 |
| Bias | 0.40 | **0.28** | 0.38 | 0.50 | 0.41 |
| Gini index | 2.16 | **2.36** | 2.21 | 2.25 | 2.26 |
| | (1.79) | (1.75) | (1.71) | (1.76) | (1.74) |

## 5. Real data analysis

This section applies the proposed premium calculation method to a real-world dataset and compares the TSPER with the TSQR (Heras et al., 2018; Baione and Biancalana, 2019), TSPQR (Baione and Biancalana, 2021), and TSER methods based on the Gini index (Frees et al., 2011).

### 5.1. Dataset

We apply the proposed method for the analysis of the 'dataCar' dataset in R package *insuranceData*. The dataset provides 67856 vehicle insurance policies from 2004 to 2005 from an Australian automobile insurance company. Table 1 presents the notations used for this dataset. For simplification, we only consider the effects of the two most significant categorical rating factors, i.e., the driver's age ($A$) and the age of the vehicle ($V$), on the log of the aggregate claim amount submitted by each policyholder. There are 6 levels for the driver's age ($A = 1, \dots, 6$) and 4 levels for the age of the vehicle ($V = 1, \dots, 4$), resulting in $L = 6 \times 4 = 24$ risk classes.

### 5.2. Model estimation

In the following, we aim to calibrate the total risk premium and the risk premiums for all the policies.

We first calculate the total risk premium $\pi(Y)$ for the portfolio to ensure that the probability of the total loss being less than the total risk premium is sufficiently large. To obtain such a total risk premium, we draw $B = 10,000$ bootstrap samples for the aggregate claim amounts from the original dataset. Based on the Solvency II (EIOPA, 2009), we set the 99.5% quantile of the bootstrap samples as the total risk premium, i.e., $\pi(Y) = \text{VaR}_{99.5\%}(Y) = 10,031,705$. Such a total premium ensures the probability of the total loss being greater than the total risk premium for the portfolio to be less than 0.5%.

To allocate the total risk premium to policies such that the same quantile level is used for all the policies, we apply the TSPER method to obtain the building blocks ($\hat{p}_1, \dots, \hat{p}_n$, and $\hat{\boldsymbol{\gamma}}$) for individual premium calculation in (16). Table 2 reports the coefficient estimates of the logistic regression in the first stage of the TSPER method. These estimates are the same for all six two-stage methods mentioned above, as they share the same logistic regression in the first stage. The regression parameters for the risk classes $V = 2$ and $V = 3$ in logistic regression are not significant (similar results are also derived for Gamma regression in Table 5), i.e., their p-values are higher than 0.05, suggesting that these levels could be aggregated to the base level ($V = 1, A = 1$). However, we leave these levels separated for our analysis.

We apply the PER method with the form of $\mathbf{b}(\tau)$ selected from a list of commonly used candidates for parametric modeling of the coefficient function (Frumento and Bottai, 2016; Baione and Biancalana, 2021), which includes $\mathbf{b_1}(\tau) = (1, \tau)^\top$, $\mathbf{b_2}(\tau) = (1, \tau, \tau^2)^\top$, $\mathbf{b_3}(\tau) = (1, \tau, \tau^2, \tau^3)^\top$, and the shifted Legendre polynomials (Abramowitz and Stegun, 1965) of degree 2 and 3, i.e., $\mathbf{b_4}(\tau) = (1, 2\tau - 1, 6\tau^2 - 6\tau + 1)^\top$ and $\mathbf{b_5}(\tau) = (1, 2\tau - 1, 6\tau^2 - 6\tau + 1, 20\tau^3 - 30\tau^2 + 12\tau - 1)^\top$. Table 3 reports MSE, Bias, and the Gini index with the benchmark model chosen as the TSER (the bigger the Gini index is, the better the competing model performs compared with the

**Table 4**
The parameter estimates of the PER method for the automobile insurance dataset.

$$\hat{\boldsymbol{\gamma}} = \begin{pmatrix} \gamma_{00} & \dots & \gamma_{80} \\ \gamma_{01} & \dots & \gamma_{81} \\ \gamma_{02} & \dots & \gamma_{82} \end{pmatrix}^{\top} = \begin{pmatrix} 5.692 & 0.036 & 0.090 & 0.124 & -0.034 & -0.044 & -0.060 & -0.073 & -0.047 \\ 1.978 & 0.089 & 0.141 & 0.245 & -0.378 & -0.372 & -0.407 & -0.459 & -0.318 \\ 0.530 & -0.026 & -0.061 & -0.110 & 0.132 & -0.066 & 0.124 & 0.077 & -0.064 \end{pmatrix}^{\top}$$

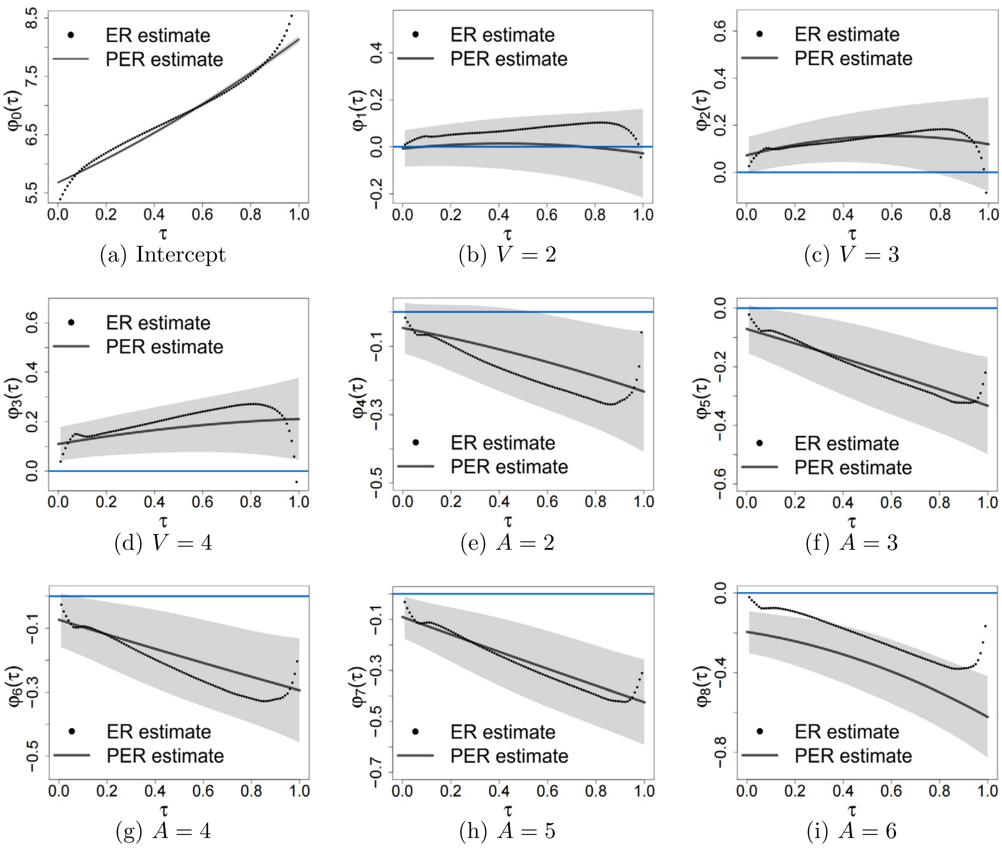


**Fig. 2.** ER parameter estimates (black dots) of $\varphi_j(\tau)$, $j = 0, \dots, 8$ for each of the expectile level $\tau = 0.01, 0.02, \dots, 0.99$, PER parameter estimates (black line) of $\varphi_j(\tau|\boldsymbol{\gamma})$, $j = 0, \dots, 8$, and the latter's 95% CIs (gray band) for $\tau \in (0, 1)$. The blue horizontal line represents $\hat{\boldsymbol{\varphi}}(\tau) = 0$. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

benchmark; we will discuss the Gini index in detail in Section 5.3) for different choices of $\mathbf{b}(\tau)$. It is not surprising to find that the Gini indices for different choices of $\mathbf{b}(\tau)$ in Table 3 are nearly indistinguishable. In fact, this could happen even for very different model specifications (we refer the reader to Section 3.5 of Frumento et al. (2021) for a detailed discussion). As pointed out by Frumento et al. (2021), this suggests parsimony and interpretability should be taken into account when selecting models. For simplicity reasons and by noticing that the TSPER with $\mathbf{b_2}(\tau)$ also outperforms other models in terms of MSE and Bias, we adopt $\mathbf{b_2}(\tau)$ for the TSPER as well as for the TSPQR method in the following analysis.

Table 4 reports the estimated $\hat{\boldsymbol{\gamma}}$ of the PER method and results for the QR, PQR, and ER methods are relegated to Appendix A. Based on the estimated $\hat{\boldsymbol{\gamma}}$, we find that the signs of the covariate effects $\hat{\boldsymbol{\varphi}}(\tau|\boldsymbol{\gamma})$ estimated by the PER and ER methods are the same except for a few very high expectiles. The parameters for all the levels of $V$ and $A$ are positive and negative, respectively.

Fig. 2 shows how the covariate effects $\hat{\boldsymbol{\varphi}}(\tau|\boldsymbol{\gamma})$ estimated by the PER method (black lines) vary across the expectile level $\tau \in (0, 1)$ with a comparison of the ER estimates (dots) obtained at multiple expectile levels $\tau = 0.01, 0.02, \dots, 0.99$. The blue line represents $\hat{\boldsymbol{\varphi}}(\tau) = 0$. We also obtain the asymptotic standard errors of PER estimates by using the asymptotic covariance matrix in Theorem 3.2 and plot their asymptotic CIs with gray bands in Fig. 2. It shows that the estimates obtained by the two methods are very close to each other and the CIs of the PER estimates cover most of the ER estimates. Although ER estimates exhibit a steady increase (decrease) for moderate expectile levels, they suffer from high volatility in the tails. Whereas, the PER method yields a more smooth trend across all expectile levels, especially in the extremes, which facilitates the interpretation of the covariate effects. In addition, the PER estimates are computed with more computational efficiency. From Fig. 2(a), we observe an increasing trend of the estimated intercept, i.e., the PER estimate for the base risk class $V = 1, A = 1$ increases as the expectile level raises. Fig. 2(b) shows the difference between the PER estimates for risk class $V = 2, A = 1$ and those for the base class across the expectile levels, and the subsequent subplots can be explained in the same fashion. We also find that the regression coefficients of all levels of $A$ are bigger than those of $V$, suggesting that the drivers' age has a stronger effect on the risk premiums than the age of the vehicle.

Fitting the QR and PQR methods can be easily realized using the R package *quantreg* and *qrcm*, respectively. Coefficient estimates via the QR are also similar to those via the PQR methods.

**Table 5**
The parameter estimates of Gamma regression for the calculation of the loading factor.

| Amount model | Parameter | Estimation | S.E. | $t$ value | $p$ value | |
|---|---|---|---|---|---|---|
| | Intercept | 7.808 | 0.095 | 81.851 | 0.000 | *** |
| | $V=2$ | 0.051 | 0.079 | 0.646 | 0.518 | |
| | $V=3$ | 0.078 | 0.078 | 1.005 | 0.315 | |
| | $V=4$ | 0.169 | 0.080 | 2.106 | 0.035 | * |
| Gamma regression | $A=2$ | -0.224 | 0.098 | -2.288 | 0.022 | * |
| | $A=3$ | -0.335 | 0.095 | -3.517 | 0.000 | *** |
| | $A=4$ | -0.320 | 0.095 | -3.353 | 0.001 | *** |
| | $A=5$ | -0.439 | 0.106 | -4.124 | 0.000 | *** |
| | $A=6$ | -0.355 | 0.122 | -2.918 | 0.004 | ** |

Notes: Signif. codes: '***' 0.001, '**' 0.01, '*' 0.05, '.' 0.1, ' ' 1.

To apply the premium principles, we obtain the loading factors, i.e., $\alpha_{\text{EVPP}} = 8.32\%$ for the EVPP and $\alpha_{\text{SDPP}} = 1.61\%$ for the SDPP (see Heras et al. (2018) and Appendix A for details), and the unique quantile levels $\bar{\theta}$ that balance $\pi(Y)$ and the sum of the risk premiums for each policyholder, i.e., 96.32% for the TSQR, 96.12% for the TSPQR, 99.52% for the TSER, and 99.71% for the TSPER.

Individual premium calculation via the TSER or TSQR method requires fitting the 24 times of the ERs or QRs at different quantile levels in the second stage with $4 + 6 - 1 = 9$ unknown parameters for a single run and a total of $24 \times 9 = 216$ parameters to be estimated. However, this onerous and computationally demanding task can be greatly simplified by applying the TSPER or the TSPQR method, for which we only need to fit a single PER or PQR given $\mathbf{b_2}(\tau)$ in the second stage with $3 \times 9 = 27$ parameters. The parametric approaches will be more parsimonious and efficient as the number of risk classes increases.

Table 6 summarizes the probabilities of having at least one claim and the risk premiums for each risk class calculated by the six two-stage methods, respectively. The probabilities of having claims for the risk classes $V=2$ are higher than those of other classes and their risk premiums are the highest among all the classes. We estimate the pure premium $\text{E}(Y_i), i = 1, \ldots, n$ via Gamma regression with logarithmic link function and the corresponding parameter estimates are reported in Table 5. The six approaches utilize different features of the data and yield different premium calculations. The TSPER-based risk premiums are significantly higher for the risk classes $A=1$ and lower for $A=5$ and $A=6$ than other classes, which indicates again that these drivers' age groups have a stronger influence on the risk premiums than other rating factors. The lowest risk premiums calculated by the six methods are different. EVPP and SDPP select $V=3, A=5$ as the least risky class, TSQR and TSPQR select $V=4, A=5$, and TSER and TSPER select $V=1, A=5$. TSPQR-based risk premiums have the biggest range (145.22, 845.05), SDPP-based risk premiums have the smallest range (217.23, 559.75), and the ranges of expectile-related risk premiums are in the middle.

To assess the relative risk across different risk classes, Table 6 reports the pure premiums and the risk class-specific loading factors calculated for the four two-stage methods (the loading factors of the EVPP and SDPP methods are fixed for all risk classes). The risk classes $A=1$ or $A=3$ have higher positive loading factors than other classes for the four methods, indicating higher volatility in the aggregate claim amount for policyholders belonging to these classes. Most loading factors calculated by TSPER are positive, except for $V=1, A=6$. The negative values also exist for other methods. The expectations of the aggregate claim amounts for these risk classes are higher than the extreme quantiles/expectiles, suggesting that the policyholders belonging to these risk classes are much more right-skewed than the others.

We find that the pure premium for class $V=1, A=1$ is similar to that for class $V=3, A=1$ (484.58 versus 484.98), but the latter has a larger TSPER-based risk premium and a larger loading factor (517.62 versus 560.21, 6.94% versus 15.64%, respectively). This implies that there is higher volatility in the aggregate claim amount, i.e., higher uncertainty involved in the claim process, arising from policyholders belonging to class $V=3, A=1$ compared with class $V=1, A=1$. We also observe that though policyholders belonging to the classes $V=3, A=2$ and $V=3, A=4$ have similar risk margins (i.e., 26.88 and 29.70, respectively), the pure premium for the former class is higher (328.44 versus 273.17) and thus a higher TSPER-based risk premium is collected (355.32 versus 302.87). This indicates that policyholders belonging to class $V=3, A=2$ have larger average aggregate claim amounts with relatively lower uncertainty involved in the claim process compared with class $V=3, A=4$, which results in a lower loading factor for class $V=3, A=2$. Compared with the loading factor (8.30%) for class $V=3, A=2$, the higher loading factor (10.99%) for $V=3, A=4$ suggests that these policyholders have to pay a larger percentage of the average aggregate claim amount for the higher underlying risk exhibited in their claim process.

### *5.3. Model comparison using the Gini index*

We compare the performance of the proposed TSPER with those of EVPP, SDPP, TSQR, TSER, and TSPQR using the ordered Lorenz curve and the Gini index in the empirical study (Frees et al., 2011), which are commonly used criteria for the comparison of premium pricing strategies (Frees et al., 2014; Shi and Yang, 2018; Verschuren, 2021). The calculation of the Gini index requires choosing a model as the benchmark model and setting other models as the competing models. We calibrate the risk premiums using the six methods and set each of these six methods as the benchmark model in turn. By calculating the ratio of the risk premium of each competing model to the benchmark model, we yield the 'relativity' whose order provides information about the vulnerability of the benchmark model relative to the competing model and is used to generate the ordered Lorenz curve (Frees et al., 2011).

Fig. 3 displays four ordered Lorenz curves with TSQR, TSPQR, TSER, and TSPER being the benchmark model, respectively. When the estimated ordered Lorenz curve is below (above) the line of equality (the 45-degree line in black), the corresponding competing model is better (worse) than the benchmark. In the first subplot of Fig. 3, we set the TSQR method as the benchmark, in which case we observe that it is not as competitive as the other three methods for the premium calculation. The TSPQR method is shown to have a similar performance to the TSQR method and is no better than the TSER and TSPER methods. From Fig. 3, we find the proposed TSPER method is the best premium calculation method among these four methods.

We then calculate two times the area between the line of equality and the Lorenz curves, which is known as the Gini index, and provide the results in Table 7. The positive Gini index in Table 7 suggests that the competing model performs better than the benchmark

**Table 6**
The probabilities of submitting at least one claim, the probability/expectile levels, the pure premiums, and the risk premiums calculated by EVPP, SDPP, TSQR, TSPQR, TSER, TSPER, and their corresponding loading factors for 24 risk classes.

| Risk class | Probability of claims $1-\hat{p}_i$ | Probability/expectile level | | | | Pure premium $\mathrm{E}(Y_i)$ | Risk premium | | | | | | Loading factor | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | QR $\hat{\hat{\theta}}_i$ | PQR $\hat{\hat{\theta}}_i$ | ER $\hat{\hat{\tau}}_i$ | PER $\hat{\hat{\tau}}_i$ | | EVPP $\alpha=8.32\%$ | SDPP $\alpha=1.61\%$ | TSQR $\bar{\theta}=96.32\%$ | TSPQR $\bar{\theta}=96.12\%$ | TSER $\bar{\theta}=99.52\%$ | TSPER $\bar{\theta}=99.71\%$ | TSQR | TSPQR | TSER | TSPER |
| $V=1, A=1$ | 19.69% | 0.813 | 0.803 | 0.864 | 0.891 | 484.58 | 524.92 | 519.20 | 652.09 | 684.29 | 531.36 | 516.92 | 34.57% | 41.21% | 9.65% | 6.94% |
| $V=2, A=1$ | 20.19% | 0.818 | 0.808 | 0.865 | 0.892 | 522.88 | 566.41 | 559.75 | 853.30 | 845.05 | 606.50 | 585.60 | 63.19% | 61.61% | 15.99% | 12.12% |
| $V=3, A=1$ | 18.23% | 0.798 | 0.787 | 0.859 | 0.887 | 484.98 | 525.35 | 521.07 | 805.71 | 756.06 | 575.53 | 560.21 | 66.13% | 55.90% | 18.67% | 15.64% |
| $V=4, A=1$ | 16.86% | 0.782 | 0.770 | 0.854 | 0.883 | 491.20 | 532.09 | 529.26 | 753.78 | 711.24 | 569.46 | 558.80 | 53.46% | 44.80% | 15.93% | 13.89% |
| $V=1, A=2$ | 16.73% | 0.780 | 0.768 | 0.853 | 0.883 | 329.07 | 356.47 | 354.68 | 318.07 | 340.80 | 329.82 | 329.49 | -3.34% | 3.56% | 0.23% | 0.23% |
| $V=2, A=2$ | 17.17% | 0.786 | 0.774 | 0.855 | 0.884 | 355.42 | 385.00 | 382.70 | 405.28 | 417.48 | 377.31 | 373.05 | 14.03% | 17.46% | 6.16% | 5.07% |
| $V=3, A=2$ | 15.45% | 0.762 | 0.749 | 0.848 | 0.878 | 328.44 | 355.78 | 355.09 | 379.77 | 368.06 | 356.71 | 355.32 | 15.63% | 12.06% | 8.61% | 8.30% |
| $V=4, A=2$ | 14.25% | 0.742 | 0.728 | 0.843 | 0.874 | 331.81 | 359.43 | 359.87 | 374.54 | 341.51 | 351.92 | 353.27 | 12.88% | 2.92% | 6.06% | 6.58% |
| $V=1, A=3$ | 15.90% | 0.768 | 0.756 | 0.850 | 0.880 | 279.83 | 303.12 | 302.19 | 276.27 | 295.05 | 294.11 | 293.69 | -1.27% | 5.44% | 5.10% | 5.07% |
| $V=2, A=3$ | 16.32% | 0.774 | 0.762 | 0.852 | 0.881 | 302.31 | 327.48 | 326.14 | 346.50 | 360.40 | 336.57 | 332.54 | 14.62% | 19.21% | 11.33% | 10.12% |
| $V=3, A=3$ | 14.66% | 0.749 | 0.735 | 0.845 | 0.876 | 279.08 | 302.31 | 302.34 | 323.96 | 316.28 | 318.23 | 316.47 | 16.08% | 13.33% | 14.03% | 13.52% |
| $V=4, A=3$ | 13.52% | 0.728 | 0.713 | 0.839 | 0.871 | 281.74 | 305.19 | 306.23 | 323.14 | 292.14 | 313.97 | 314.46 | 14.69% | 3.69% | 11.44% | 11.73% |
| $V=1, A=4$ | 15.33% | 0.760 | 0.747 | 0.848 | 0.878 | 274.05 | 296.86 | 296.36 | 236.18 | 265.21 | 278.31 | 281.38 | -13.82% | -3.22% | 1.55% | 2.79% |
| $V=2, A=4$ | 15.75% | 0.766 | 0.754 | 0.849 | 0.880 | 296.12 | 320.77 | 319.90 | 299.86 | 323.46 | 318.77 | 318.70 | 1.26% | 9.23% | 7.65% | 7.74% |
| $V=3, A=4$ | 14.14% | 0.740 | 0.726 | 0.842 | 0.874 | 273.17 | 295.91 | 296.37 | 268.92 | 282.27 | 300.66 | 302.87 | -1.56% | 3.33% | 10.06% | 10.99% |
| $V=4, A=4$ | 13.03% | 0.718 | 0.702 | 0.836 | 0.869 | 275.65 | 298.59 | 300.06 | 257.57 | 259.40 | 296.14 | 300.62 | -6.56% | -5.89% | 7.43% | 9.18% |
| $V=1, A=5$ | 12.58% | 0.708 | 0.692 | 0.834 | 0.867 | 199.67 | 216.29 | 217.68 | 153.49 | 158.40 | 198.99 | 203.92 | -23.13% | -20.67% | -0.34% | 2.24% |
| $V=2, A=5$ | 12.93% | 0.715 | 0.700 | 0.836 | 0.868 | 215.94 | 233.92 | 235.14 | 188.09 | 191.15 | 228.27 | 231.10 | -12.90% | -11.48% | 5.71% | 7.14% |
| $V=3, A=5$ | 11.57% | 0.682 | 0.665 | 0.828 | 0.862 | 198.53 | 215.05 | 217.23 | 165.06 | 162.20 | 214.86 | 218.69 | -16.86% | -18.30% | 8.23% | 10.28% |
| $V=4, A=5$ | 10.64% | 0.654 | 0.635 | 0.821 | 0.857 | 199.86 | 216.49 | 219.51 | 150.11 | 145.22 | 211.30 | 216.36 | -24.89% | -27.34% | 5.72% | 8.37% |
| $V=1, A=6$ | 12.55% | 0.707 | 0.691 | 0.834 | 0.867 | 216.63 | 234.66 | 236.19 | 155.37 | 167.55 | 205.76 | 212.24 | -28.28% | -22.66% | -5.01% | -1.92% |
| $V=2, A=6$ | 12.90% | 0.715 | 0.699 | 0.836 | 0.868 | 234.28 | 253.78 | 255.14 | 191.44 | 201.50 | 236.02 | 240.50 | -18.29% | -13.99% | 0.74% | 2.77% |
| $V=3, A=6$ | 11.54% | 0.681 | 0.664 | 0.828 | 0.862 | 215.38 | 233.31 | 235.69 | 170.76 | 173.16 | 222.25 | 227.72 | -20.72% | -19.60% | 3.19% | 5.85% |
| $V=4, A=6$ | 10.61% | 0.653 | 0.634 | 0.821 | 0.856 | 216.81 | 234.86 | 238.16 | 160.74 | 156.43 | 218.65 | 225.41 | -25.86% | -27.85% | 0.85% | 4.07% |

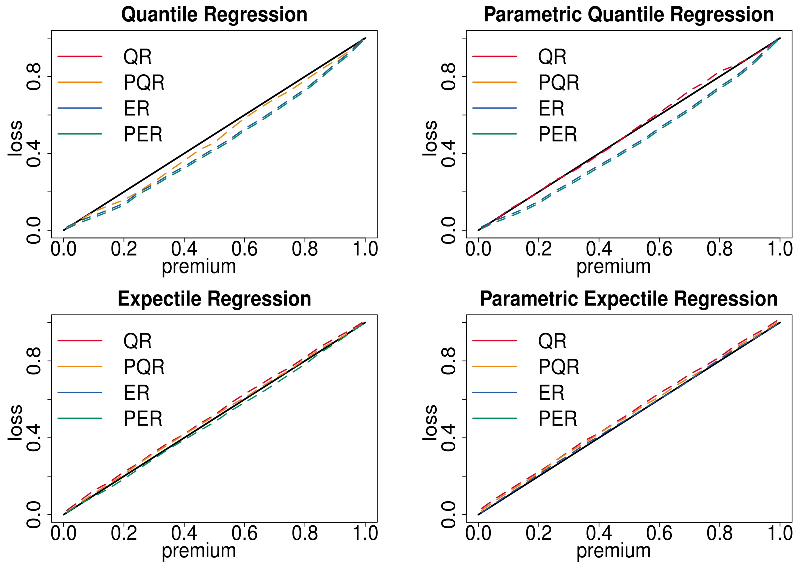


**Fig. 3.** The ordered Lorenz curves, a graph of distributions of loss versus premium ordered by relativities, with the QR, PQR, ER, and PER methods as the benchmark, respectively.

**Table 7**
The Gini indices for comparison of premium calculation methods, including EVPP, SDPP, TSQR, TSPQR, TSER, and proposed TSPER. The corresponding standard errors are given in parentheses.

| Benchmark model | Competing model | | | | | | Min-max strategy |
|---|---|---|---|---|---|---|---|
| | EVPP | SDPP | TSQR | TSPQR | TSER | TSPER | |
| EVPP | 0.00<br>(0.00) | 1.26<br>(1.79) | 0.04<br>(1.71) | 0.01<br>(1.73) | 1.28<br>(1.69) | **1.55**<br>(1.74) | 1.55<br>(1.74) |
| SDPP | -0.43<br>(1.79) | 0.00<br>(0.00) | 0.35<br>(1.72) | 0.29<br>(1.73) | 1.23<br>(1.69) | **1.55**<br>(1.70) | 1.55<br>(1.70) |
| TSQR | 12.12<br>(1.70) | 12.08<br>(1.70) | 0.00<br>(0.00) | 4.82<br>(1.69) | **12.36**<br>(1.77) | 12.33<br>(1.77) | 12.36<br>(1.77) |
| TSPQR | 11.15<br>(1.73) | 11.18<br>(1.73) | -1.02<br>(1.69) | 0.00<br>(0.00) | 11.55<br>(1.76) | **11.58**<br>(1.76) | 11.58<br>(1.76) |
| TSER | 1.79<br>(1.69) | 2.00<br>(1.70) | -2.08<br>(1.77) | -2.04<br>(1.76) | 0.00<br>(0.00) | **2.36**<br>(1.75) | 2.36<br>(1.75) |
| TSPER | 0.98<br>(1.74) | **1.06**<br>(1.71) | -1.18<br>(1.77) | -1.18<br>(1.76) | -0.95<br>(1.74) | 0.00<br>(0.00) | **1.06**<br>(1.71) |

model, and the negative value suggests the opposite. The bigger the Gini index is, the more profitable the competing model is in the sense of risk discrimination compared with the benchmark model. The maximum Gini index in each row is in bold and listed in the last column, and the subplots in Fig. 3 correspond to expectile-based methods and quantile-based ones as the benchmark, respectively. Then, we apply the min-max strategy (Frees et al., 2014) and find that the TSPER method is superior to other methods. Since TSPER is more general than TSER and the former takes more information (information across all expectile levels rather than that on a specific expectile level) of the portfolio into account, it is not surprising to find that TSPER yields better results than TSER in Table 7. It can also be seen from Table 7 that the expectile-based methods (TSER and TSPER) outperform the quantile-related ones (TSQR and TSPQR). Additionally, TSPER also performs better than EVPP and SDPP. We also conduct simulation studies under different scenarios to compare the performance of the six methods in Appendix A and they yield similar comparison results as mentioned above.

## 6. Conclusion

In this paper, we proposed a parametric expectile regression method to simultaneously estimate multiple expectiles of the response in a single run. When applied to premium calculation, this parametric modeling approach provides more parsimony and efficiency compared with the expectile regression method, especially when the number of risk classes is large. The parametric functions in the PER method can be selected from a wide range of forms, and the PER estimator can also be viewed as an extension of the ER estimator. Besides, the PER method can alleviate the crossing problem by choosing a proper structure for $\mathbf{b}(\tau)$.

Given that the magnitude of the extreme values in the right tail of the distribution may result in a huge loss for the insurance company, we introduce the coherent and elicitable expectile into the premium calculation and adapt the idea of the two-stage model to obtain the two-stage parametric expectile regression for premium calculation, which is preferable to the existing methods for heavy-tailed distributions. We propose the Expectile Premium Principle, which can be viewed as applying the EVPP for each risk class with a set of chosen loading factors, to allocate the total risk premium to each policy efficiently. As far as we know, we are the first to apply the expectile regression in the actuarial literature. The actuarial meaning of the expectile-based risk premium is also discussed in this paper. Finally, the proposed method is applied to an automobile insurance dataset. Comparison results based on the Gini index show that the proposed TSPER method outperforms the EVPP, SDPP, QR, ER, and PQR counterparts.

## Declaration of competing interest

The authors declare no conflict of interest.

## Data availability

Data will be made available on request.

## Acknowledgements

We would like to thank the editor and the anonymous reviewer for valuable comments and suggestions which have greatly helped us improve the initial version. Dr. Yu's research is supported by the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (Grant No. 21XNH155). This research was supported by Public Computing Cloud, Renmin University of China.

## Appendix A. Supplementary material

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.insmatheco.2023.05.004.

## References

Abramowitz, M., Stegun, I.A. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables.
Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge.
Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. Mathematical Finance 9 (3), 203–228. https://doi.org/10.1111/1467-9965.00068.
Asmussen, S., Møller, J.R., 2003. Risk comparisons of premium rules: optimality and a life insurance study. Insurance. Mathematics & Economics 32 (3), 331–344. https://doi.org/10.1016/S0167-6687(02)00208-1.
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 (4), 573–590. https://doi.org/10.1080/10920277.2019.1604238.
Baione, F., Biancalana, D., 2021. An application of parametric quantile regression to extend the two-stage quantile regression for ratemaking. Scandinavian Actuarial Journal 2021 (2), 156–170. https://doi.org/10.1080/03461238.2020.1820372.
Barigou, K., Bignozzi, V., Tsanakas, A., 2022. Insurance valuation: a two-step generalised regression approach. ASTIN Bulletin 52 (1), 211–245. https://doi.org/10.1017/asb.2021.31.
Bellini, F., Di Bernardino, E., 2017. Risk management with expectiles. European Journal of Finance 23 (6), 487–506. https://doi.org/10.1080/1351847X.2015.1052150.
Bellini, F., Klar, B., Müller, A., Rosazza Gianin, E., 2014. Generalized quantiles as risk measures. Insurance. Mathematics & Economics 54, 41–48. https://doi.org/10.1080/1351847X.2015.1052150.
Björkwall, S., Hössjer, O., Ohlsson, E., 2009. Non-parametric and parametric bootstrap techniques for age-to-age development factor methods in stochastic claims reserving. Scandinavian Actuarial Journal 2009 (4), 306–331. https://doi.org/10.1080/03461230903239738.
Björkwall, S., Hössjer, O., Ohlsson, E., 2010. Bootstrapping the separation method in claims reserving. ASTIN Bulletin 40 (2), 845–869. https://doi.org/10.2143/AST.40.2.2061138.
Bottai, M., Orsini, N., 2019. qmodel: a command for fitting parametric quantile models. Stata Journal 19 (2), 261–293. https://doi.org/10.1177/1536867X19854002.
Bottai, M., Orsini, N., Geraci, M., 2015. A gradient search maximization algorithm for the asymmetric Laplace likelihood. Journal of Statistical Computation and Simulation 85 (10), 1919–1925. https://doi.org/10.1080/00949655.2014.908879.
Cai, J., Weng, C., 2016. Optimal reinsurance with expectile. Scandinavian Actuarial Journal 2016 (7), 624–645. https://doi.org/10.1080/03461238.2014.994025.
Daouia, A., Girard, S., Stupfler, G., 2021. ExpectHill estimation, extreme risk and heavy tails. Journal of Econometrics 221 (1), 97–117. https://doi.org/10.1016/j.jeconom.2020.02.003.
De Jong, P., Heller, G., 2008. Generalized Linear Models for Insurance Data. Cambridge University Press, Cambridge.
de Lourdes Centeno, M., 2002. Measuring the effects of reinsurance by the adjustment coefficient in the Sparre Anderson model. Insurance. Mathematics & Economics 30 (1), 37–49. https://doi.org/10.1016/S0167-6687(01)00095-6.
Efron, B., 1991. Regression percentiles using asymmetric squared error loss. Statistica Sinica 1 (1), 93–125.
Eilers, P.H.C., 2013. Discussion: The beauty of expectiles. Statistical Modelling 13, 317–322. https://doi.org/10.1177/1471082X13494313.
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.
England, P., 2002. Addendum to "Analytic and bootstrap estimates of prediction errors in claims reserving". Insurance. Mathematics & Economics 31 (3), 461–466. https://doi.org/10.1016/S0167-6687(02)00161-0.
England, P., Verrall, R., 1999. Analytic and bootstrap estimates of prediction errors in claims reserving. Insurance. Mathematics & Economics 25 (3), 281–293. https://doi.org/10.1016/S0167-6687(99)00016-5.
Frees, E.W., 2010. Regression Modeling with Actuarial and Financial Applications. Cambridge University Press, New York.
Frees, E.W., Jin, X., Lin, X., 2013. Actuarial applications of multivariate two-part regression models. Annals of Actuarial Science 7 (2), 258–287. https://doi.org/10.1017/S1748499512000346.
Frees, E.W., Meyers, G., Cummings, A.D., 2011. Summarizing insurance scores using a Gini index. Journal of the American Statistical Association 106 (495), 1085–1098. https://doi.org/10.1198/jasa.2011.tm10506.
Frees, E.W., Meyers, G., Cummings, A.D., 2014. Insurance ratemaking and a Gini index: insurance ratemaking and a Gini index. The Journal of Risk and Insurance 81 (2), 335–366. https://doi.org/10.1111/j.1539-6975.2012.01507.x.
Frumento, P., Bottai, M., 2016. Parametric modeling of quantile regression coefficient functions: quantile regression coefficients modeling. Biometrics 72 (1), 74–84. https://doi.org/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. https://doi.org/10.1111/biom.12675.
Frumento, P., Bottai, M., Fernández-Val, I., 2021. Parametric modeling of quantile regression coefficient functions with longitudinal data. Journal of the American Statistical Association 116 (534), 783–797. https://doi.org/10.1080/01621459.2021.1892702.
Frumento, P., Salvati, N., 2021. Parametric modeling of quantile regression coefficient functions with count data. Statistical Methods and Applications 30 (4), 1237–1258. https://doi.org/10.1007/s10260-021-00557-7.
Fusco, E., Benedetti, R., Vidoli, F., 2022. Stochastic frontier estimation through parametric modelling of quantile regression coefficients, Empirical Economics. https://doi.org/10.1007/s00181-022-02273-x.
Girard, S., Stupfler, G., Usseglio-Carleve, A., 2020. Nonparametric extreme conditional expectile estimation. Scandinavian Journal of Statistics, 1–38. https://doi.org/10.1111/sjos.12502.
Gu, Y., Zou, H., 2016. High-dimensional generalizations of asymmetric least squares regression and their applications. The Annals of Statistics 44, 2661–2694. https://doi.org/10.1214/15-AOS1431.

Heras, A., Moreno, I., Vilar-Zanón, J.L., 2018. An application of two-stage quantile regression to insurance ratemaking. Scandinavian Actuarial Journal 2018 (9), 753–769. https://doi.org/10.1080/03461238.2018.1452786.

Hou, Y., 2022. A two-stage model for high-risk prediction in insurance ratemaking: asymptotics and inference. Insurance. Mathematics & Economics 104, 283–301. https://doi.org/10.1016/j.insmatheco.2022.03.003.

Hou, Y., Kang, S.K., Lo, C.C., Peng, L., 2022. Three-step risk inference in insurance ratemaking. Insurance. Mathematics & Economics 105, 1–13. https://doi.org/10.1016/j.insmatheco.2022.03.005.

IAA, 2009. Measurement of liabilities for insurance contracts: current estimates and risk margins. https://www.actuaries.org/LIBRARY/Papers/IAA_Measurement_of_Liabilities_2009-public.pdf.

IFRS, 2017. Insurance contracts. https://www.ifrs.org/supporting-implementation/supporting-materials-by-ifrs-standards/ifrs-17/.

Jones, M., 1994. Expectiles and M-quantiles are quantiles. Statistics & Probability Letters 20 (2), 149–153. https://doi.org/10.1016/0167-7152(94)90031-0.

Kaluszka, M., 2001. Optimal reinsurance under mean-variance premium principles. Insurance. Mathematics & Economics 28 (1), 61–67. https://doi.org/10.1016/S0167-6687(00)00066-4.

Kang, S.K., Peng, L., Golub, A., 2021. Two-step risk analysis in insurance ratemaking. Scandinavian Actuarial Journal 2021 (6), 532–542. https://doi.org/10.1080/03461238.2020.1863856.

Kang, S.K., Peng, L., Xiao, H., 2020. Risk analysis with categorical explanatory variables. Insurance. Mathematics & Economics 91, 238–243. https://doi.org/10.1016/j.insmatheco.2020.02.007.

Khorshidi, E., Ghezavati, V.R., 2019. Application of mathematical modeling value-at-risk (var) to optimize decision making in distribution networks. SN Applied Sciences 1, 1671. https://doi.org/10.1007/s42452-019-1715-6.

Kim, M., Lee, S., 2016. Nonlinear expectile regression with application to value-at-risk and expected shortfall estimation. Computational Statistics & Data Analysis 94, 1–19. https://doi.org/10.1016/j.csda.2015.07.011.

Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50. https://doi.org/10.2307/1913643.

Kuan, C.-M., Yeh, J.-H., Hsu, Y.-C., 2009. Assessing value at risk with CARE, the conditional autoregressive expectile models. Journal of Econometrics 150 (2), 261–270. https://doi.org/10.1080/03461238.2018.1452786.

Kudryavtsev, A.A., 2009. Using quantile regression for rate-making. Insurance. Mathematics & Economics 45 (2), 296–304. https://doi.org/10.1016/j.insmatheco.2009.07.010.

Li, D., Fukushima, M., 1999. A globally and superlinearly convergent Gauss-Newton-Based BFGS method for symmetric nonlinear equations. SIAM Journal on Numerical Analysis 37 (1), 152–172. https://doi.org/10.1137/17M1142338.

Newey, W.K., Mcfadden, D., 1994. Large Sample Estimation and Hypothesis Testing, vol. 4. Cambridge University Press, North-Holland, Amsterdam, pp. 2111–2245.

Newey, W.K., Powell, J.L., 1987. Asymmetric least squares estimation and testing. Econometrica 55 (4), 819–847. https://doi.org/10.2307/1911031.

Olivieri, A., Pitacco, E., 2011. Introduction to Insurance Mathematics: Technical and Financial Features of Risk Transfers. Springer, Berlin.

Peremans, K., Segaert, P., Aelst, S.V., Verdonck, T., 2017. Robust bootstrap procedures for the chain-ladder method. Scandinavian Actuarial Journal 2017 (10), 870–897. https://doi.org/10.1080/03461238.2016.1263236.

Pitera, M., Schmidt, T., 2022. Estimating and backtesting risk under heavy tails. Insurance. Mathematics & Economics 104, 1–14. https://doi.org/10.1016/j.insmatheco.2022.01.006.

Risk Margin Working Group, 2009. Measurement of liabilities for insurance contracts: current estimates and risk margins. Technical Report. International Actuarial Association.

Schnabel, S.K., Eilers, P.H., 2013. A location-scale model for non-crossing expectile curves. Stat 2 (1), 171–183. https://doi.org/10.1002/sta4.27.

Shi, P., Yang, L., 2018. Pair copula constructions for insurance experience rating. Journal of the American Statistical Association 113 (521), 122–133. https://doi.org/10.1080/01621459.2017.1330692.

Sottile, G., Frumento, P., 2021. Parametric estimation of non-crossing quantile functions. Statistical Modelling. https://doi.org/10.1177/1471082X211036517.

Sottile, G., Frumento, P., Chiodi, M., Bottai, M., 2020. A penalized approach to covariate selection through quantile regression coefficient models. Statistical Modelling 20 (4), 369–385. https://doi.org/10.1177/1471082X19825523.

Szynal, D., Teugels, J.L., 1993. A stop-loss experience rating scheme for fleets of cars, part ii. Insurance. Mathematics & Economics 13 (3), 255–262. https://doi.org/10.1016/0167-6687(93)90406-F.

Verschuren, R.M., 2021. Predictive claim scores for dynamic multi-product risk classification in insurance. ASTIN Bulletin 51 (1), 1–25. https://doi.org/10.1017/asb.2020.34.

Waltrup, L.S., Sobotka, F., Kneib, T., Kauermann, G., 2015. Expectile and quantile regression—David and Goliath? Statistical Modelling 15 (5), 433–456. https://doi.org/10.1177/1471082X14561155.

Yao, Q., Tong, H., 1996. Asymmetric least squares regression estimation: a nonparametric approach. Journal of Nonparametric Statistics 6 (2–3), 273–292. https://doi.org/10.1080/10485259608832675.

Young, V.R., 2004. Premium Principles, vol. 3. John Wiley and Sons, Ltd.

Zhao, J., Chen, Y., Zhang, Y., 2018. Expectile regression for analyzing heteroscedasticity in high dimension. Statistics & Probability Letters 137, 304–311. https://doi.org/10.1016/j.spl.2018.02.006.

Ziegel, J.F., 2014. Coherence and elicitability. Mathematical Finance 26 (4), 901–918. https://doi.org/10.1111/mafi.12080.